UNIVERSITY OF CALIFORNIA, SAN DIEGO

Ligand mediated interactions between ER alpha enhancers on chromosome 21 results in the formation of a “distributed estradiol super enhancer”.

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

in

Biology

By

Lu Yang

Committee in charge:

Professor Michael Geoff Rosenfeld, Chair
Professor Christopher Glass
Professor James Kadonaga
Professor Mark Kemps
Professor Cornelis Murre

2017

The Dissertation of Lu Yang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

Chair

University of California, San Diego

2017

DEDICATION

Dedicated to my wife Wang Qian.

She gave up much to be with me,
I hope that my achievements will be sufficient to justify this.

EPIGRAPH

If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe that it is the atomic hypothesis…that all things are made of atoms – little particles that move around in perpetual motion.

-Richard Feynmann

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ChIP seq = **Ch**romatin **I**mmuno**P**recipitation **Seq**uencing. A technique which combines chromatin immunoprecipitation with genome wide sequencing to elucidate the genome wide binding sites of any given transcription factor.

3C = **C**hromatin **C**onformation **C**apture. A technique which allows for interactions between pairs of loci to be detected with fixation and primer amplification.

4C = **C**ircularized **C**hromatin **C**onformation **C**apture. A technique which allows the capture of interactions between one locus, "the viewpoint" and in theory all other loci.

5C = **C**hromatin **C**onformation **C**apture **C**arbon **C**opy. A technique which allows the capture of interactions between pools of donor and acceptor oligonucleotides.

3D FiSH = 3-Dimensional **F**luorescence **i**n **S**itu **H**ybridization.

Hi-C = **Hi**gh Throughput Sequencing **C**hromatin Conformation Capture.

E2 = 17β-estradiol, an agonist for estrogen receptor protein

ERα = **E**strogen **R**eceptor α, a ligand dependent sex steroid regulated transcription factor

ICI = ICI 182,780 A high affinity estrogen receptor antagonist.

TF = **T**ranscription **F**actor

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Michael Geoff Rosenfeld for his support as the chair of my graduate committee. Though many drafts of this paper and through the many days and nights of experiments, writing and editing his advice and experience have been crucial.

I would also like to acknowledge everyone who worked in the Rosenfeld lab, including our talented lab manager Kenny Ohgi, and those who do much of the work behind the scenes to keep a large lab fully stocked and functional.

Chapter 1, in part, is currently being prepared for submission for publication of the material. Nair, Sreejith; Yang, Lu; Meluzzi, Dario; Oh, Soohwan; Gamliel, Amir; Suter, Thomas; Yang, Feng; Jayani, Ranveer; Zhang, Jie; Ohgi, Kenny; Rosenfeld, Michael Geoff. "Chromosomal Enhancer Syntax: Spatially-Distributed Super Enhancers and Subnuclear Structural Associations Dictate Enhancer Robustness". The dissertation author was the primary investigator and author of this material.

VITA

2009	Bachelor of the Arts, Cornell University
2010-2012	Teaching Assistant, Department of Biology, University of California, San Diego
2017	Doctor of Philosophy, University of California, San Diego

FIELD OF STUDY

Major Field: Biology

Studies of Molecular Biology and Molecular Genetics

Professor Geoff Rosenfeld

ABSTRACT OF THE DISSERTATION

Ligand mediated interactions between ER alpha enhancers on chromosome 21 results in the formation of a “distributed estradiol super enhancer”.

By

Lu Yang

Doctor of Philosophy in Biology

University of California, San Diego, 2017

Professor Michael Geoff Rosenfeld, Chair

Discovery of transferable distal genetic elements capable of activating gene transcription was a milestone in the study of transcriptional control. These genetic elements were termed “enhancers” because of their ability to enhance proximal coding gene transcription. Since then, it was believed that the activation potential of enhancers is entirely intrinsic. However, most prior studies of enhancer elements have removed them from their native genomic context, transplanting them onto plasmids or artificial chromosomes, thereby removing any potential effects of other elements in the genome. We sought to ask the question of whether the strongest E_2 regulated enhancers on human chromosome 21 might interact, even when located on opposite ends of the chromosome q arm. These interactions play a functional role in allowing all interacting enhancers to have synergistically high levels of robustness. We have found that a cohort of the most highly active “first tier” ERα bound enhancers on chromosome 21, distributed across a linear distance of over 30 mega-bases, exhibit induced proximity when treated with the ligand agonist estradiol-17β (E_2). Using CRISPR-Cas9 technology to delete some of these enhancers, we found, these enhancers seem to confer additional robustness onto other interacting enhancers. We believe that this occurs due to the physical proximity between “first tier” enhancers gained in response to E_2, and that these interactions are dependent upon eRNA and a host of protein factors. These individual first tier enhancers in effect form a previously unappreciated “spatially distributed super-enhancer network”. These interaction events reflect physical constraints placed upon chromosome 21 imposed by a very large heterochromatic “B” compartment, which seems to predominantly localize to the surface of the nucleolus arising from the ribosomal DNA repeat rich *p*-arm of the acrocentric chromosome 21. The euchromatin “A” compartments on either side of the “B”

compartment are thus allowed to interact dynamically. These sorts of dynamic interactions between regions of the chromosome constitute, in part, the overall structure of the chromosome. Therefore, the robustness of individual enhancers in a chromosome cannot be disentangled from the contributions of all interacting enhancers and the subnuclear structures in which these interactions can occur.

INTRODUCTION

Transcriptional programs are under the control of both cis-acting DNA factors like gene proximal promoters and distal enhancers, as well as, trans-acting factors like DNA binding transcription factors and chromatin modifying enzymes. These multitudinous combinations of DNA elements and their protein binding partners coordinate to maintain specific gene expression programs involving the repression and expression of thousands of genes in cell specific and signal specific manners. Regulatory events occurring at promoters and enhancers contribute to the determination of cellular state by maintaining specific gene expression programs. Initial studies on mechanisms of transcriptional control focused upon the DNA elements and protein factors found on or near the proximal gene promoter, and most of these studies were done in prokaryotic and budding yeast systems[1,2,3]which have a relative sparsity of non-coding DNA and distal regulatory elements[4,5].

Further studies of transcriptional control have found the importance of many different types of cis-regulatory modules of gene expression located at greater distances from the proximal gene promoter, including enhancers[6,7], insulators[8] and silencers[9,10]. Of specific interest to us are enhancers, which recent studies have shown to be mostly responsible for cell specific gene expression in metazoans[11]. One attribute of enhancers which helps to explain their role in driving cell specific gene expression is their association with nuclear receptor proteins which are key signal transducers for the various cell signaling pathways involved in development[12,13]. Therefore, it seems that enhancers act to bring dynamic regulation to transcriptional systems, basically allowing cells to modify

gene transcription in response to a variety of stimuli from developmental (hormonal) cues, to infections by pathogens, to maintenance of energy homeostasis.

Given the functional importance of enhancers to dynamic regulation of gene expression programs, it is no surprise that much attention has been paid to identifying functionally important enhancers genome-wide in various organisms and cell types. Genomic methods used to identify putative enhancer elements genome-wide have found over 400,000 distinct enhancers from a set of human cell lines, with the estimates for total number of enhancers in all cell types in the human genome numbering around one million[14]. Experiments using these genomic approaches have found illuminated the genetic, epigenetic and architectural features which define active enhancers: DNase I hypersensitivity[15], enrichment for unstable nucleosome variants H2AZ[16,17] and H3.3[18,19], specific post-translation modifications of adjacent nucleosomes including H3K4me1[20], H3K4me2[21] and H3K27ac[22], and finally by the active transcription of noncoding enhancer RNAs (eRNAs).

With the influx of genome-wide sequencing data, it has become apparent that enhancers don't exist as isolated units, with singular enhancers controlling the expression of singular coding genes; instead, many enhancers seem to cluster in groups with individual enhancers often less than 12.5kb from each other. In addition, prior research had demonstrated the increased potency of multimerized regulatory elements in proximity[23,24] ; this principle, when applied to whole enhancers suggests that groups of enhancers may have unique properties separate from their individual constituents. The various research groups involved gave different names to the phenomenon of clustered and proximal enhancers: "shadow enhancers[25]" "stretch enhancers[26]" and "super-enhancers[27]".

Regardless of the names used, one attribute of these grouped enhancers seems to be that individual enhancers have synergistic effects on one another, therefore, the whole "super-enhancer" has a higher collective activity than the sum of all component enhancers acting in isolation. Early work on this topic showed that a "shadow enhancer" distal from the primary proximal enhancer in Drosophila development gives robustness (resistance to environmental challenges like heat shock) to transcription of the snail gene, necessary for the process of gastrulation in fly development.[28]

In the mouse genome, important cell specific genes were found often associated with collections of enhancers spanning tens or hundreds of kb of linear DNA, these were called "super-enhancers"[29]. Super enhancers are mainly collections of enhancers with less than 12.5 kb of separation between them, acting to produce synergistically high transcriptional output of target coding genes. These cell fate determining super-enhancers are marked by very high binding of several key transcription factors including MED1 protein and are highly enriched for typical marks of enhancers including H3K27ac and H3K4me1. However, since the primary criterion for super enhancer classification is a higher density of the relevant transcription factor binding, approximately 15% of these "super-enhancers" contained only a single enhancer element[30]; therefore, super-enhancer can be either unusually strong singular enhancers, or collections weaker enhancers close in proximity.

Using the model system of ER positive breast cancer to study the E_2 mediated transcriptional program in MCF7 cells our lab found other independent marks of the most robust enhancers. Firstly, the most robust ERα regulated enhancers in MCF7 cells also had the highest levels of eRNA transcription[31]; this finding has been confirmed by others also

working in the MCF7 cell system[32]. Secondly, we found a strong predictor of highly robust enhancers in MCF7 cells was the ERα dependent recruitment, *in trans*, of a large complex of DNA binding transcription factors including RARα, FOXA1 and GATA3 (MegaTrans) to the ERα-bound enhancer[33]. This provided an independent criterion for identification of the most important functional enhancers in MCF7 cells, only 30% of which fit the commonly used definition of super-enhancers established by the Young lab. Active enhancers are non-uniformly distributed on any given chromosome; however, it is unknown if enhancers activated by the same regulated transcription factors directly interact in the nucleus to participate in coordinated transcriptional programs. The genome is divided into thousands of "topological domains" or "topologically associated domains" (TADs), these megabase sized units tend to confine interactions within themselves, intra-TAD interactions outnumbering inter-TAD interactions by a ratio of two to one[34,35]. Each TAD contains possibly tens of differing code genes and enhancers; therefore, it seems natural for TAD boundaries to place constrains on the looping interactions of enhancers and genes in different TADs.

Fractal-globule models of genome organization have been proposed to explain interactions up to several megabases based on Hi-C analysis[36,37], more recent data indicates that loop domains isolate chromosomal segments into regions permitting local gene activation events based upon a loop extrusion model[38,39]. In addition to models generated by Hi-C[40] and FISH[41] approaches, recent multiplexed FISH experiments have provided evidence for high cell-to-cell variability in 3D chromosomal architecture[42]. While much emphasis has been placed on the results of Hi-C and 5C data that indicates the presence of local interactions[43,44], a possibility which hasn't been fully explored is that even enhancers

in distant TADs can occasionally interact or be brought into physical proximity due to three-dimensional folding and looping at the highest level of chromosomal organization[45]; these distal enhancers thus have then potential to form a transcriptionally functional enhancer network distributed across the chromosome.

The large transcriptional program regulated by E_2 in breast cancer cells provides an accessible model to investigate potentially overlooked mechanisms of enhancer activation and control of transcriptional programs. ERα is a ligand dependent sex steroid regulated transcription factor that mediates most of the biological effects of estrogens, primarily at the level of gene transcription[46,47]. Estrogen receptor alpha binds to tens of thousands of EREs found genome wide, a subset of which harbor the identifying histone marks of enhancers. ERα does not act alone on enhancers, it recruits a large cohort of coactivators[48] which are responsible for its ability to facilitate enhancer to promoter looping and subsequently to increase the transcription of nearby coding genes. On highly active enhancers, the list of recruited co-factors includes proteins such as p300[49], SRC proteins[50,51,52], Mediator proteins including TRAP220/MED1[53], Cohesin and Condensin I and II[54] subunits as well as the MegaTrans complex[55], all of which seem to have some role in the activation of eRNA transcription and target coding gene transcription. Whether all these proteins are present on an active enhancer simultaneously, or whether these proteins are rapidly associating and disassociating from the enhancer without all being present at the same time is still a widely-debated topic.

Here, we provide evidence suggesting that a subset of the most robust “first tier” of the ERα enhancers on chromosome 21, separated by as much as 33 million base pairs of linear sequence are brought non-simultaneously into spatial proximity to form a dynamic

E_2 regulated enhancer network. Two clusters of these “first tier” enhancers are separated by an approximately 14 million base pair long B compartment. The B compartment has boundaries which exhibit high basal, E_2 independent interactions and is primarily localized to the surface of the nucleolus or the lamina. Unexpectedly, the strongest ERα bound enhancers on chromosome 21 exhibited rapid, dynamic and non-coincident alterations in topography, coming into closer proximity outside of the nucleolus in response to E_2, and acting synergistically with other strong enhancers on the cis-chromosome. These events generate a large de facto, “distributed super enhancer network” which obtains synergistically high levels of robustness from the contributions of all the individual component enhancers despite of the large linear distances separating the components of this “distributed super enhancer”. This unexpected new data seems to imply the 3D structure of the interphase chromosome 21 in MCF7 cells is set up in such a way that it is both possible and encouraged for individual enhancers located many TADs apart to gain proximity and synergy upon addition of E_2 ligand. Our data provides a largely overlooked structural regulatory strategy by which ligand-induced networks of enhancers combinatorically converge and interact in space; improving the robustness of the component regulated enhancers in the chromosomal enhancer network, without affecting the function of other unrelated enhancers.

BIBLIOGRAPHY

[1]Gruber, T. and Gross, C. (2003). Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. Annual Review Of Microbiology *57*, 441-466.

[2] Haugen, S., Ross, W., and Gourse, R. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nature Reviews Microbiology *6*, 507-519.

[3] Ghosh, T., Bose, D., and Zhang, X. (2010). Mechanisms for activating bacterial RNA polymerase. FEMS Microbiology Reviews *34*, 611-627.

[4]Mira, A., Ochman, H., and Moran, N. (2001). Deletional bias and the evolution of bacterial genomes. Trends In Genetics *17*, 589-596.

[5]Comeron, J. (2001). What controls the length of noncoding DNA?. Current Opinion In Genetics & Development *11*, 652-659.

[6] Levine, M. (2010). Transcriptional Enhancers in Animal Development and Evolution. Current Biology *20*, R754-R763.

[7] Bulger, M. and Groudine, M. (2011). Functional and Mechanistic Diversity of Distal Transcription Enhancers. Cell *144*, 825.

[8] Gaszner, M. and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nature Reviews Genetics *7*, 703-713.

[9] Petrykowska, H., Vockley, C., and Elnitski, L. (2008). Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. Genome Research *18*, 1238-1246.

[10] Ayer, S. and Benyajati, C. (1990). Conserved enhancer and silencer elements responsible for differential Adh transcription in Drosophila cell lines. Molecular And Cellular Biology *10*, 3512-3523.

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

[12]John, S., Sabo, P., Thurman, R., Sung, M., Biddie, S., Johnson, T., Hager, G., and Stamatoyannopoulos, J. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature Genetics *43*, 264-268.

[13] Carroll, J., Meyer, C., Song, J., Li, W., Geistlinger, T., Eeckhoute, J., Brodsky, A., Keeton, E., Fertuck, K., and Hall, G. (2006). Genome-wide analysis of estrogen receptor binding sites. Nature Genetics *38*, 1289-1297

[14] Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R.; ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

[15] Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Intergenic Transcription and Developmental Remodeling of Chromatin Subdomains in the Human β-globin Locus. Molecular Cell *5*, 377-386.

[16] Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant–containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nature Genetics *41*, 941-945.

[17] Barski, A., Cuddapah, S., Cui, K., Roh, T., Schones, D., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell *129*, 823-837.

[18]Jin, C. and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & Development *21*, 1519-1529.

[19] Wang, Z., Zang, C., Rosenfeld, J., Schones, D., Barski, A., Cuddapah, S., Cui, K., Roh, T., Peng, W., and Zhang, M. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nature Genetics *40*, 897-903.

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

[21] He, H., Meyer, C., Shin, H., Bailey, S., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., and Lupien, M. (2010). Nucleosome dynamics define transcriptional enhancers. Nature Genetics *42*, 343-347.

[22] Creyghton, M., Cheng, A., Welstead, G., Kooistra, T., Carey, B., Steine, E., Hanna, J., Lodato, M., Frampton, G., and Sharp, P. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings Of The National Academy Of Sciences *107*, 21931-21936.

[23] Carey, M. (1998). The Enhanceosome and Transcriptional Synergy. Cell *92*, 5-8.

[24] Chi, T., Lieberman, P., Ellwood, K., and Carey, M. (1995). A general mechanism for transcriptional synergy by eukaryotic activators. Nature *377*, 254-257.

[25] Hong, J., Hendrix, D., and Levine, M. (2008). Shadow Enhancers as a Source of Evolutionary Novelty. Science *321*, 1314-1314.

[26] Parker, S., Stitzel, M., Taylor, D., Orozco, J., Erdos, M., Akiyama, J., van Bueren, K., Chines, P., Narisu, N., and Black, B. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proceedings Of The National Academy Of Sciences *110*, 17921-17926.

[27] Hnisz, D., Abraham, B., Lee, T., Lau, A., Saint-André, V., Sigova, A., Hoke, H., and Young, R. (2013). Super-Enhancers in the Control of Cell Identity and Disease. Cell *155*, 934-947.

[28] Perry, M., Boettiger, A., Bothma, J., and Levine, M. (2010). Shadow Enhancers Foster Robustness of Drosophila Gastrulation. Current Biology *20*, 1562-1567.

[29] Whyte, W., Orlando, D., Hnisz, D., Abraham, B., Lin, C., Kagey, M., Rahl, P., Lee, T., and Young, R. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307-319.

[30] Pott, S. and Lieb, J. (2014). What are super-enhancers?. Nature Genetics *47*, 8-12.

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

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

[33] Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M., Song, X., Zhang, F., Ma, Q., and Ohgi, K. (2014). Enhancer Activation Requires trans-Recruitment of a Mega Transcription Factor Complex. Cell *159*, 358-373.

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

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

[36] Lieberman-Aiden, E., van Berkum, N., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B., Sabo, P., and Dorschner, M. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science *326*, 289-293.

[37] Mirny, L. (2011). The fractal globule as a model of chromatin architecture in the cell. Chromosome Research *19*, 37-51.

[38] Sanborn, A., Rao, S., Huang, S., Durand, N., Huntley, M., Jewett, A., Bochkov, I., Chinnappan, D., Cutkosky, A., and Li, J. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proceedings Of The National Academy Of Sciences *112*, E6456-E6465.

[39] Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Reports *15*, 2038-2049.

[40] Dekker, J., Marti-Renom, M., and Mirny, L. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature Reviews Genetics *14*, 390-403.

[41] Giorgetti, L. and Heard, E. (2016). Closing the loop: 3C versus DNA FISH. Genome Biology *17*.

[42] Wang, S., Su, J., Beliveau, B., Bintu, B., Moffitt, J., Wu, C., and Zhuang, X. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. Science *353*, 598-602.

[43] Markenscoff-Papadimitriou, E., Allen, W., Colquitt, B., Goh, T., Murphy, K., Monahan, K., Mosley, C., Ahituv, N., and Lomvardas, S. (2014). Enhancer Interaction Networks as a Means for Singular Olfactory Receptor Expression. Cell *159*, 543-557.

[44] Proudhon, C., Snetkova, V., Raviram, R., Lobry, C., Badri, S., Jiang, T., Hao, B., Trimarchi, T., Kluger, Y., and Aifantis, I. (2016). Active and Inactive Enhancers Cooperate to Exert Localized and Long-Range Control of Gene Regulation. Cell Reports *15*, 2159-2169.

