UNIVERSITY OF CALIFORNIA, SAN DIEGO

Estrogen Receptor alpha-regulated eRNAs are functionally required for enhancer:promoter looping, gene activation, and ligand-dependent relocation of genes to nuclear bodies

A Dissertation submitted in partial satisfaction of the
Requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Esperanza Núñez

Committee in charge:

Professor M. Geoffrey Rosenfeld, Chair
Professor Xiang-Dong Fu
Professor Christopher K. Glass
Professor Don W. Cleveland
Professor Bruce A. Hamilton

2012

The dissertation of Esperanza Núñez is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2012

DEDICATION

To my grandmother for teaching me that those who persevere can attain their goals.

To the special people in my life that have made a difference at one point or another.

To all of those who pursue Eureka moments because they have been an inspiration when I doubted myself.

EPIGRAPH

In the creative state a man is taken out of himself. He lets down as it were a bucket into his subconscious, and draws up something which is normally beyond his reach. He mixes this thing with his normal experiences and out of the mixture he makes a work of art.

E. M. Forster

The quest for certainty blocks the search for meaning. Uncertainty is the very condition to impel man to unfold his powers.

Erich Fromm

The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed.

Albert Einstein, What I Believe

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

To the Biomedical Sciences program for providing an excellent infrastructure for learning and interacting with elite scientists, for fostering the betterment of its students and for providing outstanding resources and attention to each individual. Thanks to Gina Butcher and Leanne Nordeman for making the program run seamlessly, even when they themselves were juggling more than they could handle.

To my advisor Dr. M. Geoff Rosenfeld who is one of the most dedicated professionals I know, and with whom I have had the privilege to work closely and experience the spell of his passion for science. Geoff is driven, resolute and extremely creative; one has to work really hard to keep up with his revolutionary ideas and capacity to embrace new technologies. I am extremely grateful for his trust and support, especially during those hard times that seemed to last forever. Thank you for your words of encouragement and for not giving up on me. Likewise I am thankful to the members of my committee: Dr. Xiang-Dong Fu, Dr. Christopher K. Glass, Dr. Bruce A. Hamilton and Dr. Don W. Cleveland for their invaluable advise, patience and support.

Just as "no man is an island, entire of itself"' (John Donne), I would not have been able to accomplish all the work presented here without the help of exceptionally talented people. Dr. Qidong Hu, Dr. Maria Dafne Cardamone, Dr. Wenbo Li and Dr. Dimple Notani for their superb collaboration resulting in insightful high quality data, useful discussion and moments of creativity. To all the members of the Rosenfeld lab, especially those that have a supporting role and remain behind the scenes managing and encouraging an effective work environment; their effort is invaluable, and at times not

appreciated enough. Among these awesome people is Marie Fisher, a former member of the lab, who is one of the nicest people I know, and a true magician when it came to making things happen. To Deanna Benson and Rachel Pardee for their assistance with daily matters, support and great capacity to improvise to accommodate impromptu situations. Also, Kenny Ohgi who deals with the wants and needs of a large group of individuals on a daily basis, and always manages to do it with a smile, politeness and much patience; Kenny's diligence towards his job is simply exemplary.

To the friends that I have met along the way for they have made my days better by providing great discussion, companionship and laughter, it is said that a friend is someone that knows all about you and loves you anyway, and I am most humbled by people's capacity to share themselves and accept others for whom they are. I am happy that throughout my graduate career I have cultivated friendships with colleagues who are a source of inspiration, support and many smiles. I am especially grateful to all the coffee drinkers with whom I have share many conversations and fun memories.

To my husband who shows me how much he loves me every single day, for his tremendous support over the years, patience, and his uncanny ability to make everything better with a hug. I am in awe of his endless pursuit of knowledge, his methodical approach to problems and the serenity he inspires in others. I love you and owe you much of what I am today, thank you for being you and loving me for being me.

To all the amazing people that in the absence of my parents have helped me reach my goals one by one; to my grandmother who cared for me with unconditional love and raised me to believe that dreams can come true. To Dr. Tom Frederick and Dr.

Robert Rothman at the Rochester Institute of Technology, for taking a personal approach to teaching and making a difference in every student's life, I am most grateful for all their help during my college years.

I am grateful to Dr. Wenbo Li who contributed Gro-seq experiments analyzed by Mr. Qi Ma, and to Dr. Dimple Notani who performed 3D-DSL experiments analyzed by Mr. Bogdan Tanasa. All these data is presented in chapter 5, which is currently being prepared for submission for publication of the material. Esperanza Núñez, the dissertation author, was the primary investigator and author of this material.

VITA

2001 Bachelor of Science, Biotechnology
Rochester Institute of Technology, Rochester, NY

2005 Teaching Assistant, University of California, San Diego
"Recombinant DNA Techniques"
BIMM 101

2012 Doctor of Philosophy, Biomedical Sciences
University of California, San Diego.

PUBLICATIONS

Nuclear organization in the 3D space of the nucleus - cause or consequence? **Nunez E**, Fu XD, Rosenfeld MG. Curr Opin Genet Dev. 2009 Oct;19(5):424-36. Epub 2009 Oct 19.PMID: 19846290.

Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. Hu Q, Kwon YS, **Nunez E**, Cardamone MD, Hutt KR, Ohgi KA, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG, Fu XD. Proc Natl Acad Sci U S A. 2008 Dec 9;105(49):19199-204. Epub 2008 Dec 3.PMID: 19052240.

A sensitive array-based assay for identifying multiple TMPRSS2:ERG fusion gene variants. Lu Q, **Nunez E**, Lin C, Christensen K, Downs T, Carson DA, Wang-Rodriguez J, Liu YT. Nucleic Acids Res. 2008 Nov;36(20):e130. Epub 2008 Sep 15.PMID: 18794177.

Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Lunyak VV, Prefontaine GG, **Núñez E**, Cramer T, Ju BG, Ohgi KA, Hutt K, Roy R, García-Díaz A, Zhu X, Yung Y, Montoliu L, Glass CK, Rosenfeld MG. Science. 2007 Jul 13;317(5835):248-51.PMID: 17626886.

Hpr (ScoC) and the phosphorelay couple cell cycle and sporulation in Bacillus subtilis. Shafikhani SH, **Núñez E**, Leighton T. FEMS Microbiol Lett. 2004 Feb 9;231(1):99-110. Erratum in: FEMS Microbiol Lett. 2004 May 1;234(1):185. PMID: 14769473.

ScoC mediates catabolite repression of sporulation in Bacillus subtilis. Shafikhani SH, **Núñez E**, Leighton T. Curr Microbiol. 2003 Oct;47(4):327-36.PMID: 14629015.

ABSTRACT OF THE DISSERTATION

Estrogen Receptor alpha-regulated eRNAs are functionally required for enhancer:promoter looping, gene activation, and ligand-dependent relocation of genes to nuclear bodies

by

Esperanza Núñez

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2012

Professor M. Geoffrey Rosenfeld, Chair

Increasing evidence reveals the rather massive transcription of non-coding RNA (ncRNA) transcripts in mammalian cells, including bidirectional transcripts at enhancer sites, referred to as eRNAs; however, it has remained unclear whether these eRNAs are functional, or merely a reflection of altered enhancer chromatin. Here, we report that estrogen receptor α (ERα) causes an upregulation of eRNAs on ERα-bound activated enhancers that is required both for specific enhancer:promoter looping, and for activation of E_2-regulated coding target genes. Estrogen regulated recruitment of the cohesin complex to enhancers, the ERα mediated relocation of induced target genes from the repressive environment of the polycomb body to the activating environment of the interchromatin granule, and the enhanced strength of specific enhancer:promoter looping, are all eRNA dependent events. Enhancer functions involve the ability of eRNAs to interact with WDR82 (a component of the COMPASS complex), as well as reading the H3K4me3 mark. Our data suggest that eRNAS are likely to exert important functions in many regulated programs of gene expression, ultimately based on their roles in altering long range chromatin interactions and the location of regulated transcription units between functionally-distinct subnuclear architectural structures.

1 Introduction

Breast cancer (BC) is the most commonly diagnosed cancer in women, with one in ten developing the disease. The National Cancer Institute (NCI) published estimates indicating that for 2011 over 200,000 new BC cases would have been diagnosed in the US, accounting for 30% of all the cancer types affecting females (1). Even with marked progress in our understanding of cancer biology, further insights into the most "distal" events underlying transcriptional regulation would be instructive to develop new approaches to modulating tumor cell proliferation and metastatic behavior. Precise regulation of a large gamut of inducible genes is central to the identity and behavior of normal and cancer cells, which are themselves influenced by specific genomic elements, particularly enhancers (2, 3).

Transcriptional enhancers are largely responsible for the expression and regulation of cell type specific factors that respond to intrinsic and extrinsic stimuli. These genomic elements are traditionally characterized as portable genomic regions that, upon binding by specific cohorts of transcription factors, can influence transcription from gene promoters (4, 5). However, it has been unclear how established enhancers might be reprogrammed to switch the transcription profile in response to specific cellular signaling pathways. While there have been substantial insights into our understanding of the mechanistic control of gene expression by core transcription factors, epigenetic cofactors and ncRNAs (6-8, 9) collective data suggests that modulation of nuclear architecture is also important to control gene expression (10, 11), suggesting the crosstalk between subnuclear architectural features and ncRNA as a

possible mechanism for controlling gene expression. Enhancers that are activated by estradiol in breast cancer cells (BCCs) are characterized by the binding of liganded ERα and the concomitant presence of mono and dimethylated H3K4 (12, 13).

Although discovered over 30 years ago, the molecular mechanisms underlying enhancer function remain rather poorly understood. It was recently reported that the methylation status of Pc2 present on growth control gene regulatory regions determines its association with two abundant ncRNAs, TUG1, and NEAT2, located in distinct subnuclear structures, polycomb bodies and interchromatin granules, respectively, and that this "switch" in methylation is required for control of cancer cell proliferation (14). Here we report comparable regulatory strategies underlie enhancer-mediated regulation by nuclear receptors, where regulated enhancer transcripts, eRNAs, serve to fine-tune the activation response by facilitating interaction with nuclear bodies, as well as enhancer:promoter looping.

1.1 References

1) http://www.cancer.org/Research/ResearchProgramsFunding/SurveillanceandHealthPolicyResearch/cancer-occurence

2) Mikkelsen, T. S., M. Ku, et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature 448(7153): 553-560.

3) Zhao, J., B. K. Sun, et al. (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." Science 322(5902): 750-756.

4) Lupien, M. and M. Brown (2009). "Cistromics of hormone-dependent cancer." Endocr Relat Cancer 16(2): 381-389.

5) Maston, G. A., S. K. Evans, et al. (2006). "Transcriptional regulatory elements in the human genome." Annu Rev Genomics Hum Genet 7: 29-59.

6) Blackshaw, S., S. Harpavat, et al. (2004). "Genomic analysis of mouse retinal development." PLoS Biol 2(9): E247.

7) Rinn, J. L., M. Kertesz, et al. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell 129(7): 1311-1323.

8) Dinger, M. E., P. P. Amaral, et al. (2008). "Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation." Genome Res 18(9): 1433-1445.

9) Costa, F. F. (2005). "Non-coding RNAs: new players in eukaryotic biology." Gene 357(2): 83-94.

10) Mercer, T. R., M. E. Dinger, et al. (2008). "Specific expression of long noncoding RNAs in the mouse brain." Proc Natl Acad Sci U S A 105(2): 716-721.

11) Wilusz, J. E., H. Sunwoo, et al. (2009). "Long noncoding RNAs: functional surprises from the RNA world." Genes Dev 23(13): 1494-1504.

12) Chen, J., H. K. Kinyamu, et al. (2006). "Changes in attitude, changes in latitude: nuclear receptors remodeling chromatin to regulate transcription." Mol Endocrinol 20(1): 1-13.

13) Felsenfeld, G. and M. Groudine (2003). "Controlling the double helix." Nature 421(6921): 448-453.

14) Marmorstein, R. (2001). "Protein modules that manipulate histone tails for chromatin regulation." Nat Rev Mol Cell Biol 2(6): 422-432.

2 Background on Methodology

2.1 MCF-7 cell line as a model system

The model system used in this work, the MCF-7 cell line, was derived from the metastatic breast tissue of a female patient. The cells retain important features of the cancer, such as the ability to respond to estrogen and its metabolites through estrogen receptor α and β, and the capacity to form clusters resembling those seen in the diseased tissue (1). It's usage in the field has allowed *in vivo* studies that have help to understand how steroid hormone receptors work and contribute to the development and differentiation of breast cancer. However, it is important to note that since the cell line was generated, it has undergone major karyotypic changes, making it a near-tetraploid. Additionally, the use of any cell line is inherently limited due to the lack of surrounding tissue and the inability of researchers to fully mimic the spatial and chemical complexity of the native extracellular environment (2). Despite these caveats, MCF-7 cells are one of the most widely used and valuable tools for the study of the mechanisms underlying tumor initiation and evolution.

2.2 Single cell nuclear microinjection

Mammalian-cell microinjection uses a glass micropipette controlled by mechanical micromanipulators to precisely deliver solutions of proteins or nucleic acids directly into the cytoplasm or nuclei of a target cell. When paired with the sensitivity of immunological staining, this technique allows the researcher to explore gene expression

and regulation at the single-cell level (3, 4). In this work specifically, intra-nuclear rearrangements were studied by injecting antibodies against nuclear myosin I, which blocks its motor activity. This was followed by immunoFISH, and is further described in chapter 4.

2.3 Fluorescence in situ hybridization (FISH)

Fluorescence *in situ* hybridization, or FISH, is a technique used to spatially localize specific nucleic acid sequences in target material. Traditionally, probes are generated by the incorporation of fluorescent or immunogenic nucleotide analogs into either DNA or RNA single-stranded fragments. In this study, biological samples were prepared on slides and fluorescently-labeled BAC (bacterial artificial chromosome) probes were applied under conditions promoting hybridization to the target sequence. Excess and non-specifically hybridized probe was then removed by stringent washing, and bound probe was imaged using widefield epifluorescence microscopy.

2.3.1 3D-FISH

FISH techniques involve relatively harsh treatment steps, which can distort the spatial complexity of the nucleus. For example, while methanol or acetone effectively fix and permeabilize material in a single step, they also dehydrate and shrink nuclei, which can change the spatial separation of nuclear targets. To address this limitation, the group led by Amanda Fisher developed a three dimensional FISH protocol in order

to more accurately determine the organization of specific intranuclear regions. This technique involves the use of fixation chemicals specifically intended to maintain spatial information, such as paraformaldehyde. There are some drawbacks to this protocol, however, such as reduced probe penetration due to crosslinked proteins, which must be addressed with further permeabilization steps. Finally, to truly take advantage of this technique the researcher must have access to either a scanning confocal microscope, or, as is the case with this work, highly accurate deconvolution software (5, 6 and 7).

2.3.2 ImmunoFISH

ImmunoFISH combines the aforementioned polynucleotide detection techniques with a standard immunofluorescence procedure targeting a specific protein or proteins of interest. When used in conjunction, these tools allow the experimenter to visualize colocalized proteins and DNA or RNA. This procedure is often used to probe for associations of genomic regions and protein complexes within the nucleus. Perhaps the greatest challenge of this technique is avoiding the destruction of target antigens during the relatively harsh FISH treatments. Various strategies can be used to overcome these problems (5, 6). In the work presented here, these problems were addressed by performing the immunofluorescence protocol first, followed by limited fixation in the FISH procedure.

2.4 Chromatin ImmunoPrecipitation (ChIP)

Chromatin immunoprecipitation, or ChIP, is an experimental technique used to isolate the nucleic acids associated with a particular protein, such as a transcription factor. Cells or tissues are treated with formaldehyde, which reacts with the primary amines of DNA bases and amino acids to form covalent bonds. This procedure is known as crosslinking, and specifically links proteins with their spatially associated genomic DNA (8, 9). The cells or tissue samples are then lysed and sonicated to break the DNA into smaller pieces, which remain covalently bound to their associated proteins. These complexes are isolated by immunoprecipitation with an antibody targeting the protein of interest. The crosslinking can be reversed via a heat-treatment, and the proteins digested with proteinase K, thus leaving a purified DNA sample amenable to further downstream manipulation, such as identification with PCR or high-throughput analysis.

2.4.1 ChIP variations

The above procedure, using formaldehyde fixation followed by sonication, is also known as X-ChIP. One weakness of this process is that excessive crosslinking reduces the yield of appropriately-sized DNA fragments. This procedure therefore requires an amplification step to generate sufficient DNA for downstream study. It can also capture transient DNA-protein interactions, confounding the search for stable associations. A variant of the procedure, referred to as N-ChIP, dispenses with the

formaldehyde and uses native, nuclease digested chromatin. Antibody binding is generally stronger against unfixed proteins, and thus the N-ChIP immunoprecipitation step is both more specific and more efficient, removing the need for post-process amplification. However, this procedure is only really useful for very tightly bound proteins, such as histones. There is also the possibility of DNA and protein rearrangements during processing, which is largely avoided in formaldehyde-treated samples (9, 10).

2.4.2 Chromatin immunoprecipitation followed by microarray chip analysis (ChIP-on-chip)

The first high-throughput analysis of the DNA isolated via ChIP was done with microarrays, a technique known as ChIP-on-chip. It can be used to determine protein binding sites, thus assisting in the search for functional non-coding genetic elements such as promoters, repressors, enhancers, or other features. Generally, probes are amplified from ChIP-enriched DNA or control DNA, and labeled with two differently colored fluorescent cyanine dyes such as Cy3 and Cy5. After hybridization, the ratio of dye intensities is used to determine the degree of enrichment of any particular genomic region in the ChIP sample. This allows the researcher to pinpoint novel potential protein binding sites across the genome (11-13). For an experimental overview of the ChIP-on-chip protocol, please see Fig 2.4.2.1.

It is important to note that this procedure suffers from the same limitations as any microarray experiment. In particular, the potential binding sites that can be discovered are limited to those present on the array. Indeed, due to the large size and complexity of the human genome, it is not feasible to obtain complete coverage with a single microarray, although there are multi-array solutions available specifically for this class of experiment. However, even these microarrays do not provide truly complete coverage or the depth of sequence information that is available with high-throughput sequencing methods. As the price of these methods has fallen dramatically in the past years, ChIP-on-chip is being quickly supplanted by ChIP-Seq. For all of these reasons, it is the latter method that was chosen for this work.

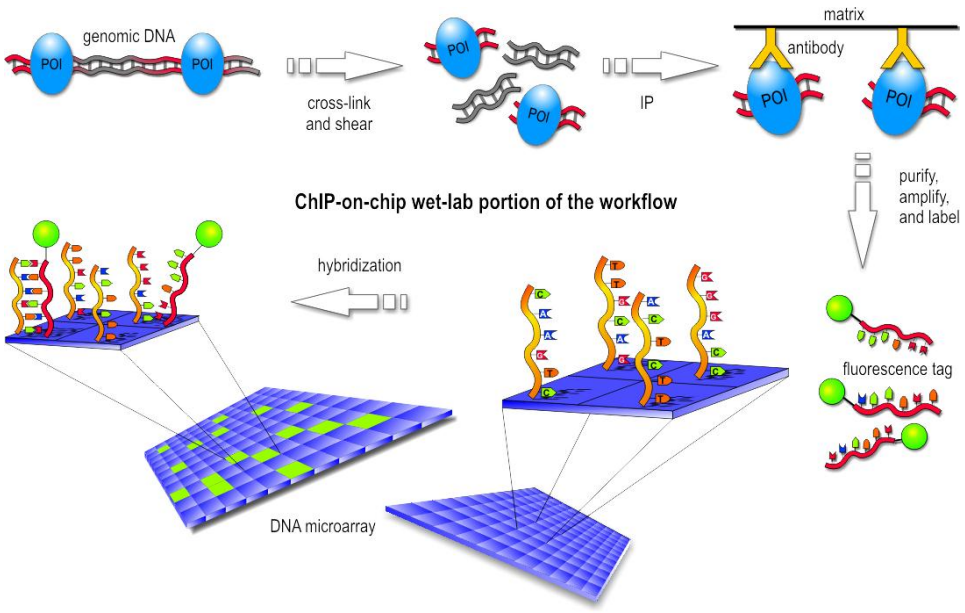


Figure 2.4.2.1: Workflow overview of a ChIP-on-chip experiment
Taken in its entirety from ref 14; this work is in the public domain.

2.4.3 ChIP followed by massive parallel deep sequencing (ChIP-Seq)

To avoid some of the limitations of the ChIP-on-chip technique, the experimenter can use massive parallel deep sequencing, a variant known as ChIP-Seq. After size selection and isolation of the fragments obtained from a ChIP experiment, the DNA pieces are ligated to adaptors specific to the sequencing platform in use. The resultant library is PCR amplified and then sequenced using one of several high-throughput sequencing machines, such as those available from Applied Biosystems or Illumina, Inc.

This technique is an open-ended approach that has the potential of providing unbiased high resolution data; however, its sensitivity depends on the depth of the particular sequencing run (the number of uniquely mapped sequence tags), the size of the genome, and the distribution of the target protein. So, for this particular approach the quality of the data is inherently correlated to cost. In this study, next generation sequencing was performed using two different Illumina sequencing platforms, the GAII and the HiSeq 2000. For a comparison with other ChIP variants, please see Table 2.4.3.1.

Table 2.4.3.1: Combinations of ChIP technologies
Constructed from refs 11-13.

ChIP-Based Techniques Used to Study Nuclear Receptor Genomic Localization								
Technique	Procedure	Assay	Identification of Genomic Regions	Binding Site Resolution	Extensive Sequencing	Bias Introduced by the Technique	Repeat Regions	Need for Data Normalization
ChIP-cloning	Linker ligation; entire ChIP fragment cloned	Sequencing	Unbiased	Low	Yes	No	No	No
ChIP-chip (PCR based array)	Linker ligation; amplification; labeling with fluorescent dye	Microarray hybridization	Biased	Low	No	Amplification	Yes	Yes
ChIP-chip (tilled array)	Linker ligation; amplification; labeling with fluorescent dye	Microarray hybridization	Array dependent	High	No	Amplification	Yes	Yes
ChIP-Seq	Linker ligation; entire ChIP library sequenced	Sequencing	Unbiased	High	Yes	Amplification	No	Yes

2.5 Chromatin Conformation Capture (3C)

Chromantin conformation capture is a technique well suited for the analysis of interactions between disparate regulatory elements. The 3C protocol, like the ChIP technique described above, starts with a fixation step. The cells or tissue samples are crosslinked using formaldehyde in order to preserve the three dimensional spatial organization of proteins and nucleic acids. The samples are then digested with a restriction enzyme to break the DNA into smaller pieces and remove the non-crosslinked fragments. A ligase is used to join the colocalized nucleic acid fragments and the crosslinking is reversed. PCR with primers designed against specific fragments can be used to identify ligated pairs, and thus study spatial interactions (15).

2.5.1 Considerations for the 3C Method

It is important to note that the presence of a ligated pair does not necessarily demonstrate specific association. Due to the flexibility of chromosomes, as well as the compact nature of the nucleus, there is a finite probability of random spatial co-localization that is inversely proportional to genomic distance. Thus, the researcher must determine the background level of interaction in order to determine the statistical significance of a ligated pair. In order to ascertain interaction frequencies, one can include primers to interaction sites separated by increasing genomic distances; this will allow a determination of the frequency of random versus specific interactions.

2.5.2 Variations of 3C

While a valuable technique, the 3C method can only be used to interrogate specific interactions, limiting its usefulness as a tool of discovery. To address this problem, several variations of the method are employed. 3C-carbon copy (5C) uses primers specific to junctions of interest, but also containing universal tails, to carry out a multiplex ligation-mediated amplification. This generates a copy of a subsection of the 3C library, which can then be analyzed via microarray or high-throughput sequencing (16, 17). However, although this method addresses 3C's single pair limit, the data generated is still specific to a particular genomic region. 3C followed by ChIP, a technique known as 4C, overcomes this limitation through the use of an exogenous bait fragment that contains a universal hybridization region for a common primer. This fragment is used during the ligation step to provide an unbiased selection, thus allowing amplification of previously unknown interacting pairs (18, 19). For a pictorial comparison of the variants, please see Fig 2.5.2.1.

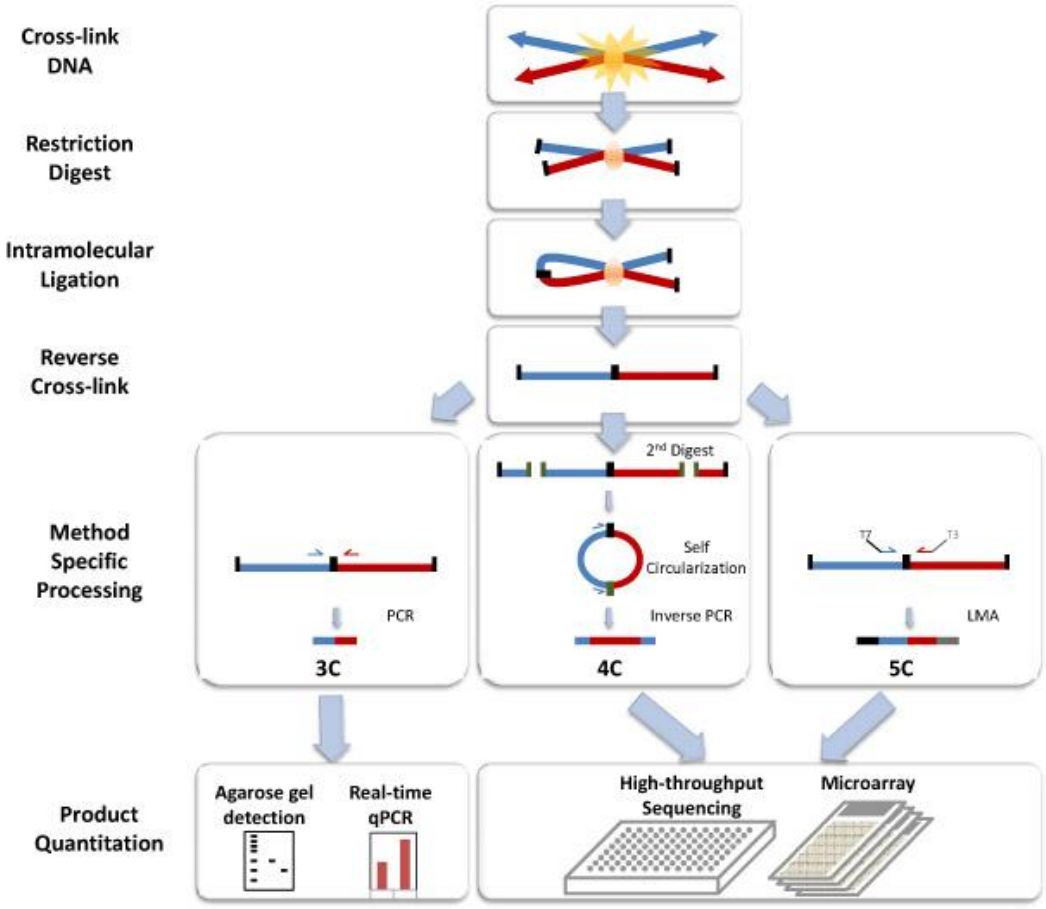


Figure 2.5.2.1: 3C variants comparison
Taken in its entirety from ref 20; this work is in the public domain.

2.5.3 Deconvolution of DNA Networks by DSL (3D-DSL)

3D-DSL is a novel combination of the 3C and ChIP-DSL protocols developed to query inter- and intra- chromosomal interactions in a high throughput manner (21). In brief, the 3C protocol is followed through the ligation step, but rather than interrogating the DNA directly at that point, an immunoprecipitation is done to enrich for a specific set of interacting regions. A biotinylated capture oligo is designed to strongly hybridize a target region of DNA, such as a promoter and/or enhancer. The digested 3C DNA

fragments are then hybridized to the capture oligo and immunoprecipitated against the biotin hapten tag by streptavidin pulldown. The ligation products are now put through a DNA selection and ligation (DSL) procedure, and the resultant DSL library is PCR amplified including linkers compatible with a sequence analyzer of choice. For a diagram of the workflow of 3D-DSL, please see Fig 2.5.3.1.

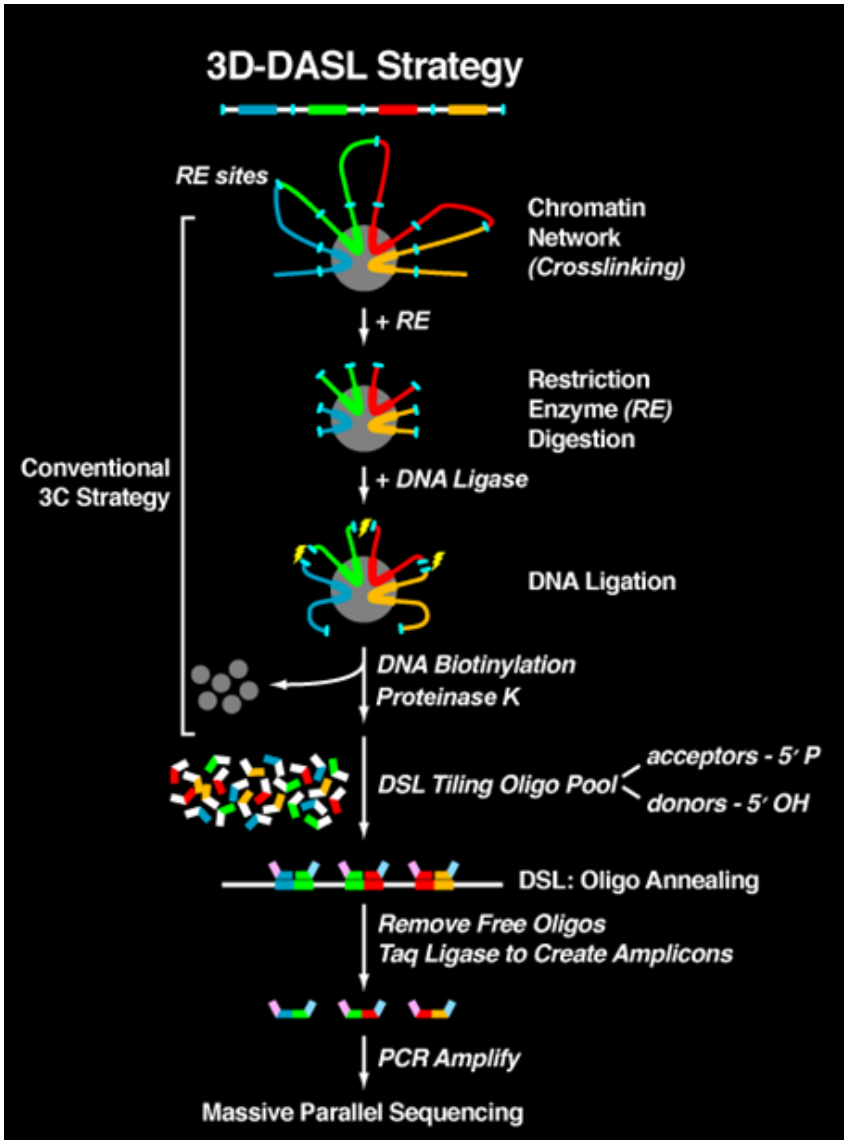


Figure 2.5.3.1: Workflow of 3D-DSL

2.6 Global run on followed by deep sequencing (Gro-Seq)

Gro-Seq is a technique that allows the researcher to profile the exact position and magnitude of transcription at a genome-wide scale. It captures, for a single point in time, all transcriptionally engaged RNA polymerases, although for the purposes of this study we refer only to RNA polymerase II (Pol II). The assay is quantitative and sensitive enough to study even lowly expressed transcripts, and can identify the location of a polymerase's active site with single nucleotide resolution. In contrast to ChIP, which can also identify the location of RNA polymerase complexes, Gro-Seq can discriminate between initiation and elongation events. These advantages allow the experimenter to search for novel transcripts and their promoters, as well as explore detailed regulatory mechanisms involving RNA turnover and polymerase pausing (22). However, it should be noted that in order to query a population of cells in an unbiased fashion, this technique may require a prohibitively costly amount of sequencing coverage to achieve the depth necessary to properly profile low abundance transcription.

2.7 References

1) Soule, H. D., J. Vazguez, et al. (1973). "A human cell line from a pleural effusion derived from a breast carcinoma." J Natl Cancer Inst **51**(5): 1409-1416.

2) Lacroix, M. and G. Leclercq (2004). "Relevance of breast cancer cell lines as models for breast tumours: an update." Breast Cancer Res Treat 83(3): 249-289.

3) Graessmann, M. and A. Graessmann (1983). "Microinjection of tissue culture cells." Methods Enzymol 101: 482-492.

4) Proctor, G. N. (1992). "Microinjection of DNA into mammalian cell in culture: Theory and practise." Methods Mol. Cell. Biol. 3: 209-231.

5) Solovei, I., A. Cavallo, et al. (2002). "Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH)." Exp Cell Res 276(1): 10-23.

6) Brown, K. E., S. S. Guest, et al. (1997). "Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin." Cell 91(6): 845-854.

7) Zirbel, R. M., U. R. Mathieu, et al. (1993). "Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries." Chromosome Res 1(2): 93-106.

8) Orlando, V., H. Strutt, et al. (1997). "Analysis of chromatin structure by in vivo formaldehyde cross-linking." Methods 11(2): 205-214.

9) Orlando, V. (2000). "Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation." Trends Biochem Sci 25(3): 99-104.

10) O'Neill, L. P. and B. M. Turner (2003). "Immunoprecipitation of native chromatin: NChIP." Methods 31(1): 76-82.

11) Buck, M. J. and J. D. Lieb (2004). "ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments." Genomics 83(3): 349-360.

12) Ren, B., F. Robert, et al. (2000). "Genome-wide location and function of DNA binding proteins." Science 290(5500): 2306-2309.

13) Wong, E. and C. L. Wei (2009). "ChIP'ing the mammalian genome: technical advances and insights into functional elements." Genome Med 1(9): 89.

14) http://en.wikipedia.org/wiki/File:ChIP-on-chip_wet-lab.png

15) Dekker, J., K. Rippe, et al. (2002). "Capturing chromosome conformation." Science 295(5558): 1306-1311.

16) Dostie, J., T. A. Richmond, et al. (2006). "Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements." Genome Res 16(10): 1299-1309.

17) Johnson, D. S., A. Mortazavi, et al. (2007). "Genome-wide mapping of in vivo protein-DNA interactions." Science 316(5830): 1497-1502.

18) Zhao, Z., G. Tavoosidana, et al. (2006). "Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions." Nat Genet 38(11): 1341-1347.

19) Simonis, M., J. Kooren, et al. (2007). "An evaluation of 3C-based methods to capture DNA interactions." Nat Methods 4(11): 895-901.

20) http://en.wikipedia.org/wiki/Chromosome_conformation_capture

21) Harismendy, O., D. Notani, et al. (2011). "9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response." Nature 470(7333): 264-268.

22) Core, L. J., J. J. Waterfall, et al. (2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters." Science 322(5909): 1845-1848.

3 Background on Biology

3.1 Estrogen receptor biology

Estrogen receptor is a steroid nuclear hormone receptor expressed in many tissue types throughout the human body, such as brain, heart, liver, breast, and others (see Fig 3.1.1 for a pictorial description). There are two isoforms, α and β, although for the purpose of this work, we will focus exclusively on ERα. In response to 17β-estradiol (E_2), the liganded receptor translocates into the nucleus, where it can oligomerize with a variety of protein complexes and activate, repress, or modulate transcriptional activity in response to changes in cellular homeostasis. Estrogen receptor is the main regulator in hormone-responsive breast cancer, where it promotes cell differentiation and tumor progression (1-4). The work presented here takes advantage of an ERα-positive breast cancer cell line, MCF-7, as a model system for hormone-dependent transcriptional regulation.

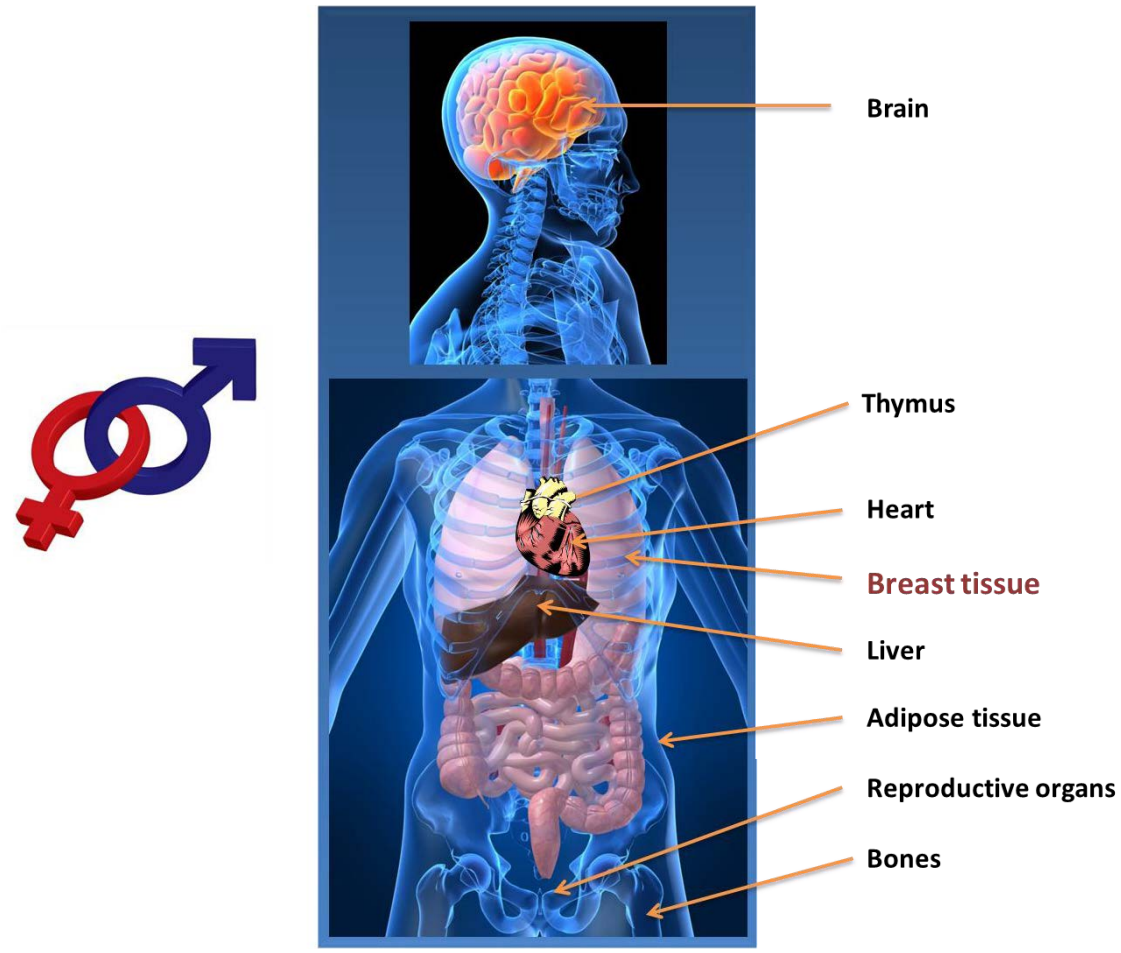


Figure 3.1.1: Tissues expressing estrogen receptor
Adapted from ref 5.

3.2 Genome-wide localization analyzes of ERα

ERα is one of the most widely studied transcription factors because of the aforementioned relevance in biology and disease. There have been several attempts to map the binding profile of ERα using the MCF7 cell line. These studies disagree as to the number of sites and their genomic distribution, likely because of differences in cell treatment (ligand, synchronization, dosage and timing), as well as the platform used

(ChIP-on-chip, ChIP-DSL, ChIP-PET or ChIP-seq). This has led to a variety of different datasets, each with their own specific bias; the current literature estimates a range of ~1,200 to ~10,000 sites with only a small subset in common (6-9).

In spite of the large discrepancy as to the absolute number of binding sites, there are some general trends common to all the studies reported thus far. ERα has a wide distribution with preferential localization outside promoters, high confidence ERα binding sites are positively and preferentially correlated to estrogen-upregulated genes rather than downregulated ones, and the majority of the upregulated genes are associated with multiple binding sites.

3.3 Noncoding RNA transcription

The falling cost of high throughput sequencing technology has accelerated the discovery of RNAs without protein coding potential (ncRNAs), further demonstrating that organisms do not solely depend on protein coding genes for all cellular functions. It is now evident that the level of transcribed ncRNAs in mammalian genomes significantly exceeds the number of genes translated into proteins. Many of the newly discovered ncRNAs have been shown to be functional and some are associated with human disease. These transcripts are generally split into two distinct groups based on function, housekeeping ncRNAs, as detailed in Table 3.3.1, and regulatory ncRNAs, discussed below (10).

Table 3.3.1: Housekeeping ncRNAs

ncRNA	Function
Ribosomal RNA (rRNA)	Translation of genetic information
Transfer RNA (tRNA)	Translation of genetic information
Small nuclear RNA (snRNA)	Pre-mRNA splicing; spliceosome components
Small nucleolar RNA (snoRNA)	RNA modifications, 2'-O-methylation and pseudouridylation
Transfer-messenger RNA (tmRNA)	Trans-translation (properties of both tRNA and mRNA)
Telomerase RNA	Telomeric DNA synthesis
SRP-Ribonuclase RNA	Signal recognition particle, RNA processing

Regulatory RNAs are implicated in a wide variety of biological processes, such as gene silencing and transcription, DNA imprinting and de/methylation, and chromatin structure dynamics, among other functions. These transcripts are often divided based on size; microRNAs are 21-25bp in length, while the remaining ncRNAs are described as either small (<200bp) or long (>200bp) (11). Lastly, ncRNAs have also been characterized based on their orientation with respect to the gene with which they are associated and the DNA strand from which they are transcribed. For a graphical explanation, please see Fig 3.3.1.

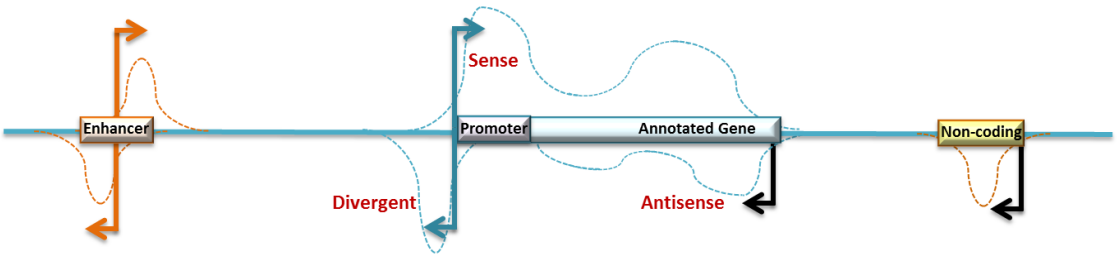


Figure 3.3.1: Classification of ncRNAs based on physical location and directionality

3.3.1 ncRNAs interact with protein complexes

There are reports in the literature that ncRNAs are sufficient to deliver protein complexes to specific genomic loci *in vivo*. For example, the recruitment of PRC2 by the ncRNA RepA is required for proper inactivation of the X-chromosome (12). Similarly, the ncRNA HOTAIR delivers PRC2 to the *HoxD* locus to silence the region in *trans* (13). Additionally, ncRNAs can modulate protein function through direct binding. The Evf-2 ncRNA forms a complex with Dlx2, which is necessary for this protein's function as a transcriptional enhancer (14). The ncRNA NRON causes accumulation of the NFAT transcription factor in the cytoplasm through its interaction with trafficking proteins (15). Lastly, ncRNAs can provide structural support or aid in the *de novo* formation of nuclear bodies. For example, the ncRNA NEAT2 is sufficient and required for paraspeckle genesis and maintenance (16-18). In this work, in particular, we are focused on the ability of eRNAs to bind protein complexes to mediate the specific associations of genes with interchromatin granules or Polycomb bodies in response to estrogen.

3.3.2 Enhancer-templated noncoding RNAs (eRNAs)

The correct response of cells to extrinsic and intrinsic signals relies on coordinated gene expression aided by distal regulatory elements. Of these, enhancers are of great interest in biology because they can augment the transcriptional response of associated genes in a distance and orientation independent manner (19). Furthermore,

an individual promoter can be targeted by one or more distinct enhancers at different times, tissues, or in response to different stimuli (20).

Enhancer-templated ncRNAs (eRNAs) are long transcripts originating from genomic enhancers. Functional studies in neurobiology (21) and immunology (22) have laid the groundwork for the characterization of eRNAs; these initial findings report that they are bidirectionally transcribed by Pol II, associated with genomic regions rich in coactivators, and responsive to cellular signaling pathways.

Prior to the discovery of eRNAs, enhancers were recognized not only for their effects on associated genes, but also for their remarkable capacity to mediate long-range chromatin interactions via looping to targeted genes (23). Thus, the question arises whether eRNAs can play structural roles, bringing the enhancer areas together with the promoter region of genes, or if they are simply innocent bystanders. The involvement of RNA could provide a three dimensional platform to recruit and organize appropriate protein complexes, many of which contain domains that bind RNA or higher-order structures containing RNA (24).

Section 3.4 in full is a reformatted reprint of the material as it appears in:

Nunez, E., X. D. Fu, et al. (2009). "Nuclear organization in the 3D space of the nucleus - cause or consequence?" Curr Opin Genet Dev 19(5): 424-436.

I, Esperanza Núñez, am the first author and my advisor Dr. M. G. Rosenfeld is the corresponding author.