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

[46] Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J., Argos, P., and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. Nature *320*, 134-139.

[47] Saji, S., Jensen, E., Nilsson, S., Rylander, T., Warner, M., and Gustafsson, J. (2000). Estrogen receptors alpha and beta in the rodent mammary gland. Proceedings Of The National Academy Of Sciences *97*, 337-342.

[48] Shang, Y., Hu, X., DiRenzo, J., Lazar, M., and Brown, M. (2000). Cofactor Dynamics and Sufficiency in Estrogen Receptor–Regulated Transcription. Cell *103*, 843-852.

[49] Hanstein, B., Eckner, R., DiRenzo, J., Halachmi, S., Liu, H., Searcy, B., Kurokawa, R., and Brown, M. (1996). p300 is a component of an estrogen receptor coactivator complex. Proceedings Of The National Academy Of Sciences *93*, 11540-11545.

[50] Halachmi, S., Marden, E., Martin, G., MacKay, H., Abbondanza, C., and Brown, M. (1994). Estrogen receptor-associated proteins: possible mediators of hormone-induced transcription. Science *264*, 1455-1458.

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

[52] Xu, J. and Li, Q. (2003). Review of the *in Vivo* Functions of the p160 Steroid Receptor Coactivator Family. Molecular Endocrinology *17*, 1681-1692.

[53] Kang, Y., Guermah, M., Yuan, C., and Roeder, R. (2002). The TRAP/Mediator coactivator complex interacts directly with estrogen receptors and through the TRAP220 subunit and directly enhances estrogen receptor function in vitro. Proceedings Of The National Academy Of Sciences *99*, 2642-2647.

[54] Li, W., Hu, Y., Oh, S., Ma, Q., Merkurjev, D., Song, X., Zhou, X., Liu, Z., Tanasa, B., and He, X. (2015). Condensin I and II Complexes License Full Estrogen Receptor α-Dependent Enhancer Activation. Molecular Cell *59*, 188-202.

[55] Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M., Song, X., Zhang, F., Ma, Q., and Ohgi, K. (2014). Enhancer Activation Requires trans-Recruitment of a Mega Transcription Factor Complex. Cell *159*, 358-373.

CHAPTER 1: A spatially distributed E_2 regulated super-enhancer network on chromosome 21 in MCF7 cells.

RESULTS

Detecting estrogen mediated chromatin alterations in MCF7 cells:

To investigate any potential relationship between robust ERα-bound enhancers located at great linear distances within a chromosome, we have focused on the set of coding target genes and ERα bound activated enhancers on chromosome 21. We have identified 179 enhancers marked by the presence of ERα, H3K27Ac, and H3K4me1[1], amidst a total of 479 enhancers on chromosome 21. Of these 179 ERα bound enhancers, 39 were particularly highly active enhancers (Table 1), or super-enhancers, characterized by binding components of the MegaTrans complex, the other 140 ERα bound enhancers had considerably less ERα binding and were also less significantly activated by addition of E_2 hormone. The relative locations of the major induced messenger RNAs are shown schematically (Figure 1), on this figure the *p* arm and the centromere of the chromosome are omitted because they are mostly devoid of E_2 responsive coding genes. A distinguishing karyotypic feature of chromosome 21 is the short *p* arm, making it an acrocentric chromosome. Much like the other acrocentric autosomes 13,14,15 and 22[2]; the short *p* arm is home to a nucleolar organizing region (NOR), which contains genes encoding the 5.8S, 18S and 28S ribosomal RNAs. The nucleolus is self-assembled onto the NORs meaning that the *p* arm of interphase chromosome 21 is imbedded in the nucleolus, linking chromosomal architecture to nuclear ribonucleoprotein (RNP) bodies, of which the nucleolus is the largest.

To explore basal and E_2 induced alterations on chromosome 21 architecture, we initially performed Hi-C, as modified by the *in-situ* fixation method[3], to increase efficacy of the experimental protocol. Hi-C was sequenced to a depth of $200x10^6$ unique reads for each condition, analyzed by generating z score of contact counts. These analyses indicated the broad area of B compartment spanning roughly 14Mb in the middle of the *q* arm of chromosome 21 (Figure 2). Treatment with E_2 ligand did not substantially alter either the A/B compartment or the chromosomal boundaries, as determined by the insulation score, analyzed across the chromosome. When analyzed at a resolution of 1 megabase, the data also indicated the presence of extremely long distance interactions (Figure 3), although not having the resolution to permit precise identification of specific, super long distance DNA interactions. Indeed, even at 3 billion assignable reads, such long-distance interactions between specific genomic regions still cannot be clearly ascertained[4].

An initial assessment of the effects of E_2 on enhancer function was provided by ATAC-seq which revealed the increased openness of the 39 strongest "first tier" ERα enhancers in response to E_2. The most robust effects of E2 ligand on the openness of chromatin were observed on a few enhancers, namely TFF1e1, NRIP1e1-3, COL18A1e1, TMPRSS2e1 and DSCAMe1-2. In contrast, E_2 did not cause further openness of the less robust ERα bound enhancers on chromosome 21, nor did it alter the openness of the eighty-two enhancers not bound by ERα (Figure 4). To characterize "first tier" ERα enhancers, we examined the distribution of several classical coactivators for estrogen receptor. Of importance amongst the canonical coactivators, the CBP/P300 histone acetyltransferase coactivator complex is well known as a haplo-insufficient factor in human development; mutations causing loss of an allele of either CBP or P300 result in Rubinstein-Taybi

syndrome[5]. P300 protein is a cofactor for estrogen receptor, therefore, we wanted to observe the dynamics of P300 binding to estrogen enhancers. The "first tier" ERα enhancers on chromosome 21 exhibited a roughly 400% increase in P300 binding post E_2 treatment, meanwhile, the weaker ERα enhancers on chromosome 21 and the non-ERα enhancers exhibited decreased association of P300 following E_2 treatment (Figure 4). This pattern of P300 redistribution is indicative of rate-limiting factor P300 being removed from relatively inactive enhancers to activate the "first tier" ERα enhancers. Analysis carried out on ERα protein, MED1 protein, FOXA1 protein and Pol II protein (Figure 5) found a similar pattern wherein most of the binding gained with E_2 treatment could be attributed to the "first tier" ERα enhancers, while weaker ERα enhancers gained little ERα, and Pol II and with no apparent increase in MED1 and FOXA1 proteins and non- ERα enhancers showed only a slight gain in Pol II tag count.

To visualize how binding of factors P300, MED1, AP2γ, GATA3 and FOXA1 changed on individual enhancers, we made schematic representations of the ChIP-seq tag counts of these factors on ten top ERα enhancers in both ICI and E_2 treated conditions (Figure 6). These experiments showed that these ten ERα enhancers gained binding of most or all cofactors in response to E_2 ligand treatment. One enhancer which stood out was the TFF1 enhancer one (TFF1e1), which had the highest binding of factors P300, MED1 and FOXA1 factors (Table 2). One unique feature of TFF1e1 is the bimodal binding distribution of factors ERα, P300, and, MED1 this bimodal binding of transcription factors may explain why this enhancer has the highest binding of P300 and MED1 (Figure 7). The proximal coding gene regulated by TFF1e1 is the TFF1 gene; this gene has been used as a prognostic marker for the diagnosis of ER+ breast cancers since the 1980s[6] although the

mechanistic contribution of TFF1 overexpression to ER+ breast cancer pathology is still debated. Another collection of enhancers which stood out as being of interest is NRIP1e1-3, which is notable for being the most active super-enhancer in the relatively isolated A compartment on the 5' side of chromosome 21 q arm. Even more surprisingly, prior research in the lab had shown that the NRIP1 gene and the TFF1 gene gained spatial proximity in FISH experiments when MCF7 cells were treated with E_2 ligand, prompting us to question whether the regulatory enhancers of these genes were involved in the process of NRIP1 and TFF1 coming closer.

E_2-induced long-distance interactions of Chr 21 regulatory interactions.

Based on Hi-C data suggesting very long distance interactions and previous suggestive FISH experiments, we evaluated the effects of E_2 ligand on chromosomal architecture and the induced proximity of "first tier" ERα enhancers using FISH. In this procedure, cells were processed for 3-D DNA FISH using probes generated from either bacterial artificial chromosomes (BACs) or fosmids. Fifteen probes spread over the length of the q arm of chromosome 21 were chosen, most probes targeting ERα enhancers expect for the Het1 probe and 43.3Mb probes which did not target any enhancers (Figure 8). Acquired images were processed using Volocity Software to calculate voxel overlaps (an overlap of at least two voxels) as a measure of genomic interaction. In addition, the 3-D coordinates generated by the software were used to estimate the spatial distance between genomic loci using custom scripts where only nuclei with the expected number of enhancer loci where considered for calculation. Pairs of enhancers which interact are generally

separated by 550nm of spatial distance or less; which, as a point of reference, is the volume occupied by a length of DNA roughly 100,000-200,000 base pairs in length[7].

After treating MCF7 cells with E_2 for 50 minutes, the colocalization rates and average distances were measured for six pairs of enhancers: DSCR3/TIAM1(5.7Mb), DSCR3/COL18A1(8.1Mb), NRIP1/DSCR3 (22.0Mb), NRIP1/DSCAM-AS1 (25.2Mb), NRIP1/TFF1 (27.0Mb), and NRIP1/COL18A1 (30.3Mb). All six pairs of enhancers had basal levels of interaction which were increased by 2 to 4-fold following the addition of E_2 ligand for 50 minutes (Figure 9), indicative of the strong effect E_2 ligand has upon facilitating long range chromosomal interactions. The closest enhancer pair DSCR3/TIAM1 (5.72Mb) had the highest rate of interaction, twelve percent, in the E_2 treated conditions; however, rather surprisingly, the most distant pair of enhancers NRIP1/COL18A1(30.3Mb) had almost as high a rate of E_2 mediated interaction at nine percent. This piece of data demonstrates the non-linear relationship between one dimensional genomic distance and E_2 mediated interaction rates for pairs of enhancers. A consequence of this startling piece of data seems to be that the NRIP1/COL18A1 (30.3Mb apart) enhancer pair are within 550nm distance in roughly one out of eleven MCF7 cells treated with E_2; in these cells, the 3D distance from NRIP1 enhancer to COL18A1 enhancer is approximately equal to the 3D distance from NRIP1 enhancer to the NRIP1 gene promoter (100kb apart), making it seem likely that, in these nine percent of cells, the two enhancers could meaningfully interact and influence each other's behavior.

In addition to looking at the rates of overlap between enhancers, we analyzed the cumulative probability distribution of the spatial distances between these enhancer pairs. We found the median spatial distance between five out of six pairs of enhancers we looked

at were also decreased in a E_2 ligand dependent manner, except for the NRIP1/COL18A1 enhancer pair which showed no significant change (Figure 10). Upon inspection of the NRIP1/COL18A1 interaction, a closer spatial distance can be seen for the E_2 condition if we look at the 10^{th} percentile cutoff instead of the 50^{th} percentile; implying that the spatially closest 10% of NRIP1/COLA18A1 enhancer pairs get even closer with ligand. An analysis of all pairs of ERα enhancer probes in MCF7 cells found that E_2 ligand caused a decrease in the median spatial distance between enhancer probes, which is indicative of a general compaction of probed ERα enhancers following E_2 treatment (Figure 11). To explore whether the compaction of ERα enhancers observed by FISH could also be captured by chromatin crosslinking techniques, we performed a 4C experiment[8] using the TFF1 enhancer as a "viewpoint". The TFF1 enhancer was chosen as a viewpoint for this experiment because of its proximity to many other strong ERα enhancers (Figure 1), this attribute is important given the limited range (1.5-3Mb) of interactions capturable with the 4C technique. Our analysis of the 4C experimental data found an E_2-induced interaction between TFF1 enhancer and the DSCAM-AS1 super-enhancer region, located approximately 2 million base pairs apart. The DSCAM-AS1 super-enhancer is located roughly 40kb from the promoter of the DSCAM-AS1 lncRNA; the most highly transcribed E_2 regulated lncRNA. While not much is known about the function of the DSCAM-AS1 lncRNA, a recent study found it to be associated with ER+ breast cancer, interact with hnRNPL protein, and to potentially play a role in mediating tumor progression and tamoxifen resistance[9]. Our 4C analysis showed a weak E_2 independent interaction between TFF1 enhancer and the promoter of the DSCAM-AS1 lncRNA, additionally, there was a strong interaction of TFF1 enhancer with the DSCAM-AS1 super-enhancer which was

dependent on E_2 (Figure 12). Thus, in concert with FISH data, the 4C results independently confirm that E_2 can induce long range enhancer-enhancer interactions.

The B compartment of Chromosome 21.

We next elected to investigate the basal and E_2-induced interactions of enhancers proximal to the extended B compartment of chromosome 21, as revealed by Hi-C, spanning the region 18 to 32Mb (Figure 2). This large B compartment had the typical epigenetic marks of repressed chromatin, enrichment for marks $H3K27Me^3$, $H3K9me^3$ and $H4K20me^3$, thus fitting the epigenetic features of B domains previously reported[10]. Two ERα bound enhancers (NCAM2, TIAM1) fall within this region. We therefore interrogated the interactions between loci in and near the B compartment; we evaluated basal and E_2 induced interactions between pairs of enhancers NRIP1/NCAM2 (6Mb), NRIP1/TIAM1 (16.5Mb) and NRIP1/DOPEY2 (21.3Mb) (Figure 13). These pairs of enhancers had high basal rates of interaction, with the basal rate of interaction decreasing as genomic distance increased. Interestingly, the rates of interaction between these pairs of enhancers either decreased with E2 treatment like NRIP1/NCAM2, or showed no significant change with E2 treatment. We next looked at the cumulative probability distribution for the spatial distances between these three pairs of enhancers to get a better idea about why they showed no increase in E2 mediated interaction. These three pairs of enhancers showed no statistically significant change to the median spatial distance in the basal and E2 treated states (Figure 14); the NRIP1/DOPEY2 spatial distance was even lower in the basal condition; suggesting that E2 treatment moved them further apart.

After probing the behavior of ERα enhancers in or near the extended B-domain on chromosome 21, we next wanted to study the behavior of the heterochromatic regions in the B-domain not marked by epigenetic marks of enhancers. We used a probe in the B compartment at megabase 25 (B compartment probe: BCP), as a representative of the B compartment. The B compartment probe behaved very differently with respect to the NRIP1(10Mb 5') and TFF1(17Mb 3') enhancers. The B compartment showed a very high basal level of interaction with NRIP1 (12%), which decreased with E_2 treatment (9.6%), on the other the TFF1 enhancer had a very low basal level of interaction with the BCP (1.5%), which did not change much with E_2 treatment (1.3%) (Figure 15). As seen before, NRIP1 and TFF1 had a moderate level of basal interaction (4%) which increased in the presence of E_2 ligand (7%). From this data, we can draw some conclusions about the overall conformation of chromosome 21, mainly, NRIP1 has constitutive proximity with the B compartment, while TFF1 is always distal to the B compartment. This data would be congruent with a model where TFF1 interacts with NRIP1 by forming a large loop, this can be seen from the FISH images showing TFF1 interacting with NRIP1 from the opposite direction of the 25mB BCP.

Dynamic Interactions of ER enhancers form a Distributed Super-Enhancer Network

FISH data from many pairs of ERα enhancers on chromosome 21 colocalizing and gaining proximity in response to E_2 treatment suggested the possibility that the interactions between distal enhancers brought into proximity may have measurable effects on the E_2 regulated transcriptional program in MCF7 cells. To visualize how a collection of enhancers change their median spatial distances in response to E_2 ligand, we collected the

median spatial distance data between NRIP1 and seven other ERα enhancers in the basal and E_2 treated conditions (Figures 10 and 14) and made them into a single figure showing the E_2 regulated median spatial distances between NRIP1 and the other enhancers (Figure 16). In the basal condition, there is a strong association between one-dimensional genomic distance and three-dimensional spatial distance; where the enhancer closest to NRIP1 in genomic distance is also closest in spatial distance, this relationship holds true for all seven enhancers. This configuration of the chromosome is what one might expect from a relaxed polymer without the presence of chromosome scale loops. On the other hand, in the E_2 treated MCF7 cells, genomic distance and spatial distance are less closely associated. In E_2 treated cells, DSCR3 becomes spatially closer to NRIP1 than DOPEY2, and TFF1 becomes closer than DSCAM-AS1 in contrast to their genomic distances. One way to understand this data would be the presence of large-scale looping interactions, wherein, DSCR3 and TFF1 are looped closer to NRIP1 at the expense of other loci being looped out.

An interesting consequence of the E_2 induced proximity of some ERα enhancer loci is the effect of this proximity on the promotion of further ERα enhancer-enhancer interactions. For example, the NRIP1 and TFF1 enhancers have a median spatial distance of 1.86μm in the ICI condition and 1.52μm in the E2 condition. This can be geometrically modeled by placing one locus at the center of a sphere with a radius R equal to the median spatial distance, in 50% of cells, the other locus with be inside of this sphere (Figure 16). After treatment of cells with E_2 ligand for 50 minutes, the volume of this sphere shrinks from 26.95μm^3 to 14.71μm^3, almost halving the volume of space between NRIP1 and TFF1. Since we believe that these ERα enhancer-enhancer interactions arise from the diffusive motion of the chromosome, by reducing the volume of space between two loci by 45%; the

rate of interaction between these loci should therefore increase by roughly 82%. Of course, not all enhancer loci gain proximity with E_2 treatment; and specific changes in distance between two loci depend upon the percentile cutoff we choose to study in the cumulative probability distribution. However, using the median spatial distance as an approximate estimate of the movement of enhancer loci can be a useful tool for understanding large-scale conformational changes induced in the chromosome. A list of the median spatial distances, volume of space, and percent changes in these stats are shown (Table 3).

From these datasets, we propose that the stochastic and dynamic interactions between random sets of "first tier" ERα enhancers could form what amounts to a "distributed super-enhancer network"; a chromosome wide enhancer network made of individual enhancers and super-enhancers. Random sets of individual enhancers distributed across the length of chromosome twenty-one converge on some three-dimensional point, and upon interaction, form a ribonucleoprotein complex which synergistically increases the robustness of the E_2 transcriptional program in MCF7 cells. Although the idea of inter-TAD interactions between ERα enhancers on the same chromosome may seem unlikely, given the large genomic distances involved, our FISH data clearly shows that in a significant fraction (9%) of MCF7 cells treated with E_2 ligand, enhancers located thirty-three megabases apart can still interact in three-dimensional space. Our data regarding long distance contacts of enhancers on the same chromosome is reminiscent of data regarding the three-dimensional convergence of several olfactory receptor enhancers on different chromosomes[11]. Our system has one advantage for study of the phenomenon of enhancer convergence, mainly, the ERα enhancer interactions in MCF7 cells are largely E_2 dependent and therefore can be studied at precise times points after E_2 treatment.

To systematically model the types of interactions which could result in the formation of a distributed superenhancer network, we represented the convergence of enhancers as states in a Markov chain, which can be used to model possible convergent states for any given number of enhancers. For example, we can build a simple two state Markov chain by using observed interaction frequencies from the NRIP1 and TFF1 enhancers to infer the steady-state distribution of the Markov chain (Figure 17). Unfortunately, without real-time video of the movement of these two loci, it will be impossible to determine the transition probabilities of the Markov chain model, this is the main downside of using formaldehyde fixed cells for experiments. Next, we built a five state Markov chain model using interaction data from three color FISH experiments with probes targeting the NRIP1, TFF1 and COL18A1 enhancers. As expected, after E_2 treatment, pairs of enhancer interactions were more frequent than the triad of enhancers colocalizing together. However, the observed frequency of the three loci interaction exceeds the expected probability of interaction calculated from the frequency of pairs of enhancers interacting independently (.039*.046*.044=.000079 vs .004 observed) by roughly 50-fold (Figure 18). The relative abundance of the three loci interaction supports a model wherein random enhancer interactions result in the formation of a ribonucleoprotein structure which, in turn, stabilizes these interactions. A three-color FISH experiment using probes for NRIP1, TIAM1 and TFF1 found different results. Here, the interaction of TFF1 and TIAM1 is increased by E2 treatment, while the NRIP1 and TIAM1 interaction was decreased; suggesting these interactions may be mutually exclusive to some degree, as evidenced by the fact that no three loci interactions were observed (Figure 19).

As the number of enhancers being observed increases, the number of possible convergent enhancer states also increases combinatorically. Therefore, in a hypothetical four-color FISH experiment there would be a total of 11 unique interactions states in the Markov chain, with a total of 121 transitions possible between these states (Figure 20). If we were to increase the number of enhancers being observed to ten top enhancers (Figure 6), where would be a total of 45 states of two enhancer interactions, 120 states of three enhancer interactions, 210 states of four enhancer interactions etc. Meanwhile, the number of transition probabilities increases with the square of the number of states. In theory, any of the enhancer interaction states is possible; although, *in vivo*, many states are either impossible or highly improbable, as we observed from the lack any three loci interactions between TFF1, NRIP1 and TIAM1. When we apply this model to the idea of a distributed super-enhancer network we can see that any number of nodes in the enhancer network could become the basis for the formation and nucleation of a ribonucleoprotein complex to assemble. However, since the probability of three and four enhancer interactions is much lower than those of two enhancer interactions, an even larger sample size of cells would be necessary to make statistically significant observations about the steady state distribution of these states. The lower probability of these interactions is offset, in part, by the larger number of possible three and four enhancer interaction states in any sufficiently large set of enhancers.

Functional significance of the enhancer network:

To test our theoretical model of an ERα enhancer network, it was critical to determine the functional importance of E_2 induced interactions between members of the

distributed super-enhancer on chromosome 21. We employed CRISPR-Cas9 based gene editing approaches[12,13] relying on pairs of guide RNAs and the non-homologous end joining repair pathway to generate deletions of two especially interesting ERα enhancers, NRIP1e3 (Figure 21) and TFF1e1 (Figure 22). To minimize potential off-target effects of gRNA used, we generated multiple cell lines for each deletion and used independent sets of gRNAs to generate the same deletion. Initial genotyping of the enhancer deletions was done by PCR, using primers outside of the deletion to amplify the whole region (Figure 23). After this, TOPO cloning was used to ligate the deletion PCR fragments into a vector, and Sanger sequencing was used to identify the deletion junctions made by the process of NHEJ; in most cases the deletion junctions correspond very closely to the expected sites of gRNA binding (Figures 24 and 25).

To characterize the effects of the TFF1e1 deletion on the gene expression program in MCF7 cells, we analyzed the expression of three highly expressed E_2 induced coding genes on chromosome 21: NRIP1, TMPRSS2 and TFF1. To reduce the chance that what we observed is due to random variations in gene expression levels from stochastic variations in single cell derived clones of MCF7 cells, four independent clones were used, each of which was analyzed in three separate experiments. In the TFF1e1 deletion cell lines, the E_2 induced expression levels of the cognate TFF1 mRNA was reduced by more than 98%, TMPRSS2 mRNA expression was reduce by roughly 50%, and NRIP1 mRNA expression was reduced by roughly 30% (Figure 26). The largest effect of the TFF1e1 deletion was on the nearby TFF1 gene which is only 5 Kb from the TFF1 enhancer, less pronounced effects were seen on TMPRSS2 1 Mb away, and a slight effect was seen on NRIP1 located 27 Mb away. In NRIP1e3 deletion cell lines E_2 induced expression levels

of NRIP1 mRNA was reduced by 50%, TMPRSS2 mRNA was reduced by 50%, and TFF1 mRNA was reduced by 30% (Figure 27). Deletion of the NRIP1e3 enhancer had a weaker effect on the expression of genes near and far, this is because the NRIP1 superenhancer is composed of three enhancers in total (Figure 21) and therefore any single enhancer deletion would be unable to completely disrupt the function of the NRIP1e1-3 superenhancer. Another interesting phenotype of the NRIP1e3 enhancer deletion was its effect on the E_2 induced transcript levels of the non-coding TFF1 eRNAs, TFF1e1 sense transcript was reduced by 50% and TFF1e1 antisense transcript was reduced by 60% (Figure 28). Interestingly, the NRIP1e3 deletion had a more pronounced effect on expression levels of TFF1 eRNA than on TFF1 mRNA, possibly implying that the NRIP1e3 deletion's effect on TFF1 mRNA is mediated by the reduction of TFF1e1 eRNA levels.

To measure the effects of the TFF1e1 and NRIP1e3 deletion genome-wide, we performed GRO-seq on two clonal lines of TFF1e1 deletion and one clonal line of NRIP1e3 and compared them to one clonal line of wild type MCF7 cells which had undergone the same transfection and puromycin selection regiments. A meta-analysis of the GRO-seq data found that both enhancer deletions had a significant effect on the transcription of E_2 induced eRNAs from the first tier ERα enhancers on chromosome 21; in contrast, the less robust ERα enhancers and non-ERα bound enhancers were not affected (Figure 29). To better visualize the effects of the TFF1e1 and NRIP1e3 deletions on the transcription of individual enhancers on chromosome 21, we plotted circles with surface areas quantitatively depicting the RPKM values for ten top ERα enhancers in the presence and absence of E_2 ligand (Figure 30). The TFF1e1 deletion affected the robustness of the E_2 induced eRNA transcription on many distal ERα enhancers on chromosome 21, most

dramatically: NRIP1, NCAM2, DSCAM1, UMODL1 and COL18A1 enhancers; for full details (Table 4). Analysis of GRO-seq data from the NRIP1e3 deletion cell line showed that the deletion diminished the eRNA transcription of the adjacent enhancers in the superenhancer cluster by roughly 75%, as well as mRNA transcription at the NRIP1 promoter by roughly 50% (Figure 31). Surprisingly, the NRIP1e3 deletion reduced transcription of the TFF1e1 eRNAs by over 50% despite being located on the far side of the chromosome 21 q arm (Figure 32); while this data conforms with results from qPCR experiments, it was still puzzling to imagine how this "action at a distance" might be mechanistically possible. We hypothesized that these enhancer deletions may disrupt E_2 mediated changes in the conformation of the chromosome, disrupting the process by which ERα enhancers converge in 3D space after ligand treatment.

Effects of enhancer deletion/interference on the distributed superenhancer network:

To find a mechanistic explanation for the "action at a distance" observed from deletions of NRIP1e3 and TFF1e1, we wished to ascertain whether long range ligand induced enhancer interactions are effected by deletions of the enhancer sequences. FISH experiments in TFF1e1 deleted cells found that E_2 induced interactions of TFF1 and NRIP1 were no longer observed, rather, the interaction frequency decreased relative to the basal condition. Further, the E_2 induced proximity (reduction in median spatial distance) between NRIP1 and TFF1 was no longer observed, implying that whatever process caused the E_2 induced enhancer convergence could no longer function without the presence of the TFF1 enhancer sequence. (Figure 33). When we looked at the induced proximity of this enhancer pair in the NRIP1e3 deletion cell line, we found an intermediate phenotype between that

of wild type and the TFF1e1 deletion; we believe that this is because of the redundancy present in the NRIP1e1-3 superenhancer cluster which is still partially active. Similarly, E_2 induced interactions between the NRIP1 and DSCAM-AS1 super-enhancers were abolished by the deletion of NRIP1e3, and the E_2 induced proximity between these super-enhancers was also largely diminished (Figure 34). Since deletion of individual robust enhancers adversely affected activation of a cohort of robust ERα enhancers on chromosome 21, our data suggests a functional significance for the "distributed super-enhancer network" consequent to the induced long-distance interactions.

Given previous observations that eRNA may have a functional effect on promoting enhancer to promoter looping interactions[14,15,16], we wanted to investigate the role of eRNAs for robust function of the ERα distributed super-enhancer network. We therefore transfected MCF7 cells with LNAs targeting the eRNA transcripts from the TFF1 and NRIP1 enhancers to degrade these non-coding RNA transcripts. Following 24 hours of LNA treatment, the effects of the eRNA knockdown on the ERα enhancer activation program was determined using GRO-seq. Meta-analysis carried out on these data revealed that, analogous to effects of the TFF1 and NRIP1 enhancer deletions, knockdown of the TFF1e1 or NRIP1e3 eRNAs by LNA caused impairment in the activation of the other 39 "first tier" enhancers (Figure 35), whilst having no effect on the other ERα bound enhancers or the non-ER enhancers. These data provide independent evidence that the robustness of all "first tier" enhancers in the ER distributed super-enhancer network is dependent upon the transcriptional activity and integrity of its member enhancers, and that deletion or interference of TFF1e1 and NRIP1e3 (and presumably the other strong enhancers perhaps to a lesser degree) affected the activity of other members of this set. When the GRO-seq

data was plotted visually to depict the quantitative effects on ten of the most robust E2-regulated enhancers, we observed that, the pattern highly resembled, but was not as dramatic as those observed following deletion of the TFF1e1 and NRIP1e3 enhancers (Figure 36) (Table 5). In accordance with this data, treatment of MCF7 cells with two independent ASOs targeting TFF1 Sense eRNA completely disrupted the induced proximity of TFF1 and NRIP1 when compared with a scrambled ASO control sequence (Figure 37). Collectively, these data suggest that eRNAs play a critical role in these extremely long range enhancer-enhancer interactions, as well as their previous reported roles in short range enhancer-promoter interactions.

Estrogen reconfigures the topography of Chr.21 relative to the nucleolus:

Chromosome 21 is one of five acrocentric chromosomes in the human genome that carries rRNA repeat sequences called nucleolar organizing regions (NORs) for the synthesis of ribosomes. The transcription of these NORs by RNA pol I spontaneously initiates the process of nucleolar formation, where rRNA transcription, processing and assembly into mature ribosomes will take place. The formation and dynamics of the nucleolus have been generalized for the study of other ribonucleoprotein (RNP) bodies, and the NOR on p-arm of Chr. 21 provides us with a unique opportunity to study the dynamics of ERα mediated chromosomal movement with respect to the nucleolus. ImmunoFISH experiments using NRIP1/TFF1 probes and Fibrillarin antibody indicate that the NOR is located within the fibrillar center (FC) of nucleolus, with the NRIP1 locus being located immediately adjacent to the outer surface of the dense fibrillar component (DFC) of the nucleolus in the basal condition. TFF1, in the basal condition, is generally located

distal to the nucleolus, as might be expected since it is on the other side of the Chr. 21 q arm. Upon addition of E_2 ligand, the topology of chromosome 21 with respect to the nucleolus changes rather significantly, 50 minutes after E_2 treatment; the TFF1 locus interacts with fibrillarin with the same frequency as the NRIP1 locus (Figure 38). In short, for the TFF1 and NRIP1 interaction to occur they must move towards each other, as NRIP1 cannot relocate away from the nucleolus since it is proximal to the NOR; TFF1 must move towards the nucleolus and joins NRIP1 on the outer surface of the dense fibrillar component of the nucleolus (Figure 39).

Discussion

In this paper, we have uncovered a currently unappreciated functional consequence of the ligand-induced interactions between robust "first tier" enhancers separated by vast linear genomic distances, dictating the robustness of chromosome wide enhancer driven transcriptional regulation programs. While enhancers are well established to have a transferable ability to activate transcription of reporter genes[17,18,19], the observations described here reveal that the physiological robustness of the regulated "first tier" enhancers, at least in part, derives from their interactions with other strong, regulated enhancers located on the same chromosome. Using the actions of liganded estrogen receptor (ERα) in breast cancer cells as a model, our data indicates that ligand induces dynamic, asynchronous increased proximity of a cohort of the most robust ERα bound enhancers located in two A compartments on chromosome 21 that are separated by a centrally located 14 mega-base long B compartment. Surprisingly these interactions caused an increased robustness of activation for the entire set of interactions enhancers.

Specifically, proximity and interactions between NRIP1e3 and TFF1e1 enhancers, and several other first tier enhancers are increased in response to E_2 in MCF7 cells and this appears to occur in a non-synchronous fashion after ligand treatment. While one might have reasonably predicted that the robustness of each enhancer represents only an intrinsic property of the DNA primary sequence, based on the opportunity to test this assumption by employing CRISPR-Cas9 gene editing, we have found that with deletion of even one of the most potent ERα bound enhancers, the robustness of other enhancers in network becomes significantly diminished. These data reveal that, in the context of the intact chromosome, robustness of some strong enhancers substantially depends on the presence

of interactions with other enhancers. In effect, constituting a ligand-induced, dynamic "spatially distributed super-enhancer". Specificity of these interactions is indicated by the fact that non E_2 regulated enhancers and weak ERα bound enhancers are apparently unaffected by these interaction dependent events. This study has thus uncovered a currently unappreciated aspect of enhancer networking that dictates programs of chromosome wide transcriptional regulation.

Much time and thought has been devoted to research about how the genome can be effectively organized three dimensionally without the benefit of having controlled or directed movements of genetic loci towards and away from each other. Much of the recent literature with regards to chromosomal organization has been focused upon the ability of proteins like cohesins or SMCs to set up loop domains and TADs by actively extruding DNA[20,21,22]. This model has many attractive qualities, one being that it is based upon an active one dimensional process controlled by the presence of CTCF associated boundary elements, rather than relying upon the stochastic nature of the Brownian random walk. Our paper, on the other hand, deals with the types of random 3D interactions which can go across TAD boundaries and even across chromosomes, this 3D diffusive process cannot be controlled by loop-extrusion since that process is naturally limited by the presence of boundary elements. TADs are defined by the preferential interactions of loci located within them compared with interactions between loci in different TADs (roughly 2 to 1)[23,24]. However, genome-wide interaction maps from Hi-C experiments show interactions between loci in different TADs and on different chromosomes[25]; these interactions reflect the sum of dissimilar configurations of the genome in a large population of cells[26] which

has been proposed to explain the relatively large cell-to-cell variations in gene expression patterns[27].

Interaction of enhancers and promoters on different chromosomes has been demonstrated in the past, once by a group which inserted an extra β-globin locus control region (LCR) into the mouse genome[28]. This group concluded that the ectopic LCR made preferential contact with genes controlled by shared transcription factors, much like our own conclusion about the ERα enhancer network's preferential effects on other ERα enhancers. Interestingly, they found that trans-activation by this LCR occurred predominantly in "jackpot" cells which showed an inter-chromosomal interaction between the ectopic LCR and the *βh1* gene. They estimated the ectopic LCR-*βh1* interaction as occurring in 5-10% of cells when measured by DNA-FISH, but those "rare" interactions caused a 100% increase in the *βh1* expression level in the cell population. The interaction frequencies we've measured for distal ERα enhancers on chromosome 21 are also in the range of 5-10% in the E_2 induced condition; however, our enhancer network has many members leading to a combinatoric increase in the possible number of interactions. As previously mentioned, there are 45 pairwise interactions between any set of ten enhancers, even if the rate of E_2 induced interactions was very conservatively estimated at 3% on average for all these interactions, there would only be 25.4% (.97^45) of cells without any pairwise enhancer interactions between these ten enhancers; in other words, in 74.6% of cells at least one pair of ERα enhancers would be interacting. Of course, not all enhancer-enhancer interactions are equally likely; therefore, the most likely interactants will also have the largest impacts on their partners' transcriptional robustness.

Of course, there is still the tricky question of what happens when enhancers meet in 3D space which allows them to have such synergistic effects on transcription. Unfortunately, not much is yet known about the types of biochemical and biophysical processes which play out when enhancers gain sufficient proximity for meaningful interactions. Another group has speculated that the aggregation of olfactory receptor (OR) enhancers around an (OR) promoter results in the formation of a stable nucleoprotein complex which is sufficient to cause a feedback-eliciting level of OR gene transcription resulting in a permanent OR choice in that specific sensory neuron[29]. This idea of enhancers meeting in 3D space to form some "stable" structure is quite attractive given what we know about the spontaneous formation of RNP bodies from highly accessible and transcribed regions like the NORs. This process may also explain the role of eRNAs in synergizing enhancer interactions, since eRNAs are known to be highly enriched for transcription factor binding motifs[30]; which is to say they have a high valency value[31]. This hypothesis is supported by our data that knockdown of TFF1e1 eRNA by LNA or ASO, reduces the transcriptional activity of a larger set of ERα enhancers as well as reducing the contacts between TFF1 and NRIP1 loci. It has been reported that the presence of highly valent RNA in addition to their protein binding partners can cause the process of liquid-liquid phase separation to occur resulting in the spontaneous formation of RNP bodies[32,33,34] which can stabilize otherwise highly transient interactions between proteins, RNA and, in our case, DNA as well.

Another interesting feature of the enhancer network in MCF7 cells is the relative abundance of enhancers proximal to coding genes and lncRNAs which have been found to be either associated with or responsible for cancer progression (TIAM1, DSCAM-AS1,

TMPRSS2, TFF1, etc.). The extreme proliferative abilities of cancer cell lines like MCF7 cells makes them uniquely compatible for experiments like generating CRISPR based single colony derived mutant cell lines that normal breast epithelium would be unsuitable for. We can only speculate how changes to the copy number of some of the most highly active ERα enhancers and genes (TFF1, TMPRSS2, DSCAM-AS1 etc.) affect the enhancer network in MCF7 cells; and whether the interactions we observe between NRIP1 and TFF1 (2 copies) or DSCAM-AS1 (5 copies) are on the *cis*-chromosome. Unfortunately, with currently available DNA FISH methodologies it is not possible to discern which copies of these duplicated oncogenic regions are on chromosome 21, and which copies are on other chromosomes resulting from translocations and duplications; while we believe the clear majority of interactions to be intrachromosomal, we cannot give exact numbers regarding the ratio of intra vs interchromosomal interactions.

Perhaps what we are observing in these MCF7 cells is an enhancer-network run amok; an uncontrolled positive feedback loop of oncogenic enhancer and gene expression which all synergize off each other. Many researchers have speculated that the competition for resources inside of a tumor microenvironment exerts strong selective pressures on the cells within[35,36,37], and since MCF7 cells are taken from a metastatic pleural effusion site[38], we know for certain that these cells have been able to survive, proliferate and escape from the initial site of the mammary tumor. It is interesting to speculate about the evolutionary process of cells inside of a tumor microenvironment, perhaps part of the oncogenic competition involves the rewiring of existing hormonal and inflammatory enhancer networks to favor survival and proliferation, at the cost of genomic integrity. The idea of an enhancer network seems to have a positive feedback loop built into it: more enhancers

contact leads to greater proximity leads to more contact (repeated for many cycles); it would not be hard to see how this sort of logic could be hijacked by an oncogenic evolutionary process ultimately leading to unchecked proliferation. Although the enhancer network in MCF7 cells may vary considerably from those in normal human cell types, we believe that the basic biochemical and biophysical mechanisms (whatever they may be) at work in the MCF7 ERα enhancer network will be broadly applicable to the study of other enhancer networks in both cancerous and normal cells.

Chapter 1, in part, is currently being prepared for submission for publication of the material. Nair, Sreejith; Yang, Lu; Meluzzi, Dario; Oh, Soohwan; Gamliel, Amir; Suter, Thomas; Yang, Feng; Jayani, Ranveer; Zhang, Jie; Ohgi, Kenny; Rosenfeld, Michael Geoff. "Chromosomal Enhancer Syntax: Spatially-Distributed Super Enhancers and Subnuclear Structural Associations Dictate Enhancer Robustness". The dissertation author was the primary investigator and author of this material.

Materials and Methods

Antibodies: The antibodies in this study were: anti-ERα (HC-20, Santa Cruz); anti-H3K4me^3 (07-473, Santa Cruz); anti-H3K4me^1 (ab8899, Abcam); anti-H3K27ac (ab4729, Active motif); anti-α-tubulin (T5168, Sigma), and anti-IgG (I5006, Sigma).