3.4 Nuclear organization in the 3D space of the nucleus - cause or consequence?

Recent evidence suggests that dynamic three-dimensional genomic interactions in the nucleus exert critical roles in regulated gene expression. Here, we review a series of recent paradigm-shifting experiments that highlight the existence of specific gene networks within the self-organizing space of the nucleus. These gene networks, evidenced by long-range intra- and inter-chromosomal interactions, can be considered as the cause or consequence of regulatory biological programs. Changes in nuclear architecture are a hallmark of laminopathies and likely potentiate genome rearrangements critical for tumor progression, in addition to potential vital contribution of non-coding RNAs and DNA repeats. It is virtually certain that we will witness an ever-increasing rate of discoveries that uncover new roles of nuclear architecture in transcription, DNA damage/repair, aging and disease.

We shape our buildings; thereafter they shape us.

Winston Churchill

In the past few years we have witnessed ever-increasing investigation and insights into the regulation of genomic organization in the nucleus in response to stimulatory signals. It is now evident that the nucleus is a complex, dynamic "organelle" with functional domains encompassing incredible biosynthetic abilities such as DNA replication and repair, transcription, splicing as well as RNA/DNA modifications (25, 26). The structural components of the nucleus are intimately linked to the genome allowing signaling and ultimately regulation of gene activity. These specialized domains [nuclear matrix (NM), inter-chromatin granules (ICGs), Cajal bodies (CBs), PML bodies, etc] are characterized by the lack of demarcating membranes, but contain a defining set of proteins that form distinct "structures" and exhibit mobility in the nucleus (see Fig 3.4.1 and Table 3.4.1 for more details). A growing body of evidence also supports the existence of intimate links between cellular architecture and human disease, with alterations in nuclear organization observed in many cancers such as solid tumors, leukemias and lymphomas (27). These perturbations usually result in "misplacement" of protein complexes at "wrong" locations, leading to alterations in gene expression (28). The question remains, however, with respect to whether variations in nuclear microenvironments are the cause for their associated diseases or the consequence of biological processes, such as tumor progression.

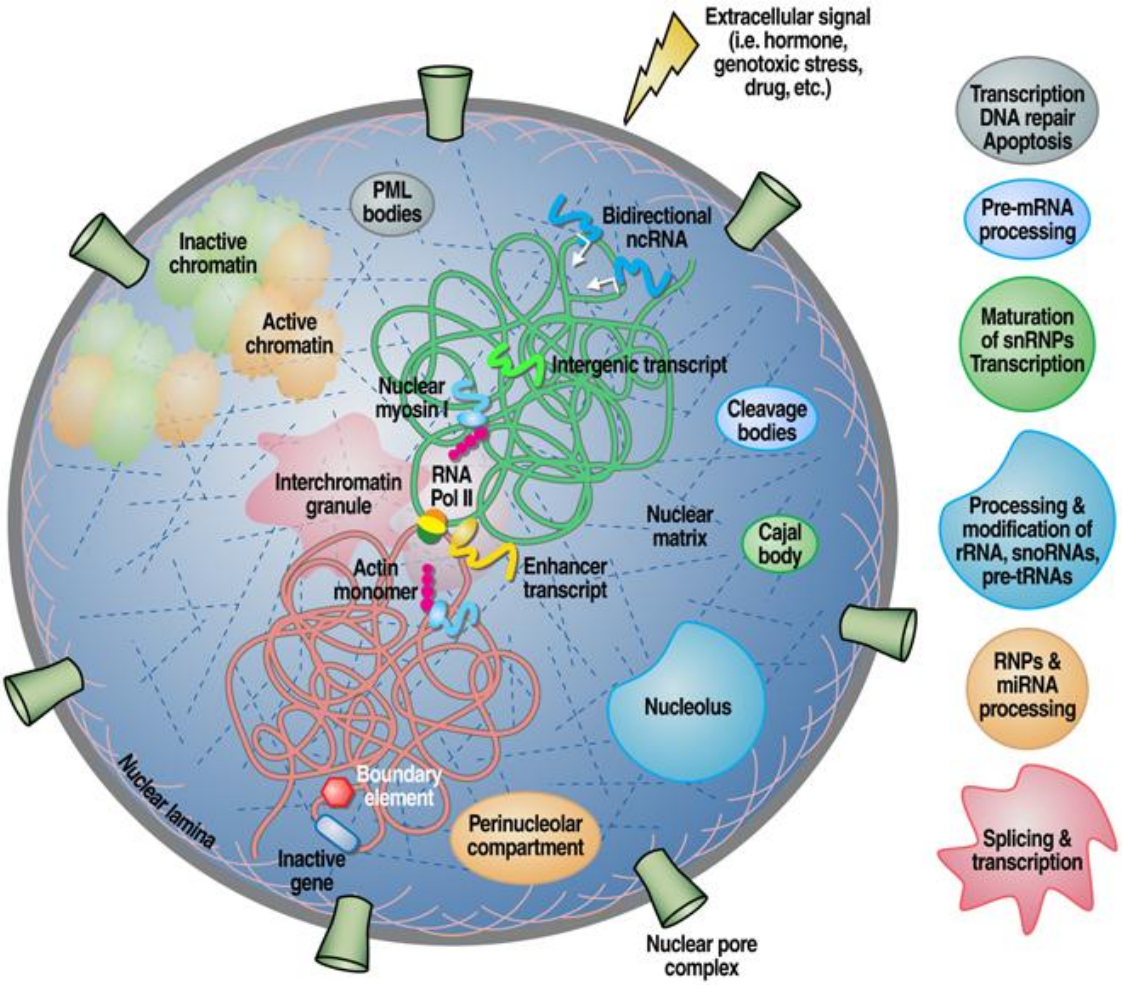


Figure 3.4.1: Cartoon depicting nuclear bodies and their organization in the nucleus. Graphical representation of the mammalian cell nucleus depicting a number of compartments and their respective functions. Also shown is the formation of interchromosomal interactions by specific genes in response to extracellular signals aided by molecular motors. Intergenic transcripts are drawn suggesting their role in regulating the epigenetic and chromatin status of the cell.

Table 3.4.1: Nuclear bodies and their reported functions

Nuclear Body	No. per cell (range)	Diameter (μm)	Major (components)	Established or proposed functions
PML Promyelocytic leukemia ND10, PODs, SP-100 and Kr	10– 30 Associated with the MHC locus on chromosome 6	0.3–1.0	Sp100, sp140, PML, SUMO-1 (PIC1, sentrin), Daxx, HAUSP, elf-4, Int-6, NDP53, NDP55, PLZF, CBP, Rb, p53, HP1, BLM, P27Kip1, RFP, ISG20, GRIP-1, nascent RNA, LYSP100, AP-1, RARα, TIF1α, Sp1	Transcription, DNA repair, viral defence, stress, cell cycle regulation, proteolysis, apoptosis.
IGCs Interchromatin granule clusters Splicing speckles, SC35	25– 30	0.8–1.8	pre-mRNA splicing factors, snRNPs, SR proteins, RNA pol II, PBA2, HnRNPs, Clk/STY, hPRP4, forms of protein phosphatase I	Storage of mRNA splicing factors, transcription , SR protein storage,
Paraspeckle	10-30 Often found adjacent to IGCs	0.2–1.5	PSP1, PSP2, p54/nrb, CPSF6, BCL11A, BCL6	Transcription, pre-mRNA splicing
CB-GEM Cajal Body/Gems	0-10 Often found adjacent to nucleoli	0.1–2.0	p80 coilin, fibrillarin U1-U8, U11, U12 snRNPs, Nopp 140, NAP57 (Cbf5, dyskerin), Gar1, Pol IIo LS, PTF (SNAPc), TFIIF (RAP74), TBP, topoI, histone SLBP, TFIIH (cyclin H/cdk7/Mat1/ p62), U2B/U1A, cyclin E/cdk2, trimethylguanosine cap, PKA (cAMP-dependent kinase), Sm proteins, PKR, (DAI kinase), SMN/SIP1, MEQ oncoprotein, TLS/FUS (hnRNP P2, pigpen), lamin A, RP S6, scaRNAs (U85, U87, U88, U89, U90, U91, U92) ELL, EAF1, gemin2/SIP1, FLASH, nuclear actin	snRNP biogenesis/modification; trafficking of snoRNPs and snRNPs, to nucleoli or to IGCs.
Nucleolus	1– 5	0.5–5.0	p80 coilin, fibrillarin Nopp 140; ARF, MDM2, p53, components of telomerase, UBF, MRP subunits, Rpp29, B23, ribosomal subunit proteins (S5, L9), RNA polymerase I, nucleolin, Nop52 RENT (regulator of nuclear silencing and telophase exit), UBTS, BLM, NOL1, NOL5A, NOLC1, Nrap, TCOF1, TOP2B, DDX5, DDX21	rDNA transcription, rRNA biogenesis; rRNA metabolism, cell cycle regulation (by sequestration of proteins), cell lifespan, translation, SRP biogenesis, protein folding and primary microRNA processing.
PNC Perinucleolar compartment	Variable in transformed cells Rarely in normal primary cells	0.25–1.0	PTB/hnRNP I, CUG-BP/hNab50, KSRP, Pol III SmRNAs (RNase P, MRP RNA, hY RNA	RNA processing/metabolism
Cleavage bodies	1– 4 foci Often found adjacent to CBs Subclasses may exist (-/+ nascent RNA, cell cycle regulated)	0.3–1.0	CstF-64 (S-phase), CPSF-100 (S and G2 phases,) , DDX1, RNA Pol II	Cleavage and polyadenylation steps of mRNA processing
OPT domain Oct1/PTF/Transcription	1-3 Appears in G1, dissapears in S-phase Often found adjacent to nucleoli Often associated with chromosomes 6, 7 Often associated with PIKA domains	1.0-1.5	PTF, Oct1, TBP, SP1, RNA pol II, nascent transcripts, TBP	Transcription of certain PTF- and OCT1- dependent genes
PcG-Ring1 Polycomb group-Ring1	Variable in transformed cells Not in normal primary cells as NBs but rather uniform distribution Often close to sites of constitutive heterochromatin Cell cycle regulated	Variable 50-100 in embryonic cells	Composition likely to vary between cell types PRC1 complex (Pc1-3, Bmi, mel18, Rae28, Edr2, RING1-2, YY1) PRC2/3 complex (EZH1/2, Eed, Suz12, Pcl1-3) PRC4 complex (undifferentiated pluripotent cells and cancers) Eed2, SirT1	Histone modification, epigenetic control (imprinting, x cromosome inactivation), maintenance of stem cell identity, regulation of homeotic genes, repression
Nuclear stress bodies (nSB) Sam68/SLM Nuclear bodies (SNBs) HSF1 granules HAP granules (hnRNP A1 interacting protein)	Variable Number is related to cell ploidy Often found close to nucleoli Detectable under stress	0.3 to 3	Sam68, SLM1, SLM2, YT521-B, BRK/SIK, HSF1, HAP, SRp30c, 9G8, SF2/ASF, ncRNAs	pre-mRNA processing, RNA metabolism, splicing post-transcriptional regulation, heterochromatin spread

Intriguingly, interphase chromosomes are organized into discrete chromosome territories that exhibit a non-random distribution within sub-nuclear positions, but appear sufficiently flexible to allow interchromosomal interactions with important consequences for genome function and stability (28-30). Based on a large body of groundbreaking work, this field of investigation is rapidly moving towards developing genomic and proteonomic tools that will help elucidate the molecular principles governing the cellular organization in the 3D space of the nucleus.

3.4.1 On chromosome territories and interchromosomal interactions

The non-random positioning of chromosome territories seems to be related to chromosome size, gene density and morphology, with gene-rich chromosomes tending to be located at the center of the nucleus, while gene-poor ones tend to associate at the periphery with the nuclear lamina. In an attempt to search for general rules of the localization of chromosomal domains, two groups recently embarked on genome-wide localization studies of nuclear lamina interactions in Drosophila (31) and human cells (32). Both studies revealed that, for the most part, the association with nuclear lamina was characterized by a repressive chromatin environment, devoid of active histone marks and RNA polymerase II binding activities. While in human cells the lamina-interacting domains have been delineated by the insulator protein CTCF and CpG islands, it is important to note that not all loci with perinuclear localization are silenced; the work of Finlan et al., 2008 (33) provides evidence for the causative role of the nuclear periphery in altering gene expression in human cells. This group was able to relocate specific human chromosomes to the nuclear periphery by tethering them to a protein of the inner nuclear membrane and showed that this process could reversibly suppress the expression of some endogenous genes but not others, demonstrating that location at the nuclear periphery is not incompatible with active transcription.

A similar argument has been made with respect to chromosome size, where small chromosomes are located more internally than larger ones (34, 35). New methods of image processing suggest that non-spherical chromosome territories with high surface-to-volume ratios are more likely related to transcriptional activity, thus more

likely to engage in aberrant chromosome exchanges. The ellipsoid morphology is thought to provide more contact points with nuclear neighbors in order to favor transcriptional activity (36). Beyond the linear sequence of the genome lies the higher order chromatin folding events that bring into contact distal enhancer elements with specific promoters. Likewise these interactions can be bypassed by the action of insulators, thus mediating distinct 3D topologies. Interaction of genes with regulatory elements, both intra- and inter-chromosomal, has been demonstrated by several methods, including fluorescence in situ hybridization (FISH), RNA tagging and recovery of associated proteins (RNA-TRAP), and several versions of the chromatin conformation capture (3C) technique combined with arrays or high-throughput sequencing.

A recent example of intrachromosomal interactions is reported by Jing, et al. 2008 (37), on the c-kit gene where the importance of chromatin loop transitions is demonstrated during erythroid cell differentiation regulated by the transcription factors GATA-1 and GATA-2. Though both of these transcription factors bind to cognate DNA sequence elements, they act in a sequential fashion to allow c-kit expression during early erythropoiesis and repression later in maturation. GATA-2 first induces an activation loop that is later repressed by a downstream loop aided by GATA-1 as demonstrated by chromatin immunoprecipitation (ChIP) and 3C. This work suggests that a GATA factor switch is responsible for reconfiguring the higher order chromatin organization to allow de novo formation of a repressive domain (see Fig 3.4.1.1).

Also interesting are the recently-described interchromosomal interactions and the implications of their biological consequences, (for reviews see 38, 39). Fundamentally, it is not mechanistically clear how distal regulatory elements actually "find" their targets in the complex space of the nucleus; and apparently, not all answers reflect functional compartmentalization or the physical constraints imposed by molecular crowding. An illustrative example is the active multi-step mechanism that regulates interactions mediated by nuclear receptor (NR) activated loci, both within the same and between different chromosomes (40). These studies, however, do not address whether long-distance interactions reflect random movement and subsequent "high affinity" interactions between the involved regions, or an ordered movement towards specific complexes or "structures". However, there is now an integrative view of nuclear architecture and genomic function linked to an initial molecular requirement for the formation of "transcriptional hubs" or "factories". In the case of estrogen-dependent intrachromosomal interactions, siRNA or antibody nuclear injection experiments coupled with immuno-FISH demonstrated a central role for specific chromatin remodeling factors, molecular motors that include nuclear myosin I (NMI), and a number of transcriptional co-activators (41). These authors provide evidence that the interchromosomal interactions that they described were required for ligand-induced gene expression (RNA-FISH). Furthermore, these loci move into interchromatin granules, thus coupling the processes of transcription and mRNA processing (see Fig 3.4.1.1). Alternatively, the idea that phosphorylation of Pol II at serine-5 marks the location of what has been referred to as "transcriptional factories" has been presented (42). Therefore, questions regarding the precise locations of these chromosomal

interactions and their relationship to sites of active gene transcription remain to be elucidated, undoubtedly requiring further technological advances, to provide definitive answer. The list of unanswered questions is lengthy, and includes: how are the NMI-actin complexes or Pol II complexes recruited to chromatin regions? Does this complex take advantage of the molecular motor properties of RNA polymerase II? Although it is clear that both actin and myosin play significant roles in transcription and chromosomal architecture, the precise biochemical/structural details of how they orchestrate chromosomal movements in the nucleus remains an unsolved question.

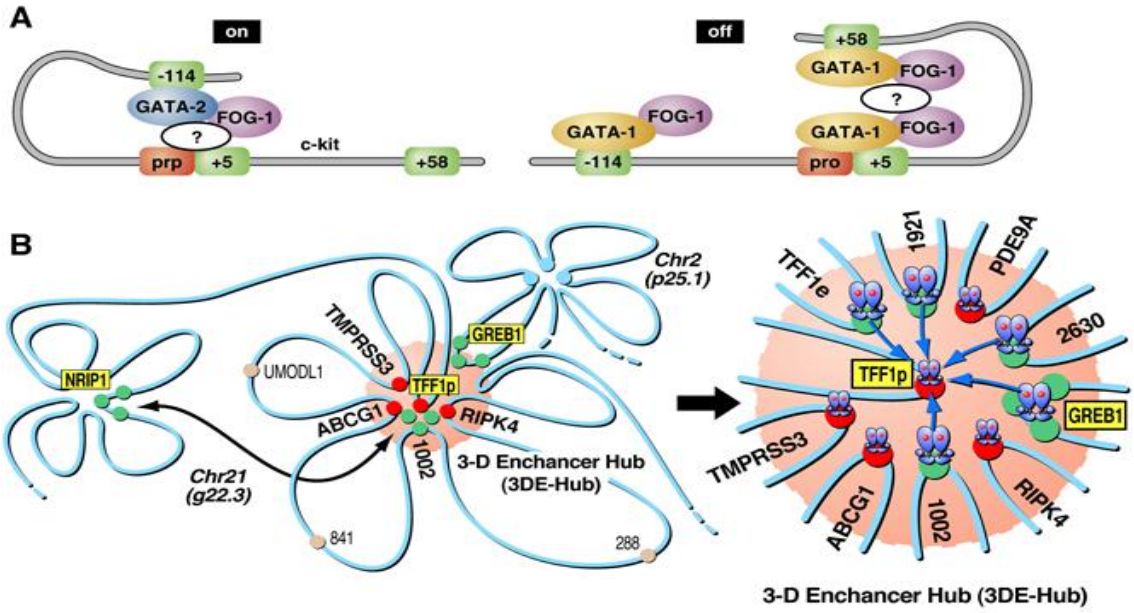


Figure 3.4.1.1: Interchromosomal interactions reported in the literature
(a) Model of the chromosomal configuration of the c-kit gene as adapted from Ref. [13]. The cartoon shows GATA-2 binding to the -114 enhancer in immature erythroid cells resulting in the activation of c-kit. During maturation, GATA-1 replaces GATA-2, allowing the contact with downstream elements, blocking accessibility to the enhancer, and ultimately leading to repression of the gene. As noted in the illustration, the effects of both GATA factors require FOG-1. (b) Proposed model of E2-induced, actin/myosin1/DLC1-mediated chromosomal movement, and LSD1-dependent interactions with interchromatin granules, creating a three-dimensional enhancer hub in the nucleus. TFF1 (chromosome 21):GREB1 (chromosome 2) interactions are depicted, indicating chromosomal movement and long-distance DNA looping.

3.4.2 The role of non-coding RNAs in nuclear architecture

The general character of interchromosomal interactions as a potential mechanism for gene expression control was exposed in the work of Apostolou and Thanos, 2008 (43) describing how viral infection brings about interchromosomal interactions between the interferon-β (*IFN-β*) gene enhancer and DNA binding sites for the NF-κB transcription factor, resulting in a nuclear rearrangement that initiates an

antiviral response. Remarkably, all of the reported interacting sequences in these events contained Alu repeats that harbored the NF-κB binding sites, suggesting the regulatory function of repetitive sequences in nuclear organization and genomic regulation (see Figure 3.4.2.1). Indeed, Zuckerkandl and Cavalli, 2007 (44) as well as Lunyak et al., 2007 (45) suggest that transcription of interspersed repetitive sequences and their regulation by epigenetic mechansims represent a strategy for the establishment of functionally distinct chromatin domains. The contribution of such DNA repeats, previously deemed "junk DNA", to regulated gene expression is still poorly understood, though there are increasing reports of non-coding RNAs derived from repeats and intergenic regions (46). These RNAs often contribute to the determination of chromatin structure, as well as the transcriptional and posttranscriptional control of gene expression.

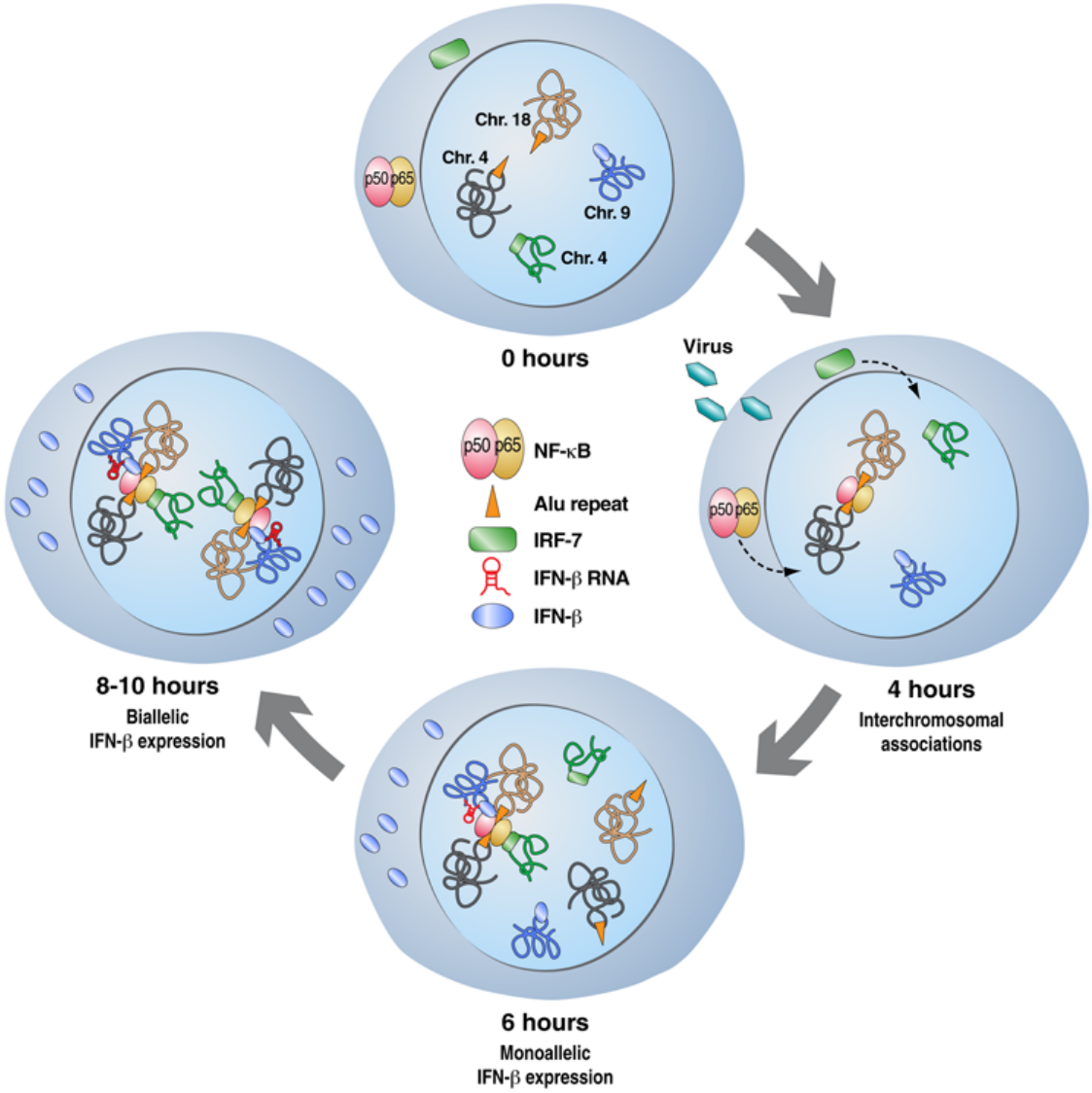


Figure 3.4.2.1: Model of the bivalent expression of the IFN-β gene.
Virus infection induces the translocation of NF-kB and IRF-7 into the nucleus. An interchromosomal interaction is brought about by NF-kB binding to Alu repeats in chromosomes 4 and 8, respectively, along with IRF-7. Monoallelic expression of IFN-b takes place in the first six hours of this event, while accumulation of IRF-7 and binding to the gene results in its biallelic expression.

One of the most detailed studies of a non-coding RNA involved in interchromosomal interactions concerns the two X-chromosomes in mammalian female cells. The *Xist* and *Tsix* genes are antisense to each other and are transcribed at low

levels prior to X-chromosome inactivation (XCI). At the initial stage of XCI, *Xist* is upregulated and its RNA transcripts paint the entire inactive X (X_i) chromosome where *Tsix* is repressed. Reciprocally, an increase in *Tsix* transcription represses *Xist* on the active X (X_a) chromosome (47). Elegant strides using differentiating mouse ES cells by Bacher, CP et al., 2006 (48) showed that, in a significant number of nuclei, the two X-chromosomes come together transiently at the onset of the inactivation process, prior to X_i and X_a being targeted to different nuclear compartments. Recently, Zhao et al., 2008 (13) discovered another non-coding RNA involved in XCI, which is transcribed from a short repeat within the *Xist* locus; they named this new ncRNA *RepA* and found PRC2 (a HMTase responsible for H3K27 trimethylation at Polycomb target genes) to its intangible target. PRC2 was found to be recruited to the X chromosome by *RepA*, with EZH2 functioning as the RNA binding subunit; conversely, *Tsix* inhibits the interaction. The authors showed that *RepA* depletion abrogates the induction of *Xist* along with the H3K27 trimethylation mark. In a similar fashion, PRC2 deficiency prevents *Xist* upregulation. Altogether, these findings suggest that *RepA* and PRC2 are needed for the spread of XCI by recruiting components of the Polycomb complex (see Fig 3.4.2.2).

Another well-known case of nuclear re-organization regulated by a non-coding RNA is represented by the *HOX* gene cluster during embryonic development. Rinn et al., 2007 (12), by means of high-resolution tiling array on human fibroblasts, characterized 231 non-coding RNAs involved in *HOX* gene expression, of which *HOTAIR* was identified as a *trans* regulator. *HOTAIR* is localized to a regulatory boundary in the *HOXC* cluster; however, its knockdown showed no local effect but rather a striking de-repression of the *HOXD* cluster along with the dismissal of PRC2.

Pull-down experiments demonstrated a direct interaction between *HOTAIR* and components of the PRC2 complex. These data suggests that *HOTAIR* transcription demarcates chromosomal domains of gene silencing at a distance, raising a series of unsolved mechanistic questions (see Fig 3.4.2.2). Does this reflect an epigenetic effect or an indirect consequence of *HOX* gene regulation? How many other genes are transcriptionally affected? Does this really reflect a direct regulation in *trans*? Are there additional functional, *trans*-acting non-coding RNAs, and do they serve as key components of the differentiation process in many regions?

The principle that the nucleus and its major components are self-organized combined with the discovery of ncRNAs that colocalize with nuclear bodies, suggest that there may be an architectural role for these RNA species. A preamble to this idea was presented by Hutchinson et al., 2007 (49) who identified two ncRNAs NEAT1 and NEAT2, associated with two related nuclear domains. The question evidently became whether these RNAs performed a unique role in nuclear body organization, or its localization simply reflected trafficking to an area of activity/storage. Both NEAT1 and NEAT2 are highly conserved long non-coding RNAS; this study presents evidence that while NEAT2 localizes within nuclear speckles, NEAT1 is found at the periphery in paraspeckles. Follow up work by several groups (16-18), further characterizes the role of both transcripts, demonstrating that NEAT1 is sufficient and required for paraspeckle formation and maintenance. In contrast, NEAT2 knockdown showed little effect on the integrity of nuclear speckles. Altogether, these studies pave the way for a new class of functional ncRNAs with structural capabilities.

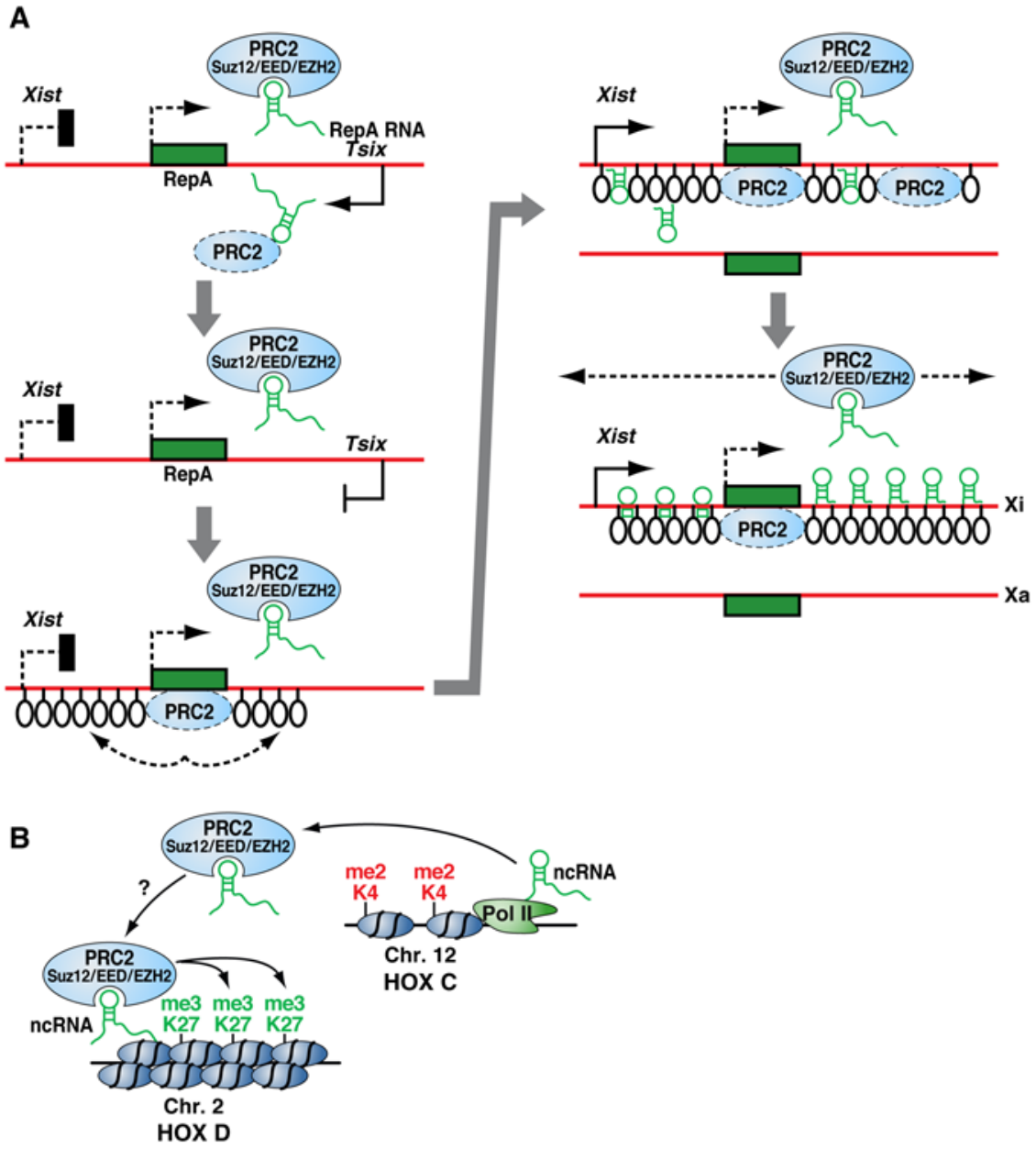


Figure 3.4.2.2: Model of X-chromosome inactivation
(a) The inactivation process starts with the expression of Xist and the newly discovered ncRNA RepA. The RepA transcript recruits the polycomb complex PRC2, which in turn deposits the H3K27me3 mark along Xi establishing and maintaining repression. (b) Model of the HOTAIR ncRNA transcribed at the HOXC locus mediating epigenetic silencing of HOXD. HOTAIR recruits the H3K27 HMTase PRC2 to HOXD resulting in a *trans* silencing effect.

3.4.3 On interchromosomal interactions, genome rearrangements and instability

The concept of chromosome territories intermingling (30), the non-random spatial positioning of genes within the nucleus (26) and the idea of interaction centers defined as transcription factories (42) have given rise to a transcription-based, contact-first model of chromosomal aberrations. This model suggests that a double strand break (DSB) induced in a loop from one chromosome becomes associated with a topoisomerase-I (topo-I) at a stalled RNA polymerase site. The trapped DNA-topo-I then interacts with another topo-I bound to a transcription unit on another chromosome (50). This model assumes that the two DNA-topo-I cleavages are reversed, causing the two chromosomes to misjoin. DSBs are the ultimate lesions for the formation of chromosomal aberrations (51), but the role of chromatin structure in their formation and processing in damage is not very well defined.

Soutoglou et al., 2007 (52) have shed some light on the dynamics of single DSBs via live microscopy imaging of tagged chromosomes in mammalian cells. In this system, each side of the break is separately marked by CFP-lac-repressor and YFP-lac-repressor, thus allowing tracking of the broken DNA ends in real time. Tracking of the break revealed that broken ends are maintained in a somewhat stable position by the DNA binding protein Ku80. This period of reduced motion is believed to allow the repair process to take place; however, if the DSB is not repaired, the opened ends are free to participate in rearrangement events. These results support a contact-first model

in which chromosome translocations predominantly form among spatially proximal DSBs.

Following on the contact-first model, observations on interchromosomal translocation partners have revealed that translocation events usually do not alter the position of the intervening chromosomes, even during cancerous transformation (53). These results are in agreement with the idea that the spatial orientation of chromosome territories is in direct relationship with the frequency of translocations, and is determined by the spatial proximity of the interacting partners. Boyle et al., 2001 (35) proposes that the positional properties of chromosomes may be related to chromosomal instability based on the fact that highly eccentric, ellipsoidal chromosome territories have a high surface-to-volume ratio to allow more contact points between chromosome territories. By this definition, heterologous neighborhoods may protect against loss of heterozygosity (LOH) caused by interhomologue recombination. In principle heterologous chromosome territory neighborhoods may act as a buffer zone to prevent inappropriate homologous interactions. Under this paradigm, prevention of homologue interactions is a cancer-protective mechanism given that loss of tumor suppressor function by LOH can directly promote tumorigenesis. A corollary to translocations is the phenomenon of *trans*-splicing, which selectively joins exons on separate pre-mRNAs in order to produce mature messages potentially encoding proteins with distinct structures and functions (54). It is logical to presume that this event might be aided by the presence of the interacting partners in the same transcription factory, provided that all the steps of mRNA processing take place co-transcriptionally. Furthermore, splicing

factors associated with the C-terminal domain of RNA polymerase II ensures the correct joining of exons. Unneberg and Claverie, 2007 (55) have used bioinformatic approaches to query public databases for the presence of chimeric EST and mRNA sequences, with a hope to map interchromosomal interactions. This study reports a total of 5,614 chimeric ESTs and 587 chimeric mRNAs; future studies will differentiate the cause of these fusion gene products as a result of chromosome translocation or trans-splicing or both.

It is now well accepted that modifications in nuclear architecture, epigenetics and gene expression are hallmarks of cancer and aging cells; therefore, understanding the governing factors of cellular organization may provide a platform for novel approaches to diagnosis and therapy.

3.4.4 CTCF: a common denominator in mediating long-range interactions and genomic reorganization at nuclear pore complexes.

The emergence of a three-dimensional model of gene regulation brings about a number of questions regarding the key players responsible for controlling nuclear organization and long-range interactions (56). Recent findings point to the ubiquitously-expressed transcription factor CTCF as a major organizer of chromatin domains in addition to other well characterized functions. CTCF is an 11-zinc-finger protein highly conserved in vertebrates that has been implicated in enhancer blockage, boundary establishment, and both transcriptional activation and repression (see Fig 3.4.4.1 [A, B])

(57-62). The mechanism by which CTCF helps mediate long-range interactions is not completely understood.

A classical example of the involvement of CTCF in long-range interaction is exemplified by T-helper cells where CTCF mediates intrachromosomal interactions resulting in the monoallelic expression of interferon-γ and interleukin genes (63). CTCF has also been shown to mediate interchromosomal interactions between the *Igf2/H19* locus on mouse chromosome 7 and an intergenic sequence between the *Wsb1* and *Nf1* genes on mouse chromosome 11 (58). The question of whether other loci also interact in *trans* in a CTCF-dependent manner remains unanswered. A more recent example is presented in random X-inactivation involving counting and choice mechanisms, where *trans* X-X interaction and possibly X-autosome interactions might be required (64). CTCF is proposed to mediate the physical association between X-chromosomes (62). It remains to be elucidated whether CTCF is the major factor regulating this phenomenon in the context of long-range interactions (see Fig 3.4.4.1 [C]).

Two studies report the genome-wide binding program of CTCF in the mouse (65) and human genomes (61). It has been estimated that there are about 4,000 CTCF binding sites in the mouse genome, while the corresponding number is 13,804 in the human genome. In both cases, most of the sites are methylation sensitive and map away from the transcriptional start site, with their distribution strongly correlated with gene density. In the human study, 46% of the CTCF binding sites are located in inter- and intragenic regions, with some sites containing Alu-like repeat elements (66), possibly acting as regulated insulator elements.

Because CTCF acts upon the topological organization of the genome inducing the formation of long-range chromatin loops, Cuddapah et al., (67) used ChIP-seq to map CTCF binding sites in three different cell lines, revealing that CTCF binding sites are significantly enriched at the boundaries between the H3K27me3 and H2AK5ac domains, supporting its role as an insulator (see Fig 3.4.4.1 [D]). Analysis of nucleosome positioning in the vicinity of non-promoter CTCF binding sites indicated that CTCF binds to the linker region between nucleosomes, and the nucleosomes surrounding the functional binding sites are well positioned. Since cohesin shares the consensus motif and colocalizes extensively with CTCF, it has been suggested that cohesin may also function as a transcriptional insulator (68). It is thus of importance to investigate the role of cohesin in the barrier action of CTCF. To complicate matters, the behavior of CTCF seems to be regulated at several levels including binding to the target site, binding to interacting partners (Sin3 (69), YB-1 (70), nucleophosmin (71), Kaiso (72), cohesin (68), Pol II (73)), and post-translational modifications. Additional regulation by post-transcriptional modifications is likely as CTCF is known to be phosphorylated (74), subject to poly(ADP-ribosylation) *in vivo* (75), and sumoylated (76).

Following on the function of covalent histone modifications in nuclear architecture, and their association with transcriptional activity, Brown et al., (77) examined the relationship between the mammalian nuclear pore and the human genome by generating high-resolution, chromosome-wide binding maps of human nucleoporin 93 (Nup93) in the presence and absence of a potent histone deacetylase inhibitor

(HDACI). This analysis defined regions of functional interactions between the nuclear pore and the human genome, emphasizing the role of the nuclear pore as a boundary element. Surprisingly, Trichostatin-A (TSA) treated Nup-93 BSs were enriched in CTCF-associated regions, suggesting that CTCF interacts with the nuclear periphery and that these regions of CTCF-mediated regulation are repositioned proximal to the nuclear pores following TSA treatment. The implication is that the balance of histone modifications and CTCF are important components of boundary elements at the mammalian nuclear periphery.

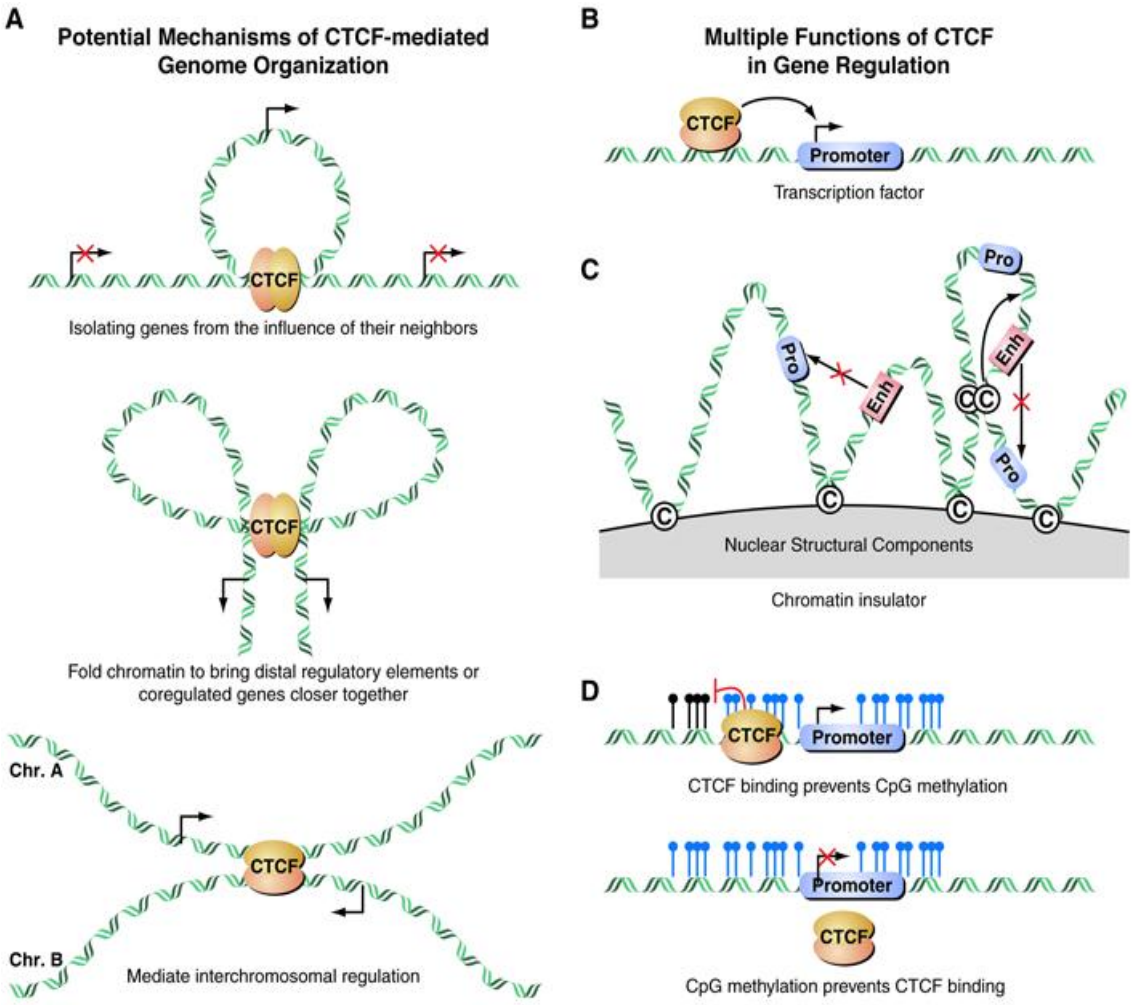


Figure 3.4.4.1: Mechanisms of CTCF-mediated genome organization
(a) CTCF can demarcate chromatin domains of activation/repression; create long-distance interaction loops; or mediate interchromosomal interactions. (b) CTCF can function as a classical transcription factor to regulate gene expression at its target sites. (c) CTCF can act as an insulator preventing promoter–enhancer interactions and establishing topologically distinct nuclear domains. (d) CTCF can determine the epigenetic status of a chromatin segment upon its binding profile because CpG methylation can oppose the binding of CTCF and vice versa.

3.4.5 Perspectives

The emerging three dimensional view of genome organization and nuclear architecture indicates that the location of a gene within a chromosome territory seems to

influence its ability to access the machinery responsible for specific nuclear functions, such as transcription and splicing. It is important to emphasize that this topological regulation is indeed quite dynamic and responsive to the ever changing status of the cell; thus understanding the molecular principles of its establishment and maintenance may help discover its role in both normal biological processes and disease states. A careful and exhaustive analysis of the biophysical properties of long-range interactions is much needed in order to understand the kinetic and thermodynamic requirements of specific networks formed as a response to the multiple stimuli that cells are exposed to. This brings the question of whether the nuclear matrix plays an integral role in nuclear architecture by providing a supporting platform for tethering various molecules, which is a titanic question to address given the number of proteins possibly involved in this phenomenon. For even as distinct a component of the nuclear matrix as the nuclear lamina, we are yet to unveil all its rules in genomic regulation. We can anticipate the identification of proteins required for "correct" positioning of chromosome territories and nuclear compartments as well as possible affecting mutations. Several studies to date have noted the energy requirement for chromosomal movement. Although the idea of nuclear motors has been met with some skepticism, it is evident that chromatin motility is governed by energy-requiring enzymatic events. Thus a more careful look is necessary to understand the processes by which the cell overcomes the spatial constrains imposed by nuclear substructure. A fascinating problem has also arisen from the observation of physical associations of distant loci mediated by repetitive elements, suggesting that long-range repeat interactions may be a determining component in interphase nuclear organization. Moreover, there is also a possible role of repeat

elements as enhancers or insulators of gene expression in certain specific cellular events during development, cell proliferation or under certain disease conditions. It is now clear that clustering of simple repeats recruits chromatin repressors to ensure that they act exclusively on sequences in their vicinity and not in a promiscuous fashion. Very little is known about the contribution and regulation of various repeat elements found in the genome, but being in the era of deep sequencing and proteomics we are ever closer to unravel the secrets associated with the architectural framework of the nucleus. This review has focused on several of many recent developments in understanding nuclear organization and genome function in the context of intra- and interchromosomal interactions, illustrating some of the remarkable progress in the field over the past few years. These findings raise the question of how the non-random organization of the nucleus contributes to the encoding of epigenetic information that impacts genomic regulation. With increasing studies reporting examples of interacting loci, a central question is how the expression and maintenance of important components within the nucleus are dynamically distributed to specific compartments in a temporal and concentration-dependent fashion. How are multiple signals that cells are constantly receiving modulated and integrated into discrete responses, rather than turning into an organizational mayhem? How many of these interactions are driven by the proximal location to topological domains of activation/repression, and how many have measurable physiological consequences? Are genes or chromosomes re-localized transiently upon specific signals or is this a long lasting phenomenon upon decisive events such as differentiation and development? More importantly, do regulatory

networks exist, and if so how are they established, maintained, and induced to change throughout the life cycle of the cell?

These very complex questions require new creative approaches and technological development; however, advancement in genome-wide platforms, visualization technologies and computational algorithms will pave the way to the understanding of nuclear organization principles, and reveal the elusive underlying molecular strategies.

3.5 References

1) Sommer, S. and S. A. Fuqua (2001). "Estrogen receptor and breast cancer." Semin Cancer Biol **11**(5): 339-352.

2) Turner, R. T., B. L. Riggs, et al. (1994). "Skeletal effects of estrogen." Endocr Rev **15**(3): 275-300.

3) Farhat, M. Y., M. C. Lavigne, et al. (1996). "The vascular protective effects of estrogen." FASEB J **10**(5): 615-624.

4) Beato, M., P. Herrlich, et al. (1995). "Steroid hormone receptors: many actors in search of a plot." Cell **83**(6): 851-857.

5) Mueller, S. O. and K. S. Korach (2001). "Estrogen receptors and endocrine diseases: lessons from estrogen receptor knockout mice." Curr Opin Pharmacol **1**(6): 613-619.

6) Carroll, J. S., C. A. Meyer, et al. (2006). "Genome-wide analysis of estrogen receptor binding sites." Nat Genet 38(11): 1289-1297.

7) Lin, C. Y., V. B. Vega, et al. (2007). "Whole-genome cartography of estrogen receptor alpha binding sites." PLoS Genet 3(6): e87.

8) Hua, S., C. B. Kallen, et al. (2008). "Genomic analysis of estrogen cascade reveals histone variant H2A.Z associated with breast cancer progression." Mol Syst Biol 4: 188.

9) Welboren, W. J., M. A. van Driel, et al. (2009). "ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands." EMBO J 28(10): 1418-1428.

10) Morey, C. and P. Avner (2004). "Employment opportunities for non-coding RNAs." FEBS Lett 567(1): 27-34.

11) Carrington, J. C. and V. Ambros (2003). "Role of microRNAs in plant and animal development." Science 301(5631): 336-338.

12) Rinn, J. L., M. Kertesz, et al. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell 129(7): 1311-1323.

13) Zhao, J., B. K. Sun, et al. (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." Science 322(5902): 750-756.

14) Feng, J., C. Bi, et al. (2006). "The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator." Genes Dev 20(11): 1470-1484.

15) Willingham, A. T., A. P. Orth, et al. (2005). "A strategy for probing the function of noncoding RNAs finds a repressor of NFAT." Science 309(5740): 1570-1573.

16) Clemson, C. M., J. N. Hutchinson, et al. (2009). "An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles." Mol Cell 33(6): 717-726.

17) Sasaki, Y. T. F., T. Ideue, et al. (2009). "MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles." Proc Natl Acad Sci U S A 106(8): 2525-2530.

18) Sunwoo, H., M. E. Dinger, et al. (2009). "MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles." Genome Res 19(3): 347-359.

19) Banerji, J., S. Rusconi, et al. (1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." Cell 27(2 Pt 1): 299-308.

20) Maston, G. A., S. K. Evans, et al. (2006). "Transcriptional regulatory elements in the human genome." Annu Rev Genomics Hum Genet 7: 29-59.

21) Kim, T.-K., M. Hemberg, et al. (2010). "Widespread transcription at neuronal activity-regulated enhancers." Nature 465(7295): 182-187.

22) Santa, F. D., I. Barozzi, et al. (2010). "A large fraction of extragenic RNA pol II transcription sites overlap enhancers." PLoS Biol 8(5): e1000384.

23) Miele, A. and J. Dekker (2008). "Long-range chromosomal interactions and gene regulation." Mol Biosyst 4(11): 1046-1057.

24) Mattick, J. S., P. P. Amaral, et al. (2009). "RNA regulation of epigenetic processes." Bioessays 31(1): 51-59.

25) Lamond, A. I. and D. L. Spector (2003). "Nuclear speckles: a model for nuclear organelles." Nat Rev Mol Cell Biol 4(8): 605-612.

26) Misteli, T. (2005). "Concepts in nuclear architecture." Bioessays 27(5): 477-487.

27) Zaidi, S. K., S. Pande, et al. (2007). "Runx2 deficiency and defective subnuclear targeting bypass senescence to promote immortalization and tumorigenic potential." Proc Natl Acad Sci U S A 104(50): 19861-19866.

28) Cremer, T., M. Cremer, et al. (2006). "Chromosome territories--a functional nuclear landscape." Curr Opin Cell Biol 18(3): 307-316.

29) Heard, E. and W. Bickmore (2007). "The ins and outs of gene regulation and chromosome territory organisation." Curr Opin Cell Biol 19(3): 311-316.

30) Branco, M. R. and A. Pombo (2006). "Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations." PLoS Biol 4(5): e138.

31) Pickersgill, H., B. Kalverda, et al. (2006). "Characterization of the Drosophila melanogaster genome at the nuclear lamina." Nat Genet 38(9): 1005-1014.

32) Guelen, L., L. Pagie, et al. (2008). "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions." Nature 453(7197): 948-951.

33) Finlan, L. E., D. Sproul, et al. (2008). "Recruitment to the nuclear periphery can alter expression of genes in human cells." PLoS Genet 4(3): e1000039.

34) Croft, J. A., J. M. Bridger, et al. (1999). "Differences in the localization and morphology of chromosomes in the human nucleus." J Cell Biol 145(6): 1119-1131.

35) Boyle, S., S. Gilchrist, et al. (2001). "The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells." Hum Mol Genet 10(3): 211-219.

36) Khalil, A., J. L. Grant, et al. (2007). "Chromosome territories have a highly nonspherical morphology and nonrandom positioning." Chromosome Res 15(7): 899-916.

37) Jing, H., C. R. Vakoc, et al. (2008). "Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus." Mol Cell 29(2): 232-242.

38) de Laat, W. and F. Grosveld (2007). "Inter-chromosomal gene regulation in the mammalian cell nucleus." Curr Opin Genet Dev 17(5): 456-464.

39) Schneider, R. and R. Grosschedl (2007). "Dynamics and interplay of nuclear architecture, genome organization, and gene expression." Genes Dev 21(23): 3027-3043.

40) Hu, Q., Y. S. Kwon, et al. (2008). "Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules." Proc Natl Acad Sci USA 105(49): 19199-19204.

41) Fan, H. Y., X. He, et al. (2003). "Distinct strategies to make nucleosomal DNA accessible." Mol Cell 11(5): 1311-1322.

42) Cook, P. R. (1999). "The organization of replication and transcription." Science 284(5421): 1790-1795.

43) Apostolou, E. and D. Thanos (2008). "Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression." Cell 134(1): 85-96.

44) Zuckerkandl, E. and G. Cavalli (2007). "Combinatorial epigenetics, "junk DNA", and the evolution of complex organisms." Gene 390(1-2): 232-242.

45) Lunyak, V. V., G. G. Prefontaine, et al. (2007). "Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis." Science 317(5835): 248-251.

46) Kawaji, H. and Y. Hayashizaki (2008). "Exploration of small RNAs." PLoS Genet 4(1): e22.

47) Shibata, S. and J. T. Lee (2004). "Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice." Curr Biol 14(19): 1747-1754.

48) Bacher, C. P., M. Guggiari, et al. (2006). "Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation." Nat Cell Biol 8(3): 293-299.

49) Hutchinson, J. N., A. W. Ensminger, et al. (2007). "A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains." BMC Genomics 8: 39.

50) Radford, I. R. (2002). "Transcription-based model for the induction of interchromosomal exchange events by ionizing irradiation in mammalian cell lines that undergo necrosis." Int J Radiat Biol 78(12): 1081-1093.

51) Obe, G., P. Pfeiffer, et al. (2002). "Chromosomal aberrations: formation, identification and distribution." Mutat Res 504(1-2): 17-36.

52) Soutoglou, E., J. F. Dorn, et al. (2007). "Positional stability of single double-strand breaks in mammalian cells." Nat Cell Biol 9(6): 675-682.

53) Parada, L. A., P. G. McQueen, et al. (2002). "Conservation of relative chromosome positioning in normal and cancer cells." Curr Biol 12(19): 1692-1697.

54) Horiuchi, T. and T. Aigaki (2006). "Alternative trans-splicing: a novel mode of pre-mRNA processing." Biol Cell 98(2): 135-140.

55) Unneberg, P. and J.-M. Claverie (2007). "Tentative mapping of transcription-induced interchromosomal interaction using chimeric EST and mRNA data." PLoS One 2(2): e254.

56) Williams, A. and R. A. Flavell (2008). "The role of CTCF in regulating nuclear organization." J Exp Med 205(4): 747-750.

57) Ciavatta, D., S. Kalantry, et al. (2006). "A DNA insulator prevents repression of a targeted X-linked transgene but not its random or imprinted X inactivation." Proc Natl Acad Sci U S A 103(26): 9958-9963.

58) Ling, J. Q., T. Li, et al. (2006). "CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1." Science 312(5771): 269-272.

59) Szabó, P. E., L. Han, et al. (2006). "Mutagenesis in mice of nuclear hormone receptor binding sites in the Igf2/H19 imprinting control region." Cytogenet Genome Res 113(1-4): 238-246.

60) Donohoe, M. E., L.-F. Zhang, et al. (2007). "Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch." Mol Cell 25(1): 43-56.

61) Kim, T. H., Z. K. Abdullaev, et al. (2007). "Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome." Cell 128(6): 1231-1245.

62) Xu, N., M. E. Donohoe, et al. (2007). "Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein." Nat Genet 39(11): 1390-1396.

63) Spilianakis, C. G., M. D. Lalioti, et al. (2005). "Interchromosomal associations between alternatively expressed loci." Nature 435(7042): 637-645.

64) Xu, N., C.-L. Tsai, et al. (2006). "Transient homologous chromosome pairing marks the onset of X inactivation." Science 311(5764): 1149-1152.

65) Mukhopadhyay, R., W. Yu, et al. (2004). "The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide." Genome Res 14(8): 1594-1602.

66) Vetchinova, A. S., S. B. Akopov, et al. (2006). "Two-dimensional electrophoretic mobility shift assay: identification and mapping of transcription factor CTCF target sequences within an FXYD5-COX7A1 region of human chromosome 19." Anal Biochem 354(1): 85-93.

67) Cuddapah, S., R. Jothi, et al. (2009). "Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains." Genome Res 19(1): 24-32.

68) Wendt, K. S., K. Yoshida, et al. (2008). "Cohesin mediates transcriptional insulation by CCCTC-binding factor." Nature 451(7180): 796-801.

69) Lutz, M., L. J. Burke, et al. (2000). "Transcriptional repression by the insulator protein CTCF involves histone deacetylases." Nucleic Acids Res 28(8): 1707-1713.

70) Klenova, E., A. C. Scott, et al. (2004). "YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders." J Neurosci 24(26): 5966-5973.

71) Yusufzai, T. M., H. Tagami, et al. (2004). "CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species." Mol Cell 13(2): 291-298.

72) Defossez, P.-A., K. F. Kelly, et al. (2005). "The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso." J Biol Chem 280(52): 43017-43023.

73) Chernukhin, I., S. Shamsuddin, et al. (2007). "CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide." Mol Cell Biol 27(5): 1631-1648.

74) Yu, W., V. Ginjala, et al. (2004). "Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation." Nat Genet 36(10): 1105-1110.

75) Klenova, E. M., I. V. Chernukhin, et al. (2001). "Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF." Mol Cell Biol 21(6): 2221-2234.

76) MacPherson, M. J., L. G. Beatty, et al. (2009). "The CTCF insulator protein is posttranslationally modified by SUMO." Mol Cell Biol 29(3): 714-725.

77) Brown, C. R., C. J. Kennedy, et al. (2008). "Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes." Genes Dev 22(5): 627-639.

4 Materials and Methods

4.1 Antibodies

The antibodies for this study were purchased as follows: ERα antibodies (HC-20 and H-184, Santa Cruz Biotechnology); anti-H3K4me3 (07-473, Santa Cruz Biotechnology); anti-RING1 (H-110, Santa Cruz Biotechnology); anti-H3K4me1 (ab8899, Abcam); anti-H4K27ac (39133, Activemotif); anti-WDR82 (AP4812a; Abgent); anti-SC35 (ab88720; abcam); anti-Myosin I beta (Nuclear Myosin I, NMI) (M3567; Sigma) and anti-IgG (I5006; Sigma).

4.2 Cell culture

MCF-7 cells were cultured in MEM media supplemented with 10% FBS in a 7% CO_2 humidified incubator. Once the cells reached 60% confluency they were hormone-deprived for 3 days by culturing in in phenol-free media plus charcoal-depleted FBS. In addition to the starvation step, the cells were synchronized for 2 hours by treatment with 2.5nM α-amanitin, and then induced with 100 nM 17β-estradiol (E2) (Sigma) for 1 hour.

4.3 Single cell microinjections

Single-cell anti-NMI microinjection experiments were performed as described in (1). Wild type NMI construct was purchased from Open Biosystems (Clone ID 14246),

while the NMI mutants construct (S497L) was generated in-house by site directed mutagenesis.

4.4 ImmunoFISH

The cells were processed for DNA ImmunoFISH essentially as described in (2), except that labeled BAC probes were commercially obtained from Empire Genomics and listed in table 4.4.1. MCF7 cells were grown onto acid-washed polylysine coated coverslips as described in the cell culture section. Cells were treated with vehicle (EtOH) and 17β-estradiol (E2) for 1 hour, respectively, washed with 1X PBS and immediately fixed with freshly made 4% paraformaldehyde/PBS for 10 mins. Permeabilization was achieved by incubating in ice-cold cytoskeletal buffer [CSK: 10 mM Pipes, pH 6.8; 300mM sucrose; 100mM NaCl; 3mM MgCl2; 1mM EGTA; 20mM vanadyl ribonucleoside complex and 1mM 4-(2-aminoethyl)benzenesulfonyl fluoride] containing 0.5% Triton X-100 for 10 min. FISH pre-hybridization treatments included incubating the coverslips in 0.1N HCl for 5 min at room temperature, followed by digestion with 0.01N HCl/ 0.002% pepsin for 5mins at 37^{0}C, stopped by 50mM MgCl2/PBS and equilibrated in 50% formamide/2XSSC 2hrs prior to hybridization. 5µl of probe/hybridization buffer mix (Empire genomics) was used per coverslip, with a hybridization program of 76^{0}C for 3mins followed by overnight hybridization at 37^{0}C in a humidified dark chamber. The coverslips were then washed with pre-warmed WS1 (0.4xSSC/0.3% NP-40) buffer to 72^{0}C for 2mins, and then transferred to WS2 (2xSSC/0.1% NP-40) buffer at room temperature for 1min. One last wash with 1X PBS

was performed, excess liquid was aspirated and the coverslips were then mounted with prolong gold-DAPI antifade mounting reagent (Invitrogen).

Table 4.4.1: BAC probes used in this study

↑eRNA - ↑gene		
Gene	**Coordinates (including eRNA/s)**	**FISH BAC**
FOXC1	chr6:1,548,833-1,590,281	RP11-13F18
P2RY2	chr11:72,568,712-72,653,158	RP11-352J15
STARD10	chr11:72,106,004-72,300,488	RP11-1087J1
CA12	chr15:61,386,448-61,521,763	RP11-100N8
SMAD7	chr18:44,690,601-44,781,579	RP11-956P18
KCNK5	chr6:39,238,511-39,357,405	RP11-136C6
PGR	chr11:100,359,782-100,591,701	RP11-599H8
SIAH2	chr3:151,910,129-152,072,555	RP11-103G8
NRIP1	chr21:15,228,005-15,513,603	RP11-22D1
TFF1	chr21:42,651,544-42,672,721	RP11-619I15
GREB1	chr2:11,530,334-11,710,942	RP11-50E1

4.5 Image acquisition and data analysis

Images were acquired with a Zeiss Axioplan 2MOT Epifluorescent microscope (Carl Zeiss, Inc) and a Hamamatsu ORCA ER black/white CCD camera with a 63X oil immersion objective, with each channel recorded once and all channels overlapped to generate merged images. These were randomly selected for analysis, and all nuclei existing entirely within the field were used. 3D-images were produced by taking Z-stacks and performing deconvolution for each channel individually using Volocity (PerkinElmer), a software package specialized in the analysis of 3D images.

Quantification of colocalization was determined by Pearson's correlation coefficient using the "colocalization" and "measurements" modules of the software, using only those colocalization events reported in both the XZ and the YZ axis (3). The Pearson's correlation coefficients obtained with individual cells from multiple separate experiments were subjected to statistical analysis using SigmaPlot 11.0 for Windows by Systat Software, Inc. For significance of colocalization, a p-value is provided to ascertain the strength of the observation, when two conditions are compared, a two-tailed unpaired Student's t test is used and the t value is used to determine the p-value. When more than two conditions are compared by means of ANOVA (Analysis of Variance) the Holm-Sidak test is used, as it is considered to be more powerful than the Tukey and Bonferroni tests.

4.6 RNAi & locked nucleic acid (LNA) transfections

One day before transfection, MCF7 cells were seeded in OPTI-MEM medium. Six hours after siRNA transfection (40nM) with Lipofectamine 2000 (Invitrogen), cells were washed twice with 1X PBS and then maintained in hormone-deprived phenol-free supplemented stripped media for 3days, and then treated with EtOH or E2 for 1hr. LNAs were obtained from Exiqon (Woburn, MA), and were designed to trigger RNAse-H cleavage of the target sequences. LNA transfections (40nM) were performed 2 days after starvation in stripped media and the LNA treatment lasted 6 or 24hrs, after which cells were treated as described above. All siRNAs and LNAs are listed in Table 4.6.1.1 and Table 4.6.1.2.

Table 4.6.1.1: siRNAs used in this study

Name (Sense)	Sequence (Sense) 5' to 3'	Name (Antisense)	Sequence (Antisense) 5' to 3'	Overhangs
TFF1e_S_832	CAGAGUCAGAGAGUCAGAGAGAGAU	TFF1e_S_832	AUCUCUCUCUGACUCUCUGACUCUG	None
TFF1e_S_1014	GAGUUUGGACCUGUGACCUUCCUAA	TFF1e_S_1014	UUAGGAAGGUCACAGGUCCAAACUC	None
TFF1e_S_1458	AAUCUCCUGGGAGGAUGAAGCUGUU	TFF1e_S_1458	AACAGCUUCAUCCUCCCAGGAGAUU	None
TFF1e_AS_680	GAAGUUAGUGGGAGCGACCAGCUUU	TFF1e_AS_680	AAAGCUGGUCGCUCCCACUAACUUC	None
TFF1e_AS_1526	CAAACACAGGAGGUCCUCACUUAAU	TFF1e_AS_1526	AUUAAGUGAGGACCUCCUGUGUUUG	None
TFF1e_AS_1610	AGCAACGUCCUUGCUUGCAAUGUAA	TFF1e_AS_1610	UUACAUUGCAAGCAAGGACGUUGCU	None
GREB1e1_S_1115	ACCACUGUUUCUGACUGCUUUCUCA	GREB1e1_S_1115	UGAGAAAGCAGUCAGAAACAGUGGU	None
GREB1e1_S_2864	GGAUUGAGAGUGACCAGGACAUUUA	GREB1e1_S_2864	UAAAUGUCCUGGUCACUCUCAAUCC	None
GREB1e1_S_3336	CGCCCAGCCUAAUUGUAGUACUUUA	GREB1e1_S_3336	UAAAGUACUACAAUUAGGCUGGGCG	None
GREB1e1_AS_732	GAGGAUCGAGAAGUGGAAUUCUAAU	GREB1e1_AS_732	AUUAGAAUUCCACUUCUCGAUCCUC	None
GREB1e1_AS_3878	CAACUUCCUGUGUGUUCCUGUUUCA	GREB1e1_AS_3878	UGAAACAGGAACACACAGGAAGUUG	None
GREB1e1_AS_4119	GGUAUAGCUCAUUUCAGCCACUCUA	GREB1e1_AS_4119	UAGAGUGGCUGAAAUGAGCUAUACC	None
GREB1e2_S_722	UGGCCCUUGGAUGCUCCCUUCUAAU	GREB1e2_S_722	AUUAGAAGGGAGCAUCCAAGGGCCA	None
GREB1e2_S_1949	GCCAGCAUUUGAUGGUGUCACUGUU	GREB1e2_S_1949	AACAGUGACACCAUCAAAUGCUGGC	None
GREB1e2_S_2860	CCACAAAGAGUGUUCAGCUUCGGGA	GREB1e2_S_2860	UCCCGAAGCUGAACACUCUUUGUGG	None
GREB1e2_AS_460	CAGGGCAAAGGACAUGGCCAGAUAA	GREB1e2_AS_460	UUAUCUGGCCAUGUCCUUUGCCCUG	None
GREB1e2_AS_1058	CAACUGAAAUCAGUCAGCAGUUUCG	GREB1e2_AS_1058	CGAAACUGCUGACUGAUUUCAGUUG	None
GREB1e2_AS_1565	GAACAACUGGGUAUCUACAGGCAAA	GREB1e2_AS_1565	UUUGCCUGUAGAUACCCAGUUGUUC	None
GREB1e3_S_782	CAGAGAGAUUCAAGCUUGACGGAAU	GREB1e3_S_782	AUUCCGUCAAGCUUGAAUCUCUCUG	None
GREB1e3_S_1965	CAGAGAGAUUCAAGCUUGACGGAAU	GREB1e3_S_1965	UAUGAUUCAUCAUUGUCUGCUGCGG	None
GREB1e3_S_2391	CAGUAGAUACCAGGCACCCAUAAAU	GREB1e3_S_2391	AUUUAUGGGUGCCUGGUAUCUACUG	None
GREB1e3_AS_35	CAGACCCAAACAUGCUGCUGCAAAU	GREB1e3_AS_35	AUUUGCAGCAGCAUGUUUGGGUCUG	None
GREB1e3_AS_1411	CAGCAGGCAAGUCACUUAACCUCUA	GREB1e3_AS_1411	UAGAGGUUAAGUGACUUGCCUGCUG	None
GREB1e3_AS_1990	AGGUUCUGGUGAAGGUGCACUUCAU	GREB1e3_AS_1990	AUGAAGUGCACCUUCACCAGAACCU	None
PGRe_S	GCAAAUUCUUUCAUGACAA	PGRe_S	UUGUCAUGAAAGAAUUUGC	UU
PGRe_AS	GCAAAGAUGGAUAGAGAUA	PGRe_AS	UAUCUCUAUCCAUCUUUGC	UU
SIAH2e1_S	GCACAUACCUCAUUAGAGA	SIAH2e1_S	UCUCUAAUGAGGUAUGUGC	UU
SIAH2e1_AS	CCAGAGAGCUGAACUGAUA	SIAH2e1_AS	UAUCAGUUCAGCUCUCUGG	UU
SIAH2e2_S	GGUAUUAAUAGCUCUGAAA	SIAH2e2_S	UUUCAGAGCUAUUAAUACC	UU
SIAH2e2_AS	CCAGAAAGUUUACCAGUUA	SIAH2e2_AS	UAACUGGUAAACUUUCUGG	UU
NRIP1e1_S	GGGAGAGGGUCUACAAUUA	NRIP1e1_S	UAAUUGUCGACCCUCUCCC	UU
NRIP1e1_AS	GCAAGAAGGAAGAGGGUUU	NRIP1e1_AS	AAACCCUCUUCCUUCUUGC	UU
NRIP1e3_S	GGCCAGAUCUCCUGUGAUA	NRIP1e3_S	UAUCACAGGAGAUCUGGCC	UU
NRIP1e3_AS	GCAUUAGGGUUGAAGUAUA	NRIP1e3_AS	UAUACUUCAACCCUAAUGC	UU
FOXC1e_S	GCUCCAUUCUGCUGCUCAA	FOXC1e_S	UUGAGCAGCAGAAUGGAGC	UU
FOXC1e_aS	CUAACGUGACAGUGACAUA	FOXC1e_aS	UAUGUCACUGUCACGUUAG	UU
P2RY2e_S1	GCAAAAGGUAGGAGGGUUU	P2RY2e_S1	AAACCCUCCUACCUUUUGC	UU
P2RY2e_S2	GGAGAUGAAUUGAUAGAGA	P2RY2e_S2	UCUCUAUCAAUUCAUCUCC	UU
P2RY2e_AS1	GGAUAAAGCUGGAGUGAGU	P2RY2e_AS1	ACUCACUCCAGCUUUAUCC	UU
P2RY2e_AS2	GCAUAAAGCCUCAGUGACA	P2RY2e_AS2	UGUCACUGAGGCUUUAUGC	UU
CA12e_S	CAGAAGAGCUAUUUGGUAU	CA12e_S	AUACCAAAUAGCUCUUCUG	UU
CA12e_AS	GAGUGGACUUCACAAGAAA	CA12e_AS	UUUCUUGUGAAGUCCACUC	UU
SMAD7e1_S	AGAGAAGAAUGAAGGUGAA	SMAD7e1_S	UUCACCUUCAUUCUUCUCU	UU
SMAD7e1_AS	ACUUAAAGGUUCCAGUGUU	SMAD7e1_AS	AACACUGGAACCUUUAAGU	UU
SMAD7e2_S	CCACAGGUGAGCAGAAAUU	SMAD7e2_S	AAUUUCUGCUCACCUGUGG	UU
SMAD7e2_AS	GCUGAUGAAAGGAAGGAAA	SMAD7e2_AS	UUUCCUUCCUUUCAUCAGC	UU
SMAD7e3_S	CCUAAUUCCCAGAAGCAGA	SMAD7e3_S	UCUGCUUCUGGGAAUUAGG	UU
SMAD7e3_AS	CGGCAGGAAUAGAGGCUCA	SMAD7e3_AS	UGAGCCUCUAUUCCUGCCG	UU
KCNK5e1_S	CGAAAUGGCCUAAAGAUGA	KCNK5e1_S	UCAUCUUUAGGCCAUUUCG	UU
KCNK5e1_AS	GCAAAGAGCUGGACUUACA	KCNK5e1_AS	UGUAAGUCCAGCUCUUUGC	UU
KCNK5e2_S	ACACAAAGGUGGAAGGAAA	KCNK5e2_S	UUUCCUUCCACCUUUGUGU	UU
KCNK5e2_AS	CCACACAGCUGUCCAUAAA	KCNK5e2_AS	UUUAUGGACAGCUGUGUGG	UU
KCNK5e3_S	GGAAGAACCUGCAGAGAUG	KCNK5e3_S	CAUCUCUGCAGGUUCUUCC	UU
KCNK5e3_AS	GGGACAGGUUGGAAGAGUA	KCNK5e3_AS	UACUCUUCCAACCUGUCCC	UU

Table 4.6.1.2: LNAs used in this study

LNAs use in this study 5'-3'	
LNA_Ctrl	C*A*C*G*T*C*T*A*T*A*C*A*C*C*A*C
TFF1E_AS_01	G*A*A*T*T*A*A*C*G*C*C*T*G*A*G*G
TFF1E_AS_02	G*A*A*C*T*G*A*C*A*A*A*G*G*T*G*G
TFF1E_S_01	A*T*C*T*C*C*C*C*A*C*T*C*A*A*G*G
TFF1E_S_02	C*A*T*T*T*T*T*C*T*G*C*T*G*A*C*C
CA12E_AS_01	A*C*A*A*G*A*C*A*G*A*G*G*C*A*G*A
CA12E_AS_02	T*C*A*G*T*T*G*G*A*G*G*A*C*A*G*T
FOXC1E_AS_01	G*A*A*G*G*A*G*C*A*G*G*T*G*A*A*A
FOXC1E_AS_02	G*G*T*A*T*T*T*C*C*G*C*T*T*C*A*C
ARHGAP12E_AS_01	T*A*C*A*C*C*T*A*C*T*A*C*G*G*A*C
ARHGAP12E_AS_02	A*T*T*T*T*C*T*G*C*T*G*G*G*T*G*C
NRIP1E3_AS_01	A*G*G*A*T*A*C*C*A*G*G*A*C*A*C*A
NRIP1E3_AS_02	T*G*A*T*A*A*A*A*G*C*A*G*G*G*T*C
(*) = phosphorothioate backbone	

4.7 RT-QPCR

RNA was isolated using Trizol (Invitrogen), and total RNA was reverse-transcribed using SuperScript® III Reverse Transcriptase (Invitrogen) as per manufacturer's instructions. Quantitative PCRs were performed in MX3000P (Stratagene) using Q-PCR master mix (Agilent Inc.). For normalization, ΔCt values were calculated using the formula: ΔCt = (Ct Target−Ct input) where input corresponds to the level of ACTB transcript. Fold differences in normalized gene expression were calculated by dividing the level of expression of the treated sample with the untreated sample or between siRNA/LNA and siControl/control-LNA transfected cells. The results are presented as the average of 3 replicates and when provided, p-values were

obtained using a two-tailed Student's t-test. A list of primers used for QPCR is provided in Table 4.7.1 [1, 2, 3, 4].

Table 4.7.1.1: eRNA associated oligos

eRNA Associated	
Name	**Sequence 5' to 3'**
TFF1e_AS_F	CTCAACATAAGATCTCCCAGCG
TFF1e_AS_R	CTCAAACTTCCCTCTCCTGTG
TFF1e_S_F	TCAGTTCCCAGCATTCTCATC
TFF1e_S_R	TTGAGCCTTGGAGACAGAAAG
GREB1e1_AS_F	CGGAACACAGGTGAGGAGAT
GREB1e1_AS_R	AGCAGAACGCCTCATGAACT
GREB1e1_S_F	TCCAAAGCATCCCATTCCTG
GREB1e1_S_R	TGAGCAAAACAAGACAAACCG
GREB1e2_AS_F	ACACAGGAGACAATGGCAAG
GREB1e2_AS_R	CAGAGCGTCCTTTCATCCAG
GREB1e2_S_F	GACCAGCGGAACAAAAGATC
GREB1e2_S_R	GTTTGACACTGCTAACCATGC
GREB1e3_AS_F	GAAGGGCAGAGCTGATAACG
GREB1e3_AS_R	GACCCAGTTGCCACACTTTT
GREB1e3_S_F	ATGACCCAGTTGCCACACTT
GREB1e3_S_R	GAAGGGCAGAGCTGATAACG
ARHGAP12e_AS_F	AGGACTTACAGCCGATTTCAC
ARHGAP12e_AS_R	TTACAGACACCAAATCCAGCA
ARHGAP12e_S_F	CCCGAGGTCATCTGCTATTAAG
ARHGAP12e_S_R	ATTCCCAAACCTTCCCTGTC
PGRe_AS_F	ACGACTCAGTCTCAGTTTTAGC
PGRe_AS_R	GTATACAGGCCCAGAGTCAC
PGRe_S_F	TTATGTTGCTCTTGATAGACTCCC
PGRe_S_R	GCTAGGTGCTGTCTGAGATTC
SIAH2e1_AS_F	ACAGGGAAGAATGCAGCCTA
SIAH2e1_AS_R	TGAGAAAGCTGAAGGCAGAGA
SIAH2e1_S_F	TTCAAGCAAAGATTATAGCCATGTG
SIAH2e1_S_R	ATCCAGTGCAGAGTAACATCAG
SIAH2e2_AS_F	GGATGGGCCTTATTCCAAAT
SIAH2e2_AS_R	ATTTGCCCAAATCAAGGAAG
SIAH2e2_S_F	AGATGCCTCTGCATACTGGTT
SIAH2e2_S_R	CAGACCATATTGGGCCACAG
NRIP1e1_AS_F	ACTCCAAGACAGTGACACAAG
NRIP1e1_AS_R	TCTCCAAGGCACTTTTCAGTC
NRIP1e1_S_F	CCACAGCAGAAAACCACTGA
NRIP1e1_S_R	TTCCCTCTGCACTGACTCCT
NRIP1e3_AS_F	TCTGGCTTTATTGGCTGTAGG

Table 4.7.1.1: eRNA associated oligos, continued

eRNA Associated	
Name	**Sequence 5' to 3'**
NRIP1e3_AS_R	GGAGTGGCTAAGGAAAGGTG
NRIP1e3_S_F	CGTCTTTTCCCACTGACACA
NRIP1e3_S_R	CCCCTCCCCAGAAGAAAATA
CCNG2e_AS_F	ACCCCTTACCATCTCTACCTC
CCNG2e_AS_R	CAAGGCCCTCTGATAACTCAC
CCNG2e_S_F	GGGCAGCTAATCTATTCCAGG
CCNG2e_S_R	GTTCTCTCCCATTAGGCTCTG
ZNF217e1_AS_F	TGCCTCCAAGTTTCAAGTCG
ZNF217e1_AS_R	CATTCTCATCTGTATCCCATCTCTG
ZNF217e1_S_F	GGAAGCCATTTTACACTAATACCAG
ZNF217e1_S_R	AGCTCCTGGTCTGTTTTCAC
ZNF217e1_BAS_F	GACAGGGAGAGGGTTACAAATC
ZNF217e1_BAS_R	ATTAGGACGTGAGTTGCTGAC
ZNF217e1_BS_F	ATTAGGACGTGAGTTGCTGAC
ZNF217e1_BS_R	GACAGGGAGAGGGTTACAAATC
ZNF217e2_AS_F	TATGCTTGGCCACAAAATGA
ZNF217e2_AS_R	CCTGACCTTGTGATCCACCT
ZNF217e2_S_F	AGAGGTCATCTGTGGGATGG
ZNF217e2_S_R	TATGTGGGTGTCAGCATCGT
ZNF217e3_AS_F	ATGTTGAGCACCTCCAGAAG
ZNF217e3_AS_R	TCCTGAATTATCTTGTGAGCCC
ZNF217e3_S_F	TTACTGACTAACCTCGGCTTTG
ZNF217e3_S_R	TGGGAGTGCAATGTTCAAGAG
ZNF217e3_BAS_F	TCCTAAAGAAAGGAAAAGCCATC
ZNF217e3_BAS_R	TTGATGTGGCTCAAAGAAGC
ZNF217e3_BS_F	TTGATGTGGCTCAAAGAAGC
ZNF217e3_BS_R	TCCTAAAGAAAGGAAAAGCCATC
ZNF217e4_AS_F	TAAGCAATCCTCCCACCTTG
ZNF217e4_AS_R	CCCTGATGGAGCAGACATTT
ZNF217e4_S_F	AGCAGCTGGGATGTTTCAAT
ZNF217e4_S_R	CCCCAGCACATCCTTTCTTA
SIX4e1_AS_F	GAGAAACGATGGCCTTAATGC
SIX4e1_AS_R	GTATGGGCACCAGATCAAAATG
SIX4e1_S_F	AAGAGCCCCTCCACGTTACT
SIX4e1_S_R	ACAGCCCTCAGCTCTGCTTA
SIX4e2_AS_F	GCCATCCCTCCATTCCAAAAG
SIX4e2_AS_R	ATCCAGAGCGACCTTTCAAG

Table 4.7.1.1: eRNA associated oligos, continued

eRNA Associated	
Name	**Sequence 5' to 3'**
SIX4e2_S_F	CCAGCACCCTCTACCCAGTA
SIX4e2_S_R	AGGATCCAGATTTTCGCAGA
RHOBTB3e_AS_F	GGGTTTCGGAGGTGGATACT
RHOBTB3e_AS_R	CCGGATGAGGACATTTTTCT
RHOBTB3e_S_F	TTTTCTGTACTTCGCCTACGG
RHOBTB3e_S_R	GACGGTCTGGATCTGCATG
FOXC1e_AS_F	CATGAAAGGTGAAGCGGAAATAC
FOXC1e_AS_R	TGAAGGAGCAGGTGAAACG
FOXC1e_S_F	CTGAGGAACACAAGACTAGCC
FOXC1e_S_R	ACTGGACTCATTTTGGGACATC
P2RY2e_AS1_F	TGAGGAAGTGGTTTGAGTGG
P2RY2e_AS1_R	CACTTCGACAATTACCAACACAG
P2RY2e_AS2_F	ACACAGTATTTCCCATCTATCAGC
P2RY2e_AS2_R	AAGAGTCTGCATTTCCAGGG
P2RY2e_S1_F	ATTGTGCATGGCTCTTACCC
P2RY2e_S1_R	CTTGGTGCATGTGAGCTTGT
P2RY2e_S2_F	AGCTTCTGGTTCCAAGGTCA
P2RY2e_S2_R	CATGTGCTGTTGTTGCTGTG
CA12e_AS_F	ATTCTCACTGCATACCTGACAC
CA12e_AS_R	AGCTGTTATTCCCCACTCAAC
CA12e_S_F	TGAAAGGGAAGACGCAGATG
CA12e_S_R	TTGTATCCTTTGACTGGGCAG
SMAD7e1_AS_F	AGTAGAACCCGGGAGAAGGA
SMAD7e1_AS_R	TGCGAGCAAATTGACTGAAC
SMAD7e1_S_F	AAAGAAGGCAGGGGAACAAT
SMAD7e1_S_R	CACTTGGGCAATCCAGAAAT
SMAD7e2_AS_F	GCAGTTCATGCACCATTGTC
SMAD7e2_AS_R	GGATGACCACAGGCATTCAT
SMAD7e2_S_F	TCACCTGTGGAAAGAGACAAC
SMAD7e2_S_R	AGAACCTTTTGCTCCCTAGTG
SMAD7e3_AS_F	CTCAAGCTCGGGTCAGACAT
SMAD7e3_AS_R	TCAGTGAAGAAGGGGACACC
SMAD7e3_S_F	TTAAACGAGCCTGGAGTTGG
SMAD7e3_S_R	AAATTCCTCAGAGCCCAGTG
KCNK5e1_AS_F	TGCTTTGAACCAGAGTGCAG
KCNK5e1_AS_R	CTCTCCCTGGTCTTTTGCAG
KCNK5e1_S_F	GGCTCAGAGAGGCCAAAA

Table 4.7.1.1: eRNA associated oligos, continued

eRNA Associated	
Name	**Sequence 5' to 3'**
KCNK5e1_S_R	TGGACCCTATCATCTCCTTTAACT
KCNK5e2_AS_F	CCTTTTGTGAGTCCATTTCCATTG
KCNK5e2_AS_R	TCTTACACATGGCACTGTTAGG
KCNK5e2_S_F	GGAAAGGAATTGCTGGATCA
KCNK5e2_S_R	GTGCAACCACTTGGGAAACT
KCNK5e3_AS_F	CGTAAAGTGGGATCTAGGAGTG
KCNK5e3_AS_R	TGGGTCAAAATGTCTTAATGTGC
KCNK5e3_S_F	CAGAGATGAGGAAAGGTTTGC
KCNK5e3_S_R	ATCTGCTTCACGGTCTCATG
PLEKHF2e1_AS_F	CAGAATAACGTCCTGTCTACCC
PLEKHF2e1_AS_R	CATTACCTGTCGCTATCCTCAC
PLEKHF2e1_S_F	TGGCTGAGGTTTTCTATGACC
PLEKHF2e1_S_R	AGGAATTGTGGGCTGTCTTC
PLEKHF2e2_AS_F	GGGACTTTATTGGTTACTTGGTTG
PLEKHF2e2_AS_R	CTCTGGAAGTTGGGAGACAAG
PLEKHF2e2_S_F	AGCCCTCTAGACCAAACATTTC
PLEKHF2e2_S_R	GTGAACAATTTAACTTGAAAGCAGG
PLEKHF2e3_AS_F	AGCTTATGATACAGTGAGGTGC
PLEKHF2e3_AS_R	GAAGAGTCAATGCAGGAAAGC
PLEKHF2e3_S_F	AGGTCAAATCCACAGGTTCTG
PLEKHF2e3_S_R	AGACTAAGAGTGCAAAGGCC
TM4SF1e1_BAS_F	TCATACTCCACCTAGACCTCATG
TM4SF1e1_BAS_R	TAGGGCAAAATCAACTGTAACTTG
TM4SF1e1_BS_F	TAGGGCAAAATCAACTGTAACTTG
TM4SF1e1_BS_R	TCATACTCCACCTAGACCTCATG
TM4SF1e2_BAS_F	GCCTTTTCCAACACCTTTAGC
TM4SF1e2_BAS_R	CACCAATGCCTAGTTGACTTG
TM4SF1e2_BS_F	CACCAATGCCTAGTTGACTTG
TM4SF1e2_BS_R	GCCTTTTCCAACACCTTTAGC
TM4SF1e3_AS_F	GCCAGTTTCAGCATTTGAGAG
TM4SF1e3_AS_R	TTTTCAAGGGTGTTTGTGTAGC
TM4SF1e3_S_F	AGGGCTTTGAAGACCACAGA
TM4SF1e3_S_R	CAGTGCAAAAGAGTGATGCAA
EFEMP1e1_AS_F	CTAGCCTGGCACTTGACACA
EFEMP1e1_AS_R	GGCACAAAAGGACTGGAAAA
EFEMP1e1_S_F	TGGCAAGAAACAAGGGTAGAG
EFEMP1e1_S_R	GCAGCCATAAACAAGGTGATG
EFEMP1e2_AS_F	TGCTATCAGTGGCTATTCATGG
EFEMP1e2_AS_R	ACTTTCCTGGCTTTTCATTGTG
EFEMP1e2_S_F	GCACACTTCAATTCTGGGTG
EFEMP1e2_S_R	GTGGGATAGATCAGTGTGAAGTG

Table 4.7.1.2: Promoter associated oligos

Promoter Associated	
Name	**Sequence 5' to 3'**
TFF1p_AS_F	GCATAGGAGGGGAGGAACTC
TFF1p_AS_R	GCAGTCCCTGATAGCAAAGC
TFF1p_S_F	CAAAAGCAGGTGGAGAGTAAC
TFF1p_S_R	TGAAACATGACACTTGGGAGG
GREB1p_AS_F	TTCTGCCAAGGATCCAATTC
GREB1p_AS_R	GCTTCTGGCCTGTTCTATGC
GREB1p_S_F	TGCTTTCAAGCTGTGTGACC
GREB1p_S_R	ACAGCCGTGTAAGGTGGTTC
ARHGAP12p_AS_F	ACCAGTCCGTCAGCATTTAG
ARHGAP12p_AS_R	CCATCAAAGACCTGCCAAAG
ARHGAP12p_S_F	GTTTTGAGAGCCCAAAGCAG
ARHGAP12p_S_R	TTTCCCCCACAGAGAAACAC
PGRp_AS_F	ACCAACATGTCCTGCTCCTC
PGRp_AS_R	TGGCACACAACATCCAAACT
PGRp_S_F	GTACGGAGCCAGCAGAAGTC
PGRp_S_R	GTACGGAGCCAGCAGAAGTC
SIAH2p_AS_F	CTAACGCCACACCCCTACAC
SIAH2p_AS_R	AAGCACTCGTGGGAGAAGAA
SIAH2p_S_F	GACCTCAGGGACTCCGGTAT
SIAH2p_S_R	GGAGGGTGGACGAAGTGTAA
NRIP1p_AS_F	AACACCTCCCCAGATGTCAG
NRIP1p_AS_R	TCAGCAACAACACTCGCTTC
NRIP1p_S_F	CCCCAACATTAAAGCAGGAC
NRIP1p_S_R	TCTACCTCCTCCCCTTCACC
CCNG2p_AS_F	GAGGTTGGTACCATGCGAGT
CCNG2p_AS_R	CCCCACGAGCATCTATCTGT
CCNG2p_S_F	GGCGCTTCTCCTAACAGCTA
CCNG2p_S_R	GGTGGGTTATTGCAAAATGG
ZNF217p_AS_F	TTGTGTTTTTACCGGCCTTC
ZNF217p_AS_R	AGCTCCGGATCGTTTCTGTA
ZNF217p_S_F	TTGTCAGTAGTTCCAAGGCAGA
ZNF217p_S_R	CTTTCCTCATCACGCCAAAT
SIX4p_AS_F	TCCCCATTCTCCCTCTTTTT
SIX4p_AS_R	AATGCTTGCCCCTTAGAACA
SIX4p_S_F	AGATACAGCCGAAACCAGGA
SIX4p_S_R	TCACTCCCTCGCACTCTTTT

Table 4.7.1.2: Promoter associated oligos, continued

Promoter Associated	
Name	**Sequence 5' to 3'**
RHOBTB3p_AS_F	TGCTTAGAGGAGGAGGAGCA
RHOBTB3p_AS_R	AAGTGGACGCAGCTGACC
RHOBTB3p_S_F	ATAGCTCCGCAGACCAACTG
RHOBTB3p_S_R	GTGGATCCCAGTGTTCGTTT
FOXC1p_AS_F	CGAGAGACCGAGAAAAGGTG
FOXC1p_AS_R	AGGAGACTCTGCCCTGAGC
FOXC1p_S_F	ATTTTGAGGGCTCGGTCAC
FOXC1p_S_R	CACCCCTTGCCTTCATTTC
P2RY2p_AS_F	GTCTGTGTGTGCTCCGATGT
P2RY2p_AS_R	AGTGACCCTGGGGTTAGGAC
P2RY2p_S_F	GAGACCTTGACCCAGTTCCA
P2RY2p_S_R	GGCAAGTTCCTTGGCTACAG
CA12p_AS_F	TAGCCTTGACTTCAGCAGCA
CA12p_AS_R	ATCCCTAGACGCACCCTTCT
CA12p_S_F	GAGCCTCCTCTCTTGGAGGT
CA12p_S_R	CGCTTCTGTTTCTCCCTGAC
SMAD7p_AS_F	CTGCCTAGGGCATTCATTTG
SMAD7p_AS_R	CGCCAGCTTGGGTATATGTT
SMAD7p_S_F	GGGGAAAGACCAGAGACTCC
SMAD7p_S_R	CGCTGGTCTTCCTCTCCTTT
KCNK5p_AS_F	CCGTGTCTCTCTTCCACACA
KCNK5p_AS_R	GGGTGGCTGGATTAGTTCCT
KCNK5p_S_F	TTGGCCAAGCAACTCTGTC
KCNK5p_S_R	CCCTCCAGCCTCTGAAAAC
PLEKHF2p_AS_F	CGACCTCACACATGGACAAG
PLEKHF2p_AS_R	GCCCTTTCTTGCCAGTTCC
PLEKHF2p_S_F	GCAATCCTTTCCGTCACACT
PLEKHF2p_S_R	GGTGGAGCACGTAGGGTATG
TM4SF1p_AS_F	AGACGTCTCCCATGGATTTG
TM4SF1p_AS_R	GGGACGCTGAACACTGAAAT
TM4SF1p_S_F	AAGGGGGAGAAAACCTAGCA
TM4SF1p_S_R	CCCATTGGGAAAGTAAAGCA
EFEMP1p_AS_F	TGTGAGGTGGGGTTTGTTTT
EFEMP1p_AS_R	TCCACCAACAGCATACAAGC
EFEMP1p_S_F	CTTTGCCCATAAACGTGGTT
EFEMP1p_S_R	TCAGCGTCCCAGCTTCTAAC