Cell culture: MCF7 obtained from ATCC were cultured in DMEM media supplemented with 10% FBS in a 5% CO_2 incubator. They were hormone-stripped for 3 days in phenol-free media with charcoal-stripped FBS. Two days into stripping, 10 nM ICI (ERα antagonist) for 24h was added into the media. One hour before harvesting the cells, 100nM 17β-estradiol (E_2) (Sigma) in ICI untreated or ethanol in ICI treated cells was given to induce estrogen signaling.

RT-QPCR: RNA was isolated using Trizol (Invitrogen) or RNeasy column (Qiagen), and total RNA was reverse-transcribed using SuperScript® III Reverse Transcriptase (Invitrogen). Quantitative PCRs were performed mostly with StepOne Plus (Applied Biosystem). For normalization, ΔCt values were calculated relative to the levels of ACTB/GAPDH transcripts. The results were repeated for at least 3 times and one representative plot is shown in figures; most p-values were obtained using a two-tailed Student's t-test.

ChIP-seq: Briefly, approximately 10^7 cells were cross-linked with 1% formaldehyde at room temperature for 10 min and neutralized with 0.125M glycine. After sonication, ~75µg soluble chromatin was incubated with 1-5µg of antibody at 4^0C overnight. Immunoprecipitated complexes were collected using Dynabeads A/G (Invitrogen). Subsequently, immuno-complexes 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 per the manufacturer's instructions. Usually, the first 48bp for each sequencing read was aligned to the hg18 assembly using BFAST or Bowtie2. Only uniquely mapped tags were selected for further analysis. The data was visualized by preparing custom tracks on the University of California, Santa Cruz, (UCSC) genome browser using HOMER (http://biowhat.ucsd.edu/homer). The total number of mapable reads were normalized to 10^7 for each experiment presented in this study.

Identification of ChIP-seq Peaks: The ChIP-seq peaks were identified by HOMER. Given different binding patterns of transcription factors and histones, parameters were optimized for the narrow tag distribution characteristic of transcription factors by searching for high read enrichment regions within a 200bp sliding window. Regions of maximal density exceeding a given threshold were called as peaks, and adjacent peaks were set to be >500bp away to avoid redundant detection. The common artifacts from clonal amplification were circumvented by considering only one tag from each unique genomic position. The threshold was set at a false discovery rate (FDR) of 0.001 determined by peak finding using randomized tag positions in a genome with an effective size of 2×10^9 bp. For ChIP-seq of histone marks, seed regions were initially found using a peak size of 500bp (FDR<0.001) to identify enriched loci. Enriched regions separated by <1kb were merged and considered as blocks of variable lengths. All called peaks were then associated with genes by cross-referencing with the RefSeq TSS database. Peaks from individual experiments were considered overlapping if their peak centers were located within 200bp (for some analysis may extend to 1kb). The peaks within ±1kb apart from RefSeq gene TSS site were considered as promoter-bound

GRO-seq: GRO-seq experiments were performed as previously reported[39]. Briefly, MCF7 cells were swollen in swelling buffer (10mM Tris-Cl pH7.5, 2mM MgCl2, 3mM CaCl2) for 5 min on ice and then lysed in lysis buffer (swelling buffer with 0.5% IGEPAL and 10% glycerol), before being re-suspended in 100µl of freezing buffer (50mM Tris-Cl pH8.3, 40% glycerol, 5mM MgCl2, 0.1mM EDTA). For the run-on assay, re-suspended 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, 500µM ATP, GTP, and Br-UTP, 2µM CTP) and incubated for 5 min at 30^0C. The nuclear-run-on RNA (NRO-RNA) was then extracted with TRIzol LS reagent (Invitogen) following manufacturer's instructions. After base hydrolysis on ice for 40min and followed by treatment with DNase I and antarctic phosphatase, the Br-UTP labeled NRO-RNA was purified by an anti-BrdU argarose beads (Santa Cruz Biotech) in binding buffer (0.5XSSPE, 1mM EDTA, 0.05% tween) for 3hr at 4°C while rotating. Then T4 PNK (NEB) was used to repair the end of NRO-RNA. Subsequently, cDNA synthesis was performed as reported[9,33] with few modifications. The RNA fragments were subjected to poly-A tailing reaction by poly-A polymerase (NEB) for 30 min at 37°C. Reverse transcription was then performed using superscript III (Invitrogen) with oNTI223 primer. The cDNA products were separated on a 10% polyacrylamide TBE-urea gel with right product (~100-500bp) being excised and recovered by gel extraction. After that, the first-strand cDNA was circularized by CircLigase (Epicentre) and re-linearized by Ape1 (NEB). Re-linearized single strand cDNA were separated by TBE gel and the products of desired size was excised (~120-320bp) for gel extraction. Finally, cDNA template was amplified by PCR using the Phusion High-Fidelity enzyme (NEB) with primers oNTI200 and oNTI201 for deep sequencing.

Computational analysis of GRO-seq: The sequencing reads were aligned to hg18 using Bowtie2. For analyzing estrogen effects and enhancer deletion effects on gene transcription, we counted the reads from the first 30kb (assuming a RNA polymerase speed of ~0.5 kb/min during 1hr E_2 treatment) of entire gene body, excluding the promoter-proximal region on the sense strand with respect to the gene orientation by using BED Tools or HOMER. EdgeR (http://www.bioconductor.org/) was used to compute the significance of the differential gene expression (FC≥1.5, FDR≤0.01). Additionally, a read density threshold (*i.e.* GRO-seq normalized read counts/kb) was used in order to exclude lowly expressed genes.

***De novo* identification of GRO-seq transcripts:** GRO-seq read densities were analyzed in a similar manner to ChIP-seq. Provided GRO-seq generates strand-specific data, separate tracks were uploaded onto the UCSC genome browser; tag-enriched sites were identified using a sliding window of 250bp. 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% as compare to that of TSS or when another transcript's start was identified on the same strand. Individual high-density peaks spanning a region less than 250bp were considered artifacts and removed from the analysis. Transcripts were defined as putative eRNAs if their *de novo* called start sites was located distal to RefSeq TSS (≥3kb) and were associated with ERα and H3K27ac co-bound regions.

Bioinformatics characterization of ERα enhancers: The ERα-H3K27ac co-bound regions are defined as that the distance from the center of an ERα peak to the H3K27ac peak-occupied region is ≤1kb. Overall, two methods were used to assign the ERα bound

enhancers to E_2 upregulated genes: 1) identifying the E_2-upregulated coding genes from GRO-seq and coupling each of them to their closest ERα-H3K27ac co-bound enhancer within first distance (200kb) (a "gene-centric" view); 2) characterizing the ERα-H3K27ac co-bound enhancers first and then coupling each of them to their closest TSS that belongs to 1,309 E_2 upregulated coding genes (an "enhancer-centric" view).

FISH and imaging: Ethanol or E_2 treated MCF7 cells were grown onto acid-washed poly-lysine coated coverslips, were washed with 1XPBS and immediately fixed with freshly made 4% paraformaldehyde/PBS for 10min. Permeabilization was achieved by incubating in PBS containing 0.5% Triton X-100 for 15min. FISH pre-hybridization treatments include incubating the cover-slips in 0.1N HCl for 5min at room temperature, followed by digestion with 0.01N HCl/ 0.002% pepsin for 5min at 37ºC, stopped by 50mM $MgCl_2$/PBS and equilibrated in 50% formamide/2XSSC 2hrs prior to hybridization. 5µl of probe/hybridization buffer mix was used per coverslip, with a hybridization program of 76ºC for 3mins followed by overnight hybridization at 37ºC in a humidified dark chamber. The coverslips were then washed with pre-warmed WS1 (0.4xSSC/0.3% NP-40), WS2 (2xSSC/0.1% NP-40) and PBS before being finally mounted with prolong gold-DAPI anti-fade mounting reagent (Invitrogen).

For FISH Image acquisition and data analysis, images were acquired using the Perkin Elmer Spinning Disk Confocal Microscope (100X objective lens). Z-stack data acquisition was set up with 0.15 µm step size. The 3D images were then generated in Volocity (v6.0.1). The FISH-positive gene loci were identified using the "Find Object Using % Intensity" (generally >20%) function in combination with "Exclude Objects by Size" (generally >0.1 μm^3). The overlap between two FISH-positive gene loci was

calculated by function "Intersect" with size exclusion (> 0.03 μm^3). For these experiments, MCF-7 cells are treated with ethanol or E_2, grown on acid-washed poly-lysine-coated coverslips, were washed with 1× PBS and immediately fixed with freshly made 4% paraformaldehyde/PBS for 10 min. Three-dimensional images were generated in Velocity (v.6.0.1). The overlap between two FISH-positive gene loci was calculated by the function "Intersect" with size exclusion (>0.03 μm^3). The cells were counted ($n > 100$ for each group) were from eight images/fields; the percentage of overlapping or proximity) events from each one was calculated separately, which together generates the mean and s.d. A second method of analysis was utilized. Digital image stacks form 3D-FISH experiments using two or three different fluorescently labeled BAC or fosmid probes were acquired on a Perkin Elmer UltraView Vox spinning disk confocal microscope, and were analyzed using Volocity software version 6.2 (Perkin Elmer) to determine the spatial locations of the probed genomic loci within each nucleus. To calculate spatial distances between the probed loci, the coordinates of centroids of the corresponding fluorescent spots found by Volocity were analyzed using custom software implemented with Python, NumPy, and SciPy. In particular, for each imaged nucleus, clusters of nearest centroids of two or three differently colored spots were identified. The diameter of each spot cluster was determined by calculating the diameter of the smallest circle bounding the centroids in the cluster. Up to two smallest clusters were retained per nucleus and clusters with diameter greater than 8 μm were discarded as possible outliers. Distributions of diameters for the remaining spot clusters were compared using the Wilcoxon rank sum test. The diameter of a spot cluster containing only two spots equals the distance between those spots. To investigate the changes in spatial proximity of genomic loci under different conditions, fractions of spot

clusters having inter-spot distances less than a pre-defined threshold were determined and p-values for changes in such fractions were estimated using a Bootstrap method. The distance threshold used for each pair of loci was calculated as a fraction of the median distance expected for the genomic separation of those loci. The expected median distance was in turn o btained by fitting a power law to the median inter-spot spatial distances observed under -E_2/ ICI treatment for pairs of loci with various genomic separations.

Chromatin Conformation Capture (3C): Briefly, 25x10^6 MCF7 cells were fixed by adding 1% formaldehyde at room temperature for 10 minutes, and the reaction was stopped by glycine. Lysis buffer (500µl 10mM Tris-HCl pH8.0, 10mM NaCl, 0.2% Igepal CA630; protease inhibitors was added and cells were incubated on ice. Next, cells were lysed with a Dounce homogenizer, and the suspension was spun down at 5,000 rpm at 4°C. The supernatant was discarded and the pellet was washed twice with 500µl ice-cold 1x NEBuffer 3 (NEB, Ipswich, MA). The pellet was then re-suspended in 1X NEBuffer 3 and split into five separate 50µl aliquots. The extracted chromatin was then digested overnight by 400 Units BamHI (NEB). Each digested chromatin mixture was ligated by T4 DNA Ligase (800U) in 20 times of initial volume for 4hrs at 16°C. The ligase step was omitted in one chromatin aliquot from the five mentioned above as the unligated control. The chromatin was subsequently 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 re-suspending in 250µl 1XTE 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. The digestion and ligation efficiencies were checked and normalized before 5C.

Probe design for 5C: Donor (Forward) primers were designed on all BamHI sites on Chr21 except for sites that were giving multiple hits. Acceptors (Reverse) oligos were designed on all ERα-enhancers and all promoters. Custom Perl scripts (available upon request) was used for designing these oligonucleotides. The uniqueness of the probe sequences was verified by Bowtie alignment to the human genome hg18 assembly. Universal adaptor sequences that are compatible with HiSeq 2000 flow cell design were added to the probe ends for bridge amplification of the ligation products and for direct sequencing. Acceptors were phosphorylated and both acceptors and donors were pooled individually in equimolar amounts for 5C reaction.

3D-DSL: Briefly, after 3C efficiency estimation, equal amount of 3C chromatin was biotinylated using the Photoprobe Kit (Vector Lab). Donor and acceptor probe pools (20fmol per probe) were annealed to the biotinylated 3C samples at 45°C for 2 hours followed by 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 1hr. 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.

ATAC-seq: ATAC-seq was performed as previously described (Buenrostro et al., 2015). Briefly, wild-type or CRIPR deletion clones of MCF7 cells were harvested after indicated time of ICI/estradiol treatment, washed once with PBS and 50,000 cells were used for each ATAC-seq sample preparation. Transposition reaction was performed at 37ºC for 30 min

using Nextera DNA Library Preparation Kit (Illumina Cat No. 15028212) in 50µl volume (2.5 µl Transposae enzyme, 25 µl 2X TD buffer and 22.5 µl nuclease-free water). The reaction mixture was immediately purified using MinElute PCR purification Kit (Qiagen, Cat No. 28006) and eluted in 10 µl nuclease-free water. The tagmented DNA was then PCR amplified using KAPA Real-Time Library Amplification Kit (Kapa Biosystems, Cat No. KK2701) in a 50 µl reaction (10 µl tagmented DNA, 2.5 µl25 µM PCR primer 1, 2.5 µl 25 µM barcoded PCR primer 2, 25 µl KAPA HiFi HotStart ReadyMix, 10 µl nuclease-free water), using the following program (One cycle at 72ºC: 5 min, 98ºC: 30sec; 8-12 cycles at 98ºC: 10 sec, 63ºC: 30 sec and 72ºC: 1min). PCR standards, supplied in the kit, were included in separate wells and the reaction was stopped between standard 2 and 3. The reaction mixture was subjected to size selection (to select fragments ranging 150-800bp) using 1.6 volume of Agencourt AMPure XP magnetic beads (Beckman Coulter, Cat No. A63880), washed once with 80% ethanol, air-dried and eluted in 15 µl of nuclease-free water. The resulting ATAC-seq library was quantified using Qubit Flurometer (Thermo Fisher) and the quality was analyzed by resolving 1 µl on Tapestation (Agilent). Ten nmol of the library was sequenced on HiSeq 4000 sequencer (Illumina).

CRISPR MCF7 Enhancer Deletion Line Generation: Guide RNA sequences were designed to target the 5’ and 3’ of enhancer sequences by entering these sequences into the website crispr.mit.edu. Top results from the website were cloned into the pX459-puro vector. Pairs of pX459 gRNA targeting plasmids were transfected into wild type MCF7 cells using lipofectamine 3000 reagent. 72 hours after transfection, MCF7 cells were treated with TrypLE Select reagent and trypsinized for 30 minutes and then pipetted up and down 50 times through a filtered P100 tip to break up any large clusters of adherent

cells. The single cell suspension was then plated onto many 10cm plates at a density of roughly 50,000 to 100,000 cells per plate. 12 hours after plating, puromycin selection reagent was added to a final concentration of 400-600ng/mL in multiple 10cm plates in increments of 50ng/mL. The puromycin selection continued for 72 hours, after which the puromycin media was removed, and non-selective media with pen/strep and Normocin antibiotics was added. The remaining cells are regrown for between 14-28 days, until colonies made up of 50-100 cells could be observed with a microscope. At this point, individual colonies of MCF7 cells were lifted from the plate with a P100 pipette under the observation of a microscope inside of a sterile tissue culture hood. Colonies were then moved to individual wells of a 96-well or 48-well plate. After reaching confluence within these wells, the cells were again trypsinized and 50% of the cells were taken for genotyping while the remaining 50% could regrown. After genotyping, any wild type, heterozygous and homozygous cell lines deemed fit for further study were systemically expanded until enough cells were available for freezing stocks at -80C or in liquid nitrogen.

BIBLIOGRAPHY

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

[2] Henderson, A., Warburton, D., and Atwood, K. (1972). Location of Ribosomal DNA in the Human Chromosome Complement. Proceedings Of The National Academy Of Sciences *69*, 3394-3398.

[3] Nagano, T., Várnai, C., Schoenfelder, S., Javierre, B., Wingett, S., and Fraser, P. (2015). Comparison of Hi-C results using in-solution versus in-nucleus ligation. Genome Biology *16*.

[4] Rao, S., Huntley, M., Durand, N., Stamenova, E., Bochkov, I., Robinson, J., Sanborn, A., Machol, I., Omer, A., and Lander, E. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665-1680.

[5] Roelfsema, J., White, S., Ariyürek, Y., Bartholdi, D., Niedrist, D., Papadia, F., Bacino, C., den Dunnen, J., van Ommen, G., and Breuning, M. (2005). Genetic Heterogeneity in Rubinstein-Taybi Syndrome: Mutations in Both the CBP and EP300 Genes Cause Disease. The American Journal Of Human Genetics *76*, 572-580.

[6] Prud'humme, J., Fridlansky, F., Le Cunff, M., Atger, M., Mercier-Bodart, C., Pichon, M., and Milgrom, E. (1985). Cloning of a Gene Expressed in Human Breast Cancer and Regulated by Estrogen in MCF-7 Cells. DNA 4, 11-21.

[7] Dekker, J. and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. Cell *164*, 1110-1121.

[8] van de Werken, H., Landan, G., Holwerda, S., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Öz, Y., and Bouwman, B. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nature Methods *9*, 969-972..

[9] Niknafs, Y., Han, S., Ma, T., Speers, C., Zhang, C., Wilder-Romans, K., Iyer, M., Pitchiaya, S., Malik, R., and Hosono, Y. (2016). The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. Nature Communications *7*, 12791.

[10] Rao, S., Huntley, M., Durand, N., Stamenova, E., Bochkov, I., Robinson, J., Sanborn, A., Machol, I., Omer, A., and Lander, E. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665-1680.

[11] Markenscoff-Papadimitriou, E., Allen, W., Colquitt, B., Goh, T., Murphy, K., Monahan, K., Mosley, C., Ahituv, N., and Lomvardas, S. (2014). Enhancer Interaction Networks as a Means for Singular Olfactory Receptor Expression. Cell 159, 543-557.

[12] Cong, L., Ran, F., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P., Wu, X., Jiang, W., and Marraffini, L. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science *339*, 819-823.

[13] Ran, F., Hsu, P., Wright, J., Agarwala, V., Scott, D., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature Protocols *8*, 2281-2308.

[14] Hsieh, C., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C., Lee, G., and Chen, S. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proceedings Of The National Academy Of Sciences 111, 7319-7324.

[15] Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A., and Rosenfeld, M. (2013). Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. Cell *155*, 1581-1595.

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

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

[18] Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Research *9*, 6047-6068.

[19] Mercola, M., Wang, X., Olsen, J., and Calame, K. (1983). Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. Science *221*, 663-665.

[20] Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Reports *15*, 2038-2049.

[21] Bouwman, B. and de Laat, W. (2015). Getting the genome in shape: the formation of loops, domains and compartments. Genome Biology *16*.

[22] Sanborn, A., Rao, S., Huang, S., Durand, N., Huntley, M., Jewett, A., Bochkov, I., Chinnappan, D., Cutkosky, A., and Li, J. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proceedings Of The National Academy Of Sciences *112*, E6456-E6465.

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

[24] Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell *148*, 458-472.

[25] Rao, S., Huntley, M., Durand, N., Stamenova, E., Bochkov, I., Robinson, J., Sanborn, A., Machol, I., Omer, A., and Lander, E. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665-1680.

[26]Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2011). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nature Biotechnology *30*, 90-98.

[27] Krijger, P. and de Laat, W. (2013). Identical cells with different 3D genomes; cause and consequences? Current Opinion In Genetics & Development *23*, 191-196.

[28] Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., Eussen, B., de Klein, A., Singer, R., and de Laat, W. (2011). Variegated gene expression caused by cell-specific long-range DNA interactions. Nature Cell Biology *13*, 944-951.

[29] Markenscoff-Papadimitriou, E., Allen, W., Colquitt, B., Goh, T., Murphy, K., Monahan, K., Mosley, C., Ahituv, N., and Lomvardas, S. (2014). Enhancer Interaction Networks as a Means for Singular Olfactory Receptor Expression. Cell *159*, 543-557.