Table 4.7.1.3: Internal oligos

Internal	
Name	**Sequence 5' to 3'**
FOXC1_Sint_F	AGTCAGCTTGCTTTGAGGCTA
FOXC1_Sint_R	AGGCATCACCGTGGTAAGAC
P2RY2_Sint_F	GTGTTGGGGTAGAGGATGGA
P2RY2_Sint_R	CACACTGCACCCCATTGTAG
CA12_ASint_F	GCAAAGCCTGGAGAAGTGAG
CA12_ASint_R	AACAGTGCACACCAGATCCA
SMAD7_ASint_F	TGTTCTGGGCCTCAGTTTTC
SMAD7_ASint_R	CACCCCTCAAATCCAAGAGA
KCNK5_ASint_F	TAGCTCTCCTGTTGGCCTGT
KCNK5_ASint_R	TGCAGCATCAAAGTGGAAAG
PGR_ASint_F	GCAGTTCCCCAAAGTTTTCTT
PGR_ASint_R	CTGCTGGGAAAACTGGCTAC
SIAH2_ASint_F	TGGCCGTTCTTATTCAAACC
SIAH2_ASint_R	TACACAGTGGCTCCAAATGC
NRIP1_ASint_F	GCTTGGGAAGGGTTAAGAGG
NRIP1_ASint_R	CGAACTAATCTGGGCTGCTC
TFF1_ASint_F	CCTCACTGGACAGTTGCTGA
TFF1_ASint_R	GGGCCACTGTACACGTCTCT
GREB1_Sint_F	GCTGGTGGGAAGAGACAGAC
GREB1_Sint_R	CTACGAGATGTGGCCTGGAT
PLEKHF2_Sint_F	AAGCAGATAGCCCAGAAGCA
PLEKHF2_Sint_R	TCAGCAAGCAGTCAACGAGT
TM4SF1_ASint_F	CACTGAACCCAAGCACATTG
TM4SF1_ASint_R	AAGCCACATATGCCTCCAAG
EFEMP1_ASint_F	ACCAAGCCAAACTGCTGAAT
EFEMP1_ASint_R	GGAAGATGTCTGATGGCACA
CCNG2_Sint_F	TTGCTATATGGGTGCCAAGA
CCNG2_Sint_R	GCCCCATTCTTATGGCTGTA
ZNF217_ASint_F	ACCCTACCCCTGCTTACCTG
ZNF217_ASint_R	CATCCGGAGGAGGAGTAACA
SIX4_ASint_F	TGAGGATAAACAGCCCACCT
SIX4_ASint_R	GCATGAGATGATGGCAATTAGA
ARHGAP12_ASint_F	CTGGATGTTCCCTGACCTCT
ARHGAP12_ASint_R	ATCCCATGCATTCCAAGTTC
RHOBTB3_Sint_F	AAAAGGAAAGTCCCAGCCTTA
RHOBTB3_Sint_R	GCTGATTGGGCTGTTTTCAT

Table 4.7.1.4: mRNA associated oligos

mRNA	
Name	**Sequence 5' to 3'**
FOXC1_F	AGTAGCTGTCAAATGGCCTTC
FOXC1_R	TTAGTTCGGCTTTGAGGGTG
P2RY2_F	GGGGACCTGTTTTTCCTGTT
P2RY2_R	GACTTGGATCTGGACCTGGA
CA12_F	TCTGTCTGCCAACAAGCAGT
CA12_R	GCACTGTAGCGAGACTGGAG
SMAD7_F	TCCTGCTGTGCAAAGTGTTC
SMAD7_R	AAATCCATCGGGTATCTGGA
KCNK5_F	TGCCAAGAGACTAGGGCAGT
KCNK5_R	GAATACGAAGGGTGGGATCA
PGR_F	CATCACAGGGAACCAGACCT
PGR_R	CACCCCGAAGAGACCATAGA
SIAH2_F	TCAGGAACCTGGCTATGGAG
SIAH2_R	GGCAGGAGTAGGGACGGTAT
NRIP1_F	GCCAGAAGATGCACACTTGA
NRIP1_R	CAAGCTCTGAGCCTCTGCTT
TFF1_F	CACCATGGAGAACAAGGTGA
TFF1_R	TGACACCAGGAAAACCACAA
GREB1_F	GGCAGGACCAGCTTCTGA
GREB1_R	CTGTTCCCACCACCTTGG
PLEKHF2_F	ACAGCACACACCCTGGAAG
PLEKHF2_R	GCGATCCACCATCTTTCACT
TM4SF1_F	TTTCTGGCATCGTAGGAGGT
TM4SF1_R	CTGCAATTCCAATGAGAGCA
EFEMP1_F	CAGGACACCGAAGAAACCAT
EFEMP1_R	GTTTTCGGAAGGCAGAGGTA
CCNG2_F	ACTTCTCGGGTTGTTGAACG
CCNG2_R	GGCATTTCTCAATCCTGGAC
ZNF217_F	GAAGACGGATCTGAGGATGG
ZNF217_R	GGCTGCAGCATATTCACAAA
SIX4_F	CAGGTCAGCAACTGGTTCAA
SIX4_R	CTCATGTCCCTTGCTGGATT
ARHGAP12_F	TTAGCAGCAGCTCCACTGAA
ARHGAP12_R	AACGCCCTGAGCTGTCTTTA
RHOBTB3_F	GTCCGGTGTTCACCGAGTAT
RHOBTB3_R	GCCCCCAATTAGATTTCGAG

4.8 ChIP-Seq

ChIP was performed as previously described (4). Briefly, approximately 10^7 treated cells were crosslinked with 1% formaldehyde at room temperature for 10 min. After sonication, the soluble chromatin was incubated with 1–5 μg of antibody at 40^0C overnight. Immunoprecipitated complexes were collected using Dynabeads A/G (Invitrogen). Subsequently, immunocomplexes were washed, DNA extracted and purified by QIAquick Spin columns (Qiagen). For ChIP-seq, 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 was aligned to the hg18 assembly (National Center for Biotechnology Information, build 36.1) using BFast allowing up to two mismatches. Only uniquely mapped tags were selected for further analysis. The data was visualized by preparing custom tracks on the University of California, Santa Cruz, genome browser using HOMER (5). The total number of mappable reads was normalized to 10^7 for each experiment presented in this study.

4.8.1 Identification of ChIP-seq peaks

The identification of ChIP-seq peaks (genomic regions enriched in the protein of interest) was performed using HOMER (http://biowhat.ucsd.edu/homer). Given that the peak distribution of transcription factors and histone marks are markedly different, parameters were optimized for the narrow tag distribution characteristic of transcription factors by searching for high read density regions with a 200bp sliding window.

Regions of maximal density exceeding a given threshold were called as peaks, and it was required for adjacent peaks to be at least 500 bp away to avoid redundant detection. The common artifacts derived from clonal amplification were circumvented by considering only one tag from each unique genomic position as determined from the mapping data. The threshold for the number of tags that determined a valid peak was selected at a false discovery rate of 0.001 determined by peak finding using randomized tag positions in a genome with an effective size of 2×10^9 bp. It was also required for peaks to have at least fourfold more tags (normalized to total count) than input control samples. In addition, it was required to obtain fourfold more tags relative to the local background region (10 kb) to avoid identifying regions with genomic duplications or non-localized binding.

In the case of histone marks, the parameters were modified this time to search for enrichment in wide genomic segments as unlike transcription factors, they can occupy large segments in the magnitude of several kb. Seed regions were initially found using a peak size of 500 bp at a false discovery rate of 0.001 to identify enriched loci. Enriched regions separated by 1kb were merged and considered as blocks of variable lengths. All called peaks meeting the criteria established for transcription factors and histone marks were then associated with genes by cross-referencing the RefSeq TSS database as available in the UCSC genome browser. Peaks from individual experiments were considered equivalent if their peak centers were located within 200 bp of each other.

4.9 Gro-Seq

Gro-Seq experiments were performed as previously reported (6). Briefly, MCF7 cells were washed 3 times with cold 1X PBS buffer and then swelled in swelling buffer (10mM Tris-Cl pH7.5, 2mM MgCl2, 3mM CaCl2) for 5min on ice and harvested. Cells were first re-supended and lysed in lysis buffer (swelling buffer with 0.5% IGEPAL and 10% glycerol). Nuclei were washed one more time with 10mL lysis buffer and finally re-suspended in 100uL of freezing buffer (50mM Tris-Cl pH8.3, 40% glycerol, 5mM MgCl2, 0.1mM EDTA).

For the run-on assay, resuspended nuclei were mixed with an equal volume of reaction buffer (10mM Tris-Cl pH 8.0, 5mM MgCl2, 1mM DTT, 300mM KCL, 20 units of SUPERase In, 1% sarkosyl, 500uM ATP, GTP, and Br-UTP, 2uM CTP) and incubated for 5 min at 300C. The nuclear-run-on RNA (NRO-RNA) was then extracted with TRIzol LS reagent (Invitogen) following manufacturer's instructions. NRO-RNA was then subjected to base hydrolysis on ice for 40min and followed by treatment with DNase I and antarctic phosphatase. To purify the Br-UTP labeled nascent RNA, the NRO-RNA was immunoprecipitated with an anti-BrdU argarose beads (Santa Cruz Biotech) in binding buffer (0.5XSSPE, 1mM EDTA, 0.05% tween) for 1hour at 40°C while rotating. To repair the end, the immunoprecipitated BrU-RNA was re-suspended in 50uL reaction (45uL DEPC water, 5.2uL T4 PNK buffer, 1uL SUPERase In and 1uL T4 PNK [NEB]) and incubated at 37°C for 1hr. The RNA was extracted and precipitated using acidic phenol-chloroform.

The cDNA synthesis was performed as in Ingolia et al., (2009) with few modifications. The RNA fragments were subjected to poly-A tailing reaction by poly-A polymerase (NEB) for 30 min at 37°C. Subsequently, reverse transcription was performed using oNTI223 primer (for sequence see Table 4.9.1). Second, tailed RNA (8.0uL) was subjected to reverse transcription using superscript III (Invitrogen). The cDNA products were separated on a 10% polyacrylamide TBE-urea gel. The extended first-strand product (100-500bp) was excised and recovered by gel extraction. After that, the first-strand cDNA was circularized by CircLigase (Epicentre) and relinearized by Ape1 (NEB). Relinearized single strand cDNA (sscDNA) was separated in a 10% polyacrylamide TBE gel as described above and the product of needed size was excised (~120-320bp) for gel extraction. Finally, sscDNA template was amplified by PCR using the Phusion High-Fidelity enzyme (NEB) according to the manufacturer's instructions. The oligonucleotide primers oNTI200 and were used to generate DNA for deep sequencing, for sequence see Table 4.9.1.

Table 4.9.1: Gro-Seq Oligos

Gro-Seq	
Name	**Sequence 5' to 3'**
oNTI223	pGATCGTCGGACTGTAGAACTCT;CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTTTVN
	p = 5' phosphorylation
	";" = abasic dSpacer furan
	VN = degenerate nucleotides
oNTI200	CAAGCAGAAGACGGCATA
oNTI201	AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGACG
Illumina small RNA-seq	CGACAGGTTCAGAGTTCTACAGTCCGACGATC

4.9.1 De novo identification of Gro-Seq transcripts

Transcript identification and assignemnt to genomic regions, including annotated genes was accomplished using HOMER. Gro-Seq read densities were analyzed in a similar manner to ChIP-Seq, except that in this case all the Gro-Seq libraries corresponding to the same experiment were merged in order to maximize read density for transcript identification. Provided Gro-Seq generates strand-specific data, separate tracks were uploaded onto the UCSC genome browser, once tag enriched sites were identified using a sliding window of 250bp. The portion of Gro-Seq tags that mapped to repeat regions was excluded and instead, the read density for these regions was approximated with values from flanking regions to avoid having to end transcripts prematurely. Transcript initiation sites were identified as regions where the Gro-Seq read density increased threefold relative to the preceding 1kb region. Transcript termination sites were defined by either a reduction in reads below 10% of the start of the transcript or when another transcript's start was identified on the same strand. Individual high density peaks spanning a region less than 250bp were considered artifactual and thus removed from the analysis. Transcripts were defined as putative eRNAs if their TSS was located distal to RefSeq TSS (≥3kb) and were associated with ERα and H3K4me1 regions. To identify differentially regulated transcripts, strand-specific read counts from each Gro-Seq experiment were determined for each transcript using HOMER (5). EdgeR (http://www.bioconductor.org/) was then used to calculate differential genomic and non-genic expression (≥1.5-fold, ≤0.01 false discovery rate).

4.10 3C & 3D-DSL

4.10.1 Chromatin confirmation capture (3C)

3C was performed as per Lieberman-Aiden et Al., (7). Briefly $25x10^6$ MCF-7cells were fixed by adding 1% formaldehyde at room temperature for 10 minutes, and the reaction stopped by adding glycine. Lysis buffer (500 µl 10 mM Tris-HCl pH8.0, 10 mM NaCl, 0.2% Igepal CA630; protease inhibitors (Sigma) was added and cells were incubated on ice. Next, cells were lysed with a Dounce homogenizer, and the suspension spun down at 5000 rpm at 4°C. The supernatant was discarded and the pellet was washed twice with 500 µl ice-cold 1x NEBuffer 2 (NEB, Ipswich, MA). The pellet was then resuspended in 1X NEBuffer 2 and split into five separate 50 µl aliquots. The extracted chromatin was then digested overnight by adding 400 Units HindIII (NEB). Each digested chromatin mixture was ligated by adding T4 DNA Ligase (800 U) in 20 times of initial volume for 4 hrs at 16°C. The ligase step was omitted in one chromatin aliquot from the five mentioned above to use the sample as the unligated control.

After incubation at 16°C, the chromatin was decrosslinked overnight at 65ºC and purified twice with phenol and then with phenol:chloroform:IAA (25:24:1). DNA was precipitated and pellets were air-dried before resuspending in 250 µl 1X TE buffer. To degrade any carryover RNA, 1µl RNAse A (1 mg/ml) was added to each tube and incubated at 37°C for 15 minutes. DNA was further purified using Phenol:Chloroform:IAA and precipitated. This enriched fraction was used for the DSL part of the protocol.

4.10.2 Probe design

Donor and acceptor probes were designed using HindIII sites covering both strands 200kb upstream and downstream of the following genes: TFF1, GREB1, FOXC1, NRIP1, ZNF217 and CA12. The interval chosen is long enough to cover the most prominent ERα binding sites as well as enhancers. The TFF1 locus was covered by 19 donors and 20 acceptors; GREB1 had 93 donors and 74 acceptors; FOXC1 had 26 donors and 24 acceptors; NRIP1 consisted of 127 donors and 54 acceptors; ZNF217 had 58 donors and 37 acceptors and CA12 had 89 donors and 30 aceptors. All sequences were mapped to genome build hg18, and they can be viewed in Table 4.10.2 [1, 2, 3, 4]. A universal sequence was added to the probe that is compatible with HiSeq 2000 adapters for direct sequencing. Both acceptors and donors were pooled individually in equimolar amounts.

Table 4.10.2.1: 3D-DSL (DNO = donor non-overlapping)

3D-DSL (DNO= donor non-overlapping)	
Name	**Sequence 5' to 3'**
FOXC1_DNO1	AATGATACGGCGACCACCGAGATAGTCCTTAAAATTTGGTAAG
FOXC1_DNO2	AATGATACGGCGACCACCGAGATCCCAGCACCAGACAAGAAAG
FOXC1_DNO3	AATGATACGGCGACCACCGAGATAGACAACCTGGAGAGACAAG
FOXC1_DNO4	AATGATACGGCGACCACCGAGATTTACATGTGAAAGATTTAAG
FOXC1_DNO5	AATGATACGGCGACCACCGAGATGTGTTCCCAAGGAAATTAAG
FOXC1_DNO6	AATGATACGGCGACCACCGAGATCTCTGTCTTCTTGTCTGAAG
FOXC1_DNO7	AATGATACGGCGACCACCGAGATTGAAAAAAGAGAAAATAAAG
FOXC1_DNO8	AATGATACGGCGACCACCGAGATGTGTAGAGTGAATAGGAAAG
FOXC1_DNO9	AATGATACGGCGACCACCGAGATATGCCTACTTGTCAATAAAG
FOXC1_DNO10	AATGATACGGCGACCACCGAGATCAGTTTATGCCACCCCAAAG
FOXC1_DNO11	AATGATACGGCGACCACCGAGATTCAGAGGCTGCGCTGGGAAG
FOXC1_DNO12	AATGATACGGCGACCACCGAGATGGAAGGTAATTTTCCTTAAG
FOXC1_DNO13	AATGATACGGCGACCACCGAGATAAGAAACAATGACCAGGAAG
FOXC1_DNO14	AATGATACGGCGACCACCGAGATATGTACTGTCAGGAAGTAAG
GREB1_DNO1	AATGATACGGCGACCACCGAGATTCTCATGTCCCCCTAGAAAG
GREB1_DNO2	AATGATACGGCGACCACCGAGATAGAAGTCACGTGTGGAGAAG
GREB1_DNO3	AATGATACGGCGACCACCGAGATATGGGAGGAAGATAACAAAG
GREB1_DNO4	AATGATACGGCGACCACCGAGATATGGATTCATACACCTAAAG
GREB1_DNO5	AATGATACGGCGACCACCGAGATATGAACATTCATGTACAAAG
GREB1_DNO6	AATGATACGGCGACCACCGAGATAAGCCACGGGGACCAGAAAG
GREB1_DNO7	AATGATACGGCGACCACCGAGATAGAGTAGCAACTTGCAAAAG
GREB1_DNO8	AATGATACGGCGACCACCGAGATAATACATTGCTTTTTAGAAG
GREB1_DNO9	AATGATACGGCGACCACCGAGATATTAACCAGTATCCTGCAAG
GREB1_DNO10	AATGATACGGCGACCACCGAGATAGGGCAAAAATGCAAAGAAG
GREB1_DNO11	AATGATACGGCGACCACCGAGATAGGGCCCGTGGGTGCTCAAG
GREB1_DNO12	AATGATACGGCGACCACCGAGATAAAACACAATGAACACAAAG
GREB1_DNO13	AATGATACGGCGACCACCGAGATTCAAACAGAAGCAAAGGAAG
GREB1_DNO14	AATGATACGGCGACCACCGAGATAGCCTGGCATTGACTGGAAG
GREB1_DNO15	AATGATACGGCGACCACCGAGATGCTATTATCAGATTTTTAAG
GREB1_DNO16	AATGATACGGCGACCACCGAGATAAATTATTTGATCAGGTAAG
GREB1_DNO17	AATGATACGGCGACCACCGAGATGGGTGTTGTGGGGCCTGAAG
GREB1_DNO18	AATGATACGGCGACCACCGAGATCTAGTACATTAGTTATAAAG
GREB1_DNO19	AATGATACGGCGACCACCGAGATGCAAGCCTTGGAGAATAAAG
GREB1_DNO20	AATGATACGGCGACCACCGAGATGGTCCCCTACAGTTTCAAAG
GREB1_DNO21	AATGATACGGCGACCACCGAGATAGGAAGTACTGTGACAAAAG
GREB1_DNO22	AATGATACGGCGACCACCGAGATGAAGTCCAGGAATTCACAAG
GREB1_DNO23	AATGATACGGCGACCACCGAGATAGGGAAGGCTAAAGTGGAAG
GREB1_DNO24	AATGATACGGCGACCACCGAGATACAATTACAGAAGTATAAAG
GREB1_DNO25	AATGATACGGCGACCACCGAGATTTCTCCCCAGTGCCTGAAAG

Table 4.10.2.1: 3D-DSL (DNO = donor non-overlapping), continued

3D-DSL (DNO= donor non-overlapping)	
Name	**Sequence 5' to 3'**
GREB1_DNO26	AATGATACGGCGACCACCGAGATACATTAAACACCCCACAAAG
GREB1_DNO27	AATGATACGGCGACCACCGAGATCTGTCCCTGACTTGGGAAAG
GREB1_DNO28	AATGATACGGCGACCACCGAGATCCCCAAAATTTGCTCACAAG
GREB1_DNO29	AATGATACGGCGACCACCGAGATAGTAGGCACAGAAACTGAAG
GREB1_DNO30	AATGATACGGCGACCACCGAGATTTCACTCAATTCCTCCAAAG
GREB1_DNO31	AATGATACGGCGACCACCGAGATTTTTCCTCTATTAAAAAAAG
GREB1_DNO32	AATGATACGGCGACCACCGAGATAACGTGTGTATTTCCTAAAG
GREB1_DNO33	AATGATACGGCGACCACCGAGATCTAAGTAAATTTTACTAAAG
GREB1_DNO34	AATGATACGGCGACCACCGAGATATGGATCCTTTGTTGTAAAG
GREB1_DNO35	AATGATACGGCGACCACCGAGATTGAGACTCTGTGGCTCTAAG
GREB1_DNO36	AATGATACGGCGACCACCGAGATTTTTAGGTAATATTCTTAAG
GREB1_DNO37	AATGATACGGCGACCACCGAGATAAAAATGCTTTGTGGGGAAG
GREB1_DNO38	AATGATACGGCGACCACCGAGATTCCTCCAAACACTTTCAAAG
GREB1_DNO39	AATGATACGGCGACCACCGAGATTCTAAAAGTTAAATTATAAG
GREB1_DNO40	AATGATACGGCGACCACCGAGATGGCATTTGGGAGGATTGAAG
GREB1_DNO41	AATGATACGGCGACCACCGAGATAAAACGAGTCCACCTTAAAG
GREB1_DNO42	AATGATACGGCGACCACCGAGATATTATTTCACTTATAAAAAG
GREB1_DNO43	AATGATACGGCGACCACCGAGATTCTTATCAGTTTAAAAAAAG
GREB1_DNO44	AATGATACGGCGACCACCGAGATATCTTATTAATGAGGAAAAG
GREB1_DNO45	AATGATACGGCGACCACCGAGATATATTCTTTTAAACTCTAAG
GREB1_DNO46	AATGATACGGCGACCACCGAGATAGACTGGTAAGCAAGGCAAG
GREB1_DNO47	AATGATACGGCGACCACCGAGATATACCAGCACAGTTAAGAAG
GREB1_DNO48	AATGATACGGCGACCACCGAGATACTTTACAGATTCAGGAAAG
GREB1_DNO49	AATGATACGGCGACCACCGAGATACTTTCCACTTGACAATAAG
GREB1_DNO50	AATGATACGGCGACCACCGAGATACAACAGAGAAGTACACAAG
GREB1_DNO51	AATGATACGGCGACCACCGAGATAAACTTCCCTATACTAAAAG
GREB1_DNO52	AATGATACGGCGACCACCGAGATGAAAGAAATAAGTTTATAAG
GREB1_DNO53	AATGATACGGCGACCACCGAGATTGTAAAAGTCAACCATAAAG
GREB1_DNO54	AATGATACGGCGACCACCGAGATACTATTCCACTCTCCAAAAG
GREB1_DNO55	AATGATACGGCGACCACCGAGATTCAGGATATTTGCAGACAAG
GREB1_DNO56	AATGATACGGCGACCACCGAGATCAGGTGAGCAGTATGATAAG
GREB1_DNO57	AATGATACGGCGACCACCGAGATACTCCGGACTGAAATGGAAG
NRIP1_DNO1	AATGATACGGCGACCACCGAGATACTAGATCATGGCAGATAAG
NRIP1_DNO2	AATGATACGGCGACCACCGAGATCCAATAAGTAGAAAGTCAAG
NRIP1_DNO3	AATGATACGGCGACCACCGAGATGAATCTGAGACTCAGACAAG
NRIP1_DNO4	AATGATACGGCGACCACCGAGATGAATGAAAACCTTGGAGAAG
NRIP1_DNO5	AATGATACGGCGACCACCGAGATTTAAATTACTAGAAGGTAAG
NRIP1_DNO6	AATGATACGGCGACCACCGAGATTAGAAAATATACAAAACAAG
NRIP1_DNO7	AATGATACGGCGACCACCGAGATCAGCAACACAGTAATTAAAG

Table 4.10.2.1: 3D-DSL (DNO = donor non-overlapping), continued

3D-DSL (DNO= donor non-overlapping)	
Name	**Sequence 5' to 3'**
NRIP1_DNO8	AATGATACGGCGACCACCGAGATAAGTTGATATAAGAGATAAG
NRIP1_DNO9	AATGATACGGCGACCACCGAGATCTTCCCTGATAAAAGCAAAG
NRIP1_DNO10	AATGATACGGCGACCACCGAGATTTGGGGGTGGGGAATGGAAG
NRIP1_DNO11	AATGATACGGCGACCACCGAGATCCTTCGCTAGAGATCTAAAG
NRIP1_DNO12	AATGATACGGCGACCACCGAGATATGCTAAATGATGCCAAAAG
NRIP1_DNO13	AATGATACGGCGACCACCGAGATTGATTTTCCACACTAACAAG
NRIP1_DNO14	AATGATACGGCGACCACCGAGATAAAAATGACTTGTTTAAAAG
NRIP1_DNO15	AATGATACGGCGACCACCGAGATTTTAAACTAAATGTTGTAAG
NRIP1_DNO16	AATGATACGGCGACCACCGAGATGCTGTTTCAGAACATTAAAG
NRIP1_DNO17	AATGATACGGCGACCACCGAGATTCTAATGGCTCAGTATAAAG
NRIP1_DNO18	AATGATACGGCGACCACCGAGATAATGCCACCATTTTAAGAAG
NRIP1_DNO19	AATGATACGGCGACCACCGAGATCACTATGTGTATGCACTAAG
NRIP1_DNO20	AATGATACGGCGACCACCGAGATGTAAATTTACACTATAAAAG
NRIP1_DNO21	AATGATACGGCGACCACCGAGATATTTCTCACAATATTTCAAG
NRIP1_DNO22	AATGATACGGCGACCACCGAGATTATTGCTGACATACAGAAAG
NRIP1_DNO23	AATGATACGGCGACCACCGAGATGGGCAGAGTCAACTGAAAAG
NRIP1_DNO24	AATGATACGGCGACCACCGAGATAAAATACCAGTTCCATCAAG
NRIP1_DNO25	AATGATACGGCGACCACCGAGATTTTACTATTCACAGACAAAG
NRIP1_DNO26	AATGATACGGCGACCACCGAGATATTCAGAGGGAGACAGAAAG
NRIP1_DNO27	AATGATACGGCGACCACCGAGATATAAGAGAATGCTATTCAAG
NRIP1_DNO28	AATGATACGGCGACCACCGAGATAAAAAATCAGGGCTGAGAAG
NRIP1_DNO29	AATGATACGGCGACCACCGAGATGCAACTAAGGCTACTACAAG
NRIP1_DNO30	AATGATACGGCGACCACCGAGATCCATCCCTATGACTTGTAAG
NRIP1_DNO31	AATGATACGGCGACCACCGAGATAGGATTAAACCTCCACCAAG
NRIP1_DNO32	AATGATACGGCGACCACCGAGATTTTCTCAATTAGAAAAGAAG
NRIP1_DNO33	AATGATACGGCGACCACCGAGATGGCAACAGCAGGAGACAAAG
NRIP1_DNO34	AATGATACGGCGACCACCGAGATACAAAGCTTCAAGCTAGAAG
NRIP1_DNO35	AATGATACGGCGACCACCGAGATTTGACCCAACATCTCCCAAG
NRIP1_DNO36	AATGATACGGCGACCACCGAGATATACTATTTCTTCTTTCAAG
NRIP1_DNO37	AATGATACGGCGACCACCGAGATGTGATAAGGAAAAGCAAAAG
NRIP1_DNO38	AATGATACGGCGACCACCGAGATCCATTAAATTTATTCTTAAG
NRIP1_DNO39	AATGATACGGCGACCACCGAGATCTTTTTTGATCGTTTCTAAG
NRIP1_DNO40	AATGATACGGCGACCACCGAGATCAATTTTGAATACCAAGAAG
NRIP1_DNO41	AATGATACGGCGACCACCGAGATTTACCTAGATCTTCTGGAAG
NRIP1_DNO42	AATGATACGGCGACCACCGAGATAAGACAAAATGTGTGTAAAG
NRIP1_DNO43	AATGATACGGCGACCACCGAGATTATATGTGACTACAGCTAAG
NRIP1_DNO44	AATGATACGGCGACCACCGAGATTCCCTTTAAGTCCTCTAAAG
NRIP1_DNO45	AATGATACGGCGACCACCGAGATCAAGACCAAAATAAAAAAAG
NRIP1_DNO46	AATGATACGGCGACCACCGAGATCTATATTTTGGATCCTGAAG

Table 4.10.2.1: 3D-DSL (DNO = donor non-overlapping), continued

3D-DSL (DNO= donor non-overlapping)	
Name	**Sequence 5' to 3'**
NRIP1_DNO47	AATGATACGGCGACCACCGAGATACTCAGCACACTAAATGAAG
NRIP1_DNO48	AATGATACGGCGACCACCGAGATTAGATTTGGGGTTGACTAAG
NRIP1_DNO49	AATGATACGGCGACCACCGAGATTGGGAAATGGTGGGGAAAAG
NRIP1_DNO50	AATGATACGGCGACCACCGAGATAATCACCACGTTTCACCAAG
NRIP1_DNO51	AATGATACGGCGACCACCGAGATAGGGTAGTCAGATTTTAAAG
NRIP1_DNO52	AATGATACGGCGACCACCGAGATTAAAATAACAAAATATAAAG
NRIP1_DNO53	AATGATACGGCGACCACCGAGATTTGTCAATATCTCTATCAAG
NRIP1_DNO54	AATGATACGGCGACCACCGAGATGATGATTGACCTCCAAAAAG
NRIP1_DNO55	AATGATACGGCGACCACCGAGATCACACACAATGGAAGGAAAG
NRIP1_DNO56	AATGATACGGCGACCACCGAGATCTGCATTTCTCATCTGTAAG
NRIP1_DNO57	AATGATACGGCGACCACCGAGATCCCCAGACCCTGAAGCAAAG
NRIP1_DNO58	AATGATACGGCGACCACCGAGATAATTCATTTCATGAAATAAG
NRIP1_DNO59	AATGATACGGCGACCACCGAGATAACTTTTCTTCTTGTGGAAG
NRIP1_DNO60	AATGATACGGCGACCACCGAGATTGCCTATCCATTACCACAAG
NRIP1_DNO61	AATGATACGGCGACCACCGAGATTCTAGGCAAATTAACTCAAG
NRIP1_DNO62	AATGATACGGCGACCACCGAGATAACTTCTTGCTGTGTAGAAG
NRIP1_DNO63	AATGATACGGCGACCACCGAGATACAATAGGGTAGTGTAAAAG
NRIP1_DNO64	AATGATACGGCGACCACCGAGATAACAGAATCTTGGTGGGAAG
NRIP1_DNO65	AATGATACGGCGACCACCGAGATTCGCAGCTGAGAGAAACAAG
NRIP1_DNO66	AATGATACGGCGACCACCGAGATTTCCCCTACTAGACTATAAG
NRIP1_DNO67	AATGATACGGCGACCACCGAGATCTCCTTTAAAATGTGTTAAG
NRIP1_DNO68	AATGATACGGCGACCACCGAGATTCACACTTATGAAAGTTAAG
NRIP1_DNO69	AATGATACGGCGACCACCGAGATTGAAGGAGAAGTACCATAAG
NRIP1_DNO70	AATGATACGGCGACCACCGAGATGTCATGAACCAGAATTTAAG
NRIP1_DNO71	AATGATACGGCGACCACCGAGATACAGCATGCAAATAAGAAAG
NRIP1_DNO72	AATGATACGGCGACCACCGAGATTTTTTAAGAATTGACTGAAG
NRIP1_DNO73	AATGATACGGCGACCACCGAGATTCTCACCGTGTTTCTGTAAG
NRIP1_DNO74	AATGATACGGCGACCACCGAGATAAAGAAACAATACCTCTAAG
NRIP1_DNO75	AATGATACGGCGACCACCGAGATGAACCTTCATTGCCTTAAAG
NRIP1_DNO76	AATGATACGGCGACCACCGAGATTAGGTGGGAACTGAACAAAG
NRIP1_DNO77	AATGATACGGCGACCACCGAGATTTCCCTATTGACAGAAGAAG
NRIP1_DNO78	AATGATACGGCGACCACCGAGATCTGTAAAACTCTAGCTCAAG
NRIP1_DNO79	AATGATACGGCGACCACCGAGATGTCTGGTTTTCAAATGAAAG
NRIP1_DNO80	AATGATACGGCGACCACCGAGATGGCTTTGTCAGTATGACAAG
NRIP1_DNO81	AATGATACGGCGACCACCGAGATAAATTCAGTGACCCTTTAAG
NRIP1_DNO82	AATGATACGGCGACCACCGAGATATGTCTTCTAAGTCTAAAAG
NRIP1_DNO83	AATGATACGGCGACCACCGAGATTGTGCCAGGCACCTTCTAAG
NRIP1_DNO84	AATGATACGGCGACCACCGAGATGAGATTTTTACCTCTCCAAG
NRIP1_DNO85	AATGATACGGCGACCACCGAGATCAGATGAGTTCCAAAGAAAG

Table 4.10.2.1: 3D-DSL (DNO = donor non-overlapping), continued

3D-DSL (DNO= donor non-overlapping)	
Name	**Sequence 5' to 3'**
NRIP1_DNO86	AATGATACGGCGACCACCGAGATCTCATGTGGGGGGTCTAAAG
NRIP1_DNO87	AATGATACGGCGACCACCGAGATGTAACCTTTTTATTTATAAG
NRIP1_DNO88	AATGATACGGCGACCACCGAGATTTATGACCTAAAAGTATAAG
NRIP1_DNO89	AATGATACGGCGACCACCGAGATGTGTGTTCCCATCACAAAAG
NRIP1_DNO90	AATGATACGGCGACCACCGAGATTTCATCTTCTAGAAGGTAAG
NRIP1_DNO91	AATGATACGGCGACCACCGAGATACTTGATAAGTAAATGAAAG
NRIP1_DNO92	AATGATACGGCGACCACCGAGATTTAGTATTGTCTTTCCAAAG
NRIP1_DNO93	AATGATACGGCGACCACCGAGATTACCCATCCTTCTTGGAAAG
NRIP1_DNO94	AATGATACGGCGACCACCGAGATATACCGACTTGCCTCCAAAG
NRIP1_DNO95	AATGATACGGCGACCACCGAGATGGGGGAACAAACAGTGGAAG
NRIP1_DNO96	AATGATACGGCGACCACCGAGATTCAAATGCTCTGAGAACAAG
NRIP1_DNO97	AATGATACGGCGACCACCGAGATTATTTATTTTCTTCAAAAAG
NRIP1_DNO98	AATGATACGGCGACCACCGAGATCTTGTTTTGTGTCTTGAAAG
NRIP1_DNO99	AATGATACGGCGACCACCGAGATTTCCACATTTGAAAGATAAG
NRIP1_DNO100	AATGATACGGCGACCACCGAGATGATCCTTTGATCTGAAAAAG
TFF1_DNO1	AATGATACGGCGACCACCGAGATTAATGAAATGCTCCCAGAAG
TFF1_DNO2	AATGATACGGCGACCACCGAGATAAGAGGCATGAAACCAGAAG
TFF1_DNO3	AATGATACGGCGACCACCGAGATTTGTTAAAATTTTTTAAAAG
TFF1_DNO4	AATGATACGGCGACCACCGAGATTGAGCAAACCTTTTCAGAAG
TFF1_DNO5	AATGATACGGCGACCACCGAGATTGAGGAACCAAACAAATAAG
TFF1_DNO6	AATGATACGGCGACCACCGAGATCTCCTGGGAGCCCCATGAAG
TFF1_DNO7	AATGATACGGCGACCACCGAGATTTTTCCCTCTGGGTCACAAG
TFF1_DNO8	AATGATACGGCGACCACCGAGATTTTTCTGTACACATATAAAG
TFF1_DNO9	AATGATACGGCGACCACCGAGATCTTCCACACTATAACAAAAG
ZNF217_DNO1	AATGATACGGCGACCACCGAGATTTATTCCTGTGATGACAAAG
ZNF217_DNO2	AATGATACGGCGACCACCGAGATATAAAACAATAGCAGTCAAG
ZNF217_DNO3	AATGATACGGCGACCACCGAGATGAATTCTTTCTCCAGGAAAG
ZNF217_DNO4	AATGATACGGCGACCACCGAGATAGCGTTATGAACACGCTAAG
ZNF217_DNO5	AATGATACGGCGACCACCGAGATATTTCCCCATTTTAAATAAG
ZNF217_DNO6	AATGATACGGCGACCACCGAGATAAATGAGACACACTACAAAG
ZNF217_DNO7	AATGATACGGCGACCACCGAGATCAAAGAGGAAGAGAGCCAAG
ZNF217_DNO8	AATGATACGGCGACCACCGAGATTTATAATTATTCAATCCAAG
ZNF217_DNO9	AATGATACGGCGACCACCGAGATGCCATTTTAGAATTAGAAAG
ZNF217_DNO10	AATGATACGGCGACCACCGAGATTTCCTTTTTCCCTTTCTAAG
ZNF217_DNO11	AATGATACGGCGACCACCGAGATTAGGGTATAAGAATTCAAAG
ZNF217_DNO12	AATGATACGGCGACCACCGAGATATAATAATAATAATGTAAAG
ZNF217_DNO13	AATGATACGGCGACCACCGAGATTGCTTTAAAAATAAGATAAG
ZNF217_DNO14	AATGATACGGCGACCACCGAGATCCAGTATAAATTTAATTAAG
ZNF217_DNO15	AATGATACGGCGACCACCGAGATATTTGAAATGTTCCCAAAAG

Table 4.10.2.1: 3D-DSL (DNO = donor non-overlapping), continued

3D-DSL (DNO= donor non-overlapping)	
Name	**Sequence 5' to 3'**
ZNF217_DNO16	AATGATACGGCGACCACCGAGATGCACCAAATAGATGGCAAAG
ZNF217_DNO17	AATGATACGGCGACCACCGAGATTTCTTTGGTGAAACTGTAAG
ZNF217_DNO18	AATGATACGGCGACCACCGAGATTTGGACATGTATCCTTAAAG
ZNF217_DNO19	AATGATACGGCGACCACCGAGATTGAGAAAAGGCCTGTCCAAG
ZNF217_DNO20	AATGATACGGCGACCACCGAGATGAGTTGGATTAACTAACAAG
ZNF217_DNO21	AATGATACGGCGACCACCGAGATCTTTCTCATTCTGAGAAAAG
ZNF217_DNO22	AATGATACGGCGACCACCGAGATTGGCGCGGCTAGATTTGAAG
ZNF217_DNO23	AATGATACGGCGACCACCGAGATAGAGCTACCTTCATTAAAAG
ZNF217_DNO24	AATGATACGGCGACCACCGAGATGGTTCTCTTTGTCAGATAAG
ZNF217_DNO25	AATGATACGGCGACCACCGAGATCATGATCACCCTCCTTTAAG
ZNF217_DNO26	AATGATACGGCGACCACCGAGATGCCTGTCATGAGCTTGGAAG
ZNF217_DNO27	AATGATACGGCGACCACCGAGATTTCCACTAGAAATTTGGAAG
ZNF217_DNO28	AATGATACGGCGACCACCGAGATTATTCCAGCATACAGAAAAG
ZNF217_DNO29	AATGATACGGCGACCACCGAGATGTTTTTGTTTTTCCAAAAAG
ZNF217_DNO30	AATGATACGGCGACCACCGAGATATTGAGTTTCATTTTTTAAG
ZNF217_DNO31	AATGATACGGCGACCACCGAGATAATAAGGAAATATATTAAAG
ZNF217_DNO32	AATGATACGGCGACCACCGAGATTTCTCAGAAGTAAGGGCAAG
ZNF217_DNO33	AATGATACGGCGACCACCGAGATCATGATGTAGTGAACAAAAG
ZNF217_DNO34	AATGATACGGCGACCACCGAGATATAACAAATTTTGCTGAAAG
ZNF217_DNO35	AATGATACGGCGACCACCGAGATTGGCGAGCGAGTGAGCCAAG
ZNF217_DNO36	AATGATACGGCGACCACCGAGATGATTTTGGACAAACTCGAAG
ZNF217_DNO37	AATGATACGGCGACCACCGAGATGCGCCAGCAACACCTGGAAG
ZNF217_DNO38	AATGATACGGCGACCACCGAGATCACGCTGTTGGTCAATAAAG
ZNF217_DNO39	AATGATACGGCGACCACCGAGATATGGGAACCTTCTACGCAAG
ZNF217_DNO40	AATGATACGGCGACCACCGAGATGCAGAACACATGCTTGGAAG
CA12_DNO1	AATGATACGGCGACCACCGAGATTACTGATTTACTTACACAAG
CA12_DNO2	AATGATACGGCGACCACCGAGATATTACTTTCTTAATGTAAAG
CA12_DNO3	AATGATACGGCGACCACCGAGATATTGGAACTAGAAATAGAAG
CA12_DNO4	AATGATACGGCGACCACCGAGATGGGCTAAAGCCAACCAAAAG
CA12_DNO5	AATGATACGGCGACCACCGAGATTACCATTCTACAATTTTAAG
CA12_DNO6	AATGATACGGCGACCACCGAGATCATAACGTAGAATTAGGAAG
CA12_DNO7	AATGATACGGCGACCACCGAGATCTATATATGTGTCCCATAAG
CA12_DNO8	AATGATACGGCGACCACCGAGATTGCATGGCACTCATGGTAAG
CA12_DNO9	AATGATACGGCGACCACCGAGATTATTCCACATGTTTAATAAG
CA12_DNO10	AATGATACGGCGACCACCGAGATGTCTATAAGCCTGTTTAAAG
CA12_DNO11	AATGATACGGCGACCACCGAGATATAAGCAGTTGGGAGAGAAG
CA12_DNO12	AATGATACGGCGACCACCGAGATATCCTCAAAGATCCTGGAAG
CA12_DNO13	AATGATACGGCGACCACCGAGATCACCTATAATATATAAAAAG
CA12_DNO14	AATGATACGGCGACCACCGAGATCACGGTCCACTATTGAGAAG

Table 4.10.2.1: 3D-DSL (DNO = donor non-overlapping), continued

3D-DSL (DNO= donor non-overlapping)	
Name	**Sequence 5' to 3'**
CA12_DNO15	AATGATACGGCGACCACCGAGATAAGCCAAGTGGGAAGAGAAG
CA12_DNO16	AATGATACGGCGACCACCGAGATAGCCTTAGGAAAAGGGAAAG
CA12_DNO17	AATGATACGGCGACCACCGAGATGGGAGTGTCACCAAATAAAG
CA12_DNO18	AATGATACGGCGACCACCGAGATAGGGACATTCTTTCAACAAG
CA12_DNO19	AATGATACGGCGACCACCGAGATGGAGGAGAATTCTGGCCAAG
CA12_DNO20	AATGATACGGCGACCACCGAGATGGATGGGTCAGTTTCTAAAG
CA12_DNO21	AATGATACGGCGACCACCGAGATCCAAGGGGGTGGAAGAGAAG
CA12_DNO22	AATGATACGGCGACCACCGAGATCAGGTGACTCAGAGCAGAAG
CA12_DNO23	AATGATACGGCGACCACCGAGATATCTCCTATCAAATAAGAAG
CA12_DNO24	AATGATACGGCGACCACCGAGATTTAGTGAAGTACCTTCCAAG
CA12_DNO25	AATGATACGGCGACCACCGAGATAAATTCCTCCAGGGCTCAAG
CA12_DNO26	AATGATACGGCGACCACCGAGATTATCAGAGCTGGGGATCAAG
CA12_DNO27	AATGATACGGCGACCACCGAGATAGAGATGCATCATTGAGAAG
CA12_DNO28	AATGATACGGCGACCACCGAGATCAGAGAGTAAGTTTTTGAAG
CA12_DNO29	AATGATACGGCGACCACCGAGATGTCAACTGCCATTCTGTAAG
CA12_DNO30	AATGATACGGCGACCACCGAGATAGCAGAAGAGGTGGGGCAAG
CA12_DNO31	AATGATACGGCGACCACCGAGATTAGCAGCATTTCTCAAAAAG
CA12_DNO32	AATGATACGGCGACCACCGAGATGACAGGTGAAAGTCAGCAAG
CA12_DNO33	AATGATACGGCGACCACCGAGATTCCCCTGATGTACCCACAAG
CA12_DNO34	AATGATACGGCGACCACCGAGATCAAGCACTTGCACACTGAAG
CA12_DNO35	AATGATACGGCGACCACCGAGATCATTGCTTTAAGCCACTAAG
CA12_DNO36	AATGATACGGCGACCACCGAGATCTGTTATTTTCTTTGGGAAG
CA12_DNO37	AATGATACGGCGACCACCGAGATTTTTCTTTGGGAAGCTTAAG
CA12_DNO38	AATGATACGGCGACCACCGAGATAAGACAAATGAAAGAAAAAG
CA12_DNO39	AATGATACGGCGACCACCGAGATTAAATGAAAAATAAACTAAG
CA12_DNO40	AATGATACGGCGACCACCGAGATGCTTTGTGTTTCTTGTGAAG
CA12_DNO41	AATGATACGGCGACCACCGAGATAAGGATCCTCCTCCTAGAAG
CA12_DNO42	AATGATACGGCGACCACCGAGATAAAATACTGACTTGAGCAAG
CA12_DNO43	AATGATACGGCGACCACCGAGATTTCATACAAATATTCTTAAG
CA12_DNO44	AATGATACGGCGACCACCGAGATTTGCCCTGCAACACACAAAG
CA12_DNO45	AATGATACGGCGACCACCGAGATCCTCCAGTCATTTCTGAAAG
CA12_DNO46	AATGATACGGCGACCACCGAGATTACTGTATCTAGACCTAAAG
CA12_DNO47	AATGATACGGCGACCACCGAGATGTCAACATGCAAAAATGAAG
CA12_DNO48	AATGATACGGCGACCACCGAGATAAAAACCCCTAACTTAGAAG
CA12_DNO49	AATGATACGGCGACCACCGAGATACAAGCAAACAAAACAAAAG
CA12_DNO50	AATGATACGGCGACCACCGAGATTGTGGCACTTTCCTTTGAAG
CA12_DNO51	AATGATACGGCGACCACCGAGATCGTCTGCTACACACCAGAAG
CA12_DNO52	AATGATACGGCGACCACCGAGATAACATAATATAGATGCAAAG
CA12_DNO53	AATGATACGGCGACCACCGAGATACCTAAATACACTCAGGAAG