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

[31] Weber, S. and Brangwynne, C. (2012). Getting RNA and Protein in Phase. Cell *149*, 1188-1191.

[32] Zhang, H., Elbaum-Garfinkle, S., Langdon, E., Taylor, N., Occhipinti, P., Bridges, A., Brangwynne, C., and Gladfelter, A. (2015). RNA Controls PolyQ Protein Phase Transitions. Molecular Cell *60*, 220-230.

[33] Nielsen, F., Hansen, H., and Christiansen, J. (2016). RNA assemblages orchestrate complex cellular processes. Bioessays *38*, 674-681.

[34] Lin, Y., Protter, D., Rosen, M., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Molecular Cell *60*, 208-219.

[35] Anderson, A., Weaver, A., Cummings, P., and Quaranta, V. (2006). Tumor Morphology and Phenotypic Evolution Driven by Selective Pressure from the Microenvironment. Cell *127*, 905-915.

[36]Merlo, L., Pepper, J., Reid, B., and Maley, C. (2006). Cancer as an evolutionary and ecological process. Nature Reviews Cancer *6*, 924-935.

[37] Palumbo, A., de Oliveira Meireles Da Costa, N., Bonamino, M., Ribeiro Pinto, L., and Nasciutti, L. (2015). Genetic instability in the tumor microenvironment: a new look at an old neighbor. Molecular Cancer *14*.

[38] https://www.atcc.org/Products/All/HTB-22.aspx

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

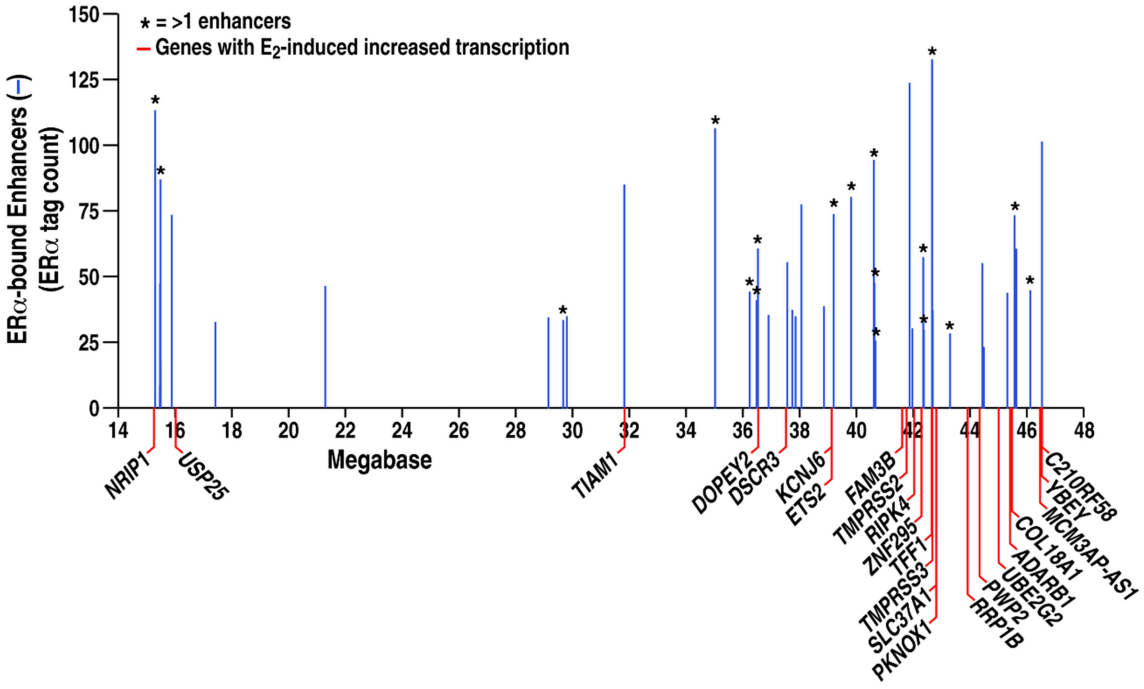


Figure 1: A map of the 39 "first tier" enhancers on Chr. 21

The 39 strongest ERα enhancers on Chromosome 21. The X axis shows the position of the enhancers on the *q* arm of chromosome 21. The Y axis shows the number of tag counts for ERα ChIP-seq experiment after E_2 treatment of 60 minutes, regions with multiple enhancers are indicated with an asterisk, only the tallest enhancer is shown in these regions. Below the graph in red are coding genes upregulated by E2 treatment in MCF7 cells.

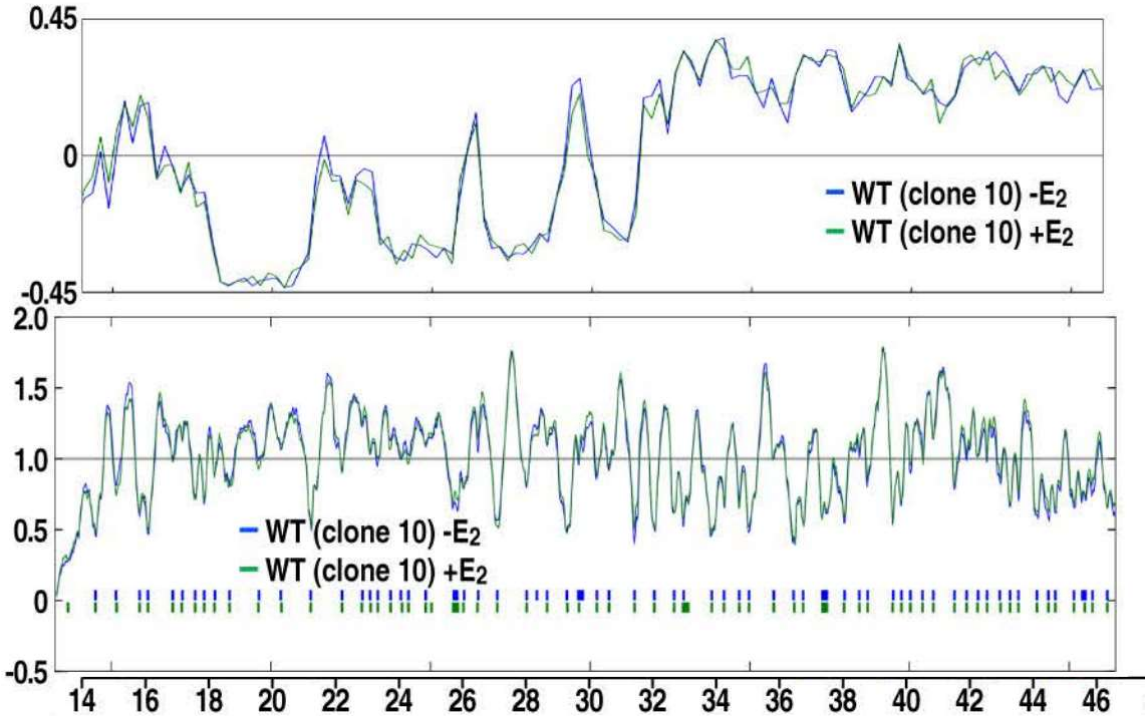


Figure 2: Maps of A and B compartments and TAD boundaries of Chr. 21

(Top Panel) Hi-C analysis permitted identification of the "A" and "B" compartments shown in the top panel. Regions above zero are in the "A" compartment, those below are in the "B" compartment.
(Bottom Panel) Calculation of the insulation scores to identify putative chromosomal boundaries. Boundaries are show as blue/green lines, from 2 replicates of the experiment.

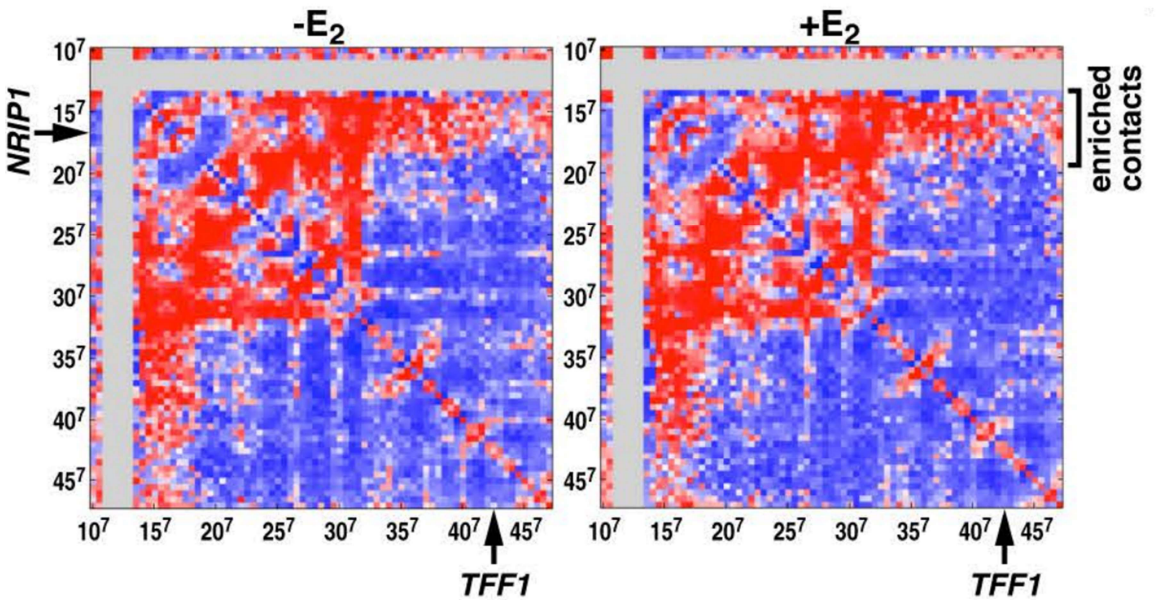


Figure 3: Hi-C contact map of Chr. 21 analyzed at a resolution of 1Mb

Contact map from Hi-C data, analyzed at a resolution of 1 megabase to detect long distance interactions across chromosome 21, indicating the presence of very long distance interactions as well as the expected "short range" interactions.

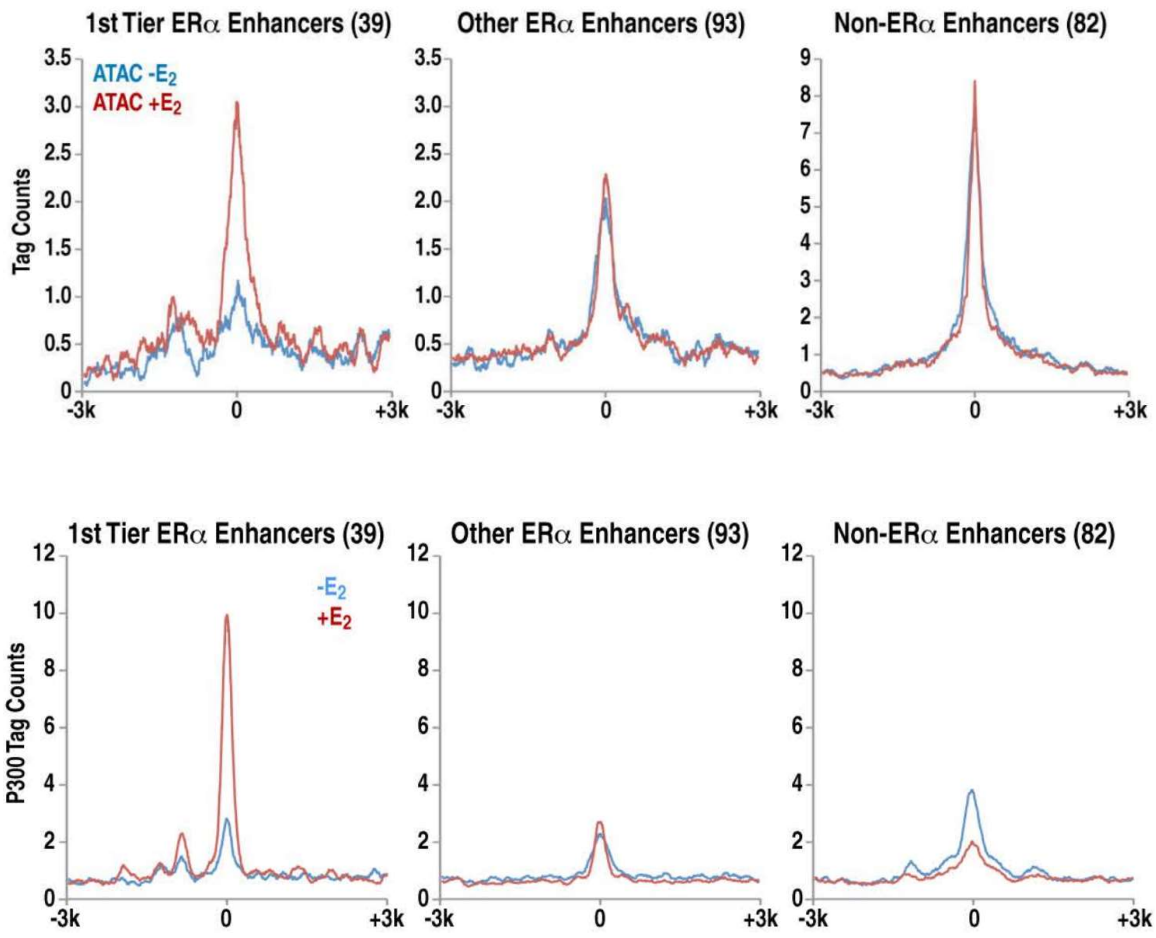


Figure 4: Meta-analysis of ATAC-seq and P300 on Chr. 21

Meta-analysis of ATAC-seq data and P300 ChIP-seq data on the 39 most robust ERα bound enhancers, compared to 93 other ERα bound enhancers and 80 randomly chosen non-ER bound active enhancers. Performed in the absence or presence of E_2 treatment for 1 hour.

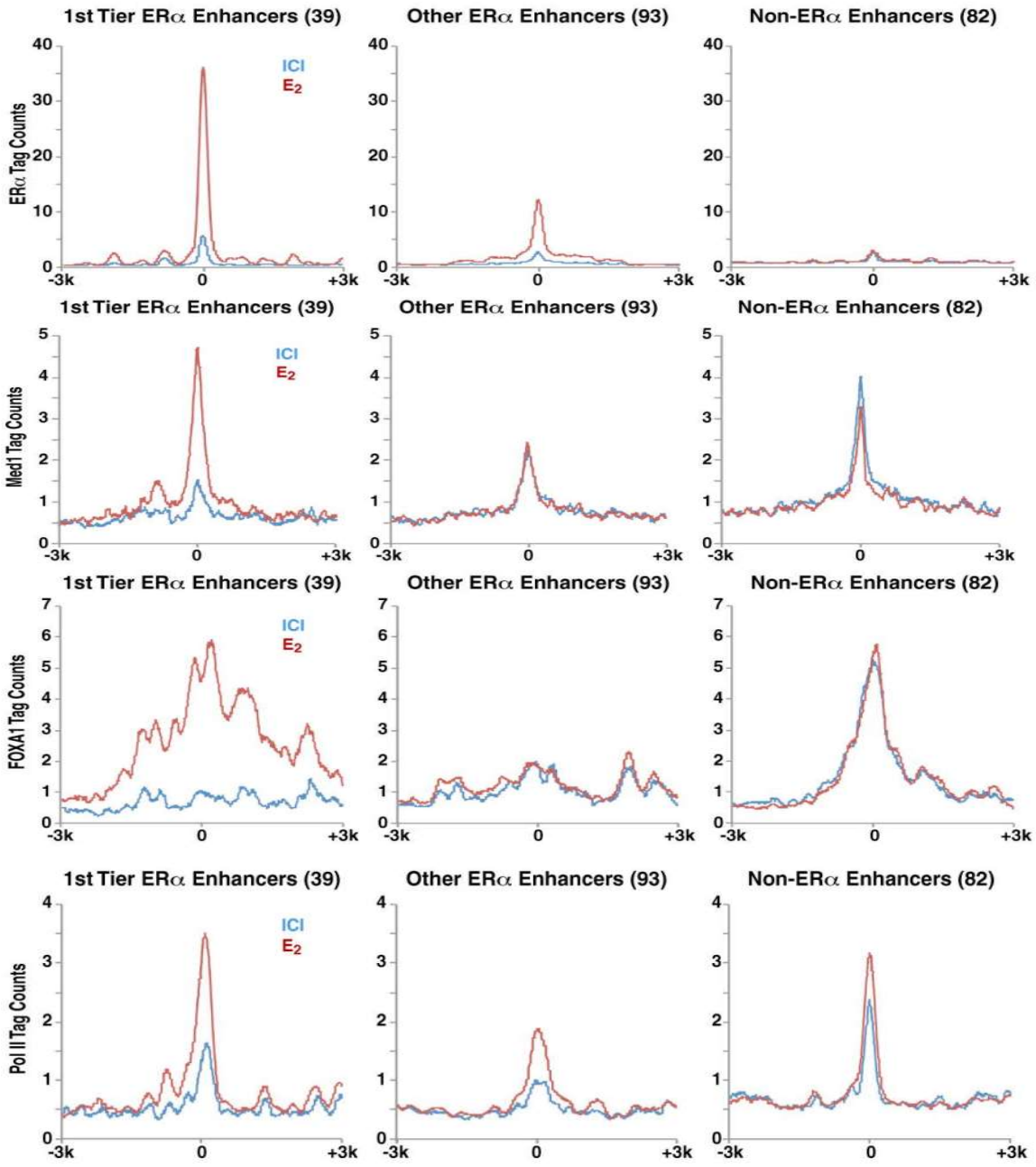


Figure 5: Meta-analysis of ERα, Med1, FoxA1 and Pol II on Chr. 21

Meta-analysis of the ChIP-seq data for ERα, Med1, FoxA1 and Pol II proteins on the 39 most robust ERα bound enhancers, compared to 93 other ERα bound enhancers and 82 randomly chosen non-ERα bound active enhancers. Performed in the absence or presence of E_2 treatment for 1 hour.

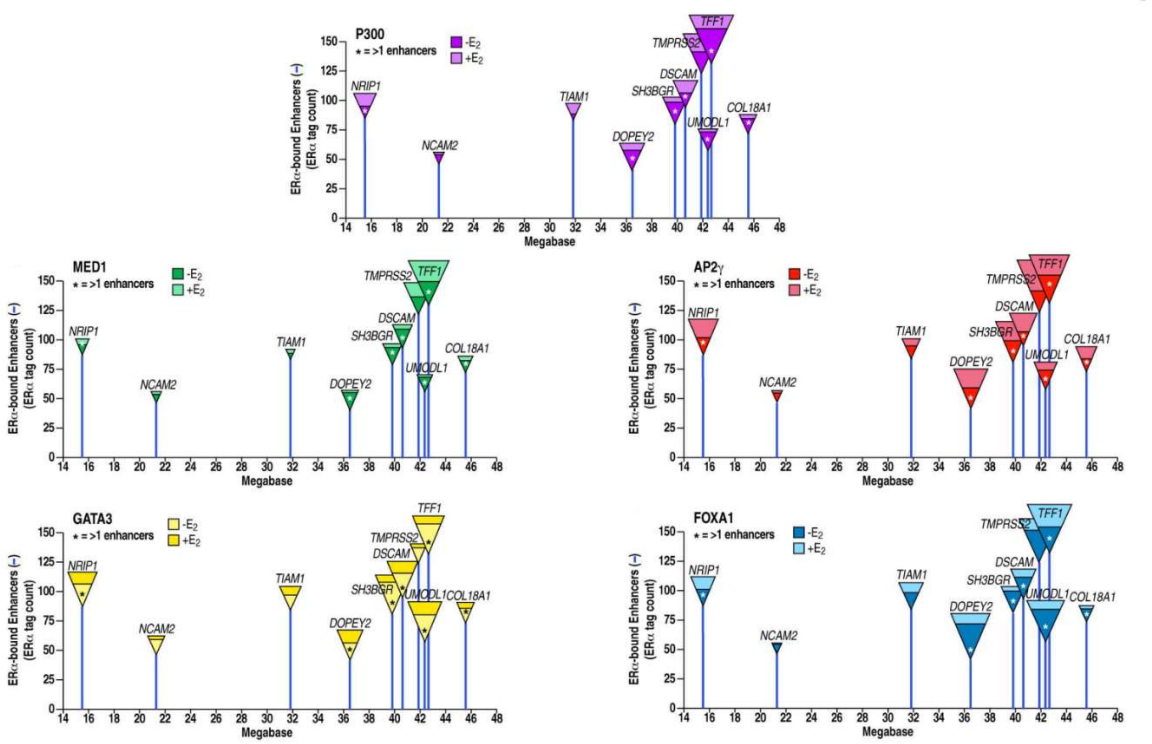


Figure 6: ChIP-seq tag counts of TFs on ten top ERα enhancers

Binding of cofactors P300, MED1, AP2γ, GATA3, and FOXA1 on ten top ERα enhancers on chromosome 21 with or without treatment of E_2 ligand. ChIP-seq tag counts are represented as the surface area of an equilateral triangle attached to the top of the line representing an ERα enhancer.

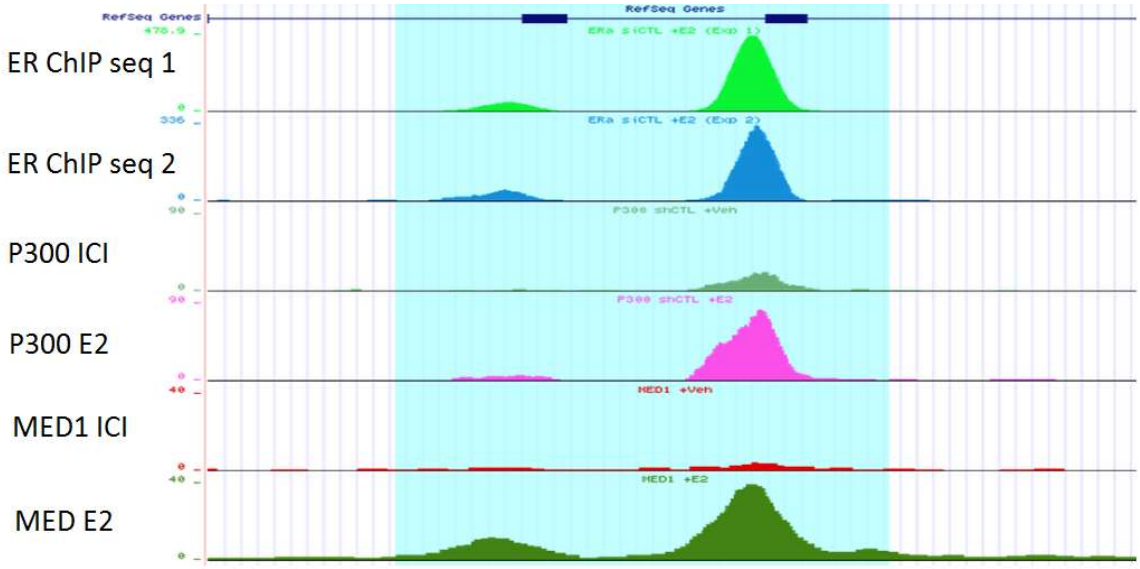


Figure 7: UCSC genome browser of TF binding on TFF1e1

Bimodal distribution of factors ERα, P300 and MED1 on the TFF1e1 locus, coordinates ch21: 42,668,281-42,670,614 on the hg18 alignment.

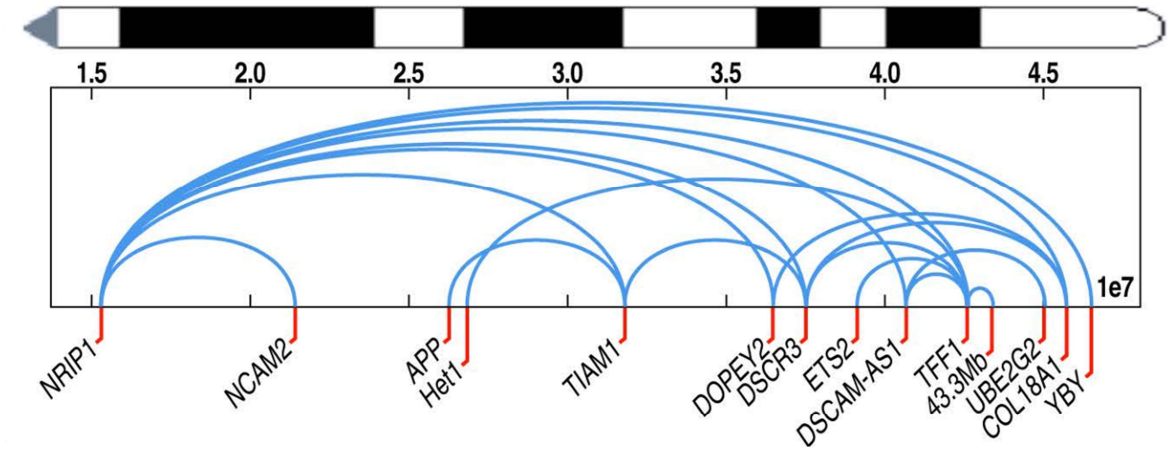


Figure 8: Map of the positions of FISH probes used in this experiments

Locations of 15 FISH probes on the *q* arm of chromosome 21. All probes are targeting ERα enhancers except for HET1 which targets the B domain heterochromatic region and a probe targeting 43.3Mb which is not a ERα enhancer.

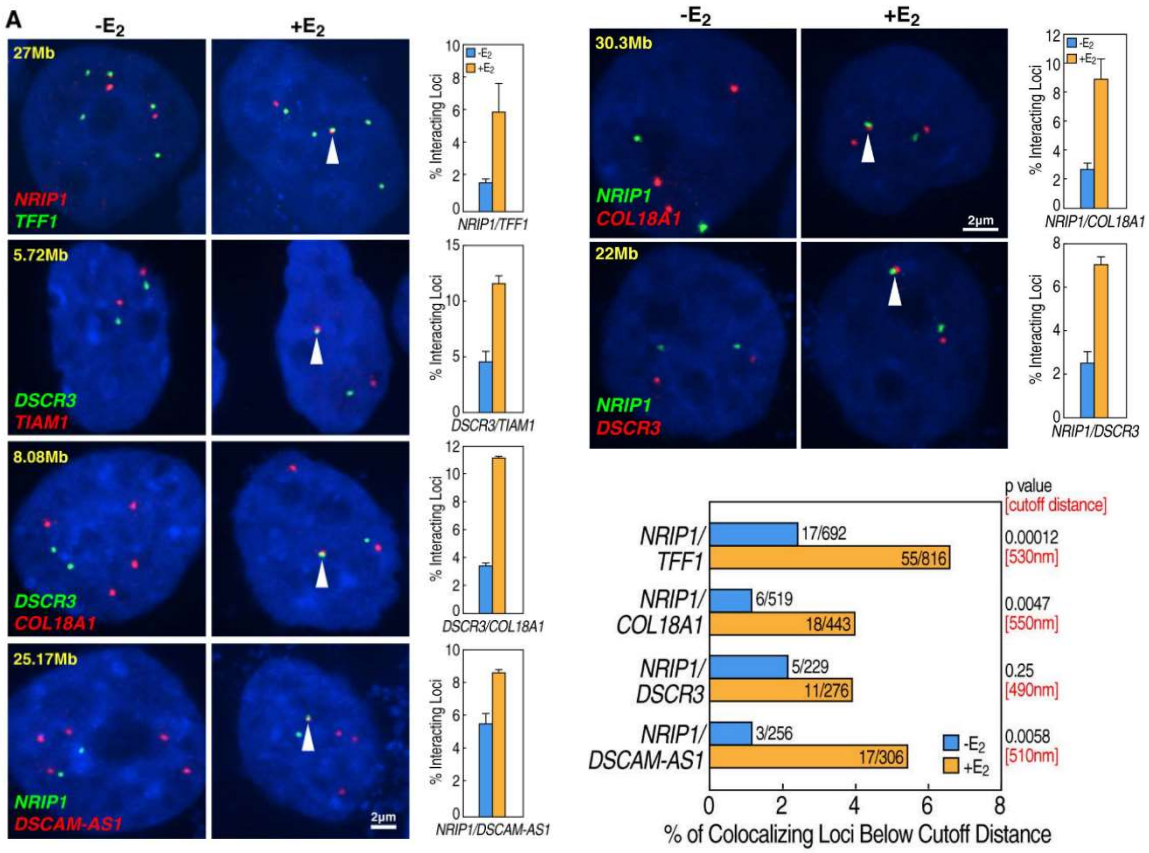


Figure 9: FISH data of interactions between enhancers in A compartments

FISH analysis of interactions between pairs of enhancers TFF1/NRIP1, DSCR3/TIAM1, DSCR3/COL18A1, NRIP1/DSCAM-AS1, NRIP1/COL18A1, and NRIP1/DSCR3 following treatment of cells with E_2 for 50 minutes. Quantification of the percent of alleles exhibiting overlaps is calculated by Volocity software shown as column plots. Genomic distance on the chromosome between enhancers is noted in yellow text. Horizontal bar plots show colocalization between pairs of enhancers with 3D distance below a certain cutoff as calculated by a custom script.

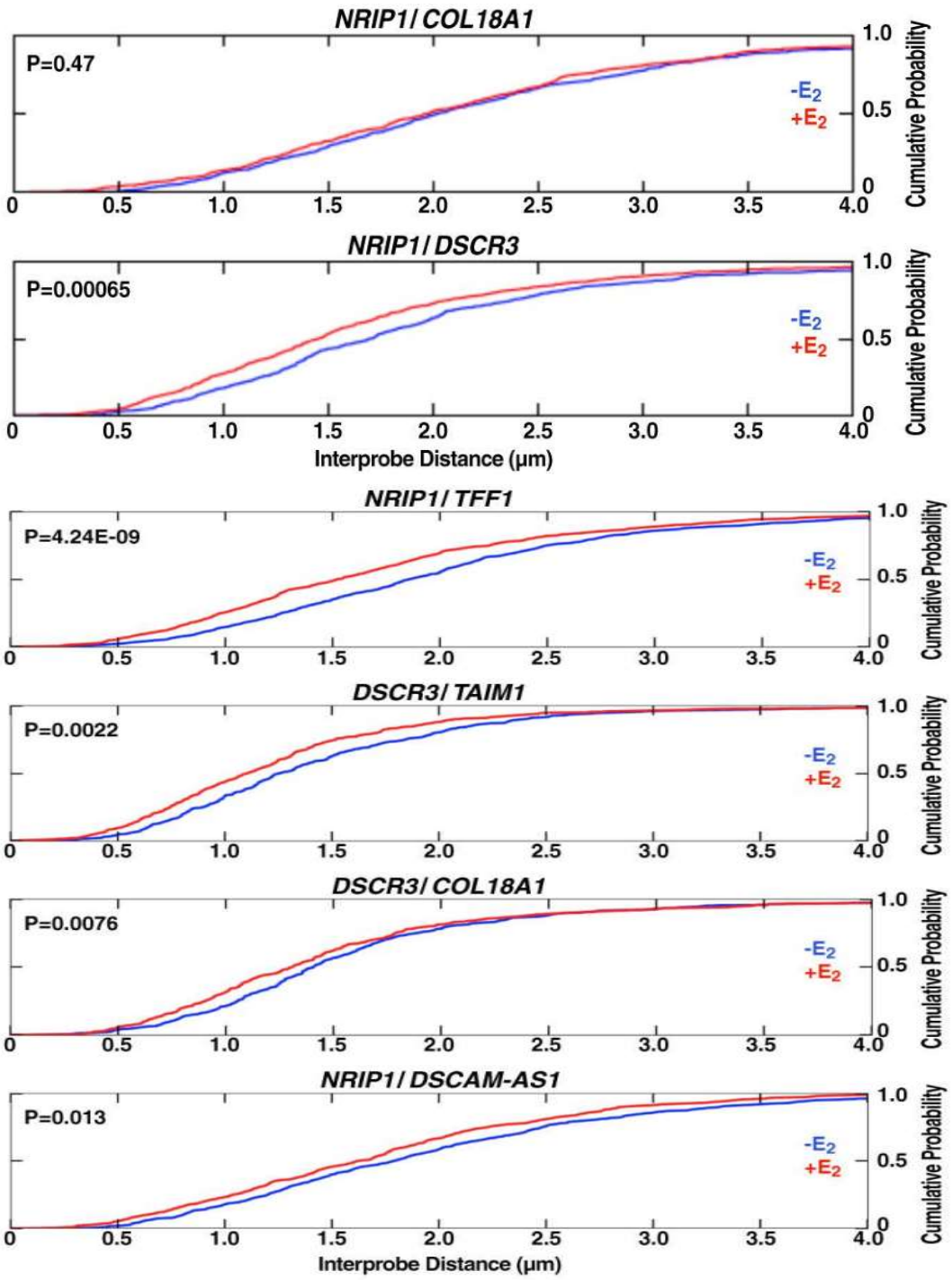


Figure 10: Cumulative distribution of distances between enhancers in Figure 2C

The cumulative distribution frequency of FISH probe distances between pairs of ERα enhancers. E_2 treatment for 50 minutes. P value indicates the probability of null hypothesis for the cumulative distribution frequencies of the two enhancers in $-E_2$ and $+E_2$ conditions.

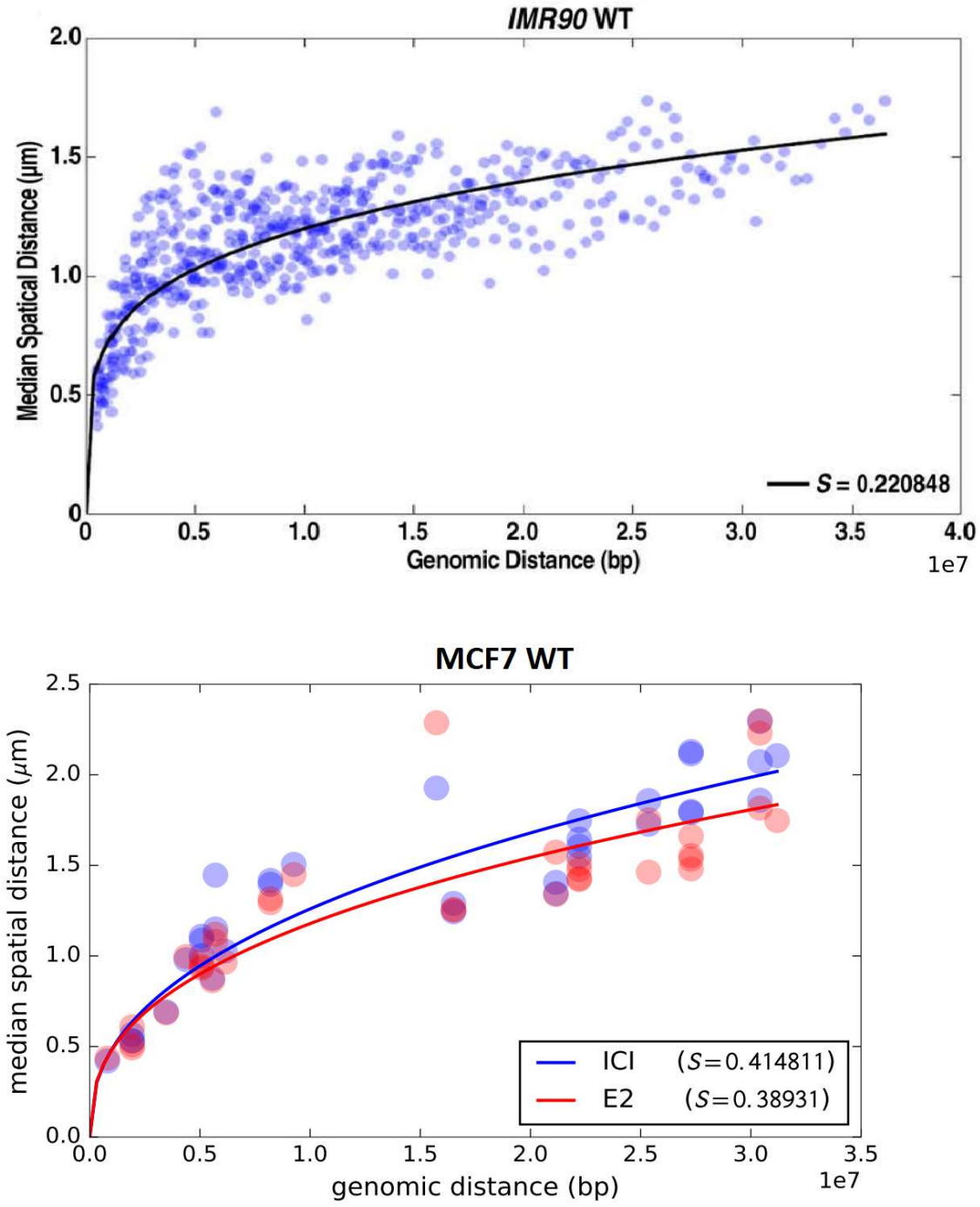


Figure 11: Plot of median inter-probe distance vs genomic distance

Plots of median inter-probe spatial distance vs genomic distance in units of 10mB. The top plot is from IMR90 human normal lung cells, the bottom plot is from MCF7 cells in ICI and E_2 treated conditions.

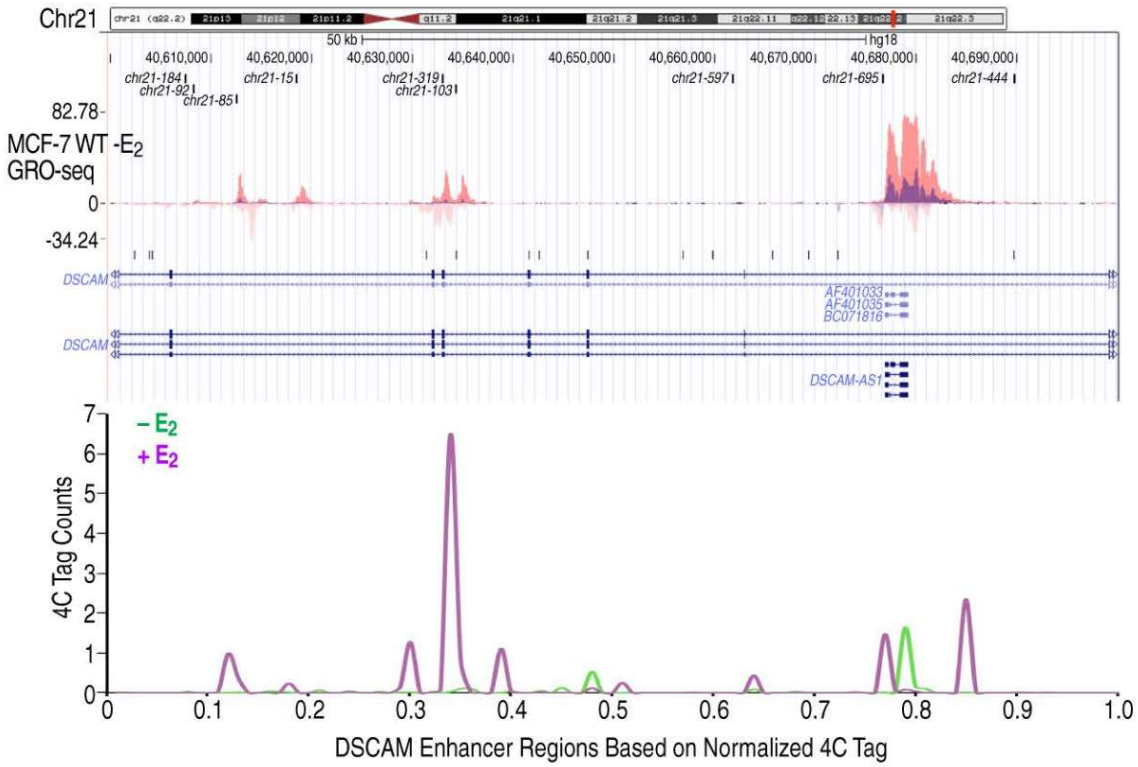


Figure 12: 4C data interaction of TFF1e1 with DSCAM-AS1e1-2

Tag count data from a 4C experiment using TFF1e1 as the "viewpoint" juxtaposed with GRO-seq data from MCF7 cells showing active transcription of the DSCAM-AS1 lncRNA, and the active eRNA transcription from the DSCAM-AS1 superenhancer region composed of individual ERα enhancers 21-319 and 21-103.