Table 4.10.2.1: 3D-DSL (DNO = donor non-overlapping), continued

3D-DSL (DNO= donor non-overlapping)	
Name	**Sequence 5' to 3'**
CA12_DNO54	AATGATACGGCGACCACCGAGATTACAGGAAAATGAACTAAAG
CA12_DNO55	AATGATACGGCGACCACCGAGATATGCTAGAATGCCTGGGAAG
CA12_DNO56	AATGATACGGCGACCACCGAGATCTGATGCTGTTTTTGTTAAG
CA12_DNO57	AATGATACGGCGACCACCGAGATACCATATGTATCCCACAAAG
CA12_DNO58	AATGATACGGCGACCACCGAGATTGGTTGGTAGCATTTCAAAG
CA12_DNO59	AATGATACGGCGACCACCGAGATTGGAGTATGATATCACTAAG
CA12_DNO60	AATGATACGGCGACCACCGAGATCACGTGGTGATGGAGCTAAG
CA12_DNO61	AATGATACGGCGACCACCGAGATTTTTAAAATAATGATAAAAG
CA12_DNO62	AATGATACGGCGACCACCGAGATGTTTCTAACACTTCACAAAG
CA12_DNO63	AATGATACGGCGACCACCGAGATCTCTGAGTGTTTTCCCCAAG
CA12_DNO64	AATGATACGGCGACCACCGAGATTTAGAACTATAGAGCTAAAG
CA12_DNO65	AATGATACGGCGACCACCGAGATGAAGTATTATAAGCAAAAAG
CA12_DNO66	AATGATACGGCGACCACCGAGATAATGCTTATAAACTATCAAG
CA12_DNO67	AATGATACGGCGACCACCGAGATGCAAGGGCAGCCCCTAAAAG
CA12_DNO68	AATGATACGGCGACCACCGAGATGAGAAAAAAATATTTTAAAG
CA12_DNO69	AATGATACGGCGACCACCGAGATTAACTGGTCTCCTCAGAAAG
CA12_DNO70	AATGATACGGCGACCACCGAGATTAGATGGAAAGAAAATTAAG
CA12_DNO71	AATGATACGGCGACCACCGAGATCCTTTGAGTTTTTATGGAAG
CA12_DNO72	AATGATACGGCGACCACCGAGATTAACGTCACAAGATTTTAAG
CA12_DNO73	AATGATACGGCGACCACCGAGATTGTCATTAGATCACCATAAG
CA12_DNO74	AATGATACGGCGACCACCGAGATATTCTAAATTGACTTAGAAG

Table 4.10.2.2: 3D-DSL (DO = donor overlapping)

3D-DSL (DO= donor overlapping)	
Name	**Sequence 5' to 3'**
FOXC1_DO1	AATGATACGGCGACCACCGAGATGTCAGCTCATTTACGGAAAG
FOXC1_DO2	AATGATACGGCGACCACCGAGATCACTGTCCCTGACCACAAAG
FOXC1_DO3	AATGATACGGCGACCACCGAGATAATCTTTAATAAAATACAAG
FOXC1_DO4	AATGATACGGCGACCACCGAGATCGAGGTATAGTTTTCTCAAG
FOXC1_DO5	AATGATACGGCGACCACCGAGATTCGTCCAGACCCAAGCCAAG
FOXC1_DO6	AATGATACGGCGACCACCGAGATCTGTAACCTGAAAATAAAAG
FOXC1_DO7	AATGATACGGCGACCACCGAGATATAGGCGTGCGCGAGCAAAG
FOXC1_DO8	AATGATACGGCGACCACCGAGATAAAATAGCCTCTGTAAAAAG
FOXC1_DO9	AATGATACGGCGACCACCGAGATGAGGCCGCCCGGCCAGCAAG
FOXC1_DO10	AATGATACGGCGACCACCGAGATCATCTGGACTGGGCAGTAAG
FOXC1_DO11	AATGATACGGCGACCACCGAGATGGGGGCTGCCCTGGGTAAAG
FOXC1_DO12	AATGATACGGCGACCACCGAGATCAGAATGGACAGGAAAAAAG
GREB1_DO1	AATGATACGGCGACCACCGAGATTGAGGTTTAGGGCAGTTAAG
GREB1_DO2	AATGATACGGCGACCACCGAGATGGCTGTTGGTCGAGGGCAAG
GREB1_DO3	AATGATACGGCGACCACCGAGATACAGCCCTGGTAATGGCAAG
GREB1_DO4	AATGATACGGCGACCACCGAGATTGCAGATATTTTCTCTGAAG
GREB1_DO5	AATGATACGGCGACCACCGAGATGAGCCTGCTTCTTGGAGAAG
GREB1_DO6	AATGATACGGCGACCACCGAGATTCATGGCTCAGAGACGGAAG
GREB1_DO7	AATGATACGGCGACCACCGAGATAAGTCGCCCTGGCTGCGAAG
GREB1_DO8	AATGATACGGCGACCACCGAGATTAAAGAGAAGATCTTGAAAG
GREB1_DO9	AATGATACGGCGACCACCGAGATGTATCTTACAGGTTAAAAAG
GREB1_DO10	AATGATACGGCGACCACCGAGATGTCTCATCTATAAAATGAAG
GREB1_DO11	AATGATACGGCGACCACCGAGATTGGAGCCCTCCCTTCTCAAG
GREB1_DO12	AATGATACGGCGACCACCGAGATAACGGGCTATTCTAGACAAG
GREB1_DO13	AATGATACGGCGACCACCGAGATGGCCAATTCTGCATTCCAAG
GREB1_DO14	AATGATACGGCGACCACCGAGATAATATTTTTTAAAGAAAAAG
GREB1_DO15	AATGATACGGCGACCACCGAGATAAAAACCAGCTTCAGGAAAG
GREB1_DO16	AATGATACGGCGACCACCGAGATCTAGCTGGCACAATTTTAAG
GREB1_DO17	AATGATACGGCGACCACCGAGATGTATTCTTCGTCAGTGAAAG
GREB1_DO18	AATGATACGGCGACCACCGAGATATCTATGTCAATGTCTTAAG
GREB1_DO19	AATGATACGGCGACCACCGAGATATCTTCAATGGTCACTCAAG
GREB1_DO20	AATGATACGGCGACCACCGAGATAGAGGAATACAGTCAACAAG
GREB1_DO21	AATGATACGGCGACCACCGAGATACTTACTTTGTATACTGAAG
GREB1_DO22	AATGATACGGCGACCACCGAGATTGGCTTGGGTAACCTAAAAG
GREB1_DO23	AATGATACGGCGACCACCGAGATTGGGCTAAATCACTAAAAAG
GREB1_DO24	AATGATACGGCGACCACCGAGATTAGGGATGAGATCATGGAAG
GREB1_DO25	AATGATACGGCGACCACCGAGATCAGTCTTCACAAGAACGAAG
GREB1_DO26	AATGATACGGCGACCACCGAGATTCTGGAGCTGAAGCGCAAAG
GREB1_DO27	AATGATACGGCGACCACCGAGATGCAGAGTACCTGGCTAAAAG

Table 4.10.2.2: 3D-DSL (DO = donor overlapping), continued

3D-DSL (DO= donor overlapping)	
Name	**Sequence 5' to 3'**
GREB1_DO28	AATGATACGGCGACCACCGAGATAGCCTACAGAGAGATTCAAG
GREB1_DO29	AATGATACGGCGACCACCGAGATCATGATTTGTCCCAGTAAAG
GREB1_DO30	AATGATACGGCGACCACCGAGATATTTATTCACAGGTGGCAAG
GREB1_DO31	AATGATACGGCGACCACCGAGATGTGAATAAAATTAAGGCAAG
GREB1_DO32	AATGATACGGCGACCACCGAGATGTTCCTTATCTTCCTTGAAG
GREB1_DO33	AATGATACGGCGACCACCGAGATGAAGGGTGGCTTGCTGGAAG
GREB1_DO34	AATGATACGGCGACCACCGAGATCCTAAACTTTTCTTTTCAAG
GREB1_DO35	AATGATACGGCGACCACCGAGATACAAAATAGAATTAATGAAG
GREB1_DO36	AATGATACGGCGACCACCGAGATGAACCACAGGCTTCTGTAAG
NRIP1_DO1	AATGATACGGCGACCACCGAGATAGAGATTTAACTTTTCAAAG
NRIP1_DO2	AATGATACGGCGACCACCGAGATATGATTCAGAGTCAGTCAAG
NRIP1_DO3	AATGATACGGCGACCACCGAGATAAGCACATTCTTTCACAAAG
NRIP1_DO4	AATGATACGGCGACCACCGAGATTCTTAAATTAAAGAAAGAAG
NRIP1_DO5	AATGATACGGCGACCACCGAGATAACCATGCCAGCTAGCAAAG
NRIP1_DO6	AATGATACGGCGACCACCGAGATTGCTCTCTCATAATTCCAAG
NRIP1_DO7	AATGATACGGCGACCACCGAGATGTCCTCTAAATGACCTAAAG
NRIP1_DO8	AATGATACGGCGACCACCGAGATAATCTTGTTAACTCTAAAAG
NRIP1_DO9	AATGATACGGCGACCACCGAGATTTGCTGTACAACAAATTAAG
NRIP1_DO10	AATGATACGGCGACCACCGAGATGAAGTGTTTTGAGATGTAAG
NRIP1_DO11	AATGATACGGCGACCACCGAGATAGACTCCTATCCTGAAAAAG
NRIP1_DO12	AATGATACGGCGACCACCGAGATTTGAATGCCAGGACAAAAAG
NRIP1_DO13	AATGATACGGCGACCACCGAGATCTTTGTTTAACTTGGGCAAG
NRIP1_DO14	AATGATACGGCGACCACCGAGATCCCTTTGATGGTAACTGAAG
NRIP1_DO15	AATGATACGGCGACCACCGAGATTTAGTGTGTTCAGACTGAAG
NRIP1_DO16	AATGATACGGCGACCACCGAGATGCACTTGAGCCTGACTGAAG
NRIP1_DO17	AATGATACGGCGACCACCGAGATCAAAAACAATTATATGGAAG
NRIP1_DO18	AATGATACGGCGACCACCGAGATTTGAGAAAAACACCAGAAAG
NRIP1_DO19	AATGATACGGCGACCACCGAGATGCTACTGAGAGGAGAAGAAG
NRIP1_DO20	AATGATACGGCGACCACCGAGATATTGAACTCACAGCCTGAAG
NRIP1_DO21	AATGATACGGCGACCACCGAGATAGAAGTGCAAGAGTAGGAAG
NRIP1_DO22	AATGATACGGCGACCACCGAGATAACAATCCAGGGAAGAAAAG
NRIP1_DO23	AATGATACGGCGACCACCGAGATGAAAGTGTATGTAAGGAAAG
NRIP1_DO24	AATGATACGGCGACCACCGAGATAGGGGACAGCATACACAAAG
NRIP1_DO25	AATGATACGGCGACCACCGAGATTATATTTAAAAAGTATTAAG
NRIP1_DO26	AATGATACGGCGACCACCGAGATATTTAATAGCAATAAGAAAG
NRIP1_DO27	AATGATACGGCGACCACCGAGATTAGTTGGTATAACCTGGAAG
TFF1_DO1	AATGATACGGCGACCACCGAGATAAGATAAGGTCTAGATCAAG
TFF1_DO2	AATGATACGGCGACCACCGAGATAGCAGGAGTCAATGGTGAAG
TFF1_DO3	AATGATACGGCGACCACCGAGATCAACACGGAAACATTCCAAG

Table 4.10.2.2: 3D-DSL (DO = donor overlapping), continued

3D-DSL (DO= donor overlapping)	
Name	**Sequence 5' to 3'**
TFF1_DO4	AATGATACGGCGACCACCGAGATACAGGGAAATTCTAGCCAAG
TFF1_DO5	AATGATACGGCGACCACCGAGATCCTTCATTTCCCCATGGAAG
TFF1_DO6	AATGATACGGCGACCACCGAGATACAGAGGTCAGAATTGGAAG
TFF1_DO7	AATGATACGGCGACCACCGAGATGGCCGTTCTGTATTTCTAAG
TFF1_DO8	AATGATACGGCGACCACCGAGATGGCGCACAAGTATGAAAAAG
TFF1_DO9	AATGATACGGCGACCACCGAGATTACAGCAGATAACAGAAAAG
TFF1_DO10	AATGATACGGCGACCACCGAGATATGAGGTGATCTTGACCAAG
ZNF217_DO1	AATGATACGGCGACCACCGAGATTCATACCAAAAATGTAAAAG
ZNF217_DO2	AATGATACGGCGACCACCGAGATATCTGCCCGTGAGCCCCAAG
ZNF217_DO3	AATGATACGGCGACCACCGAGATCTGGTGGTAGGTTCTGAAAG
ZNF217_DO4	AATGATACGGCGACCACCGAGATGTAACCTAGTTAAATACAAG
ZNF217_DO5	AATGATACGGCGACCACCGAGATCACCAAAGATTTTCTTCAAG
ZNF217_DO6	AATGATACGGCGACCACCGAGATTGATTTTATCCAGTCCGAAG
ZNF217_DO7	AATGATACGGCGACCACCGAGATCGAACTGCGTTGGCCCAAAG
ZNF217_DO8	AATGATACGGCGACCACCGAGATATCCAAACAATACAGAAAAG
ZNF217_DO9	AATGATACGGCGACCACCGAGATTTTGTGAAACCTCCCTGAAG
ZNF217_DO10	AATGATACGGCGACCACCGAGATAATACTGACAGCAATTCAAG
ZNF217_DO11	AATGATACGGCGACCACCGAGATTTTGATGTGGCTCAAAGAAG
ZNF217_DO12	AATGATACGGCGACCACCGAGATAACTTGCTCAAATAGTTAAG
ZNF217_DO13	AATGATACGGCGACCACCGAGATATGCACAAGGCAGCCCAAAG
ZNF217_DO14	AATGATACGGCGACCACCGAGATGTACTCACACAACTGATAAG
ZNF217_DO15	AATGATACGGCGACCACCGAGATAAGGTCACGATAGCATCAAG
ZNF217_DO16	AATGATACGGCGACCACCGAGATGAAAAAAAAAAAAAAGGAAG
ZNF217_DO17	AATGATACGGCGACCACCGAGATGATGAGCTGGGGAGAGAAAG
ZNF217_DO18	AATGATACGGCGACCACCGAGATGCACTATTGATATTCCAAAG
CA12_DO1	AATGATACGGCGACCACCGAGATCCTTCACTATTCTTCTCAAG
CA12_DO2	AATGATACGGCGACCACCGAGATTTTGGGAAGGGTCTGTAAAG
CA12_DO3	AATGATACGGCGACCACCGAGATGCACTTAAAACCTAGGCAAG
CA12_DO4	AATGATACGGCGACCACCGAGATCACGTGGTGAGTTTTAAAAG
CA12_DO5	AATGATACGGCGACCACCGAGATTGCACATTCTTGTCTGAAAG
CA12_DO6	AATGATACGGCGACCACCGAGATAAAAAAAAAATTTCCCTAAG
CA12_DO7	AATGATACGGCGACCACCGAGATGACCTTGACCGTGATGCAAG
CA12_DO8	AATGATACGGCGACCACCGAGATCAGCAGACTCTCTGAAGAAG
CA12_DO9	AATGATACGGCGACCACCGAGATAGAGCAGTTATTTTAGAAAG
CA12_DO10	AATGATACGGCGACCACCGAGATGTAGATATGTGTATATGAAG
CA12_DO11	AATGATACGGCGACCACCGAGATCACAAATTAAAACAATGAAG
CA12_DO12	AATGATACGGCGACCACCGAGATTTATTGAAGACTCCAAGAAG
CA12_DO13	AATGATACGGCGACCACCGAGATACTACCTAGCTCTGTCAAAG
CA12_DO14	AATGATACGGCGACCACCGAGATTTGATATTTCCTTTATAAAG
CA12_DO15	AATGATACGGCGACCACCGAGATAATATTTCAAAATAGCCAAG

Table 4.10.2.3: 3D-DSL (AF = acceptor forward)

3D-DSL (AF = acceptor forward)	
Name	**Sequence 5' to 3'**
FOXC1_AF1	/5Phos/CTTTTGGCTGATAGGAAGCACCTGTGGTCGTAGCATCAGC
FOXC1_AF2	/5Phos/CTTTAACCTGTTCCCCAAAACCTGTGGTCGTAGCATCAGC
FOXC1_AF3	/5Phos/CTTCTTAGCAACTTGCTTTACCTGTGGTCGTAGCATCAGC
FOXC1_AF4	/5Phos/CTTTATCTTCTTCCTGTCAACCTGTGGTCGTAGCATCAGC
FOXC1_AF5	/5Phos/CTTGTGTTGGGCAAGGGAGGCCTGTGGTCGTAGCATCAGC
FOXC1_AF6	/5Phos/CTTCTTCCCATGCAGCCATACCTGTGGTCGTAGCATCAGC
FOXC1_AF7	/5Phos/CTTCAGAGGTCTCTGTTCTTCCTGTGGTCGTAGCATCAGC
FOXC1_AF8	/5Phos/CTTTAGGAGAGTAAGTATTTCCTGTGGTCGTAGCATCAGC
FOXC1_AF9	/5Phos/CTTGAGGTGGAGGGAGCAGGCCTGTGGTCGTAGCATCAGC
FOXC1_AF10	/5Phos/CTTTGTTTAGAAGCAAAAGGCCTGTGGTCGTAGCATCAGC
FOXC1_AF11	/5Phos/CTTATGTTCCAATCAACTTTCCTGTGGTCGTAGCATCAGC
FOXC1_AF12	/5Phos/CTTTCACAGCAACTGGGCAGCCTGTGGTCGTAGCATCAGC
GREB1_AF1	/5Phos/CTTCCCTGTGGTTGGGTGTCCCTGTGGTCGTAGCATCAGC
GREB1_AF2	/5Phos/CTTGATGAAATAAACACTTCCCTGTGGTCGTAGCATCAGC
GREB1_AF3	/5Phos/CTTCCAGACAGATAAAACAGCCTGTGGTCGTAGCATCAGC
GREB1_AF4	/5Phos/CTTATAATACCTAACTATATCCTGTGGTCGTAGCATCAGC
GREB1_AF5	/5Phos/CTTCTTGCTTAGCATCATGTCCTGTGGTCGTAGCATCAGC
GREB1_AF6	/5Phos/CTTCCTTTCTTTTTCCTACACCTGTGGTCGTAGCATCAGC
GREB1_AF7	/5Phos/CTTCCCCCACCCCCTCACATCCTGTGGTCGTAGCATCAGC
GREB1_AF8	/5Phos/CTTTTAATGAGTTAATAAACCCTGTGGTCGTAGCATCAGC
GREB1_AF9	/5Phos/CTTGCCAGGCCAGCTTCCAGCCTGTGGTCGTAGCATCAGC
GREB1_AF10	/5Phos/CTTGAATGATTAGTGATGAGCCTGTGGTCGTAGCATCAGC
GREB1_AF11	/5Phos/CTTCCAATCACCGTGGAAGGCCTGTGGTCGTAGCATCAGC
GREB1_AF12	/5Phos/CTTGCCAAGACCAACTACGTCCTGTGGTCGTAGCATCAGC
GREB1_AF13	/5Phos/CTTATCAGTGACCTACATGTCCTGTGGTCGTAGCATCAGC
GREB1_AF14	/5Phos/CTTCGACTTCCACATGCAAACCTGTGGTCGTAGCATCAGC
GREB1_AF15	/5Phos/CTTTGGGGTATCTGCCTCTCCCTGTGGTCGTAGCATCAGC
GREB1_AF16	/5Phos/CTTTATATATAAACATTTTACCTGTGGTCGTAGCATCAGC
GREB1_AF17	/5Phos/CTTTAGGTTGCAAAAATCTACCTGTGGTCGTAGCATCAGC
GREB1_AF18	/5Phos/CTTTTCTTTAATAAACTTTTCCTGTGGTCGTAGCATCAGC
GREB1_AF19	/5Phos/CTTCCTTTGACGTTGACTTTCCTGTGGTCGTAGCATCAGC
GREB1_AF20	/5Phos/CTTACTAAAGGTTTTTAGAACCTGTGGTCGTAGCATCAGC
GREB1_AF21	/5Phos/CTTGGGGAGGCTTGGGAACTCCTGTGGTCGTAGCATCAGC
GREB1_AF22	/5Phos/CTTTTTCCTGGTGAACGGCCCCTGTGGTCGTAGCATCAGC
GREB1_AF23	/5Phos/CTTAAAAAAATGGACTTAATCCTGTGGTCGTAGCATCAGC
GREB1_AF24	/5Phos/CTTGATTTCCTAGGGTCTTTCCTGTGGTCGTAGCATCAGC
GREB1_AF25	/5Phos/CTTGACGGAATTTCATGATTCCTGTGGTCGTAGCATCAGC
GREB1_AF26	/5Phos/CTTCATATTGGTAAGAATTGCCTGTGGTCGTAGCATCAGC
GREB1_AF27	/5Phos/CTTCTGCCCTGCACTTATTCCCTGTGGTCGTAGCATCAGC
GREB1_AF28	/5Phos/CTTGAGTGACTGAGCAGCAGCCTGTGGTCGTAGCATCAGC

Table 4.10.2.3: 3D-DSL (AF = acceptor forward), continued

3D-DSL (AF = acceptor forward)	
Name	**Sequence 5' to 3'**
GREB1_AF29	/5Phos/CTTAAAATCTTTTGTTTTGCCCTGTGGTCGTAGCATCAGC
GREB1_AF30	/5Phos/CTTCCTATCGGTCCAGCCTGCCTGTGGTCGTAGCATCAGC
GREB1_AF31	/5Phos/CTTAATGGAGTTCCTGGTGTCCTGTGGTCGTAGCATCAGC
GREB1_AF32	/5Phos/CTTTGTCTTCTCTAGTGATCCCTGTGGTCGTAGCATCAGC
GREB1_AF33	/5Phos/CTTATGCTTTTCAGCTTCTACCTGTGGTCGTAGCATCAGC
GREB1_AF34	/5Phos/CTTGGTTGTCATTAAGGCTTCCTGTGGTCGTAGCATCAGC
GREB1_AF35	/5Phos/CTTCACAGTTTCACAGGAGCCCTGTGGTCGTAGCATCAGC
GREB1_AF36	/5Phos/CTTAAGACTTAAAATTAAGTCCTGTGGTCGTAGCATCAGC
GREB1_AF37	/5Phos/CTTCCCTTTCCATTGGATTTCCTGTGGTCGTAGCATCAGC
NRIP1_AF1	/5Phos/CTTTCCATTGCAGCGGTGCTCCTGTGGTCGTAGCATCAGC
NRIP1_AF2	/5Phos/CTTTCTTGTATCACTTCCTCCCTGTGGTCGTAGCATCAGC
NRIP1_AF3	/5Phos/CTTCCAGGGTCTACCATTTACCTGTGGTCGTAGCATCAGC
NRIP1_AF4	/5Phos/CTTTAAGATATTTTTGATGTCCTGTGGTCGTAGCATCAGC
NRIP1_AF5	/5Phos/CTTGCAAAAACATTTTCTCCCCTGTGGTCGTAGCATCAGC
NRIP1_AF6	/5Phos/CTTGCTTTGAGCAGTTAACGCCTGTGGTCGTAGCATCAGC
NRIP1_AF7	/5Phos/CTTTTGACCCATCTGAAACTCCTGTGGTCGTAGCATCAGC
NRIP1_AF8	/5Phos/CTTCCCTCAAGTAGCAGGAGCCTGTGGTCGTAGCATCAGC
NRIP1_AF9	/5Phos/CTTATGCCGATTATTAGTGTCCTGTGGTCGTAGCATCAGC
NRIP1_AF10	/5Phos/CTTCTAGATAAGGCAGTGAGCCTGTGGTCGTAGCATCAGC
NRIP1_AF11	/5Phos/CTTTCTGTGCAGAAGGGATACCTGTGGTCGTAGCATCAGC
NRIP1_AF12	/5Phos/CTTAGATAGCCTCCATGAAGCCTGTGGTCGTAGCATCAGC
NRIP1_AF13	/5Phos/CTTTTGATAAGCCAGGGACTCCTGTGGTCGTAGCATCAGC
NRIP1_AF14	/5Phos/CTTTGCCTGGCTTTTCAGGGCCTGTGGTCGTAGCATCAGC
NRIP1_AF15	/5Phos/CTTAGGTTAGGGTCACCAGTCCTGTGGTCGTAGCATCAGC
NRIP1_AF16	/5Phos/CTTAACAATTTCTACCTAAACCTGTGGTCGTAGCATCAGC
NRIP1_AF17	/5Phos/CTTAATATCACCGCTTTATACCTGTGGTCGTAGCATCAGC
NRIP1_AF18	/5Phos/CTTCCTCCATTTTAACCAAGCCTGTGGTCGTAGCATCAGC
NRIP1_AF19	/5Phos/CTTTCTTATAATTGACATGACCTGTGGTCGTAGCATCAGC
NRIP1_AF20	/5Phos/CTTTCACTATGACCTCAGATCCTGTGGTCGTAGCATCAGC
NRIP1_AF21	/5Phos/CTTATACTCAAAAACTGTTTCCTGTGGTCGTAGCATCAGC
NRIP1_AF22	/5Phos/CTTAGAAAGAACATGCAAGTCCTGTGGTCGTAGCATCAGC
NRIP1_AF23	/5Phos/CTTCTTTTAAGGGCCAAGAGCCTGTGGTCGTAGCATCAGC
NRIP1_AF24	/5Phos/CTTAATGGATGAAGATAACCCCTGTGGTCGTAGCATCAGC
NRIP1_AF25	/5Phos/CTTCCTTCTTTACGTTTAAACCTGTGGTCGTAGCATCAGC
NRIP1_AF26	/5Phos/CTTAAACCACAGACGTCATGCCTGTGGTCGTAGCATCAGC
NRIP1_AF27	/5Phos/CTTTCGTTTCTGCAGTAGGACCTGTGGTCGTAGCATCAGC
TFF1_AF1	/5Phos/CTTCTAGAATGTGCACGAAGCCTGTGGTCGTAGCATCAGC
TFF1_AF2	/5Phos/CTTGTTCAATGTGAGGTCTGCCTGTGGTCGTAGCATCAGC
TFF1_AF3	/5Phos/CTTGTGATAAGCTCTCACCCCCTGTGGTCGTAGCATCAGC
TFF1_AF4	/5Phos/CTTCTTCAGAAGTCACACTTCCTGTGGTCGTAGCATCAGC

Table 4.10.2.3: 3D-DSL (AF = acceptor forward), continued

3D-DSL (AF = acceptor forward)	
Name	**Sequence 5' to 3'**
TFF1_AF5	/5Phos/CTTCTGAGGCAGGCGGGTTACCTGTGGTCGTAGCATCAGC
TFF1_AF6	/5Phos/CTTTAGATAGCTAGTGATAACCTGTGGTCGTAGCATCAGC
TFF1_AF7	/5Phos/CTTTCAACTTCTTGACCAGCCCTGTGGTCGTAGCATCAGC
TFF1_AF8	/5Phos/CTTCAAGTACCTAACAGTCCCCTGTGGTCGTAGCATCAGC
TFF1_AF9	/5Phos/CTTCTAGAATTAGCAAAAAACCTGTGGTCGTAGCATCAGC
TFF1_AF10	/5Phos/CTTGTCCAACCCACATGAGGCCTGTGGTCGTAGCATCAGC
ZNF217_AF1	/5Phos/CTTCTATTTAAATTACATTTCCTGTGGTCGTAGCATCAGC
ZNF217_AF2	/5Phos/CTTTCTTGAGAACATACAATCCTGTGGTCGTAGCATCAGC
ZNF217_AF3	/5Phos/CTTCTAGCAGCTTCTGCAAGCCTGTGGTCGTAGCATCAGC
ZNF217_AF4	/5Phos/CTTACTGAAAAGTACTTCACCCTGTGGTCGTAGCATCAGC
ZNF217_AF5	/5Phos/CTTCAGCATGCCATCAATATCCTGTGGTCGTAGCATCAGC
ZNF217_AF6	/5Phos/CTTCTTCACCGCTTCCAGCGCCTGTGGTCGTAGCATCAGC
ZNF217_AF7	/5Phos/CTTGGAATTGTGCCTCCTTACCTGTGGTCGTAGCATCAGC
ZNF217_AF8	/5Phos/CTTAGCCCTGCCTTCATTTCCCTGTGGTCGTAGCATCAGC
ZNF217_AF9	/5Phos/CTTCTCACTTCCGAATGAGACCTGTGGTCGTAGCATCAGC
ZNF217_AF10	/5Phos/CTTGCTTCTTTTTCTTTTTTCCTGTGGTCGTAGCATCAGC
ZNF217_AF11	/5Phos/CTTTGCCGCACTCGGAGCAGCCTGTGGTCGTAGCATCAGC
ZNF217_AF12	/5Phos/CTTCTTTGTTACCAGCACTGCCTGTGGTCGTAGCATCAGC
ZNF217_AF13	/5Phos/CTTCAGGGAAAACTAGAAATCCTGTGGTCGTAGCATCAGC
ZNF217_AF14	/5Phos/CTTACCAGAAACTGAAGTTACCTGTGGTCGTAGCATCAGC
ZNF217_AF15	/5Phos/CTTTTCTAAATAAGACTGCACCTGTGGTCGTAGCATCAGC
ZNF217_AF16	/5Phos/CTTGAGGAAGGTCGAGAAGACCTGTGGTCGTAGCATCAGC
ZNF217_AF17	/5Phos/CTTAACCATGTTGCTGCCAACCTGTGGTCGTAGCATCAGC
ZNF217_AF18	/5Phos/CTTATTCCTGGGTCACCTCTCCTGTGGTCGTAGCATCAGC
ZNF217_AF19	/5Phos/CTTTCCATTTAATCTTATTGCCTGTGGTCGTAGCATCAGC
CA12_AF1	/5Phos/CTTCCTGATACTGTGTTTAGCCTGTGGTCGTAGCATCAGC
CA12_AF2	/5Phos/CTTAGACACATAAAGTTTTACCTGTGGTCGTAGCATCAGC
CA12_AF3	/5Phos/CTTTCATCCTCTGTAAAATGCCTGTGGTCGTAGCATCAGC
CA12_AF4	/5Phos/CTTCTGCACCTTGGAGCTCCCCTGTGGTCGTAGCATCAGC
CA12_AF5	/5Phos/CTTGCAATTATAAATCTTTCCCTGTGGTCGTAGCATCAGC
CA12_AF6	/5Phos/CTTGAATTATTGTTTTAACCCCTGTGGTCGTAGCATCAGC
CA12_AF7	/5Phos/CTTGGCAGCCAGGTGAGATCCCTGTGGTCGTAGCATCAGC
CA12_AF8	/5Phos/CTTTTAAAATTAAATTGAATCCTGTGGTCGTAGCATCAGC
CA12_AF9	/5Phos/CTTTTAAACTTTTTAGCCACCCTGTGGTCGTAGCATCAGC
CA12_AF10	/5Phos/CTTAACATTCAGTACTCCCTCCTGTGGTCGTAGCATCAGC
CA12_AF11	/5Phos/CTTTTTTTCATGTGGGTTTTCCTGTGGTCGTAGCATCAGC
CA12_AF12	/5Phos/CTTGCTATAATCATTAGAATCCTGTGGTCGTAGCATCAGC
CA12_AF13	/5Phos/CTTCCTACCTCTTCCCTGTCCCTGTGGTCGTAGCATCAGC
CA12_AF14	/5Phos/CTTGCTGTTATTACCCAAGCCCTGTGGTCGTAGCATCAGC
CA12_AF15	/5Phos/CTTGGAGAAGCAGCAGAATTCCTGTGGTCGTAGCATCAGC

Table 4.10.2.4: 3D-DSL (AR = acceptor reverse)

3D-DSL (AR = acceptor reverse)	
Name	**Sequence 5' to 3'**
FOXC1_AR1	/5Phos/CTTTGTGGTCAGGGACAGTGCCTGTGGTCGTAGCATCAGC
FOXC1_AR2	/5Phos/CTTTTATTTTCAGGTTACAGCCTGTGGTCGTAGCATCAGC
FOXC1_AR3	/5Phos/CTTTGCTCGCGCACGCCTATCCTGTGGTCGTAGCATCAGC
FOXC1_AR4	/5Phos/CTTTTTTCCTGTCCATTCTGCCTGTGGTCGTAGCATCAGC
FOXC1_AR5	/5Phos/CTTGGCTTGGGTCTGGACGACCTGTGGTCGTAGCATCAGC
FOXC1_AR6	/5Phos/CTTGTATTTTATTAAAGATTCCTGTGGTCGTAGCATCAGC
FOXC1_AR7	/5Phos/CTTGCTGGCCGGGCGGCCTCCCTGTGGTCGTAGCATCAGC
FOXC1_AR8	/5Phos/CTTTCCGTAAATGAGCTGACCCTGTGGTCGTAGCATCAGC
FOXC1_AR9	/5Phos/CTTTTTACAGAGGCTATTTTCCTGTGGTCGTAGCATCAGC
FOXC1_AR10	/5Phos/CTTTACCCAGGGCAGCCCCCCCTGTGGTCGTAGCATCAGC
FOXC1_AR11	/5Phos/CTTGAGAAAACTATACCTCGCCTGTGGTCGTAGCATCAGC
FOXC1_AR12	/5Phos/CTTACTGCCCAGTCCAGATGCCTGTGGTCGTAGCATCAGC
GREB1_AR1	/5Phos/CTTCAGAGAAAATATCTGCACCTGTGGTCGTAGCATCAGC
GREB1_AR2	/5Phos/CTTCCGTCTCTGAGCCATGACCTGTGGTCGTAGCATCAGC
GREB1_AR3	/5Phos/CTTTCAAGATCTTCTCTTTACCTGTGGTCGTAGCATCAGC
GREB1_AR4	/5Phos/CTTAACTGCCCTAAACCTCACCTGTGGTCGTAGCATCAGC
GREB1_AR5	/5Phos/CTTGAGAAGGGAGGGCTCCACCTGTGGTCGTAGCATCAGC
GREB1_AR6	/5Phos/CTTCTCCAAGAAGCAGGCTCCCTGTGGTCGTAGCATCAGC
GREB1_AR7	/5Phos/CTTTTTAACCTGTAAGATACCCTGTGGTCGTAGCATCAGC
GREB1_AR8	/5Phos/CTTGCCATTACCAGGGCTGTCCTGTGGTCGTAGCATCAGC
GREB1_AR9	/5Phos/CTTCGCAGCCAGGGCGACTTCCTGTGGTCGTAGCATCAGC
GREB1_AR10	/5Phos/CTTGCCCTCGACCAACAGCCCCTGTGGTCGTAGCATCAGC
GREB1_AR11	/5Phos/CTTCCTGAAGCCTTCACCAGCCTGTGGTCGTAGCATCAGC
GREB1_AR12	/5Phos/CTTCATTTTATAGATGAGACCCTGTGGTCGTAGCATCAGC
GREB1_AR13	/5Phos/CTTGAAAAGAAAAGTTTAGGCCTGTGGTCGTAGCATCAGC
GREB1_AR14	/5Phos/CTTACAGAAGCCTGTGGTTCCCTGTGGTCGTAGCATCAGC
GREB1_AR15	/5Phos/CTTGTCTAGAATAGCCCGTTCCTGTGGTCGTAGCATCAGC
GREB1_AR16	/5Phos/CTTGGAATGCAGAATTGGCCCCTGTGGTCGTAGCATCAGC
GREB1_AR17	/5Phos/CTTTTTCTTTAAAAAATATTCCTGTGGTCGTAGCATCAGC
GREB1_AR18	/5Phos/CTTCAGTATACAAAGTAAGTCCTGTGGTCGTAGCATCAGC
GREB1_AR19	/5Phos/CTTTTAGGTTACCCAAGCCACCTGTGGTCGTAGCATCAGC
GREB1_AR20	/5Phos/CTTTTTAGTGATTTAGCCCACCTGTGGTCGTAGCATCAGC
GREB1_AR21	/5Phos/CTTCCATGATCTCATCCCTACCTGTGGTCGTAGCATCAGC
GREB1_AR22	/5Phos/CTTCGTTCTTGTGAAGACTGCCTGTGGTCGTAGCATCAGC
GREB1_AR23	/5Phos/CTTTGCGCTTCAGCTCCAGACCTGTGGTCGTAGCATCAGC
GREB1_AR24	/5Phos/CTTTTAGCCAGGTACTCTGCCCTGTGGTCGTAGCATCAGC
GREB1_AR25	/5Phos/CTTGAATCTCTCTGTAGGCTCCTGTGGTCGTAGCATCAGC
GREB1_AR26	/5Phos/CTTGCCACCTGTGAATAAATCCTGTGGTCGTAGCATCAGC
GREB1_AR27	/5Phos/CTTCAAGGAAGATAAGGAACCCTGTGGTCGTAGCATCAGC
GREB1_AR28	/5Phos/CTTCCAGCAAGCCACCCTTCCCTGTGGTCGTAGCATCAGC

Table 4.10.2.4: 3D-DSL (AR = acceptor reverse), continued

3D-DSL (AR = acceptor reverse)	
Name	**Sequence 5' to 3'**
GREB1_AR29	/5Phos/CTTGTTGACTGTATTCCTCTCCTGTGGTCGTAGCATCAGC
GREB1_AR30	/5Phos/CTTTCACTGACGAAGAATACCCTGTGGTCGTAGCATCAGC
GREB1_AR31	/5Phos/CTTTCCTGAAGCTGGTTTTTCCTGTGGTCGTAGCATCAGC
GREB1_AR32	/5Phos/CTTTACTGGGACAAATCATGCCTGTGGTCGTAGCATCAGC
GREB1_AR33	/5Phos/CTTAAAATTGTGCCAGCTAGCCTGTGGTCGTAGCATCAGC
GREB1_AR34	/5Phos/CTTCATTAATTCTATTTTGTCCTGTGGTCGTAGCATCAGC
GREB1_AR35	/5Phos/CTTGAGTGACCATTGAAGATCCTGTGGTCGTAGCATCAGC
GREB1_AR36	/5Phos/CTTGCCTTAATTTTATTCACCCTGTGGTCGTAGCATCAGC
GREB1_AR37	/5Phos/CTTAAGACATTGACATAGATCCTGTGGTCGTAGCATCAGC
NRIP1_AR1	/5Phos/CTTCAGTTACCATCAAAGGGCCTGTGGTCGTAGCATCAGC
NRIP1_AR2	/5Phos/CTTCCTACTCTTGCACTTCTCCTGTGGTCGTAGCATCAGC
NRIP1_AR3	/5Phos/CTTTCTTATTGCTATTAAATCCTGTGGTCGTAGCATCAGC
NRIP1_AR4	/5Phos/CTTCCAGGTTATACCAACTACCTGTGGTCGTAGCATCAGC
NRIP1_AR5	/5Phos/CTTTGAAAAGTTAAATCTCTCCTGTGGTCGTAGCATCAGC
NRIP1_AR6	/5Phos/CTTGACTGACTCTGAATCATCCTGTGGTCGTAGCATCAGC
NRIP1_AR7	/5Phos/CTTGGAATTATGAGAGAGCACCTGTGGTCGTAGCATCAGC
NRIP1_AR8	/5Phos/CTTTAGGTCATTTAGAGGACCCTGTGGTCGTAGCATCAGC
NRIP1_AR9	/5Phos/CTTTTAGAGTTAACAAGATTCCTGTGGTCGTAGCATCAGC
NRIP1_AR10	/5Phos/CTTACATCTCAAAACACTTCCCTGTGGTCGTAGCATCAGC
NRIP1_AR11	/5Phos/CTTTTTCAGGATAGGAGTCTCCTGTGGTCGTAGCATCAGC
NRIP1_AR12	/5Phos/CTTTTTGTCCTGGCATTCAACCTGTGGTCGTAGCATCAGC
NRIP1_AR13	/5Phos/CTTGCCCAAGTTAAACAAAGCCTGTGGTCGTAGCATCAGC
NRIP1_AR14	/5Phos/CTTCAGTCTGAACACACTAACCTGTGGTCGTAGCATCAGC
NRIP1_AR15	/5Phos/CTTCAGTCAGGCTCAAGTGCCCTGTGGTCGTAGCATCAGC
NRIP1_AR16	/5Phos/CTTCCATATAATTGTTTTTGCCTGTGGTCGTAGCATCAGC
NRIP1_AR17	/5Phos/CTTTCTGGTGTTTTTCTCAACCTGTGGTCGTAGCATCAGC
NRIP1_AR18	/5Phos/CTTCTTCTCCTCTCAGTAGCCCTGTGGTCGTAGCATCAGC
NRIP1_AR19	/5Phos/CTTCAGGCTGTGAGTTCAATCCTGTGGTCGTAGCATCAGC
NRIP1_AR20	/5Phos/CTTTTCTTCCCTGGATTGTTCCTGTGGTCGTAGCATCAGC
NRIP1_AR21	/5Phos/CTTTCCTTACATACACTTTCCCTGTGGTCGTAGCATCAGC
NRIP1_AR22	/5Phos/CTTTGTGTATGCTGTCCCCTCCTGTGGTCGTAGCATCAGC
NRIP1_AR23	/5Phos/CTTAATACTTTTTAAATATACCTGTGGTCGTAGCATCAGC
NRIP1_AR24	/5Phos/CTTAATTTGTTGTACAGCAACCTGTGGTCGTAGCATCAGC
NRIP1_AR25	/5Phos/CTTTGCTAGCTGGCATGGTTCCTGTGGTCGTAGCATCAGC
NRIP1_AR26	/5Phos/CTTTGTGAAAGAATGTGCTTCCTGTGGTCGTAGCATCAGC
NRIP1_AR27	/5Phos/CTTCTTTCTTTAATTTAAGACCTGTGGTCGTAGCATCAGC
TFF1_AR1	/5Phos/CTTCACCATTGACTCCTGCTCCTGTGGTCGTAGCATCAGC
TFF1_AR2	/5Phos/CTTGATCTAGACCTTATCTTCCTGTGGTCGTAGCATCAGC
TFF1_AR3	/5Phos/CTTGGAATGTTTCCGTGTTGCCTGTGGTCGTAGCATCAGC
TFF1_AR4	/5Phos/CTTCCATGGGGAAATGAAGGCCTGTGGTCGTAGCATCAGC

Table 4.10.2.4: 3D-DSL (AR = acceptor reverse), continued

3D-DSL (AR = acceptor reverse)	
Name	**Sequence 5' to 3'**
TFF1_AR5	/5Phos/CTTCCAATTCTGACCTCTGTCCTGTGGTCGTAGCATCAGC
TFF1_AR6	/5Phos/CTTAGAAATACAGAACGGCCCCTGTGGTCGTAGCATCAGC
TFF1_AR7	/5Phos/CTTTTTCATACTTGTGCGCCCCTGTGGTCGTAGCATCAGC
TFF1_AR8	/5Phos/CTTGGCTAGAATTTCCCTGTCCTGTGGTCGTAGCATCAGC
TFF1_AR9	/5Phos/CTTTTCTGTTATCTGCTGTACCTGTGGTCGTAGCATCAGC
TFF1_AR10	/5Phos/CTTGGTCAAGATCACCTCATCCTGTGGTCGTAGCATCAGC
ZNF217_AR1	/5Phos/CTTGATGCTATCGTGACCTTCCTGTGGTCGTAGCATCAGC
ZNF217_AR2	/5Phos/CTTCCTTTTTTTTTTTTTTCCCTGTGGTCGTAGCATCAGC
ZNF217_AR3	/5Phos/CTTTCTCTCCCCAGCTCATCCCTGTGGTCGTAGCATCAGC
ZNF217_AR4	/5Phos/CTTGTATTTAACTAGGTTACCCTGTGGTCGTAGCATCAGC
ZNF217_AR5	/5Phos/CTTCGGACTGGATAAAATCACCTGTGGTCGTAGCATCAGC
ZNF217_AR6	/5Phos/CTTTGGGCCAACGCAGTTCGCCTGTGGTCGTAGCATCAGC
ZNF217_AR7	/5Phos/CTTTTCTGTATTGTTTGGATCCTGTGGTCGTAGCATCAGC
ZNF217_AR8	/5Phos/CTTCAGGGAGGTTTCACAAACCTGTGGTCGTAGCATCAGC
ZNF217_AR9	/5Phos/CTTGAATTGCTGTCAGTATTCCTGTGGTCGTAGCATCAGC
ZNF217_AR10	/5Phos/CTTCTTTGAGCCACATCAAACCTGTGGTCGTAGCATCAGC
ZNF217_AR11	/5Phos/CTTTCAGAACCTACCACCAGCCTGTGGTCGTAGCATCAGC
ZNF217_AR12	/5Phos/CTTTGGAATATCAATAGTGCCCTGTGGTCGTAGCATCAGC
ZNF217_AR13	/5Phos/CTTTGGGCTGCCTTGTGCATCCTGTGGTCGTAGCATCAGC
ZNF217_AR14	/5Phos/CTTATCAGTTGTGTGAGTACCCTGTGGTCGTAGCATCAGC
ZNF217_AR15	/5Phos/CTTGAAGAAAATCTTTGGTGCCTGTGGTCGTAGCATCAGC
ZNF217_AR16	/5Phos/CTTAACTATTTGAGCAAGTTCCTGTGGTCGTAGCATCAGC
ZNF217_AR17	/5Phos/CTTTTACATTTTTGGTATGACCTGTGGTCGTAGCATCAGC
ZNF217_AR18	/5Phos/CTTGGGGCTCACGGGCAGATCCTGTGGTCGTAGCATCAGC
CA12_AR1	/5Phos/CTTTATAAAGGAAATATCAACCTGTGGTCGTAGCATCAGC
CA12_AR2	/5Phos/CTTAGGGAAATTTTTTTTTTCCTGTGGTCGTAGCATCAGC
CA12_AR3	/5Phos/CTTGCATCACGGTCAAGGTCCCTGTGGTCGTAGCATCAGC
CA12_AR4	/5Phos/CTTCTTCAGAGAGTCTGCTGCCTGTGGTCGTAGCATCAGC
CA12_AR5	/5Phos/CTTTCTAAAATAACTGCTCTCCTGTGGTCGTAGCATCAGC
CA12_AR6	/5Phos/CTTCATTGTTTTAATTTGTGCCTGTGGTCGTAGCATCAGC
CA12_AR7	/5Phos/CTTGGCTATTTTGAAATATTCCTGTGGTCGTAGCATCAGC
CA12_AR8	/5Phos/CTTCATATACACATATCTACCCTGTGGTCGTAGCATCAGC
CA12_AR9	/5Phos/CTTGCCTAGGTTTTAAGTGCCCTGTGGTCGTAGCATCAGC
CA12_AR10	/5Phos/CTTTGACAGAGCTAGGTAGTCCTGTGGTCGTAGCATCAGC
CA12_AR11	/5Phos/CTTCTTGGAGTCTTCAATAACCTGTGGTCGTAGCATCAGC
CA12_AR12	/5Phos/CTTTACAGACCCTTCCCAAACCTGTGGTCGTAGCATCAGC
CA12_AR13	/5Phos/CTTGAGAAGAATAGTGAAGGCCTGTGGTCGTAGCATCAGC
CA12_AR14	/5Phos/CTTTTAAAACTCACCACGTGCCTGTGGTCGTAGCATCAGC
CA12_AR15	/5Phos/CTTTCAGACAAGAATGTGCACCTGTGGTCGTAGCATCAGC

4.10.3 3D-DSL sequencing

3C was performed as per Lieberman-Aiden et al., (7). The DSL ligation products were prepared as described in Kwon et al., (8). 3D-DSL was performed as described in Harismendy et al. (9). Briefly, equal amount of 3C chromatin was biotinylated using the Photoprobe Kit (Vector Lab). Donor and acceptor probe pools (2.5fmol per probe) were annealed to the biotinylated 3C samples at 45°C for 2 hours followed by 10min 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 5′-phosphate of acceptor probes and 3′-OH of donor probes were ligated using Taq DNA ligase at 45°C for 1h. 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.

4.11 RNA pulldown & mass spectrometric analysis

Genomic sites for eRNA transcription for CA12e, TFF1e, FOXC1e and GREB1e3 were cloned into pBluescript II KS(-) using XbaI and KpnI sites. *In vitro* transcription (IVT) reactions were carried out with the Biotin RNA Labeling Mix (Roche) and the Riboprobe® Combination System—T3/T7 RNA Polymerase (Promega). The IVT-eRNAs were purified with RNeasy® Mini Kit (QIAGEN), and RNA electrophoresis was performed using NorthernMax® kit according to

manufacturer's instructions (Ambion) to demonstrate that all the RNAs are transcribed at the right size. 50 µg of each IVT-eRNA were mixed and incubated with MCF7 nuclear extract. To isolate biotinylated RNA bound proteins, 60 µl washed BcMag™ Monomer Avidin Magnetic beads (Bioclone) were added to each binding reaction and further incubated at room temperature for 1 hour. Beads were washed one time with RNA structure buffer (10), two times with PNK buffer (20 mM Tris-HCl, pH 7.4; 10 mM MgCl2; 0.2% Tween-20), four times with high-salt wash buffer (50 mM Tris-HCl, pH 7.4; 1 M NaCl; 1 mM EDTA; 1% NP-40; 0.5% sodium deoxycholate; 0.1% SDS), two times with PBS and bound RNA-protein complexes were eluted by 2 mM biotin in PBS. The eluted material was submitted for MS analysis at the UCSD Biomolecular and Proteomics Mass Spectrometry Facility.

4.12 References

1) Perissi, V., A. Aggarwal, et al. (2004). "A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors." Cell 116(4): 511-526.

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

3) Rueden, C. T. and K. W. Eliceiri (2007). "Visualization approaches for multidimensional biological image data." Biotechniques 43(1 Suppl): 31, 33-36.

4) Garcia-Bassets, I., Y. S. Kwon, et al. (2007). "Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors." Cell 128(3): 505-518.

5) Heinz, S., C. Benner, 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.

6) Wang, D., I. Garcia-Bassets, et al. (2011). "Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA." Nature 474(7351): 390-394.

7) Lieberman-Aiden, E., N. L. van Berkum, et al. (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." Science 326(5950): 289-293.

8) Kwon, Y. S., I. Garcia-Bassets, et al. (2007). "Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters." Proc Natl Acad Sci U S A 104(12): 4852-4857.

9) Harismendy, O., D. Notani, et al. (2011). "9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response." Nature 470(7333): 264-268.

10) Tsai, M. C., O. Manor, et al. (2010). "Long noncoding RNA as modular scaffold of histone modification complexes." Science 329(5992): 689-693.

5 Results

5.1 The ERα binding program of MCF-7 cells is mainly distal and poised for transcription at enhancer sites

We initiated our investigation by performing experiments designed to determine whether liganded estrogen receptor induces eRNA transcription on ERα-bound enhancers. First a "deep" ChIP-seq analysis of ERα binding sites using vehicle and E2-treated MCF7 BCCs was completed to complement existing analyses (1-5). Even when sequencing depth was increased to $>100x10^6$ uniquely mapped reads, and the number of binding sites augmented by 3-fold, to 32,923 genome-wide; ERα was found to preferentially bind distal intergenic (63%) and intronic sites (23%), with only 1% being bound to TSSs at promoters (Fig 5.1SA). Such distal binding program raises the question of how ERα is able to exert its regulatory functions without direct binding to its target genes, and with a robust almost immediate genic and non-genic transcriptional response.

In order to define enhancer sites under the control of ERα, we also performed a "deep" ChIP-seq analysis of H3K4me1, a histone modification known to mark enhancers (6), along with H3K27ac, a mark reported to identify active enhancers (7). The final cohort of enhancers was defined as the non-redundant overlap of these three datasets, which yielded 4,727 sites (Fig 5.1SB). We then queried the transcriptional output after 1hour E2-treatment compared to vehicle by global run on sequencing [Gro-seq] (8) to provide a genome-wide catalogue of transcription units regulated by ERα. Analyses of these data uncovered uncharacterized correlations between genomic and enhancer transcription, we found that 1,168 upregulated coding genes exhibited an

ERα-dependent E2 induction of eRNAs in adjacent enhancers, while only 145 of the ERα-up-regulated coding genes exhibiting ERα binding to their promoters (Fig 5.1.1A). These data are most consistent with the initial suggestions from ChIP-chip (1-5), that ERα occupancy of enhancers is likely to be the key strategy underlying estrogen-induced gene expression.

We noted that there was often more than one ERα-bound enhancer adjacent to up-regulated coding genes, in contrast to the AR-regulated genes (37), raising the possibility that for many estrogen-regulated coding genes, more than one enhancer might be involved in upregulation events and might even regulate each other. E_2-regulated enhancers generally displayed a basal expression of bidirectional eRNAs and, as described below, those in proximity to up-regulated coding genes displayed a characteristic bidirectional activation of eRNAs, in general agreement with recent findings by the Kraus lab (9). The transcripts vary in apparent length from ~1.5 kb to several potentially long transcripts, although ~10% exhibited a marked predominance of unidirectional eRNA transcripts (Fig. 5.1SC, D). Analysis of the GRO-seq data confirmed the overall up-regulation of eRNAs in response to ligand, generally with bidirectional transcription, robust at 1h after E_2, and subsequently diminishing (Fig. 5.1SE, F).

Indeed, ~96% of all regulated enhancers exhibited the presence of detectable eRNAs. Of the 1,168 enhancers adjacent, at any distance, to ERα up-regulated coding genes, ~88% showed a corresponding up-regulation of their eRNAs (Fig. 5.1.1E); actually, >93% of those located <200kb from their upregulated target coding gene promoter exhibited E_2-dependent up-regulated eRNA transcription. The median

distance between enhancers exhibiting E_2-dependent up-regulation of their eRNAs and their closest up-regulated coding gene was ~70kb, with many <215 kb from the coding gene cap site, compared with a median distance of >400kb for enhancers exhibiting ligand-insensitive enhancer eRNAs with corresponding non-responsive coding genes (Fig. 5.1.1F). Intriguingly, examining the strength of ERα binding based on normalized ChIP-seq data on these cohorts of enhancers with upregulation of the eRNAs exhibited significantly stronger binding than the enhancers no eRNA upregulation (Fig. 5.1.1G). Based on these genome wide GRO-seq data analyses, we selected eight robustly up-regulated transcription units, each associated with enhancers exhibiting clearly increased eRNAs- *TFF1, CA12, FOXC1, GREB1, P2RY2, SMAD7, PGR, SIAH2, NRIP1,* and *KCNK5* for further experimentation. Based on these GRO-seq data, for example, there was a 4.6-11-fold increase in coding gene expression, with a corresponding 3.5-10-fold increase in eRNA expression on associated enhancers, assessed 1hr. following addition of E_2 to MCF7 cell cultures (Fig. 5.1.1B-D; Fig. 5.2.1SA-D, 5.2.1SE, F).

Even though there is increasing evidence in the literature that ncRNAs are developmentally regulated (10-12), exhibit cell type specific expression (13, 14), localize to specific subnuclear compartments (15-19), and are associated with human diseases (20-22), the question still remains whether eRNA transcripts merely reflect polymerase entry into the relaxed “open” chromatin state induced at enhancers, or whether they actively serve as key regulators of biological processes.

5.2 eRNAs are required for a proper and complete response to hormone treatment

In order to investigate the potential roles, if any, of ligand-induced eRNAs on gene activation events, two different technologies were employed to down-regulate eRNAs: First, specific siRNAs directed at several regions of each transcript were used to assess possible effects on gene expression, an approach licensed by recent evidence of nuclear effects of the RNAi regulatory machinery (38). Similarly, we used locked nucleic acid antisense oligos (ASO-LNAs) (23, 24) that were designed and synthesized by Exiqon, placing the LNA-modified bases at key positions to ensure target specificity as well as stability of the oligos, with a scrambled control LNA. For our experiments, the LNAs were designed with complete phosphorothioate backbonds to trigger RNAse H cleavage of the targeted sequences (39). For siRNAs, cells were cultured for 3 days after transfection and then treated with either vehicle or E2 for 1hr; for LNAs cells were cultured for either 6 or 24hr after transfection and exposed to the same treatment as described above. For all transcription units examined, experiments were performed with two different LNAs or siRNAs to exclude any off-target effects.

Both siRNA knockdown or LNA treatment of the *TFF1*, *FOXC1*, *CA12*, and *NRIP1* enhancers revealed that, for each transcription unit, the induction of both the eRNA and the adjacent coding gene transcript, as assessed by QPCR and GRO-seq (Fig 5.1.2H, I), respectively, was severely inhibited or fully abolished. Consistent with the observation that, even after 3 days in stripped-serum medium, MCF7 cells still exhibit some ERα dependent basal activation of a significant cohort of coding gene targets, both siRNAs and LNAs against enhancer eRNAs caused decrease in basal, as

well as E2-stimulated gene expression. In these experiments similar results were observed with either of the two LNAs or siRNAs designed for the targeted enhancer. Similar results were observed with knockdown of the eRNAs for *GREB1*, *SIAH2*, *P2RY2*, *SMAD7*, and *KCNK5* (Fig 5.2.2SG). As controls, effects of LNAs (or siRNAs) were evaluated on housekeeping genes, showing no inhibitory effects (Fig 5.3SA-D). These findings suggest that eRNA transcripts per se might be required for proper transcriptional stimulation.

FOXC1 has only one characterized enhancer and induced eRNA, and in this case, LNA-dependent knock-down of the eRNA caused a block of E_2-induced coding target gene transcription (Fig 5.3.1A). For *NRIP1*, while there are two potential enhancers the major, H3K4me1 marked unambiguous enhancer was selected as the target for siRNA. When this enhancer, or both potential enhancers simultaneously, each exhibiting induced eRNAs in response to E_2 were subjected to knockdown by siRNA or LNA, the effect was to cause a full inhibition of E_2-induced coding gene transcription (Fig 5.3.2D). Because similar knockdown results were obtained with LNA and siRNAs, and effective ERα binding persisted even after eRNA knockdown (Fig 5.3SE, F), we conclude that the induced eRNAs are important for induction of the coding gene target (Fig 5.3.1A, Fig 5.3.2D).

Once the efficacy of the siRNA/LNA knockdowns was verified, they were used for additional functional assays aiming to explore the potential role of eRNA transcripts facilitating relocation events between nuclear structures. ASO-LNA transfections were used to deplete eRNA transcripts for two genes, *FOXC1* and *NRIP1* respectively, in both no treatment and treatment conditions along with mock-scrambled LNA. In this

case, immunoFISH data from each locus individually showed a clearly discernible inhibition of the move, directly implicating the eRNA transcripts in the relocation process, most likely by preventing the activation of the associated genes (Fig 5.3.1B, C; Fig 5.3.2E, F; Fig 5.6SA, B). Our data have revealed a requirement of eRNA for target coding gene stimulation, correlated with relocation of these target genes between subnuclear architectural structures.

5.3 eRNAs facilitate relocation of target genes between nuclear substructures

We recently observed that the methylation status of Pc2 present on growth control gene regulatory regions determines its association with two abundant ncRNAs, TUG1, and NEAT2, located primarily in distinct subnuclear individual structures (Polycomb bodies and interchromatin granules, respectively) and that this "switch" in methylation is required for control of cancer cell proliferation (25). Although increasing evidence in the literature supports functional roles of ncRNAs in maintenance of cellular homeostasis (10-19), the question whether eRNA transcripts actively serve as key regulators of biological processes has remained a central, unresolved question.

To initiate investigation of this question, we first performed immunoFISH experiments to investigate the possibility that E2-regulated coding transcription units might also display alternative interactions with nuclear structures under basal conditions vs. in response to ligand, as a potential regulatory event that might require the actions of eRNAs. Using antibodies characterized for immunofluorescence marking representative

components of Polycomb (PcG) bodies (Bmi/Ring1a), PML bodies (PML), Cajal bodies (p80, coilin) and interchromatin granules (ICGs, SC35), and specific BAC probes for each genomic locus, immunoFISH analysis revealed that for all eight E2 up-regulated transcription units evaluated, there was a marked, reproducible switch in the predominant location of the transcription units from PcG bodies to ICGs. (Fig 5.2A-C; Fig 5.4.1S; 5.4.2S; Fig 5.5SA-C). Three dimensional images were produced by taking Z-stacks and performing deconvolution for each channel individually using Volocity (PerkinElmer), quantitating colocalization based on the Pearson's correlation coefficient for the ImmunoFISH signals. Statistical analyses were performed with three independent experiments using a two-tailed student's t test or analysis of variance (ANOVA) Holm-Sidak test when appropriate. These data therefore suggest an altered location from PcG bodies to ICGs of ERα target genes in response to ligand.

5.4 eRNAs mediate enhancer:promoter looping as part of the E_2-activation process

To investigate whether eRNAs mediate enhancer:promoter looping as part of the E_2-activation process, we employed an open-ended (3D-DSL) approach (27), conceptually quite analogous to 5C (28) for studying the spatial organization of genomes referred to as 3D-DSL. In this method, oligonucleotides corresponding to genomic sites that are to be analyzed for participation in a network contain a 5'-phosphate (referred to as acceptors), while oligonucleotides corresponding to genomic sites of potential interaction have a 5'-OH (referred to as donors). Equimolar amounts

of donor and acceptor probe pools are then annealed to complimentary sites of biotinylated 3C libraries for the interval to be interrogated, allowing for subsequent precipitation with streptavidin-conjugated magnetic beads. Interacting fragments are ligated using T4 DNA ligase, and the resulting amplicons are PCR amplified and coupled to adapters compatible with the HiSeq 2000 sequencing platform (Illumina) (27). This method maximizes the signal to noise ratio of 3C assay to an extent permitting interrogation of both short (<10 kb) as well as long-distance genomic interactions with specific genomic regions including promoters and enhancers. Therefore, "donor" pools of oligonucleotides spanning ~200 kb flanking the promoter of four up-regulated ERα target genes, as well as the housekeeping gene *GAPDH*, which was used as a control, were therefore designed based on the HindIII restriction site. The "acceptor" pool constituted all ERα binding sites and promoters in the interval.

In the case of the *P2RY2* transcription unit, E_2 caused an increase in the promoter:enhancer interaction compared to that observed in the control cultures, but there was also an increase in an enhancer interaction between an ERα-bound intragenic enhancer and a second transcription transcription unit (*P2RY6*), which is actually down-regulated in response to E_2 (Fig 5.7SA). Similarly, for the *KCNK5* gene locus boosted promoter:enhancer interactions were observed in response to E_2, but here there were additional interactions of the *KCNK5* promoter with a second enhancer located 3' to the termination site of the coding transcription unit, and that also exhibited induced interactions with a second ERα binding site located adjacent to the *C6ofr64* gene

promoter, which is also upregulated upon estrogen treatment (Fig 5.7SB). These observations indicate that the major effect of ligand is to enhance strength of specific promoter:enhancer interactions in parallel to induction of eRNA, but for some loci new enhancer:promoter interactions are actually established and additional interactions are also observed, further provoking the question if the induced eRNAs exert any roles in the dynamic regulation of short-range and long-range induced interactions.

We therefore investigated whether E_2-induced enhancer:promoter interactions would be inhibited by loss of specific eRNAs. In the case of the *GREB1* locus, the specific enhancer:promoter interaction induced by E_2 was markedly diminished, as was another enhancer interaction with the promoter of *NTSR2* in cells treated with enhancer-specific siRNA (Fig 5.4.1A). Another enhancer in the vicinity of *GREB1* (labeled in purple), preceding the *E2F6* gene, also interacted with another more distal enhancer preceding the *GREB1* locus, an interaction that was not significantly altered in response to knockdown against the major *GREB1* enhancer. For the *NRIP1* promoter, treatment with LNA against *eRNA,* caused a marked inhibition of the enhancer:promoter interaction in E_2-treated MCF7 cells, both by conventional 3C assay (Fig 5.7C), and 3D-DSL (Fig 5.4.1B), as well as inhibiting a second promoter interaction with another ERα-bound enhancer downstream of the *NRIP1* terminator region. The basic conclusion from these experiments is that estrogen causes quantitative, as well as some qualitative, alterations in the interactions between enhancers and promoters, and even between enhancers and other ERα-bound regions, which are highly diminished or abolished with down-regulation of the targeted eRNAs. In these gene targets, the eRNA was of functional importance for robust enhancer:promoter interactions, with knock-

down by either siRNA or LNA treatment invariantly diminishing, or even abolishing, the putative activating enhancer:promoter interactions.

In order to provide an independent documentation of ligand-induced gene relocation events histologically, it is favorable to investigate an interaction between an enhancer and promoter that spans a distance of >250kb, permitting FISH analysis. We could locate only one such known interaction suitable for study: interaction between *P2RY2* and *STARD10* at a distance of ~428 kb (32). As shown in Fig 5.4.1C, knockdown of the *P2RY2* enhancer by specific siRNAs (Fig 5.4.1D) caused downregulation of two putative target coding genes, *P2RY2* and *STAR10* which are upregulated following E_2 treatment as per GRO-seq analysis (Fig 5.4.1E). These two genomic loci exhibited E2-dependent colocalization, and upon induction relocated to SC35-positive interchromatin granules (Fig 5.4.2F). When cells transfected with siRNAs against *P2RY2e* and treated with vehicle and hormone, respectively for 1 hr, were fixed and prepared for immunoFISH as described in chapter 4, the *P2RY2e* knockdown not only blocked the previously observed *P2RY2:STARD10* co-localization but also their relocation to interchromatin granules (Fig 5.4.2G, H). Based on these data it is tempting to speculate that ligand-dependent induction of eRNAs initiates a two-step process, first involving enhancer:promoter looping and then relocation of this loop structure to the interchromatin granule, with this relocation actually proving to be required for the resultant coding gene activation events.

5.5 The ERα-guided relocation of target genes necessitates nuclear motor machinery

To explore to the intriguing question whether the induced relocation events to distinct subnuclear structures were actually required for regulated gene induction, we took advantage of previous observations indicating the putative roles of nuclear motors, including nuclear myosin I (NMI), in recruitment of gene loci to interchromatin granules (35, 36). First, we assessed by GRO-seq whether knockdown of NMI by siRNA caused loss of the regulatory eRNAs, finding that their levels after E2 treatment were equivalent to those in control siRNA-treated cells; however, NMI siRNA treatment greatly inhibited induction of the target genes, while transcription of constitutive transcription units did not appear to be affected (Fig 5.8SA-D). Therefore we employed single cell nuclear microinjection of IgG (control) and NMI-specific antibodies in MCF7 cells to assess the effects of E_2 on induced relocation of the *NRIP1* and *FOXC1* loci to interchromatin granules. As shown in Fig 5.5B and Fig 5.5D, this caused a loss of E_2-induced "relocation" of the *FOXC1* and *NRIP1* transcription units. To ascertain that this requires the motor function of NMI, we performed rescue experiments. As seen in Fig 5.5A, C and Fig 5.5D, F wild-type NMI was capable of complete rescue of the relocation of both loci in response to E2; however, a point mutant NMI (S397L), defective in ATPase function (35), was unable to rescue. We also show that the housekeeping genes *GAPDH* (Fig 5.9SA) and *ACTB* (Fig 5.9SB)do not associate with ICGs as E2 regulated genes do, supporting the specificity of this response for the ERα transcriptional program. These data are consistent with the possibility that E2-induced relocation of regulated transcription units to the interchromatin granules

requires active transport and is an eRNA-dependent event required for target gene activation.

5.6 Further mechanisms of eRNAs in transcriptional regulation

A remaining question, then, is how eRNAs might function in the induced promoter:enhancer interactions, and relocation of the gene in distinct subnuclear structures that culminate in coding gene activation events. Several studies have established a role for cohesins in enhancer:promoter looping events (42-45), and indicated protein:protein interactions between cohesin and the mediator complex protein, Med12 (46). Therefore, we first assessed the levels of cohesin recruitment after ligand treatment, and indeed, we observed increased occupancy of the cohesion subunit Rad21 after E2 treatment to ERα target genes and enhancers; this increase can be observed by conventional ChIP and analyses of ChIP-Seq data (Fig 5.6A). Depletion of specific eRNAs (*NRIP1*e, Fig 5.6B and *FOXC1*e, Fig 5.6C) resulted in a decrease of cohesin and Med12 recruitment in response to E2, but essentially no alteration of H3K4me1/ enhancer mark, without affecting ERα recruitment (Fig 5.3SE, F).

Although many alternative models of cohesin recruitment to enhancers merit consideration, we initially focused on the possibility that, while highly divergent in their primary sequences, eRNAs may help to stabilize the ERα-dependent recruitment of cohesin to the regulatory enhancers, perhaps by recruiting some common or related complexes that contribute to this event. Two approaches were taken, the first was to explore possible interactions between the cohesin complex and regulated eRNAs by

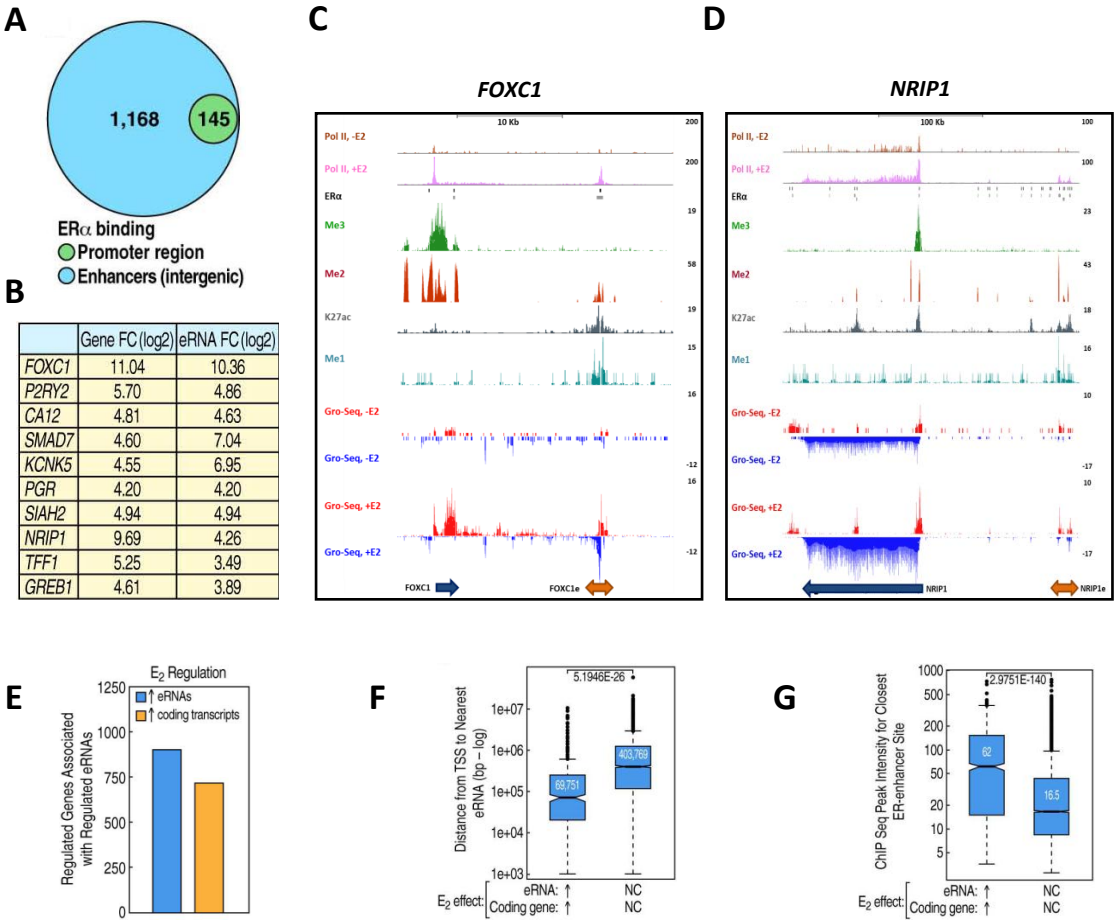
RNA immunoprecipitaion (RIP) using a Rad21 antibody. This assay revealed that the eRNAs of *FOXC1, PGR*, and *TFF1*, while divergent in primary sequence, exhibited interactions with Rad21 that were further improved after hormone treatment (Fig 5.6D). The second approach was to generate biotinylated eRNA transcripts from enhancers of the following loci: *CA12*e, *TFF1*e, *FOXC1*e and *GREB1*e (Fig 5.6E), which were used to isolate interacting protein complexes by pulldown with monoavidin beads in the presence of nuclear extracts from MCF7 cells. The eluted material was washed as described in the methods section and subjected for mass spectrometry analysis. Intriguingly, individual eRNA pulldowns identified similar or identical proteins, suggesting that in spite of sequence variability, eRNAs might actually recruit proteins based on other properties, such as secondary structure. From the list of candidates (Fig 5.6F) there were several proteins that caught our interest, including WDR82, BAF53A, and the motor-associated protein, DLC1. We also note that our results indicate that ERα itself is able to bind eRNAs from E2-regulated enhancers. WDR82 binding to *FOXC1*e and *TFF1*e was verified *in vitro* by RNA electromobility shift assay (REMSA), (Fig 5.6G).

WDR82 is a regulatory component of the SET1 complex responsible for tethering it to the promoters of active genes (40). It facilitates the methylation of histone H3 at lysine-4 (H3K4) by recruiting SET1A or SET1B to the phosphorylated serine-5 at the C-terminal domain of the large subunit of RNA polymerase II (40). RNAi mediated knockdown of WDR82 results in overall diminished H3K4me3 levels, even in the presence of the related MLL complexes (41). Furthermore, *in vitro* enzymatic assays suggest that the SET1 complex is a better H3K4 trimethylase than the MLL complexes

altogether (41). Although not tested and beyond the scope of this manuscript, it is conceivable that basal levels of eRNAs may be sufficient to anchor WDR82 to sites of active transcription, and further deliver the protein to E2-induced promoters via looping. This scenario is analogous to HOTTIP, a lncRNA transcribed from the 5' UTR of the *HoxA* gene, which recruits WDR5 to activate *HoxA* (47).

A hypothetical model summarizing our findings and possible interpretation is provided in Fig 5.7.

Dr. Wenbo Li contributed Gro-seq experiments analyzed by Mr. Qi Ma, and Dr. Dimple Notani performed 3D-DSL experiments analyzed by Mr. Bogdan Tanasa. All these data is presented in chapter 5, which is currently being prepared for submission for publication of the material. Esperanza Núñez, the dissertation author was the primary investigator and author of this material.



	Gene FC (log2)	eRNA FC (log2)
FOXC1	11.04	10.36
P2RY2	5.70	4.86
CA12	4.81	4.63
SMAD7	4.60	7.04
KCNK5	4.55	6.95
PGR	4.20	4.20
SIAH2	4.94	4.94
NRIP1	9.69	4.26
TFF1	5.25	3.49
GREB1	4.61	3.89

Figure 5.1.1: (A) Diagram showing that the majority of the ERα binding program is distal, mainly binding at enhancer sites rather than the promoters of coding genes. (B) Table listing upregulated loci with nearby enhancers chosen for further experimentation along with their corresponding fold induction based on the ratio between treated and non-treated cells. (C) UCSC genome browser (hg18) snapshot of Gro-Seq data (-/+E2) from the FOXC1 genomic locus and its associated enhancers. For reference Pol II, ERα and histone mark Chip-seq data are also plotted. (D) UCSC genome browser (hg18) snapshot of Gro-Seq data (-/+E2) from the NRIP1 genomic locus and its associated enhancers. (E) E2 upregulated eRNAs are associated with upregulated genes with statistical significance [$P \leq 0.0001$, Student's t-test]. (F) E2 induced eRNAs are closest to the gene they regulate [~70kb] in comparison to unchanged eRNAs [>400kb].

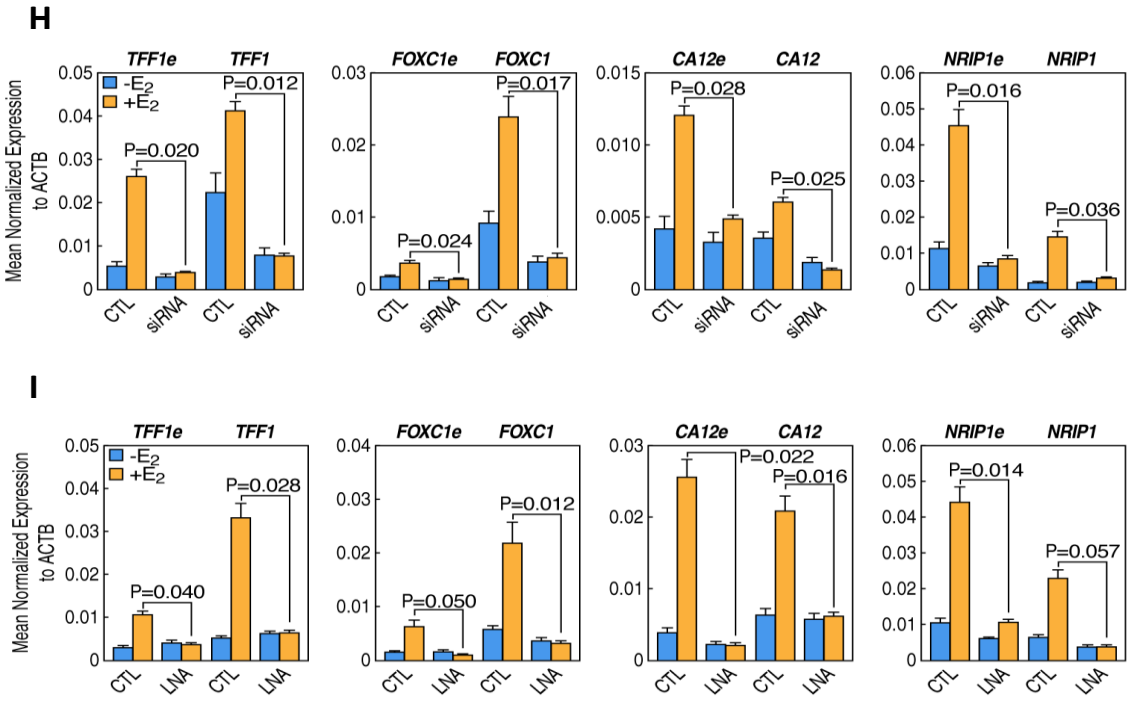


Figure 5.1.2: (H) siRNA Knockdown efficacy and effect assessed by QPCR. (I) LNA Knockdown efficacy and effect assessed by QPCR.

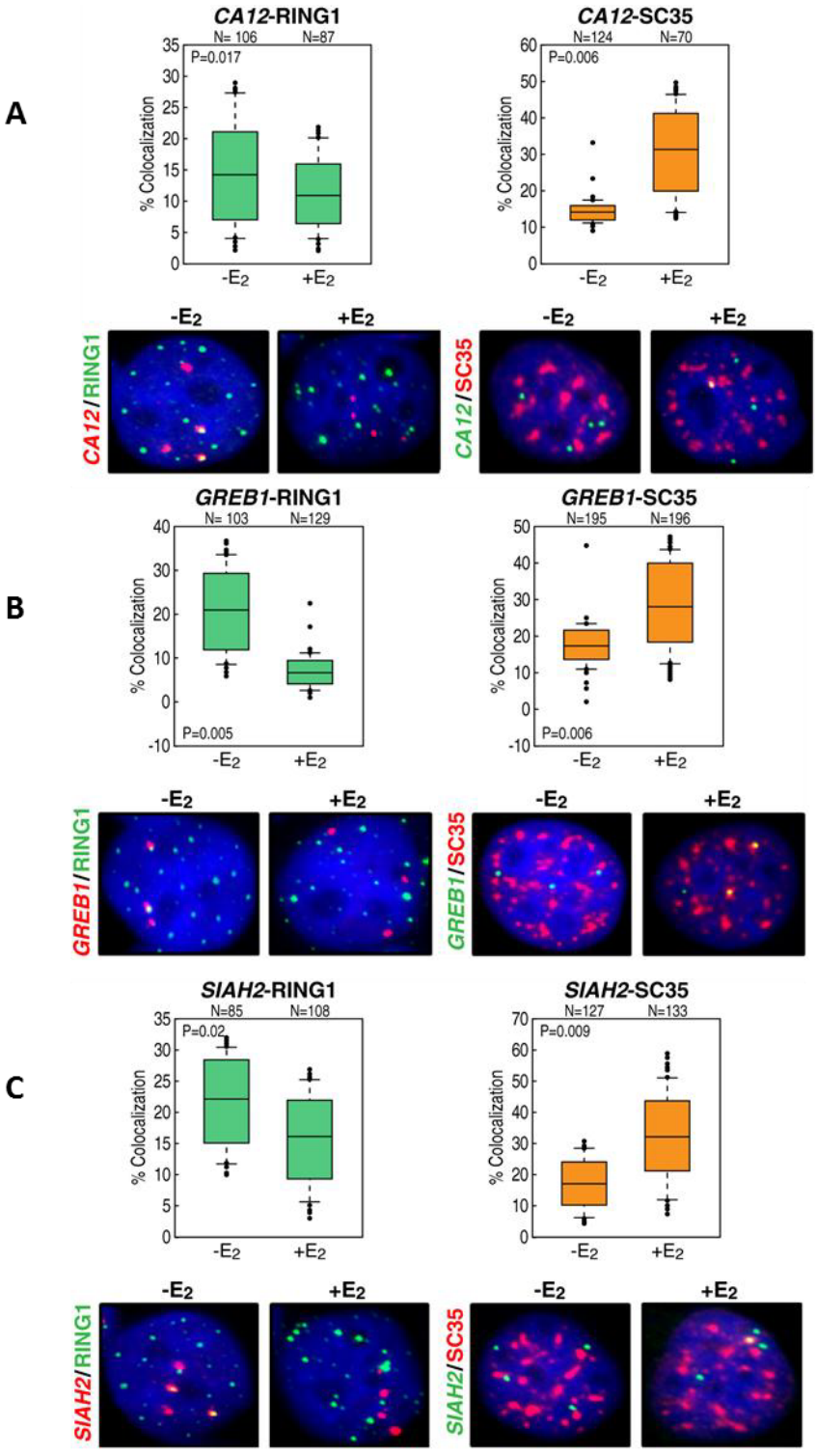


Figure 5.2: ImmunoFISH data and associated statistical analysis for three loci belonging to the ↑eRNAs-↑gene category, showing E_2 induced relocalization from PcGs to ICGs (A) *CA12*, (B) *GREB1*, (C) *SIAH2*.

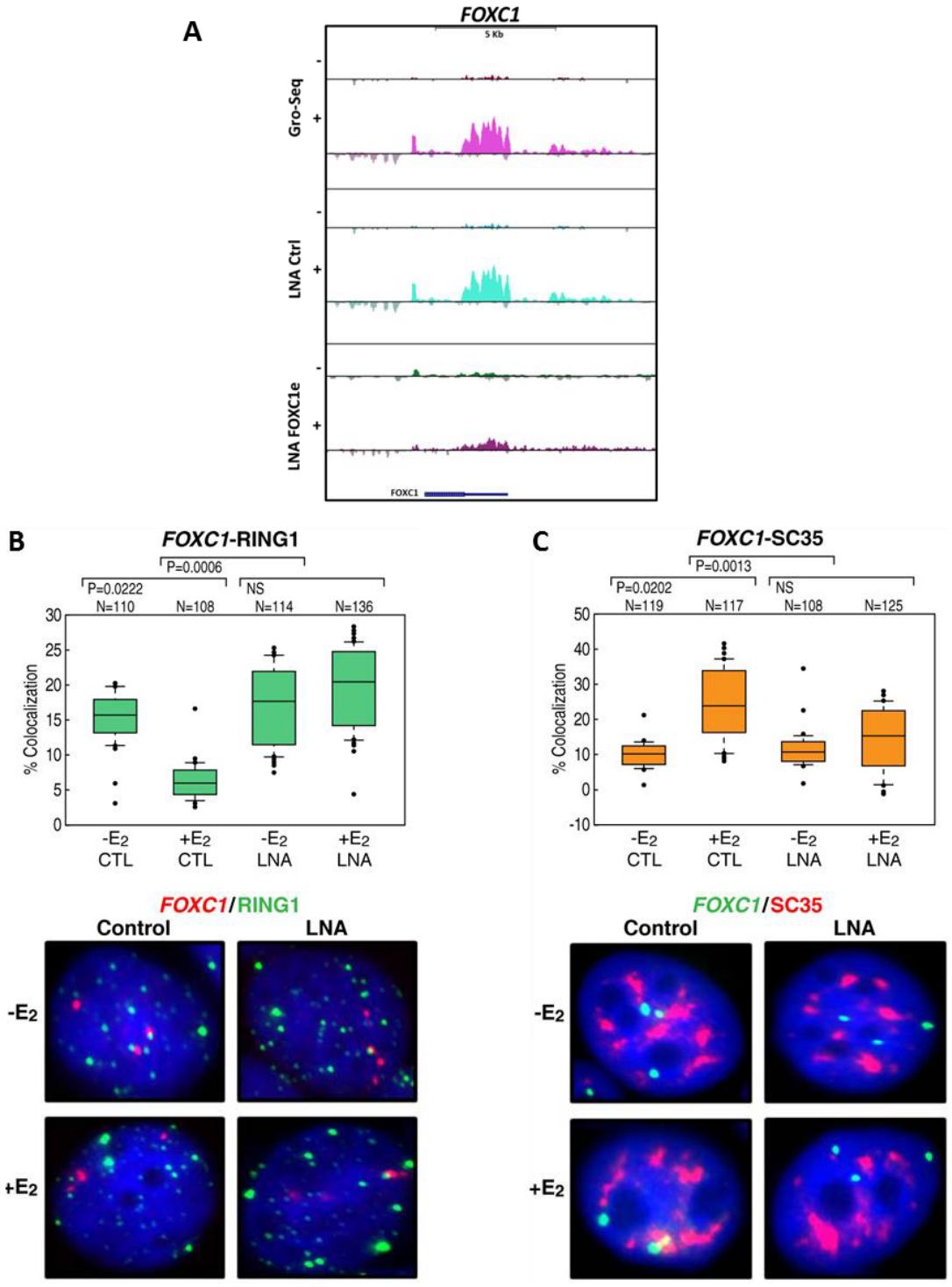


Figure 5.3.1: (A) UCSC genome browser (hg18) snapshot of Gro-Seq data from LNA treated cells showing the knockdown effect of the *FOXC1*eRNA on its associated locus. (B) RING1 immunoFISH data from cells treated with *FOXC1*e LNA in the absence and presence of ligand including statistical analysis. (C) SC35 immunoFISH data from cells treated with *FOXC1*e LNA in the absence and presence of ligand including statistical analysis.

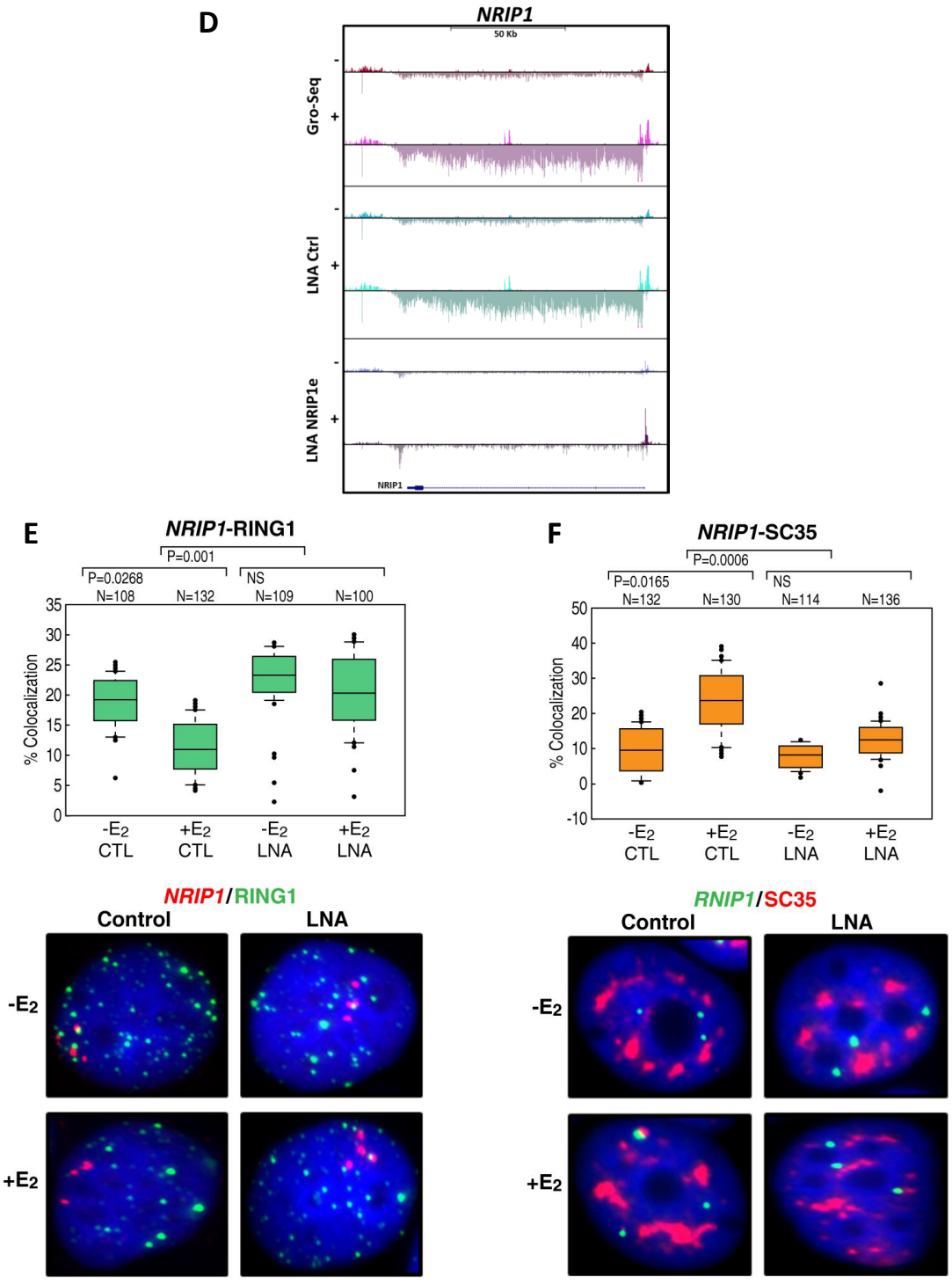


Figure 5.3.2: (D) UCSC genome browser (hg18) snapshot of Gro-Seq data from LNA treated cells showing the knockdown effect of the NRIP1eRNA and on its associated locus. (E) RING1 immunoFISH data from cells treated with *NRIP1*e LNA in the absence and presence of ligand including statistical analysis. (F) SC35 immunoFISH data from cells treated with *NRIP1*e LNA in the absence and presence of ligand including statistical analysis.