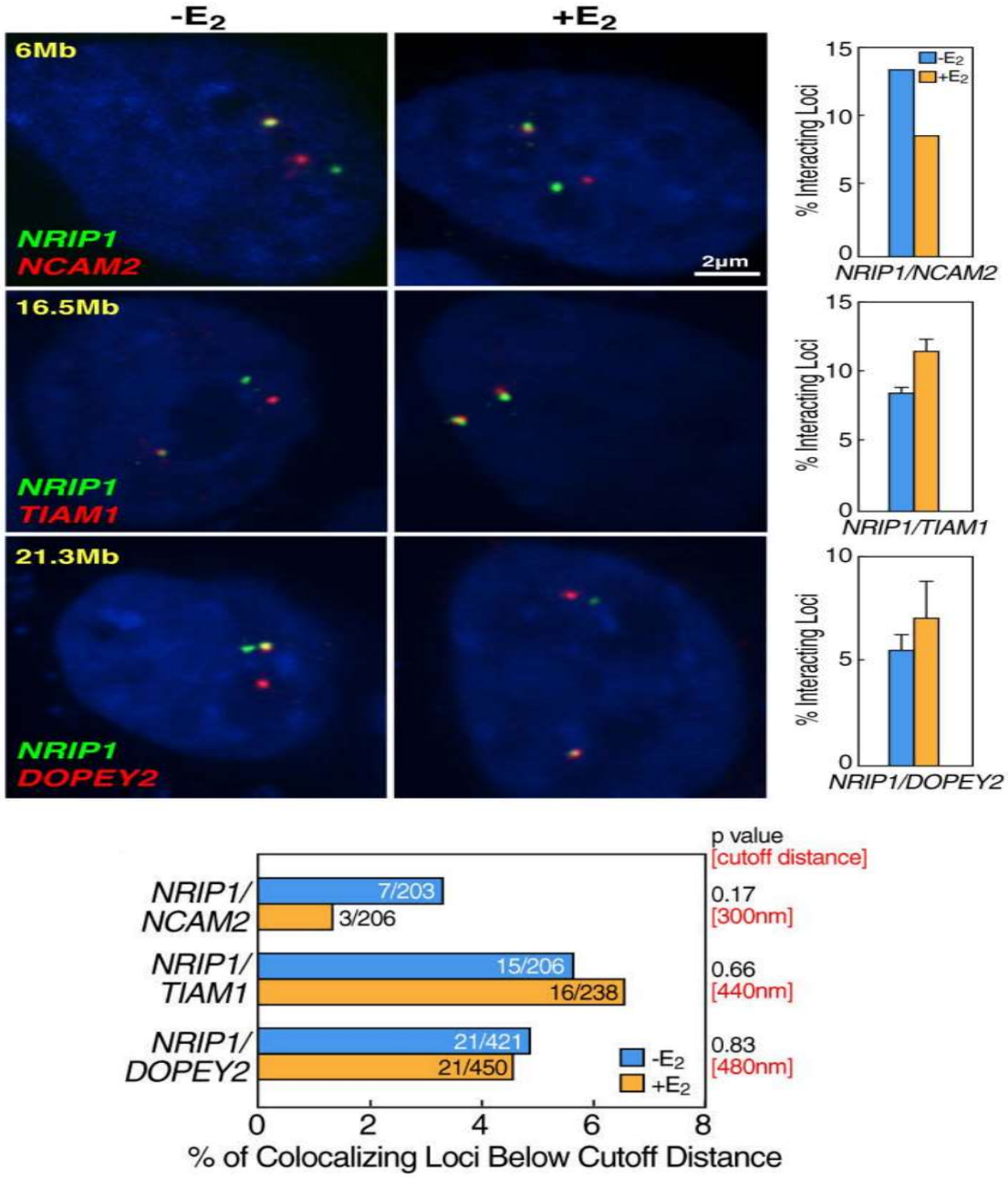


Figure 13: FISH data of interaction between NRIP1 and B enhancers

FISH analysis of interactions between pairs of enhancers located in or near the B compartment of chromosome 21 NRIP1/NCAM2, NRIP1/TIAM1, and NRIP1/DOPEY2 following treatment of cells with E_2 for 50 minutes. Quantification of the percent of alleles exhibiting overlaps is calculated by Volocity software shown as column plots. Linear distance on the chromosome between enhancers is noted in yellow text. Horizontal bar plots show colocalization between pairs of enhancers with 3D distance below a certain cutoff as calculated by a custom script.

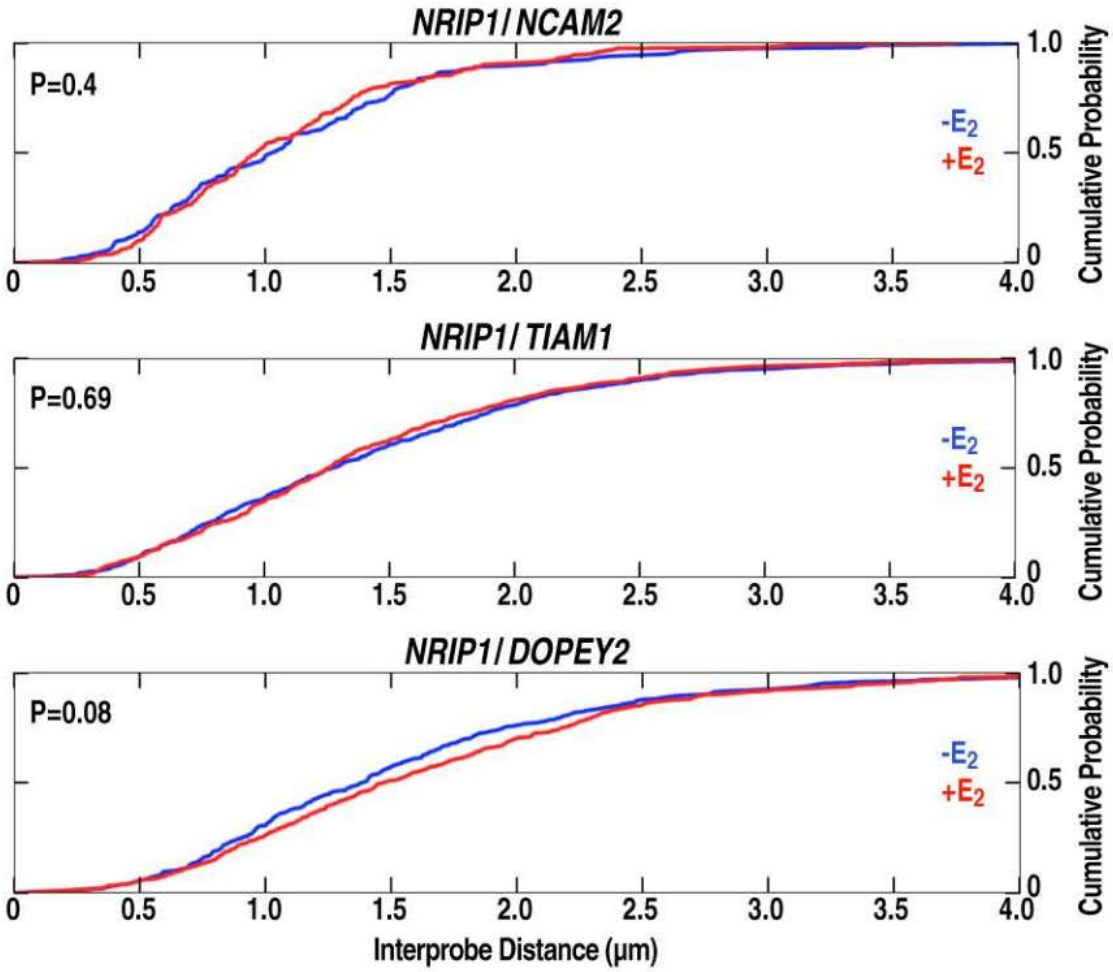


Figure 14: **Cumulative distribution of distances between enhancers in Figure 13**

The cumulative distribution frequency of FISH probe distances between pairs of ERα enhancers. E_2 treatment for 50 minutes. P value indicates the probability of null hypothesis for the cumulative distribution frequencies of the two enhancers in -E_2 and +E_2 conditions.

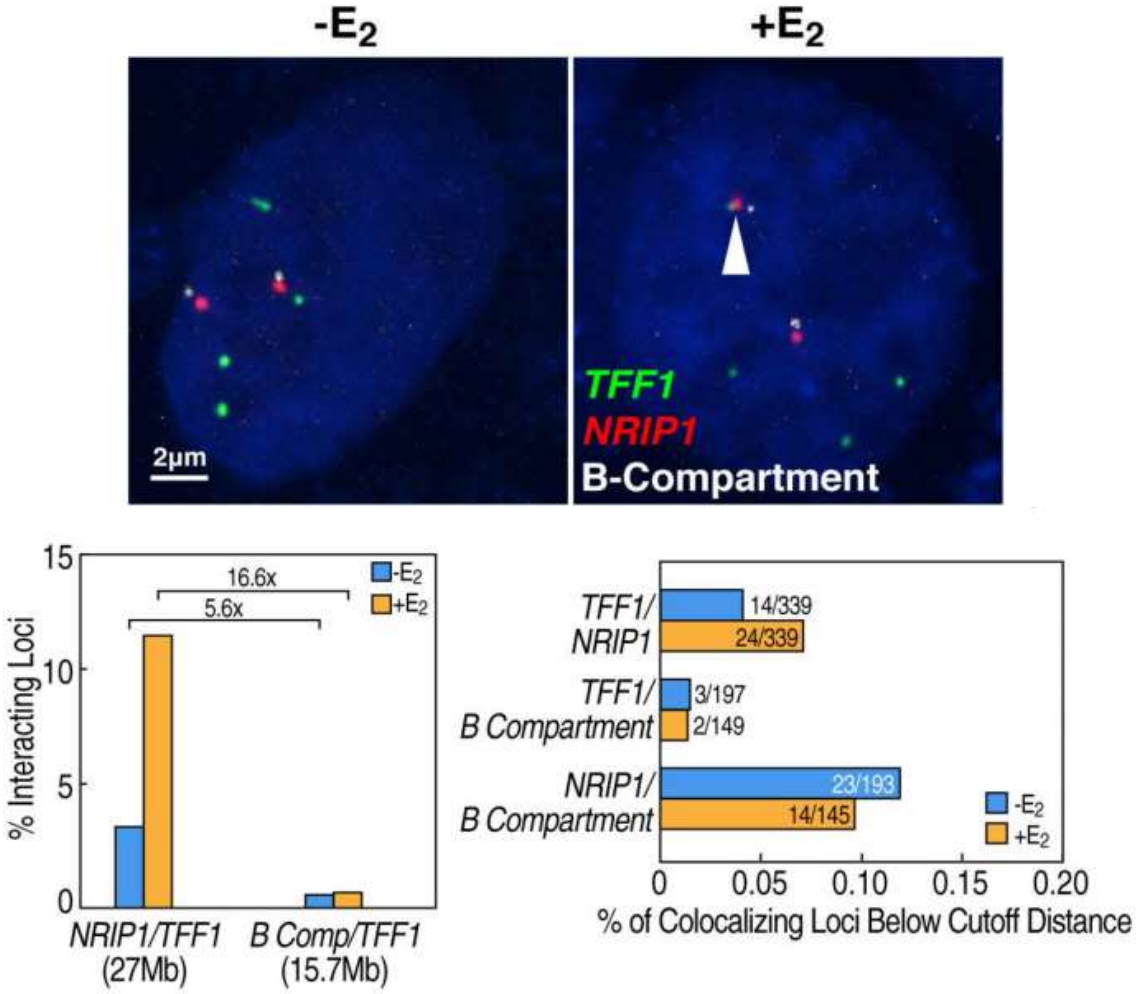
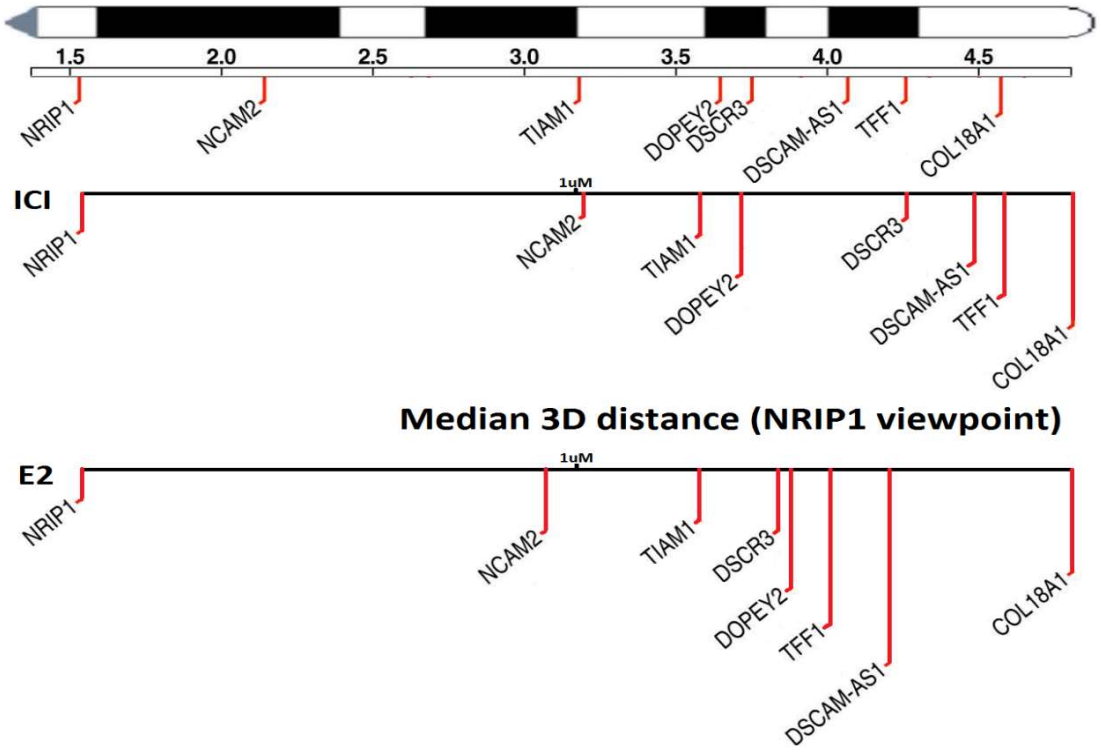


Figure 15: Interaction of NRIP1 and TFF1 with BCP probe

FISH analysis of interactions between an ERα enhancer (NRIP1 or TFF1) and a locus inside the heterochromatic B compartment (BCP) following treatment of cell with E_2 ligand for 50 minutes. Quantification of the percent of alleles exhibiting overlaps is calculated by Volocity software. Horizontal bar plots show percentage colocalization between pairs of enhancers with 3D distance below a cutoff point.

Genomic Distance and Spatial Distance



TFF1 and NRIP1

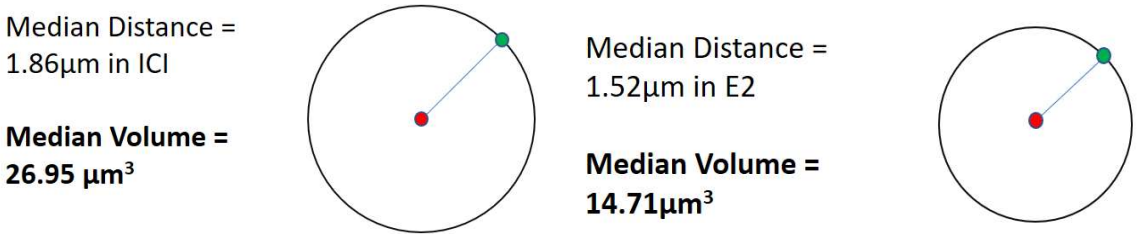


Figure 16: Comparison of 1D genomic distance and 3D spatial distance

The dynamic and E_2 regulated changes to the median 3D distance between NRIP1 and a set of other "first tier" ERα enhancers. Top panel shows a linear map of chromosome 21 with the positions of NRIP1 and seven other ERα enhancers highlighted in red. Middle panel shows the median spatial distance between NRIP1 (viewpoint) and the same enhancers in cells treated with ICI antagonist or E_2 ligand for 50 minutes. Bottom panels shows the cross section (passing through a diameter) of a sphere formed by the median distance between the TFF1 and NRIP1 loci in MCF7 cells in ICI and E2 conditions.

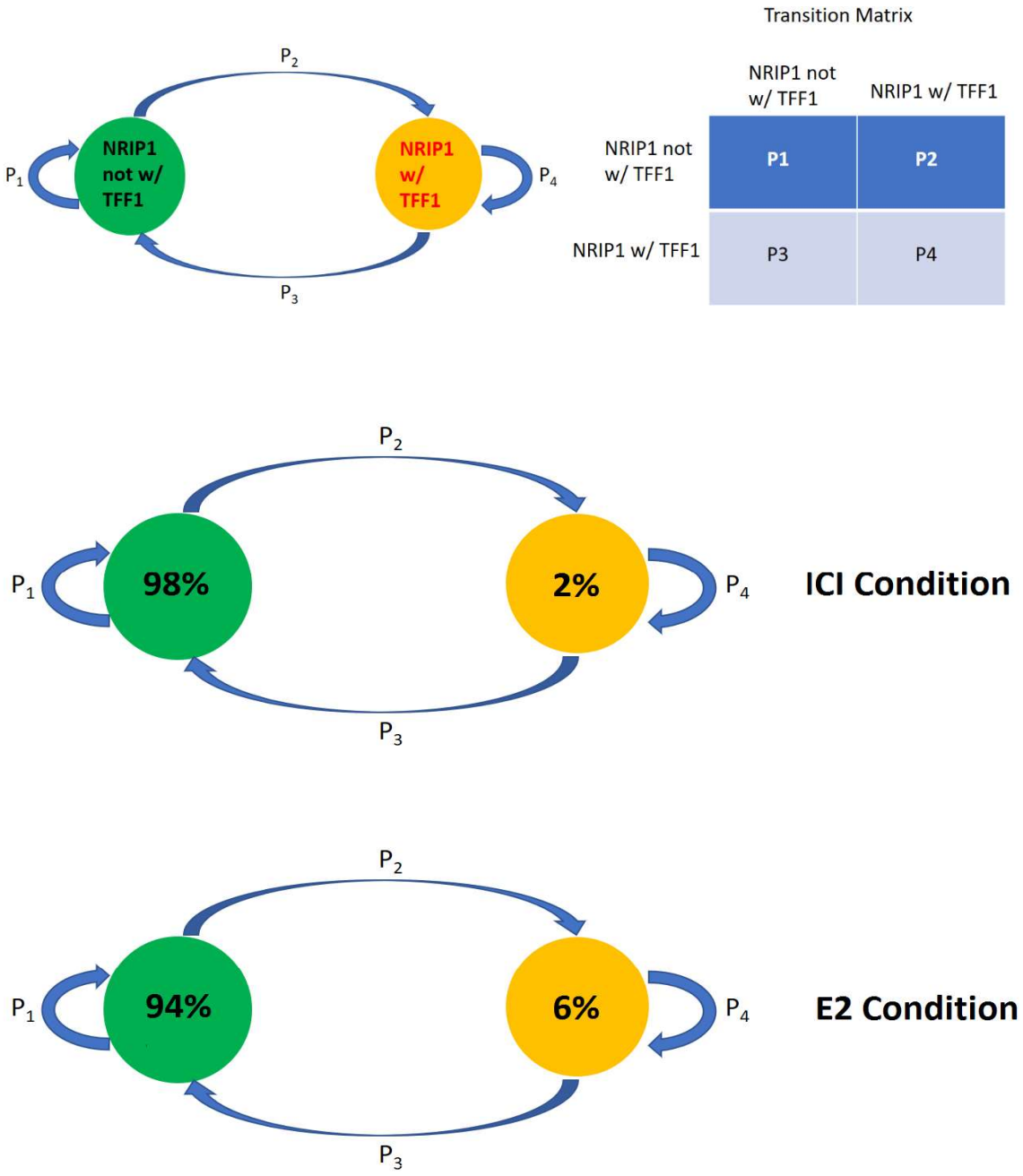


Figure 17: Markov chain representing interactions of NRIP1 and TFF1

A two state Markov chain representing the interaction status of the TFF1 and NRIP1 enhancers from a two-color FISH experiment, percentages represent steady state distributions as inferred from FISH data. Transition probabilities between these states is not observable in cells treated with formaldehyde.