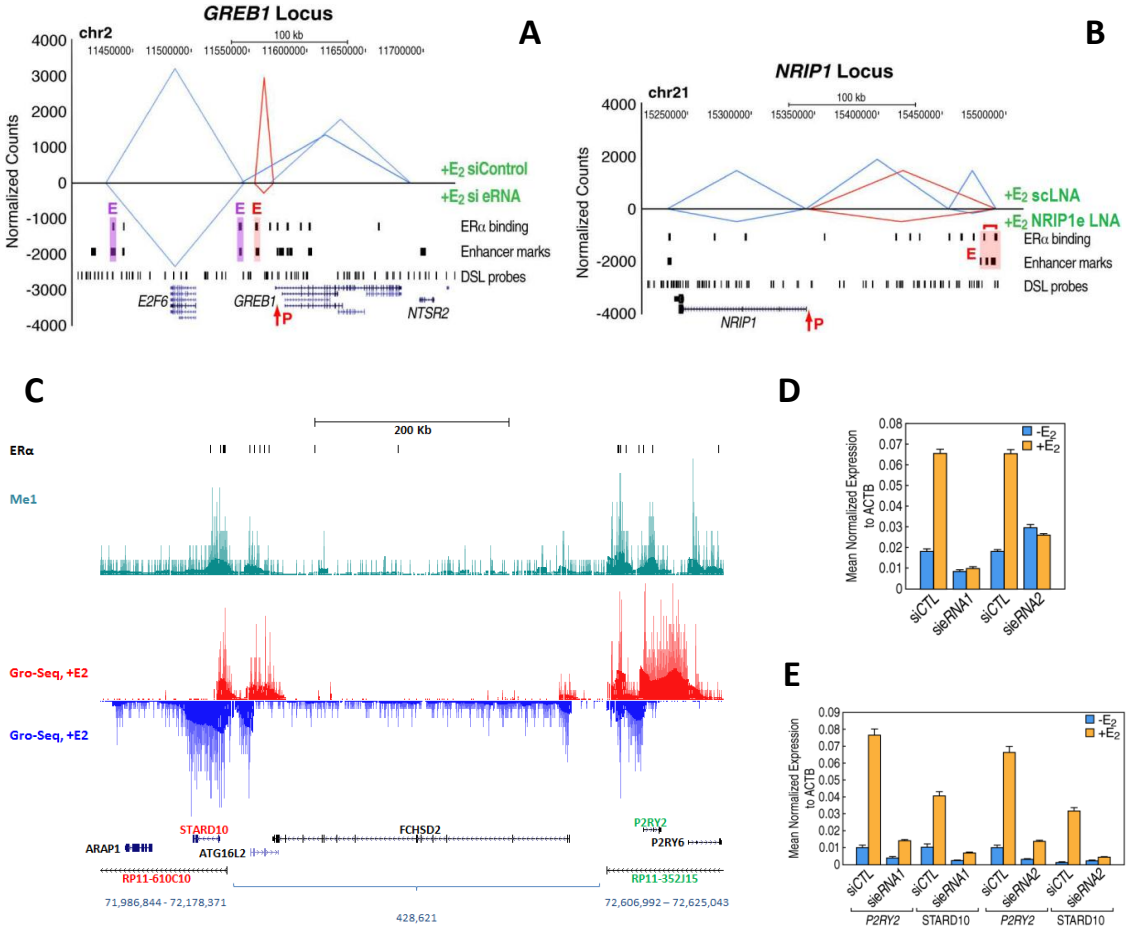


Figure 5.4.1: Graphical representation of 3D-DSL: The y-axis represents normalized tag counts, and the x-axis depicts physical length of the genomic locus (kb). The promoter:enhancer interaction of interest is shown in red and other interactions are shown in blue. (A) 3D-DSL data for *GREB1* showing that the promoter:enhancer seen after E_2 treatment is largely abrogated by siRNA, in addition to affecting other observed interactions downstream of the gene. (B) 3D-DSL data for *NRIP1* after LNA treatment against its eRNAs. The graph shows that the ligand mediated interactions are reduced upon knockdown, and other detected looping events are also affected. (C) UCSC genome browser (hg18) snapshot of Gro-Seq data for the *P2RY2* locus and its interacting partner *STARD10* after E2 treatment showing upregulation of the genes, and the intervening *P2RY2* enhancer as marked by H3K4me1. Also shown are the BAC clones used for ImmunoFISH. (D) QPCR data showing the effectiveness of the siRNAs used against the *P2RY2* enhancer. (E) QPCR data showing that depletion of the *P2RY2e* by siRNAs negatively affects the expression of *P2RY2* and *STARD10* even in the presence of E2.

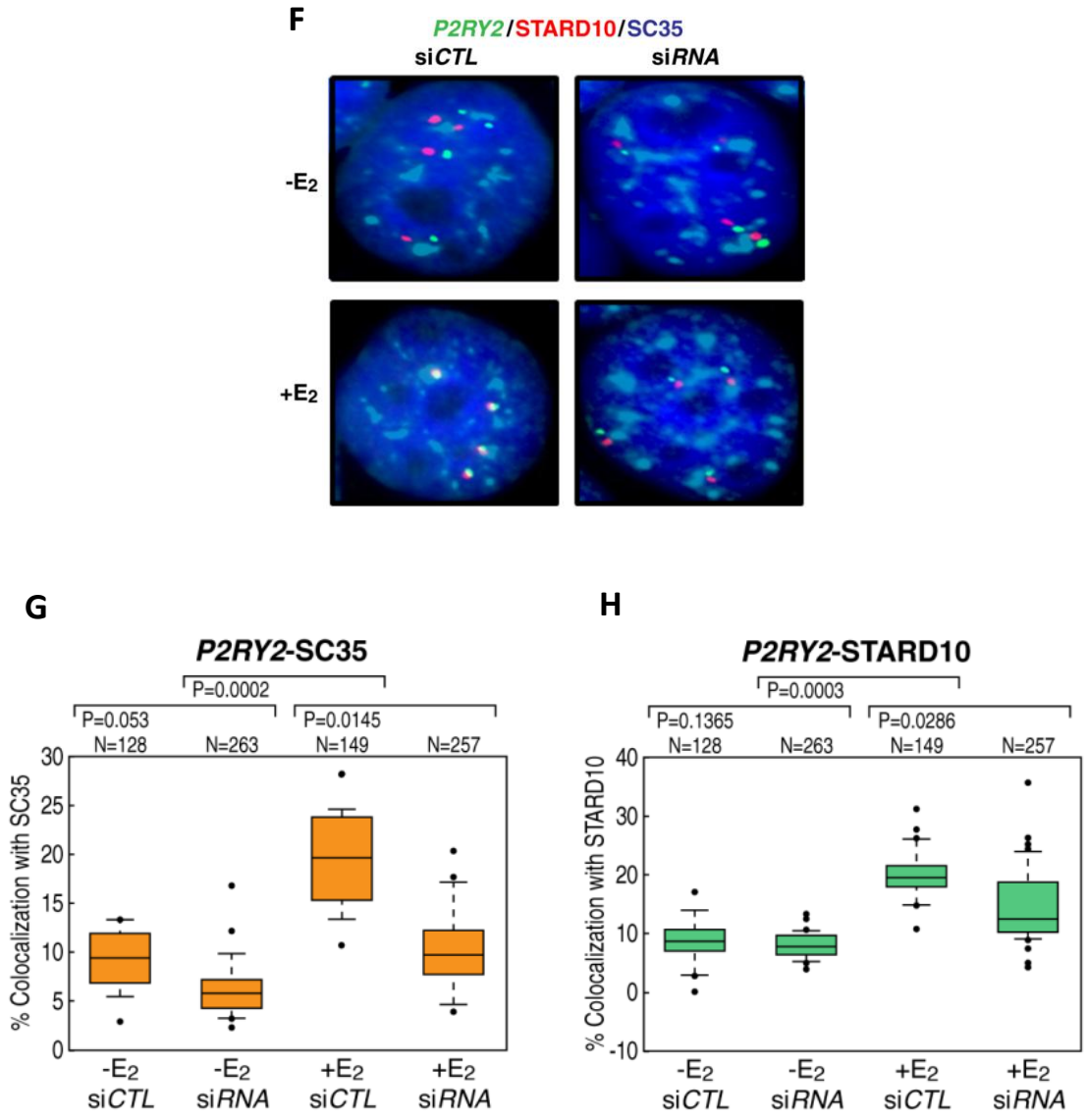


Figure 5.4.2: (F) ImmunoFISH data showing that *P2RY2* interacts with *STARD10* and ICGs upon ligand treatment, and that association is hampered by abrogating the expression from the intervening enhancer. (G, H) Statistical data associated with *siP2RY2e* immunoFISH experiment showing that the eRNA transcript is needed for relocation to ICGs as well as for the *P2RY2*:*STARD10* interaction seen in the control siRNA treated cells.

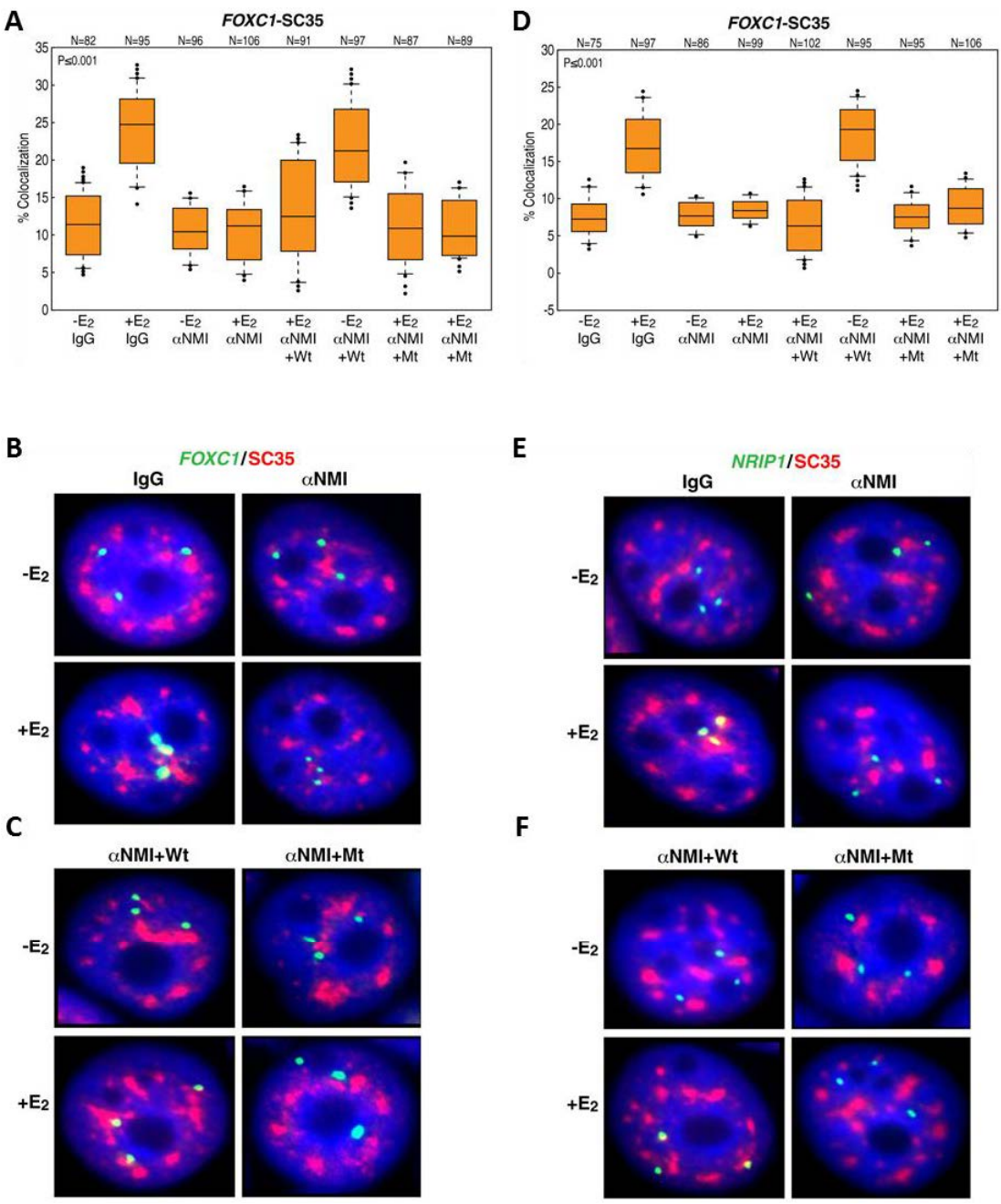


Figure 5.5: Single sell nuclear microinjection of αIgG and αNMI followed by immunoFISH (-/+E_2). (A-C) *FOXC1* and (D-F) *NRIP1*. Both panels include statistical analysis showing that the NMI antibody block causes a loss of E_2 induced relocation of the *FOXC1* and *NRIP1* transcription units ICGs, which can be impaired when a plasmid encoding the S397L NMI mutant is injected , and then rescued when a plasmid encoding wild-type NMI is co-injected.

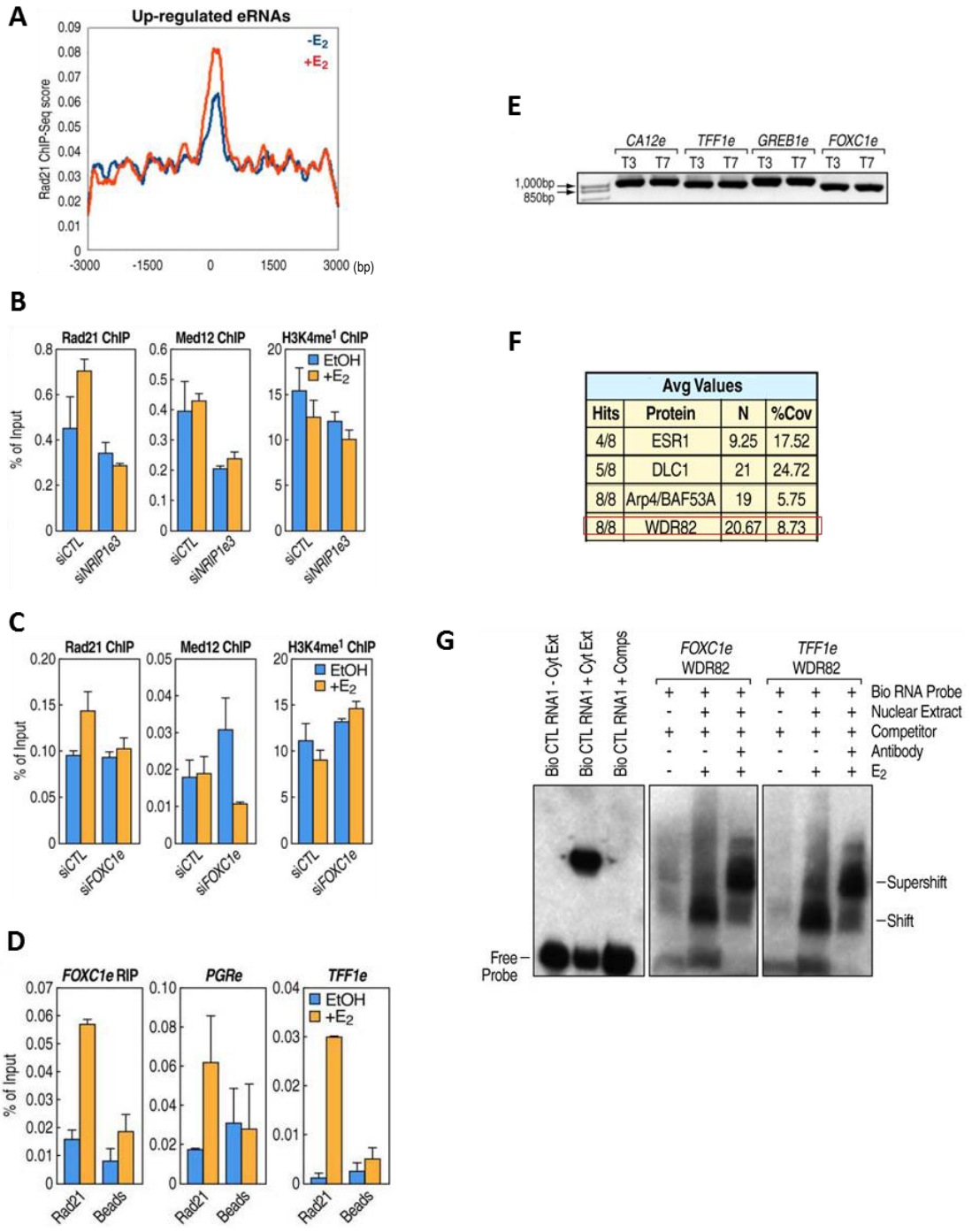


Figure 5.6: (A) Rad21 enrichment centered at upregulated eRNAs measured by ChIP-Seq. (B) si*NRIP1*e knockdown effect on the recruitment of Rad21, Med12 and H3K4Me1 (-/+ E2). (C) si*FOXC1*e knockdown effect on the recruitment of Rad21, Med12 and H3K4Me1. (D) RIP assay showing increased binding of Rad21 to selected eRNAs after hormone treatment. (E) Mass spectrometry results from pulldown assays. (F) Agarose gel showing the specific products of *in vitro* transcribed and biotynilated eRNAs [T3 = 5', T7 = 3']. (G) REMSA with *FOXC1*e and *TFF1*e probes to assay *in vitro* binding with WDR82 present in nuclear extract (shift), and with WDR82 antibody (super-shift).

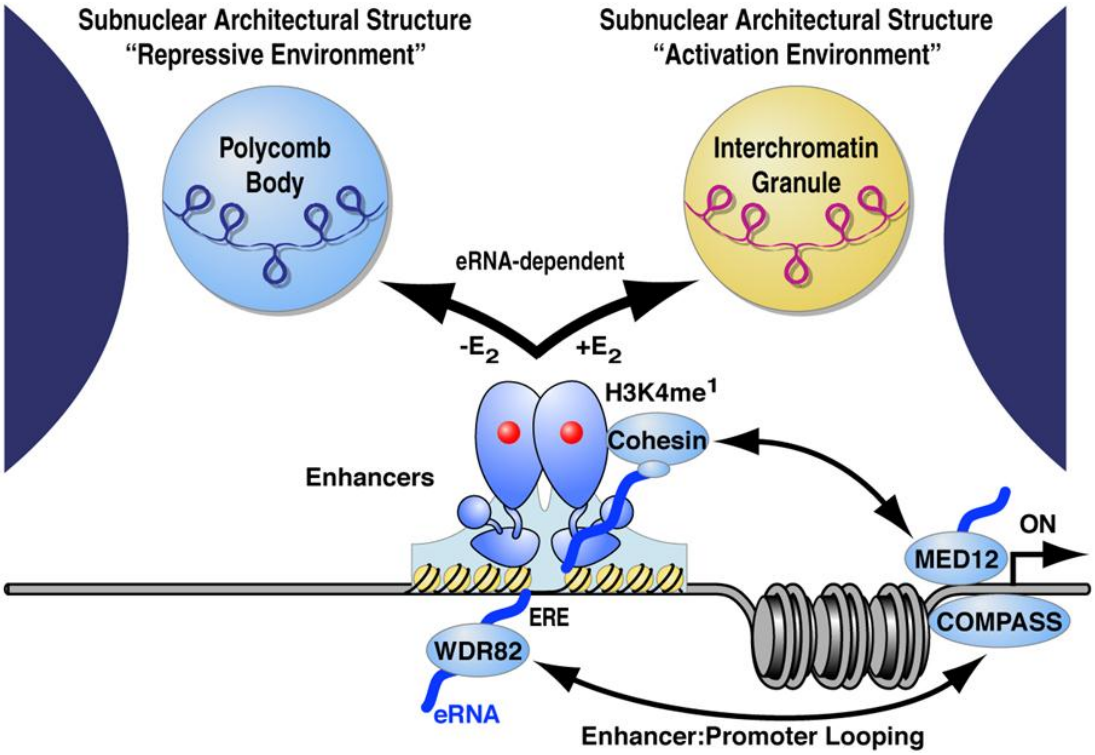


Figure 5.7: Model of eRNA regulation and function. The expression of ERα regulated genes and enhancers is kept in check by a reversible repression step at PcGs. Estradiol is provided to cultured cells that have been serum starved for 3 days, and ERα translocates into the nucleus. ERα finds its cognate binding sites at enhancers producing eRNAs. (4) eRNA transcripts serve as a beacon for Rad21-Med12 and WDR82, which delivers the COMPASS complex to the promoter of genes activated by E2 via looping. Active promoter:enhancer loops can move to ICGs to access elongation machinery and other transcription regulatory factors.

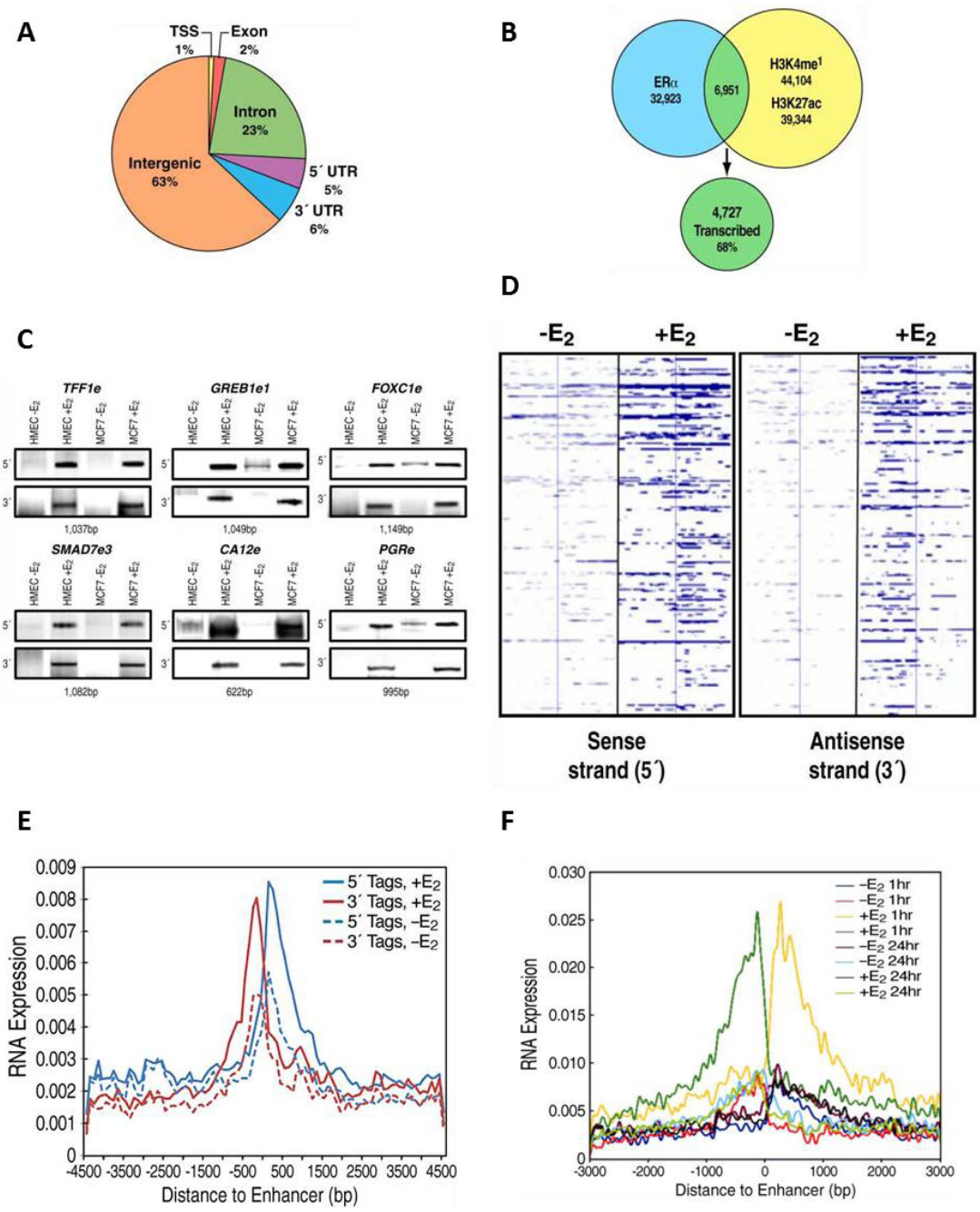


Figure 5.1S: (A) Pie chart illustrating the genomic distribution of ERα binding sites in the genome as per ChIP-Seq. (B) Enhancer definition used in this study, generated by taking the intersection of non-redundant ChIP-Seq peaks for ERα, H3K4me1 and H3K27ac, respectively. (C) Validation of bidirectional transcripts generated from upregulated enhancers after hormone treatment in both, HMEC and MCF7 cells by strand specific RT-PCR. (D) Heat map of Gro-Seq tag distribution at enhancer sites showing that they are mainly bidirectional transcription units. (E) Plot of Gro-Seq tag distribution at enhancer sites shows that they are mainly bidirectional transcription units with an average fragments length of ~3kb. (F) Plot of Gro-Seq tag distribution at enhancer sites at 1hr and 24hrs after E2 treatment illustrating that the maximal level of expression takes place at 1hr after ligand is provided and declines soon after.

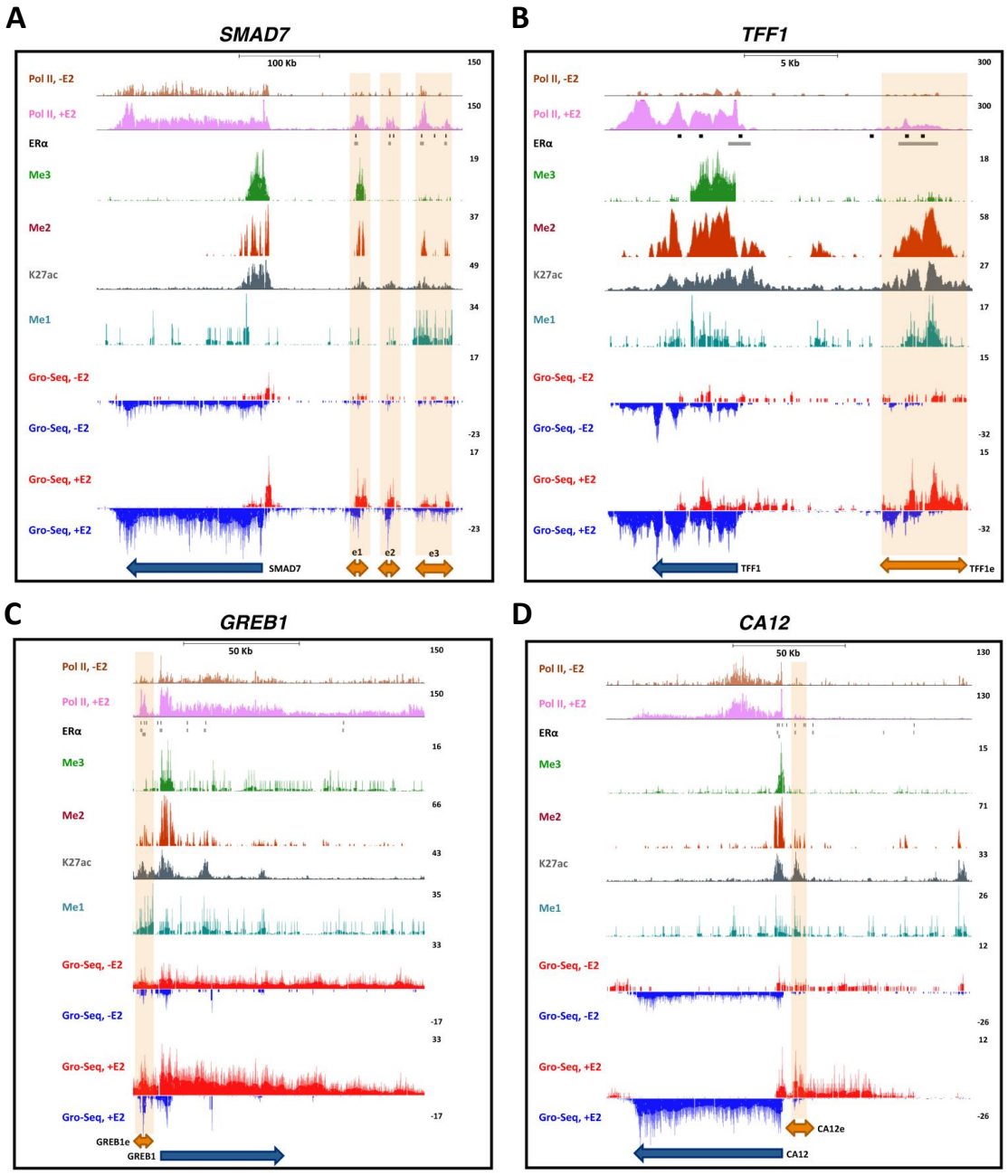


Figure 5.2.1S: UCSC genome browser (hg18) snapshot of Gro-Seq data (-/+E_2) from 4 loci. (A) *SMAD7*, (B) *TFF1*, (C) *GREB1*, (D) *CA12*.

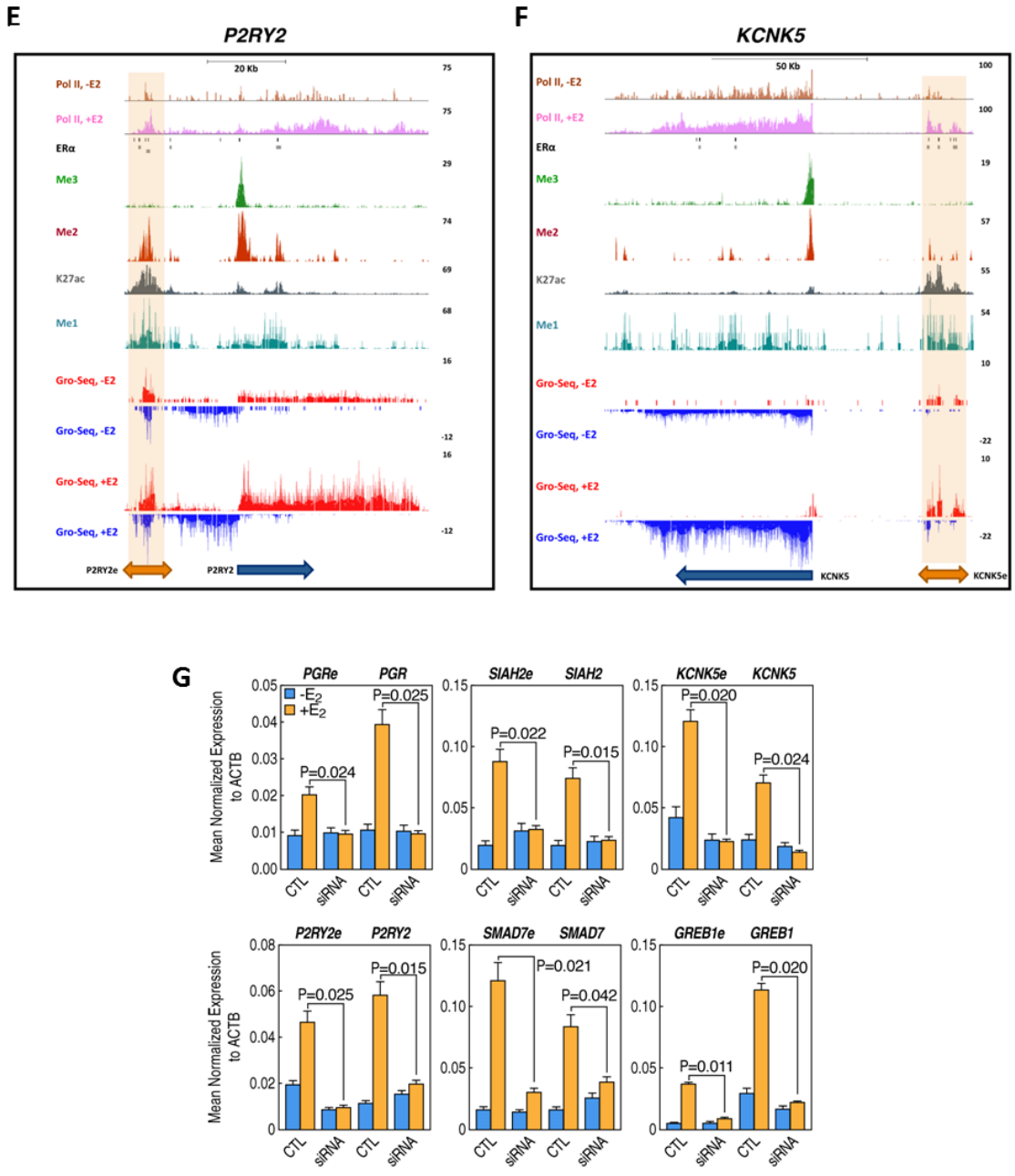


Figure 5.2.2S: UCSC genome browser (hg18) snapshot of Gro-Seq data (-/+E_2) from 2 loci. (E) *P2RY2*, (F) *KCNK5*. (G) Assessment of eRNA knockdown by QPCR from the following loci: *PGR*, *SIAH2*, *KCNK5*, *P2RY2*, *SMAD7* and *GREB1*.

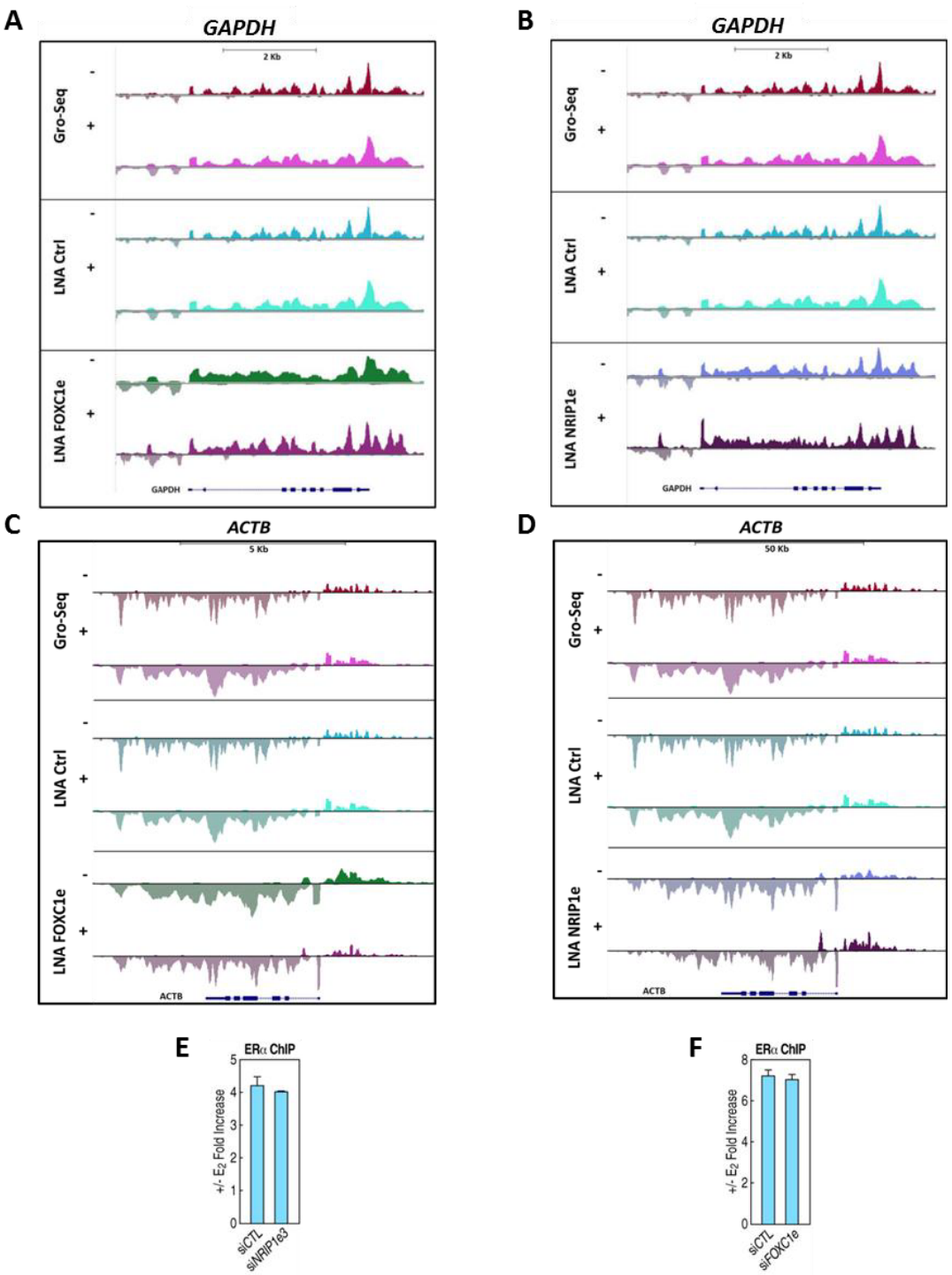


Figure 5.3S: (A, B) UCSC genome browser (hg18) snapshots of Gro-Seq data from *FOXC1e* and *NRIP1e* LNA treated cells, respectively, illustrating that there is no knockdown effect for the housekeeping gene *GAPDH* in either case. (C, D) UCSC genome browser (hg18) snapshots of Gro-Seq data from *FOXC1e* and *NRIP1e* LNA treated cells, respectively, illustrating that there is no knockdown effect for the housekeeping gene *ACTB* in either case. (E, F) Knockdown of *FOXC1* and *NRIP1* eRNA transcripts, respectively, does not affect recruitment of ERα by ChIP-QPCR.

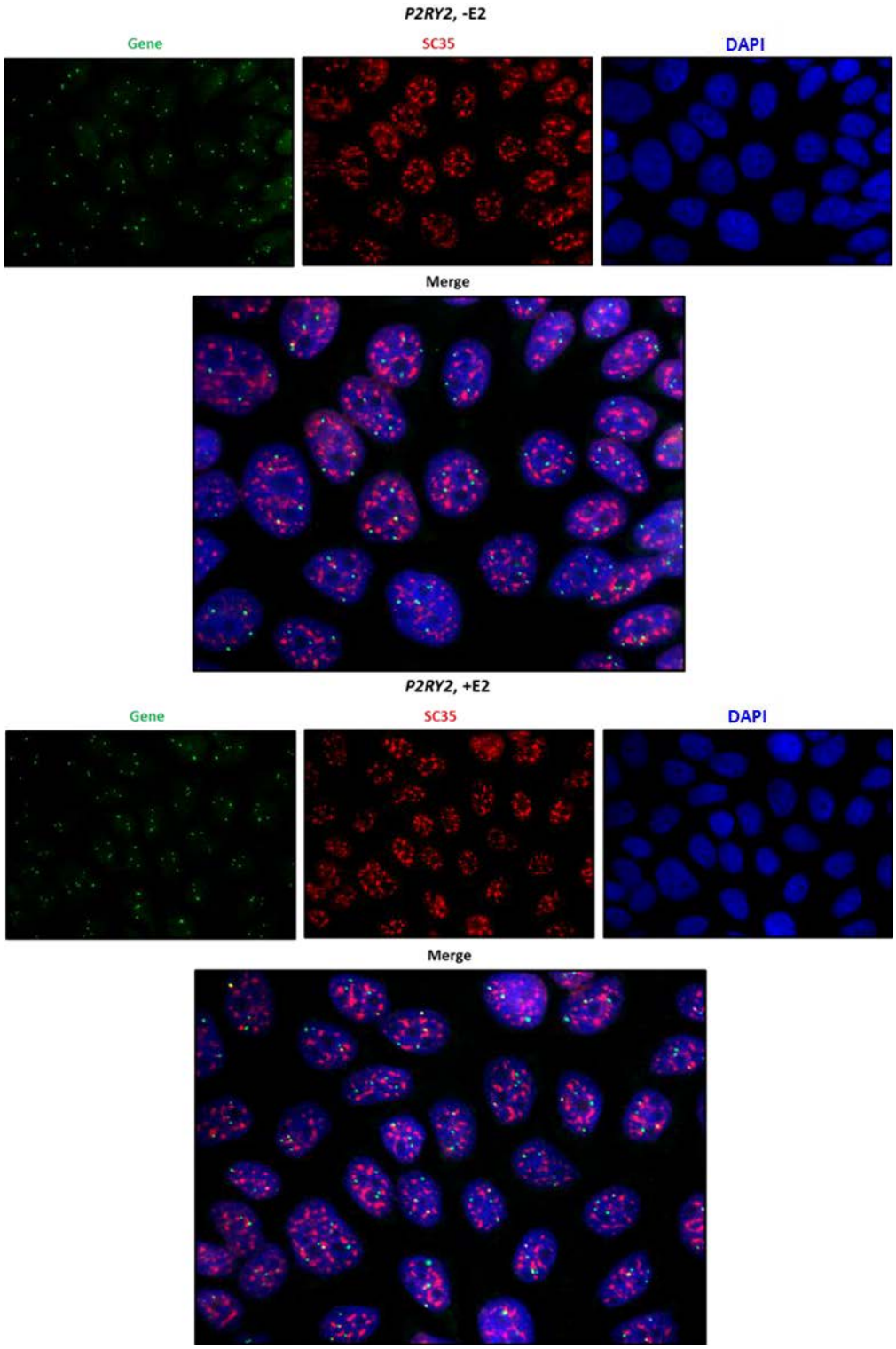


Figure 5.4.1S: Example of a field of cells as viewed with a widefield epifluorescent microscope with oil immersion at 63X. The FISH signal for *P2RY2* is seen in green, SC35 immunostaining in red, and nuclei in blue.

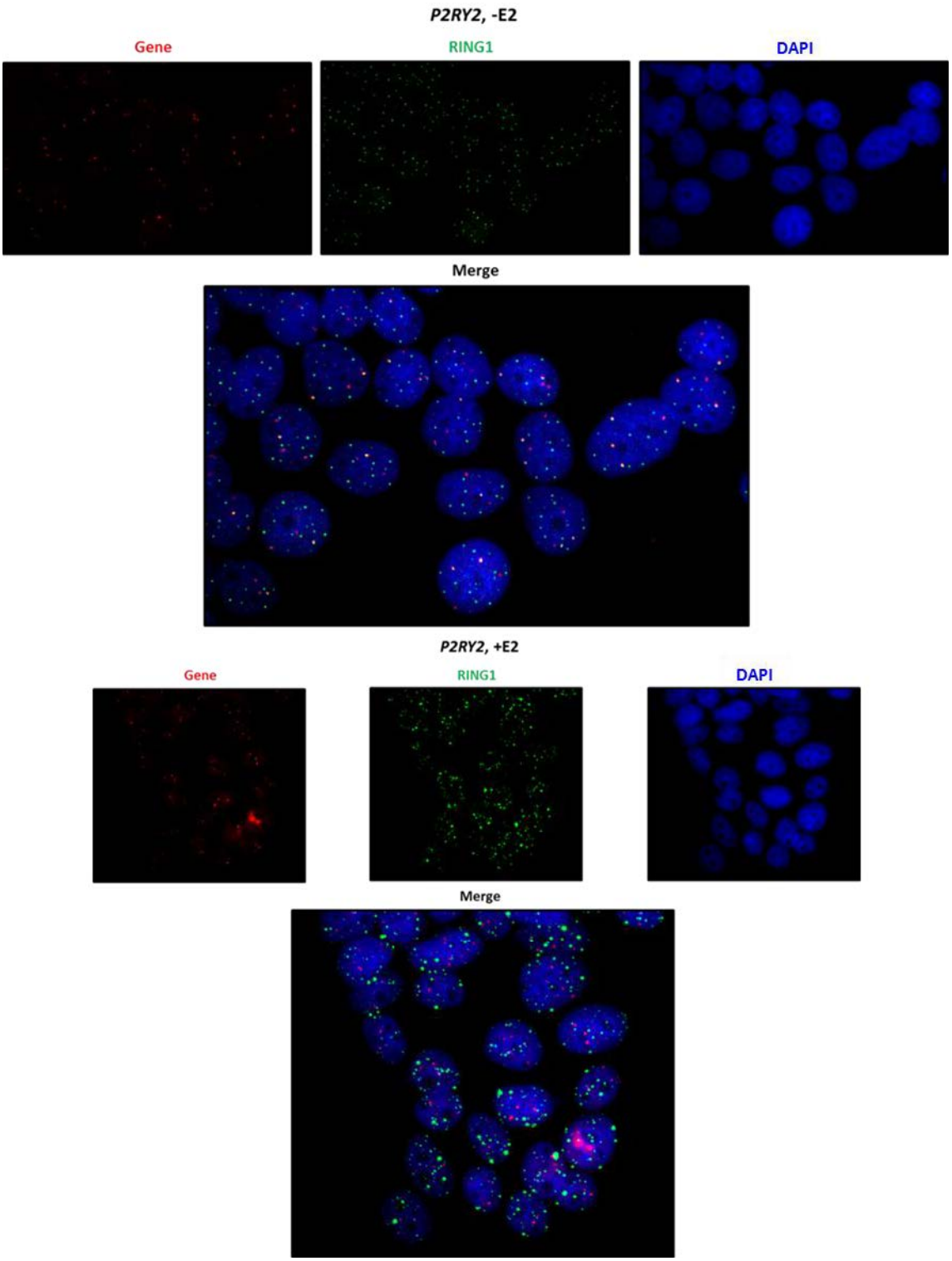
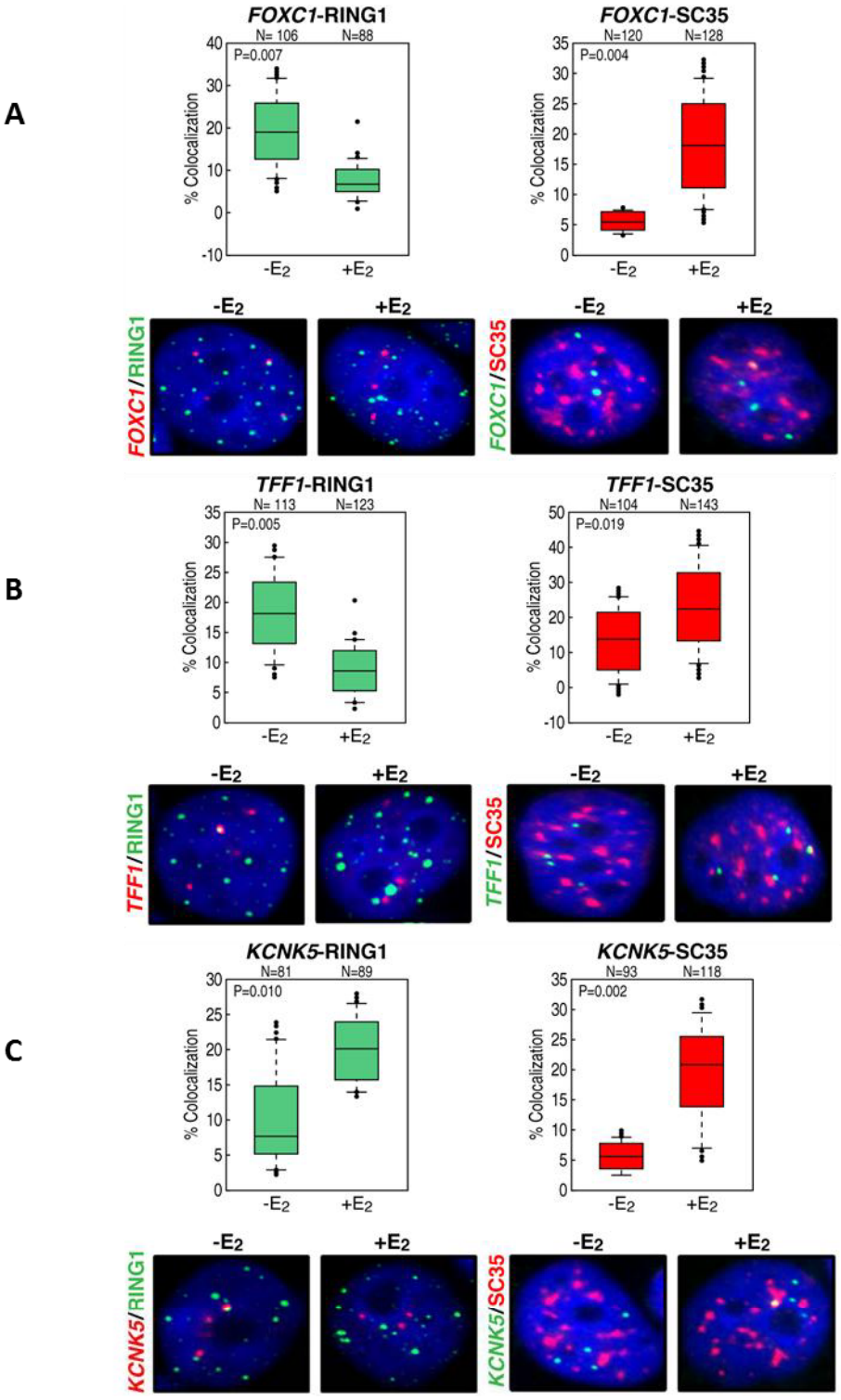


Figure 5.4.2S: Example of a field of cells as viewed with a widefield epifluorescent microscope with oil immersion at 63X. The FISH signal for *P2RY2* is seen in red, RING1 immunostaining in green, and the nuclei in blue.



Figuree 5.5S: ImmunoFISH data and associated statistical analysis for three showing E_2 induced relocalization from PcGs to ICGs. (A) *FOXC1*, (B) *TFF1*, (C) *KCNK5*.

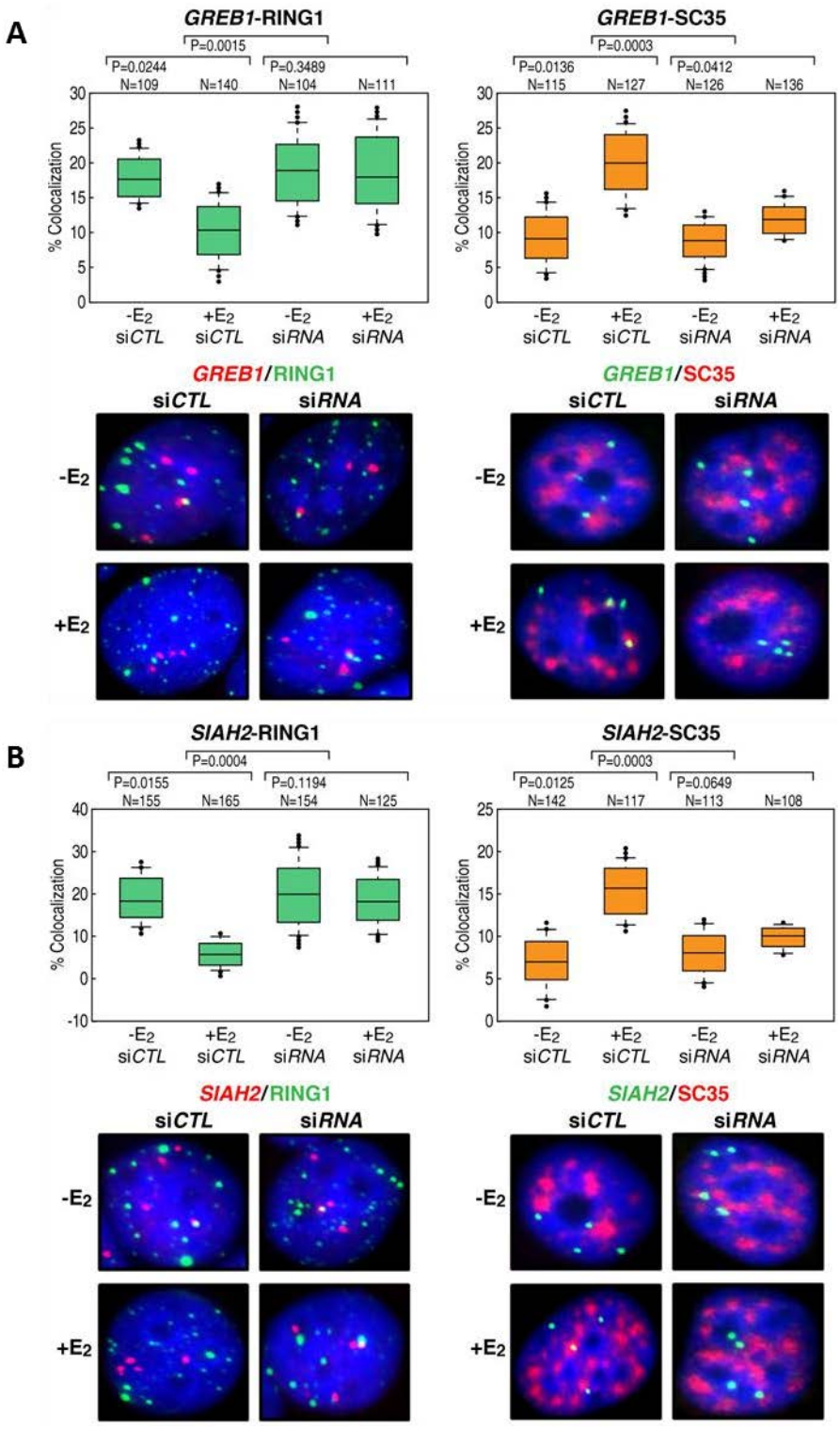


Figure 5.6S: RING1 and SC35 immunoFISH data from cells treated with siRNA against eRNAs in the absence and presence of ligand including statistical analysis. (A) *GREB1*, (B) *SIAH2*.

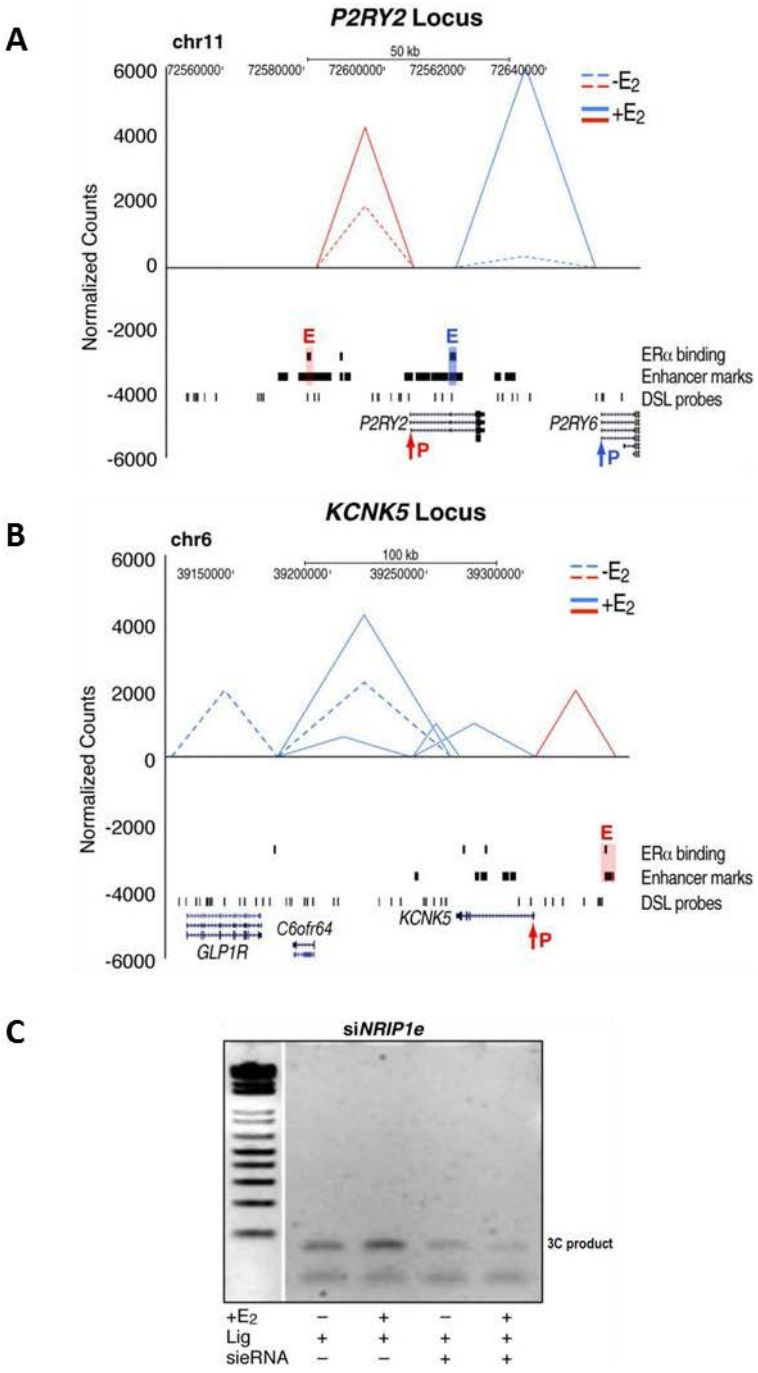


Fig 5.7S: (A) 3D-DSL data for the *P2RY2* locus note that the promoter:enhancer interaction exists in the basal level and gets strengthened after hormone treatment. Another promoter:enhancer interaction is observed for the adjacent gene *P2RY6*. (B) 3D-DSL data for the *KCNK5* locus showing that the promoter:enhancer association takes place only after ligand treatment, in this case, pre-existing interactions at the basal level disappear to give rise to new ones after treatment. (C) Agarose gel with conventional 3C data showing that the NRIP1 promoter:enhancer interaction is diminished when the cells are treated with siRNA even in the presence of ligand.

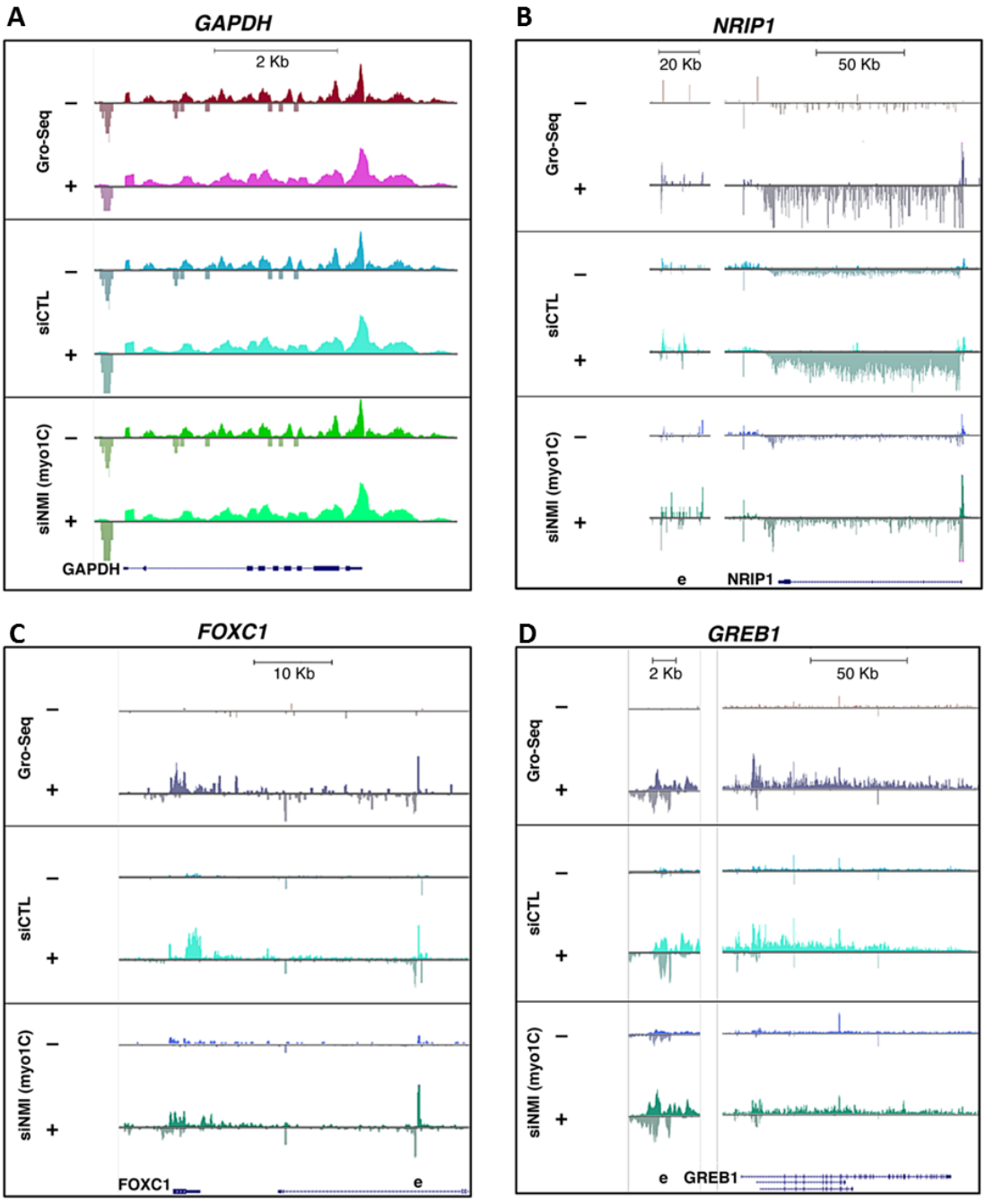


Figure 5.8S: (A) UCSC genome browser (hg18) snapshot of Gro-Seq data (-/+E_2) showing that NMI knockdown by siRNA significantly affects expression from ERα regulated target genes, but not the housekeeping gene GAPDH. The picture also shows that eRNA expression remains unaffected (B) *NRIP1*, (C) *FOXC*, (D) *GREB1*.

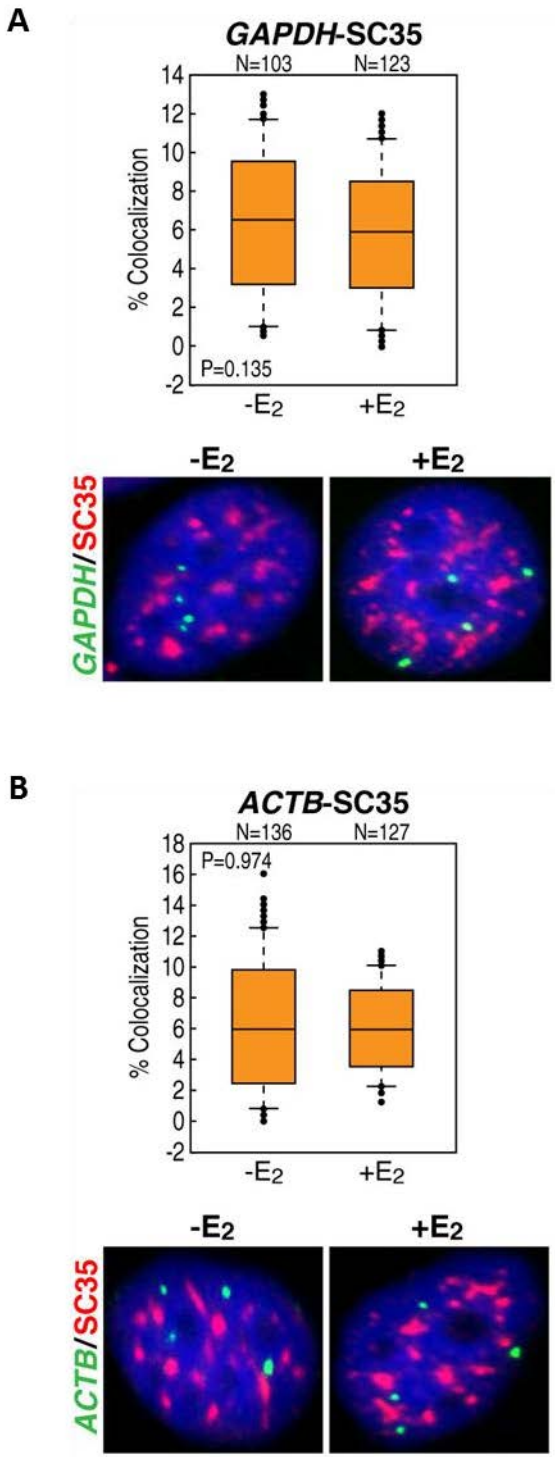


Figure 5.9S: (A) SC35 immunoFISH data showing that the housekeeping gene *GAPDH* doesn't associate with ICGs upon hormone treatment. (B) SC35 immunoFISH data showing that the housekeeping gene *ACTB* doesn't associate with ICGs upon hormone treatment.

5.7 References

1) Jin, V. X., Y. W. Leu, et al. (2004). "Identifying estrogen receptor alpha target genes using integrated computational genomics and chromatin immunoprecipitation microarray." Nucleic Acids Res 32(22): 6627-6635.

2) Carroll, J. S., C. A. Meyer, et al. (2006). "Genome-wide analysis of estrogen receptor binding sites." Nat Genet 38(11): 1289-1297.

3) Kwon, Y. S., I. Garcia-Bassets, et al. (2007). "Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters." Proc Natl Acad Sci U S A 104(12): 4852-4857.

4) Lin, C. Y., V. B. Vega, et al. (2007). "Whole-genome cartography of estrogen receptor alpha binding sites." PLoS Genet 3(6): e87.

5) Welboren, W. J., M. A. van Driel, et al. (2009). "ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands." EMBO J 28(10): 1418-1428.

6) Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet 39(3): 311-318.

7) Creyghton, M. P., A. W. Cheng, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A 107(50): 21931-21936.

8) Core, L. J., J. J. Waterfall, et al. (2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters." Science 322(5909): 1845-1848.

9) Hah, N., C. G. Danko, et al. (2011). "A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells." Cell 145(4): 622-634.

10) Blackshaw, S., S. Harpavat, et al. (2004). "Genomic analysis of mouse retinal development." PLoS Biol 2(9): E247.

11) Rinn, J. L., M. Kertesz, et al. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell 129(7): 1311-1323.

12) Dinger, M. E., P. P. Amaral, et al. (2008). "Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation." Genome Res 18(9): 1433-1445.

13) Ravasi, T., H. Suzuki, et al. (2006). "Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome." Genome Res 16(1): 11-19.

14) Mercer, T. R., M. E. Dinger, et al. (2008). "Specific expression of long noncoding RNAs in the mouse brain." Proc Natl Acad Sci U S A 105(2): 716-721.

15) Hutchinson, J. N., A. W. Ensminger, et al. (2007). "A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains." BMC Genomics 8: 39.

16) Sone, M., T. Hayashi, et al. (2007). "The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons." J Cell Sci 120(Pt 15): 2498-2506.

17) Clemson, C. M., J. N. Hutchinson, et al. (2009). "An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles." Mol Cell 33(6): 717-726.

18) Sasaki, Y. T. F., T. Ideue, et al. (2009). "MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles." Proc Natl Acad Sci U S A 106(8): 2525-2530.

19) Sunwoo, H., M. E. Dinger, et al. (2009). "MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles." Genome Res 19(3): 347-359.

20) Costa, F. F. (2005). "Non-coding RNAs: new players in eukaryotic biology." Gene 357(2): 83-94.

21) Szymanski, M., M. Z. Barciszewska, et al. (2005). "A new frontier for molecular medicine: noncoding RNAs." Biochim Biophys Acta 1756(1): 65-75.

22) Prasanth, K. V. and D. L. Spector (2007). "Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum." Genes Dev 21(1): 11-42.

23) Petersen, M. and J. Wengel (2003). "LNA: a versatile tool for therapeutics and genomics." Trends Biotechnol 21(2): 74-81.

24) Sarma, K., P. Levasseur, et al. (2010). "Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome." Proc Natl Acad Sci USA 107(51): 22196-22201.

25) Yang, L., C. Lin, et al. (2011). "ncRNA- and Pc2 Methylation-Dependent Gene Relocation between Nuclear Structures Mediates Gene Activation Programs." Cell 147(4): 773-788.

26) Abramoff, M. D., Magelhaes, P.J., Ram, S.J. (2004). "Image Processing with ImageJ." Biophotonics International 11(7): 36-42.

27) Harismendy, O., D. Notani, et al. (2011). "9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response." Nature 470(7333): 264-268.

28) Dostie, J., T. A. Richmond, et al. (2006). "Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements." Genome Res 16(10): 1299-1309.

29) Simonis, M., J. Kooren, et al. (2007). "An evaluation of 3C-based methods to capture DNA interactions." Nat Methods 4(11): 895-901.

30) Dekker, J. (2006). "The three 'C' s of chromosome conformation capture: controls, controls, controls." Nat Methods 3(1): 17-21.

31) Carroll, J. S., X. S. Liu, et al. (2005). "Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1." Cell 122(1): 33-43.

32) Fullwood, M. J., M. H. Liu, et al. (2009). "An oestrogen-receptor-alpha-bound human chromatin interactome." Nature 462(7269): 58-64.

33) Sun, J., Z. Nawaz, et al. (2007). "Long-range activation of GREB1 by estrogen receptor via three distal consensus estrogen-responsive elements in breast cancer cells." Mol Endocrinol 21(11): 2651-2662.

34) Barnett, D. H., S. Sheng, et al. (2008). "Estrogen receptor regulation of carbonic anhydrase XII through a distal enhancer in breast cancer." Cancer Res 68(9): 3505-3515.

35) Hu, Q., Y.-S. Kwon, et al. (2008). "Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules." Proc Natl Acad Sci U S A 105(49): 19199-19204.

36) Ambrosino, C., R. Tarallo, et al. (2010). "Identification of a hormone-regulated dynamic nuclear actin network associated with estrogen receptor alpha in human breast cancer cell nuclei." Mol Cell Proteomics 9(6): 1352-1367.

37) Wang, D., I. Garcia-Bassets, et al. (2011). "Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA." Nature 474(7351): 390-394.

38) Ahlenstiel, C. L., H. G. Lim, et al. (2012). "Direct evidence of nuclear Argonaute distribution during transcriptional silencing links the actin cytoskeleton to nuclear RNAi machinery in human cells." Nucleic Acids Res 40(4): 1579-1595.

39) Vester, B., A. M. Boel, et al. (2008). "Chemically modified oligonucleotides with efficient RNase H response." Bioorg Med Chem Lett 18(7): 2296-2300.

40) Lee, J. H. and D. G. Skalnik (2008). "Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes." Mol Cell Biol 28(2): 609-618.

41) Wu, M., P. F. Wang, et al. (2008). "Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS." Mol Cell Biol 28(24): 7337-7344.

42) Hadjur, S., L. M. Williams, et al. (2009). "Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus." Nature 460(7253): 410-413.

43) Mishiro, T., K. Ishihara, et al. (2009). "Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster." EMBO J 28(9): 1234-1245.

44) Nativio, R., K. S. Wendt, et al. (2009). "Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus." PLoS Genet 5(11): e1000739.

45) Hou, C., R. Dale, et al. (2010). "Cell type specificity of chromatin organization mediated by CTCF and cohesin." Proc Natl Acad Sci U S A 107(8): 3651-3656.

46) Kagey, M. H., J. J. Newman, et al. (2010). "Mediator and cohesin connect gene expression and chromatin architecture." Nature 467(7314): 430-435.

47) Burgess, D. J. (2011). "Non-coding RNA: HOTTIP goes the distance." Nat Rev Genet 12(5): 300.

6 Discussion

The nucleus is a complex structure containing genetic material packaged into chromosomes as well as subnuclear organelles with specific distribution patterns such as the nucleoli, Cajal bodies, PML bodies, polycomb bodies, SC35 speckles and paraspeckles (1). These bodies mainly occupy the interchromatin space and, as exemplified by results presented here, under certain conditions are associated with specific gene loci and/or their RNA products. Enzymatic activities that act on genes are not ubiquitously distributed throughout the nucleoplasm, but rather limited to these nuclear substructures that promote ultimate higher-order chromatin arrangements. Our data indicate that estrogen-activated transcription units either relocate from the repressive environment of the polycomb body to the putative activating environment of the interchromatin granule, or that there is a rapid de novo assembly of these architectural structures requiring the actions of eRNA transcripts from enhancers of activated coding genes.

The concept that regulated Pol II transcription units might be located in the interchromatin granule is consistent with reports indicating that heat-shock genes, erythroid specific genes, collagen genes, and muscle specific genes preferentially associate with SC35 domains when active (2-9). More sophisticated live experiments with BAC transgenes encoding *Hsp70* along with multimers of the Lac operator sequence, allowed visualization by means of GFP-LacI expression, of the transgene associating with splicing speckles within 30 min of heat shock activation (10). Interchromatin granules, do not appear to assemble only at specific gene loci, but rather

are said to be found in proximity to sites of active transcription (11), this observation led to the conclusion that although these granules form in nuclear compartments containing high transcriptional activity, they directly do not regulate transcription of genes in their vicinity (12). However, the view that interchromatin granules are relatively passive reservoirs of RNA processing machinery has been challenged by findings that demonstrate hnRNPs and RNA polymerase II that reside in the speckles are recruited to sites of active transcription, indicating that Pol II transcription and pre-mRNA processing reactions are linked spatially and temporally (13). Later proteomic analysis of interchromatin granule clusters unexpectedly identified several components involved in transcription; in addition to Pol II, they found several transcription factors, nuclear receptors, coactivators, histone methyltransferases, and actin related proteins among other interesting transcriptional regulators (14). Lastly, it has been reported that SR proteins (components of the nuclear speckles) can facilitate transcriptional elongation (15), presumably independently from their known role interpreting splicing signals, consistent with the function of SR proteins on intronless genes independent from splicing (16-19). Activation of growth control genes by serum appears to cause their location or relocation to interchromatin granules as a required aspect of their induced transcription (20). Together, these data strongly suggests that relocation to the interchromatin granule, for many classes of regulated transcription factors, constitutes an environment with high concentrations of coactivator and elongation complexes acting in concert with ncRNAs to modulate gene transcription.

These observations would also be consistent with the postulated "self-organization" model of nuclear subcompartments, in which nuclear bodies are dynamic structures that form at sites of specific activities associated with gene expression (21), and it has been proposed that the act of transcription itself is a driving force in nuclear body formation indicating that coding and noncoding RNA can prime the biogenesis of nuclear bodies (22). One suggested model argues that nuclear body formation necessitates two-steps: in the first step, nonrandom, biologically determined RNAs (or proteins bound to the template) act as a seed. In the subsequent step, nuclear body formation is driven by random and stochastic self-organization (23).

The data presented in our manuscript suggests that eRNAs are required for enhancer:promoter looping interactions, raising the intriguing question whether, in mammalian cells, relocation, as well as looping, are required for effective gene induction events. NMI (MYO1C) and monomeric G-actin are reported to be required for the ERα-activation program (24, 25). For a more extensive review on nuclear actins and myosins, please see (26). Our data indicates that siRNA knockdown of NMI does not affect eRNA transcription, but has deleterious effects on ERα regulated coding genes. Similarly, NMI depletion entirely blocked target gene relocation to ICGs. However, we cannot exclude effects of polymerase dependent transcription as an additional role of NMI function. We suggest that estrogen dependent gene activation is the consequence of two distinct, nuclear architectural regulatory events: 1) local enhancer:promoter looping and 2) NMI-facilitated association with specific nuclear substructures.

The polycomb body is enriched in polycomb group (PcG) proteins, which are epigenetic chromatin modifiers involved in gene repression. They exist in two separate protein complexes, PRC1 and PRC2, respectively. PRC2 consists of EED, EZH2, YY1 and SUZ12, and it is thought to be required at the initiating stage of silencing; while PRC1 comprises HPH, RING1, BMI1, and HPC, and it is continuously required for the stable maintenance of the initiated PcG repression on specific target loci (27). It has already been established that both complexes can associate with each other at least transiently (28). Moreover, recent studies confirm the interaction between the complexes showing that PRC2 members EED and EZH2 are required for the efficient recruitment of PRC1 proteins to silenced regions, and that the HMTase activity of EZH2 is essential for this recruiting ability (29). Unlike Drosophila PcG proteins, mammalian PcG factors are often involved in reversible gene repression, suggesting that other systems such as DNA methylation might be used to stably lock-in transcriptional silencing in these organisms. ERα functions in cycles of productive and unproductive transcription (30), during unproductive cycles, corepressor complexes and HDACs are transiently recruited until the next productive cycle, and it is not well characterized whether they are deposited by the canonical corepressor proteins or by PcG complexes. Using PcG proteins in nuclear receptor regulated programs makes a lot of sense given the rapid-transient responses they arbitrate; it is also notable that PcG binding negatively correlates with the presence of RNA pol II, suggesting that pol II is excluded from many PcG target genes as a consequence of silencing (31).

Recent work indicates that PcG complexes also exert their function by regulating the nuclear organization of their target genes, and that they can accomplish this by coopting noncoding RNAs and the RNAi machinery (32). Specifically, they have been shown to facilitate long-range interchromosomal interactions of multiple loci even when artificially inserted into ectopic sites (33, 34). The convergence of several PcG regulated genes with PcG bodies is also supported by the existence of fewer PcG bodies than PcG target genes identified by ChIP-seq (35), in support of the idea that repositioning of genomic regions with respect to nuclear compartments may be important for the regulation of gene expression. Both EZH2 and BMI are considered markers of breast cancer with worse prognosis. Interestingly, EZH2 has been implicated not only as a repressor, but also as transactivator of ERα target genes (36). The exquisite balance of nuclear architecture and chromatin structure is paramount for the regulation of gene expression, and its alteration leads to disease development.

While eRNAs are essential for gene regulation and seem to recruit the supporting required machinery for ligand-mediated responses, we cannot distinguish whether the specific association of ERα-regulated genes with nuclear bodies is a true relocation event, or whether the eRNAs nucleate *de novo* assembly of ICGs or PcGs. The need of a nuclear motor (NMI) along with actin related proteins certainly seems to support the relocation hypothesis. Ultimately, whether the association is caused by de novo aggregation of protein complexes or by relocation phenomena, the underlying event is still transcriptional activation. We suggest that the ERα-activation program

hijacks the dynamic protein turnover of these nuclear bodies to facilitate a massive transient response.

Recent mass spectrometry analysis of the ERα/β interactome yielded RNA binding as the most enriched molecular function, while the most significant biological processes besides translation, were RNA processing, chromosome organization/biogenesis and regulation of actin polymerization/depolymerization (37). Proteomic analysis revealed SC35 along with members of the PRC2 complex as main interactants. The interaction with members of the PRC2 complex, namely EZH2 and RbAp46/48 (co-IP & ChIP) is direct and detected in the absence of estradiol treatment. Additionally, siRNAs against RbAp46 showed specific de-repression of the ERα target genes PGR and TFF1, consistent with its repressive role in the basal state (38).

A largely unanswered question remains: what is the mechanism by which enhancers recognize their cognate, regulated promoters? This question is even more puzzling in the case of ERα regulation given that its genome-wide distribution revealed an overwhelming localization away from promoters, particularly overlapping enhancer sites. One idea that emerges is that enhancers and their RNA products serve to deliver the receptor to its target genes via looping, in such a case, the RNA transcripts can stabilize the interaction by providing a scaffold for proteins required to complete the activation process. One of such proteins which we found interact with eRNAs via pulldown experiments followed by mass spectrometry, is WDR82, which is a major subunit of one of the six mammalian COMPASS-related complexes that can methylate H3K4 through SET1A/B (39).

WDR82 is actually responsible for recruiting the H3K4 methyltransferase complex, SET1A/B to active genes; it does so by tethering uniquely to Ser5-p-CTD of Pol II and the RNA recognition motif (RRM) of SET1A/B, respectively (40). Interestingly, a second RRM has been described for SET1A/B (41), suggesting the possibility of its interaction with nascent RNAs (including eRNAs) to modulate the binding activity of WDR82 to either the SET1A/B complex or the Ser5-P CTD of RNAP II, as proposed for the yeast Set1/COMPASS complex (40, 41). Survey of the literature unveiled one other example of an ncRNA with activation function conferred by its recruitment of a COMPASS-like complex subunit (42). In this report, an ncRNA from the 5' UTR of the HOXA locus, HOXA transcript at the distal tip (HOTTIP), was shown to interact with WDR5, targeting WDR5–MLL complexes across HOXA to induce H3K4 trimethylation and gene activation.

We also find that eRNAs are capable of binding Rad21 (a cohesin subunit), directly or indirectly, upon hormone treatment as determined by RNA immunoprecipitation assays, and that this event is important for effective promoter:enhancer looping. There is evidence that indeed cohesin helps mediate the steroid hormone signaling program in MCF7 cells, where it concomitantly binds to target genes along with ERα (43). The finding that cohesin can regulate the response to estrogen strongly suggests that it functionally marks active target genes; one of the ways that it can perform this function is by modulating chromatin architecture, as there are reports indicating a direct role for cohesin in long-range interactions in several model systems (44-47). Along with cohesin, the transcriptional coactivator Med12 has

also been implicated in promoter:enhancer looping perhaps by helping stabilize cohesin mediated interactions, or by marking active genes by means of Pol II recruitment (48). Ultimately, our data suggests that the regulation of eRNAs rather than Rad21 recruitment alone is the determining factor of functional looping between an active enhancer and its regulated target gene promoter.

In summary, we suggest that eRNAs have functional importance in the actions of estrogen regulated gene enhancers, possibly by serving as modular scaffolds for protein complexes required both for promoter:enhancer interactions, and subnuclear structural associations that underlie regulated gene transcriptional programs.

6.1 References

1) Cremer, T. and M. Cremer (2010). "Chromosome territories." Cold Spring Harb Perspect Biol 2(3): a003889.

2) Xing, Y., C. V. Johnson, et al. (1995). "Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains." J Cell Biol 131(6 Pt 2): 1635-1647.

3) Jolly, C., C. Vourc'h, et al. (1999). "Intron-independent association of splicing factors with active genes." J Cell Biol 145(6): 1133-1143.

4) Smith, K. P., P. T. Moen, et al. (1999). "Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific." J Cell Biol 144(4): 617-629.

5) Johnson, C., D. Primorac, et al. (2000). "Tracking COL1A1 RNA in osteogenesis imperfecta. splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain." J Cell Biol 150(3): 417-432.

6) Shopland, L. S., C. V. Johnson, et al. (2002). "Evidence that all SC-35 domains contain mRNAs and that transcripts can be structurally constrained within these domains." J Struct Biol 140(1-3): 131-139.

7) Shopland, L. S., C. V. Johnson, et al. (2003). "Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods." J Cell Biol 162(6): 981-990.

8) Hall, L. L., K. P. Smith, et al. (2006). "Molecular anatomy of a speckle." Anat Rec A Discov Mol Cell Evol Biol 288(7): 664-675.

9) Smith, K. P., M. Byron, et al. (2007). "Defining early steps in mRNA transport: mutant mRNA in myotonic dystrophy type I is blocked at entry into SC-35 domains." J Cell Biol 178(6): 951-964.

10) Hu, Y., I. Kireev, et al. (2009). "Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template." J Cell Biol 185(1): 87-100.

11) Boisvert, F.-M., S. van Koningsbruggen, et al. (2007). "The multifunctional nucleolus." Nat Rev Mol Cell Biol 8(7): 574-585.

12) Wang, J., C. Shiels, et al. (2004). "Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic regions." J Cell Biol 164(4): 515-526.

13) Jiménez-García, L. F. and D. L. Spector (1993). "In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism." Cell 73(1): 47-59.

14) Saitoh, N., C. S. Spahr, et al. (2004). "Proteomic analysis of interchromatin granule clusters." Mol Biol Cell 15(8): 3876-3890.

15) Lin, S., G. Coutinho-Mansfield, et al. (2008). "The splicing factor SC35 has an active role in transcriptional elongation." Nat Struct Mol Biol 15(8): 819-826.

16) Champlin, D. T., M. Frasch, et al. (1991). "Characterization of a Drosophila protein associated with boundaries of transcriptionally active chromatin." Genes Dev 5(9): 1611-1621.

17) Champlin, D. T. and J. T. Lis (1994). "Distribution of B52 within a chromosomal locus depends on the level of transcription." Mol Biol Cell 5(1): 71-79.

18) Huang, Y. and J. A. Steitz (2001). "Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA." Mol Cell 7(4): 899-905.

19) Huang, S. and D. L. Spector (1996). "Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription." J Cell Biol 133(4): 719-732.

20) Yang, L., C. Lin, et al. (2011). "ncRNA- and Pc2 Methylation-Dependent Gene Relocation between Nuclear Structures Mediates Gene Activation Programs." Cell 147(4): 773-788.

21) Misteli, T. (2001). "The concept of self-organization in cellular architecture." J Cell Biol 155(2): 181-185.

22) Shevtsov, S. P. and M. Dundr (2011). "Nucleation of nuclear bodies by RNA." Nat Cell Biol 13(2): 167-173.

23) Dundr, M. (2011). "Seed and grow: a two-step model for nuclear body biogenesis." J Cell Biol 193(4): 605-606.

24) Hu, Q., Y.-S. Kwon, et al. (2008). "Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules." Proc Natl Acad Sci U S A 105(49): 19199-19204.

25) Ambrosino, C., R. Tarallo, et al. (2010). "Identification of a hormone-regulated dynamic nuclear actin network associated with estrogen receptor alpha in human breast cancer cell nuclei." Mol Cell Proteomics 9(6): 1352-1367.

26) de Lanerolle, P. and L. Serebryannyy (2011). "Nuclear actin and myosins: life without filaments." Nat Cell Biol 13(11): 1282-1288.

27) Lund, A. H. and M. van Lohuizen (2004). "Polycomb complexes and silencing mechanisms." Curr Opin Cell Biol 16(3): 239-246.

28) Poux, S., R. Melfi, et al. (2001). "Establishment of Polycomb silencing requires a transient interaction between PC and ESC." Genes Dev 15(19): 2509-2514.

29) Hernández-Muñoz, I., P. Taghavi, et al. (2005). "Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1." Mol Cell Biol 25(24): 11047-11058.

30) Métivier, R., G. Penot, et al. (2003). "Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter." Cell 115(6): 751-763.

31) Schuettengruber, B., D. Chourrout, et al. (2007). "Genome regulation by polycomb and trithorax proteins." Cell 128(4): 735-745.

32) Bantignies, F., C. Grimaud, et al. (2003). "Inheritance of Polycomb-dependent chromosomal interactions in Drosophila." Genes Dev 17(19): 2406-2420.

33) Vazquez, J., M. Müller, et al. (2006). "The Mcp element mediates stable long-range chromosome-chromosome interactions in Drosophila." Mol Biol Cell 17(5): 2158-2165.

34) Sexton, T., F. Bantignies, et al. (2009). "Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation." Semin Cell Dev Biol 20(7): 849-855.

35) Shi, B., J. Liang, et al. (2007). "Integration of estrogen and Wnt signaling circuits by the polycomb group protein EZH2 in breast cancer cells." Mol Cell Biol 27(14): 5105-5119.

36) Nassa, G., R. Tarallo, et al. (2011). "Comparative analysis of nuclear estrogen receptor alpha and beta interactomes in breast cancer cells." Mol Biosyst 7(3): 667-676.

37) Creekmore, A. L., K. A. Walt, et al. (2008). "The role of retinoblastoma-associated proteins 46 and 48 in estrogen receptor alpha mediated gene expression." Mol Cell Endocrinol 291(1-2): 79-86.

38) Lee, J.-H. and D. G. Skalnik (2008). "Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes." Mol Cell Biol 28(2): 609-618.

39) Trésaugues, L., P.-M. Dehé, et al. (2006). "Structural characterization of Set1 RNA recognition motifs and their role in histone H3 lysine 4 methylation." J Mol Biol 359(5): 1170-1181.

40) Gerber, M. and A. Shilatifard (2003). "Transcriptional elongation by RNA polymerase II and histone methylation." J Biol Chem 278(29): 26303-26306.

41) Hampsey, M. and D. Reinberg (2003). "Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation." Cell 113(4): 429-432.

42) Wang, K. C., Y. W. Yang, et al. (2011). "A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression." Nature 472(7341): 120-124.

43) Schmidt, D., P. C. Schwalie, et al. (2010). "A CTCF-independent role for cohesin in tissue-specific transcription." Genome Res 20(5): 578-588.

44) Hadjur, S., L. M. Williams, et al. (2009). "Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus." Nature 460(7253): 410-413.

45) Mishiro, T., K. Ishihara, et al. (2009). "Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster." EMBO J 28(9): 1234-1245.

46) Nativio, R., K. S. Wendt, et al. (2009). "Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus." PLoS Genet 5(11): e1000739.

47) Hou, C., R. Dale, et al. (2010). "Cell type specificity of chromatin organization mediated by CTCF and cohesin." Proc Natl Acad Sci U S A 107(8): 3651-3656.

48) Kagey, M. H., J. J. Newman, et al. (2010). "Mediator and cohesin connect gene expression and chromatin architecture." Nature 467(7314): 430-435.

7 Conclusion

Despite the fact that enhancers were discovered more than 30 years ago (1, 2), unveiling their biological functions and the mechanism by which they regulate gene expression has been relatively resistant to investigation. Recently, the sequential actions of cooperative DNA binding factors in initially establishing gene enhancers has been complemented by genomic definition of the epigenetic marks of enhancers and their activation (3, 4). A surprising finding was the induction of enhancer non-coding RNAs, often representing bidirectional transcripts, by various signaling pathways (5, 6), posing the question of their functional importance, if any.

Here, we have provided several lines of evidence that ERα regulated eRNAs are required for the proper and complete genomic response to hormone treatment, mediate local chromatin interactions as part of the activation process, and are required for the relocation of regulated transcription units from the repressive environment of the polycomb body to the activation environment of the interchromatin granules, apparently a pre-requisite for gene activation.

These studies also revealed that looping of E_2 regulated enhancers with their cognate promoters was not dependent on nuclear motor activity; however, relocation to ICGs was entirely blocked in the presence of defective NMI, as was estrogen-dependent gene activation. Thus, suggesting that the activation process is in fact the consequence of two distinct, nuclear architectural regulatory events: 1) local enhancer:promoter looping and 2) NMI-facilitated association with specific nuclear substructures.

Based on these findings, we are tempted to speculate that eRNAs are modular scaffolds likely to be employed in many regulated programs of gene expression based on their requirement for altering promoter:enhancer interactions and the association of regulated transcription units with specific subnuclear architectural structures. However, the precise definition of the molecular complexes and interactions that cause promoter activation continue to remain poorly understood, and presents a robust challenge for future investigation.

7.1 References

1) Mikkelsen, T. S., M. Ku, et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature 448(7153): 553-560.

2) Zhao, J., B. K. Sun, et al. (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." Science 322(5902): 750-756.

3) Chen, J., H. K. Kinyamu, et al. (2006). "Changes in attitude, changes in latitude: nuclear receptors remodeling chromatin to regulate transcription." Mol Endocrinol 20(1): 1-13.

4) Felsenfeld, G. and M. Groudine (2003). "Controlling the double helix." Nature 421(6921): 448-453.

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

6) Turner, R. T., B. L. Riggs, et al. (1994). "Skeletal effects of estrogen." Endocr Rev 15(3): 275-300.