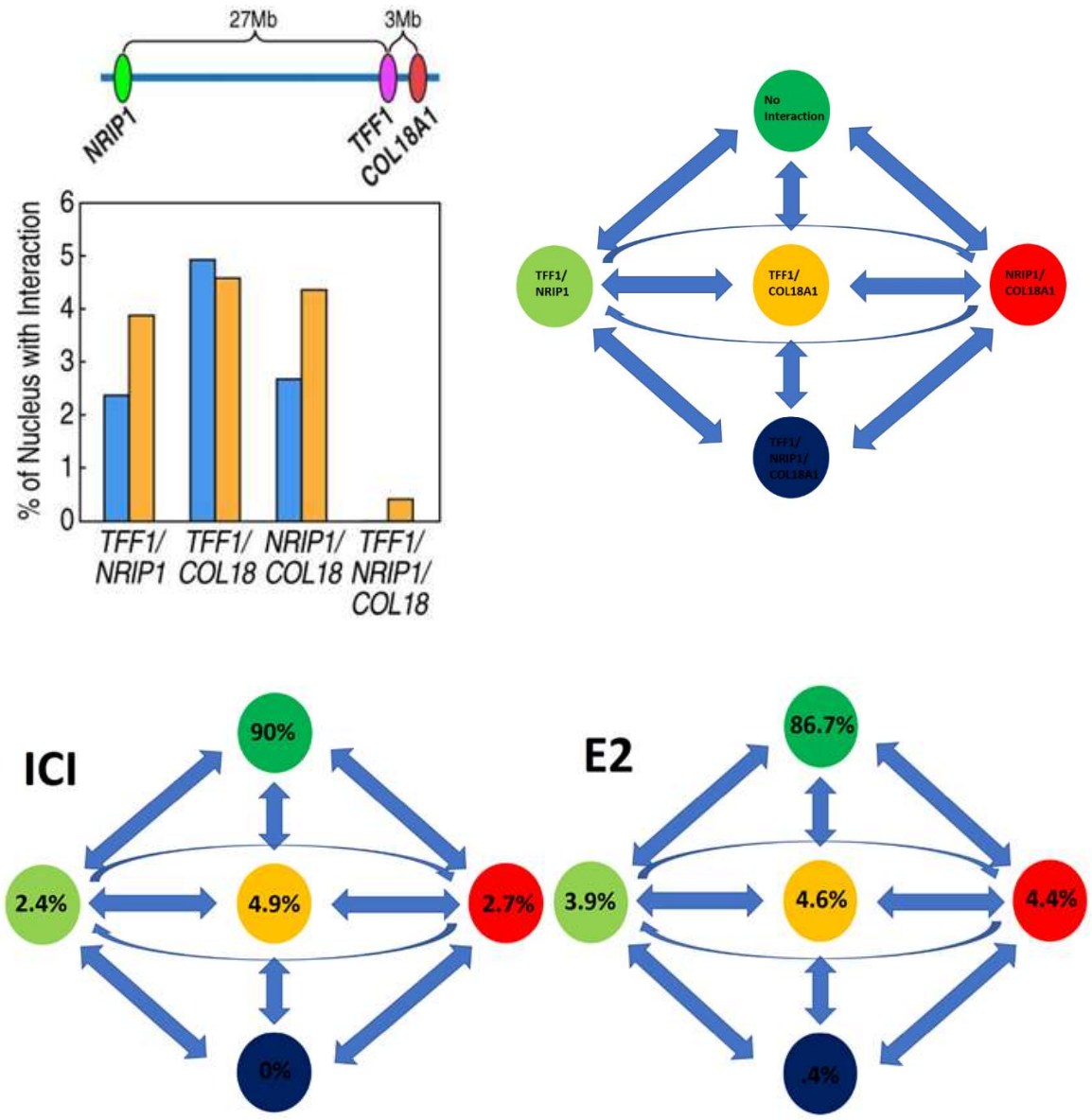


Figure 18: Markov chain representing interactions of NRIP1, TFF1, COL18A1

A five state Markov chain representing the interaction states of three ERα enhancers TFF1, NRIP1 and COL18A1. The chance of a three loci interaction is roughly 50X higher than the product of three separate two loci interactions. Transition probabilities between these states is not observable in cells treated with formaldehyde.

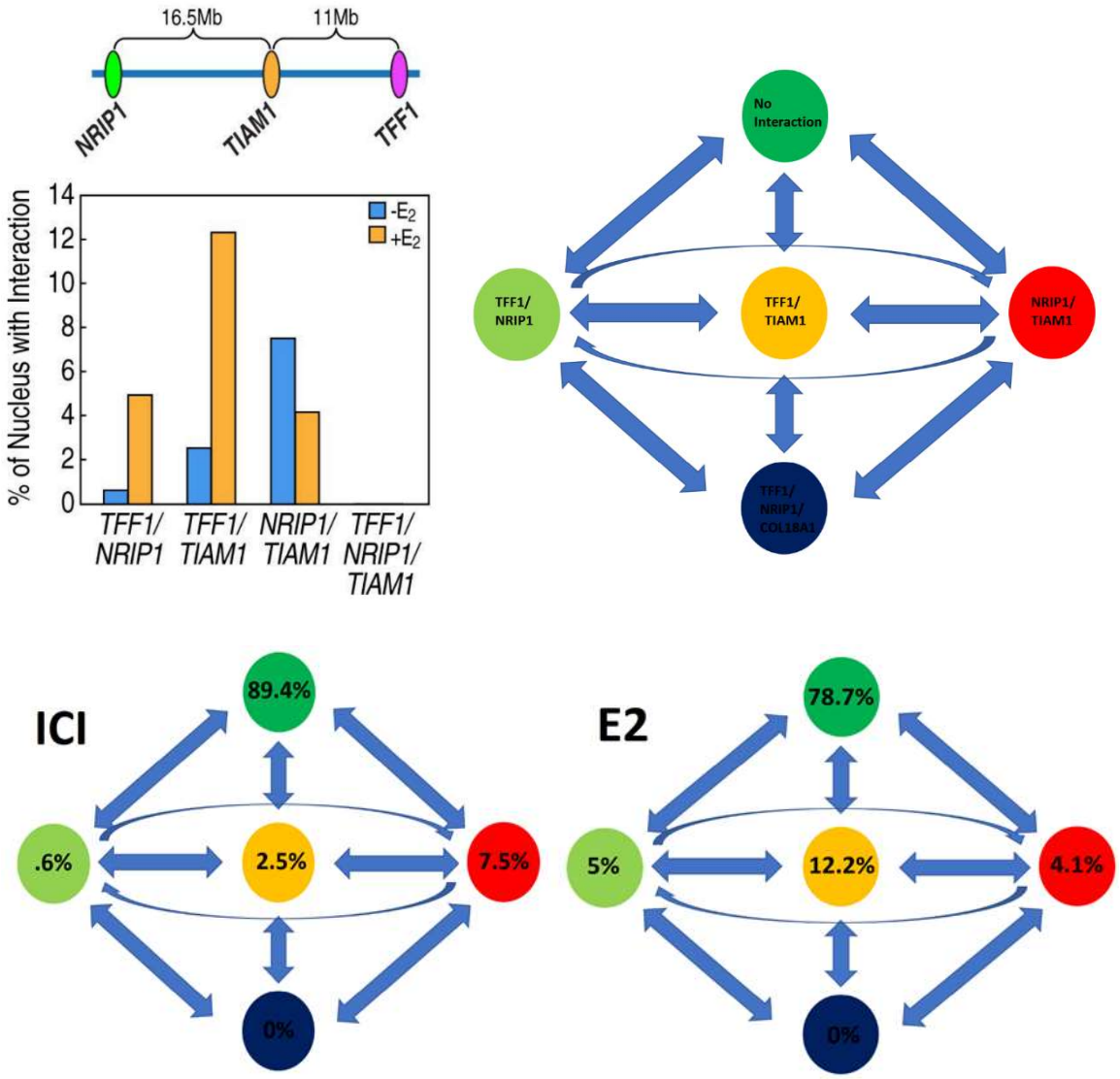


Figure 19: Markov chain representing interactions of NRIP1, TIAM1, TFF1

A five state Markov chain representing the interaction states of three ERα enhancers TFF1, NRIP1 and TIAM1. Examples of three loci interaction weren't found in the cells sampled. Transition probabilities between these states is not observable in cells treated with formaldehyde.

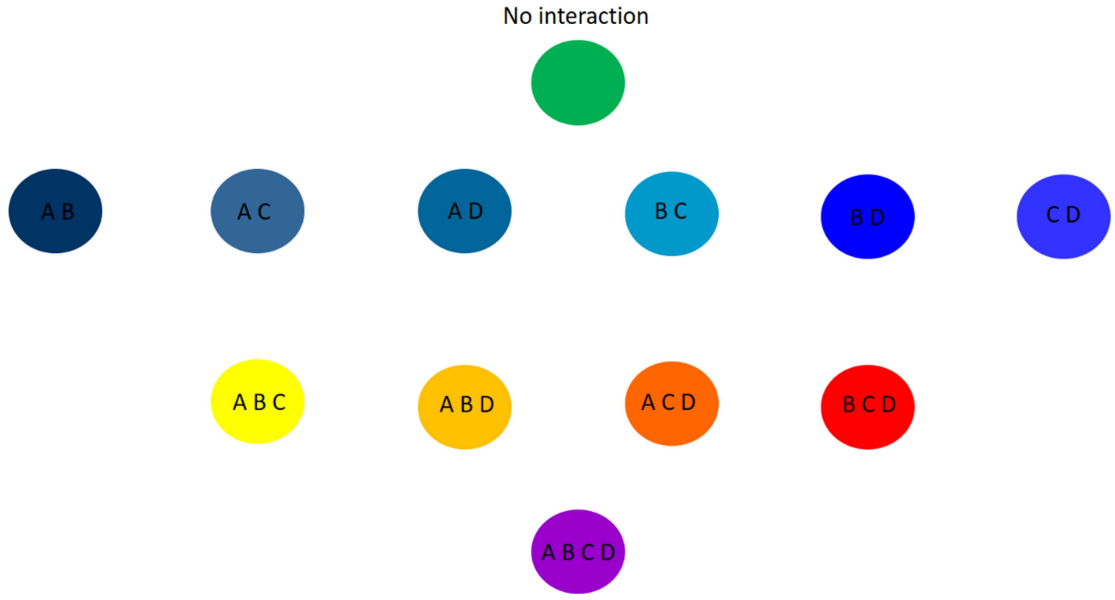


Figure 20: **Markov chain representing interactions of 4 hypothetical enhancers**

An eleven state Markov chain representing the interaction states of four hypothetical enhancers A, B, C, and D. Arrows representing the one hundred and twenty-one transition probabilities between these states have been omitted.

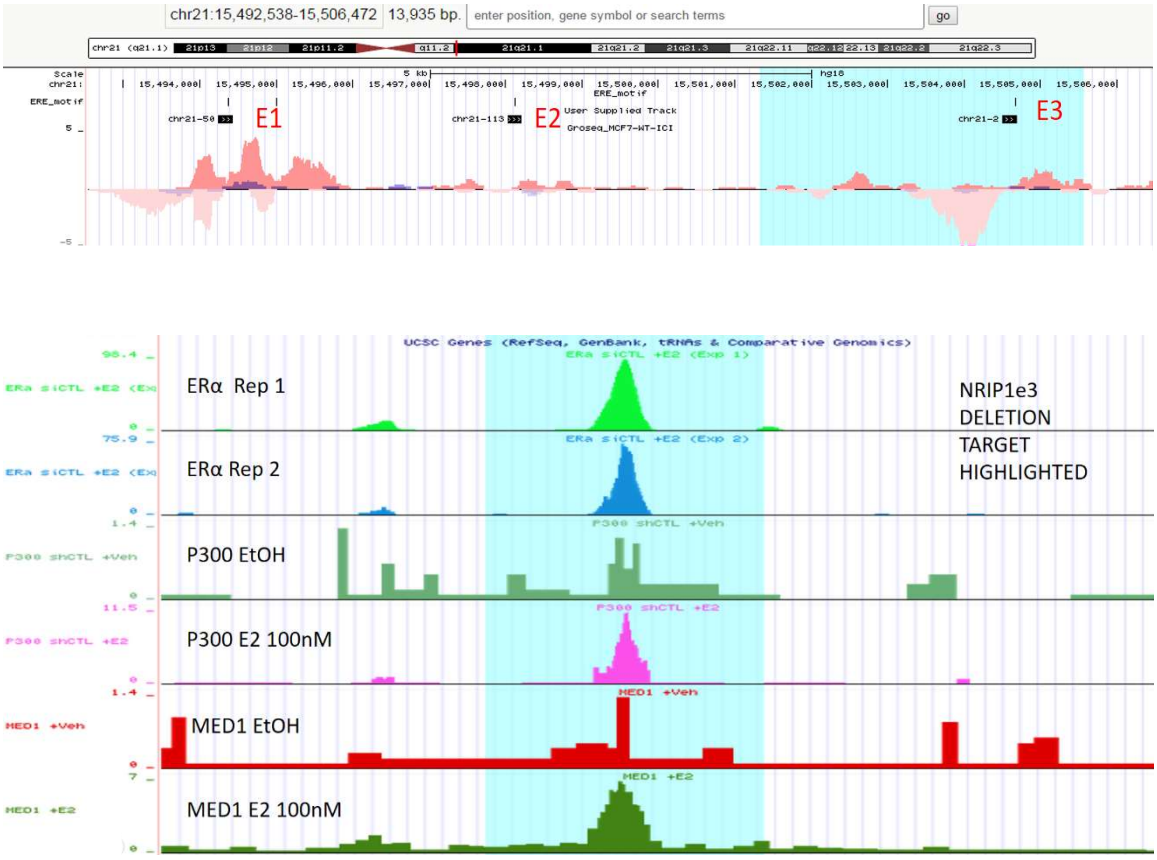


Figure 21: Overview of the NRIP1e3 superenhancer region and deletion

Overview of the NRIP1e3 deletion in MCF7 cell lines. Top panel is GRO-seq tag counts mapped on the UCSC genome browser for the NRIP1e1-3 superenhancer, deleted region of NRIP1e3 is shown with light blue highlighting. Bottom panel shows the binding of ERα, P300 and MED1 to NRIP1e3.

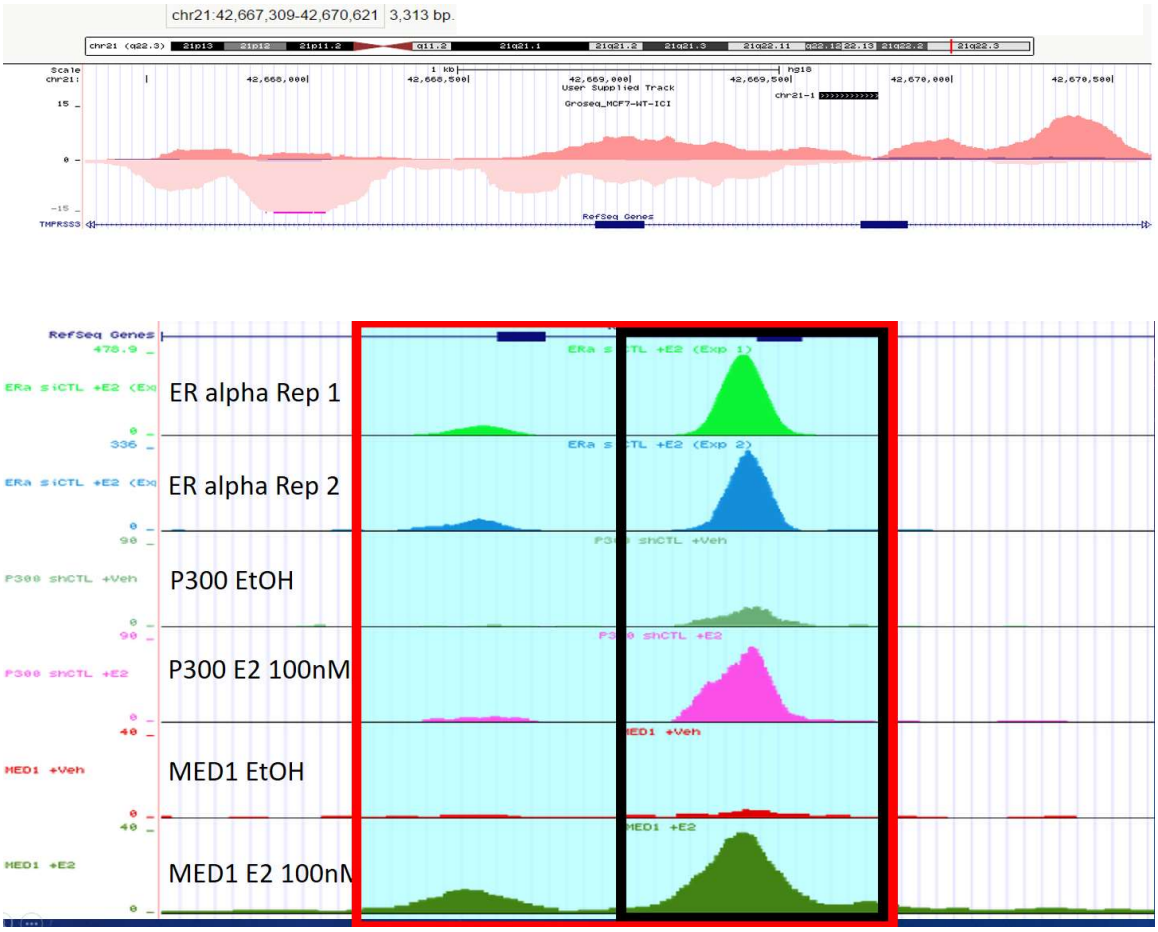
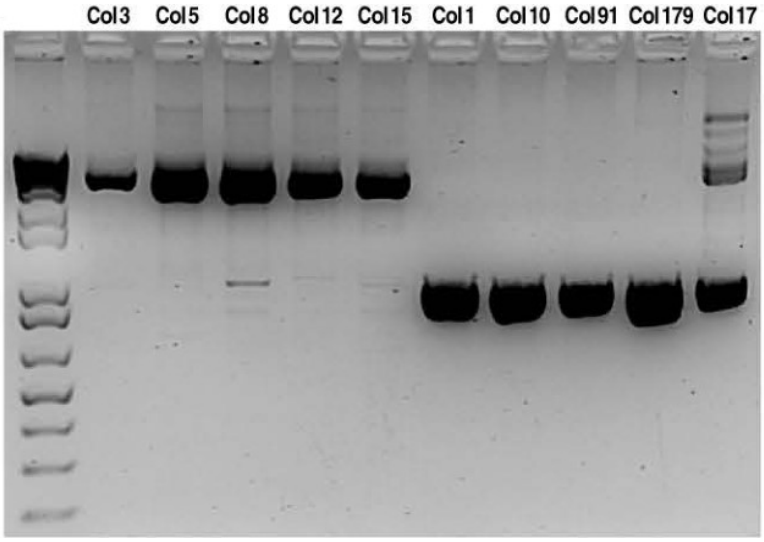
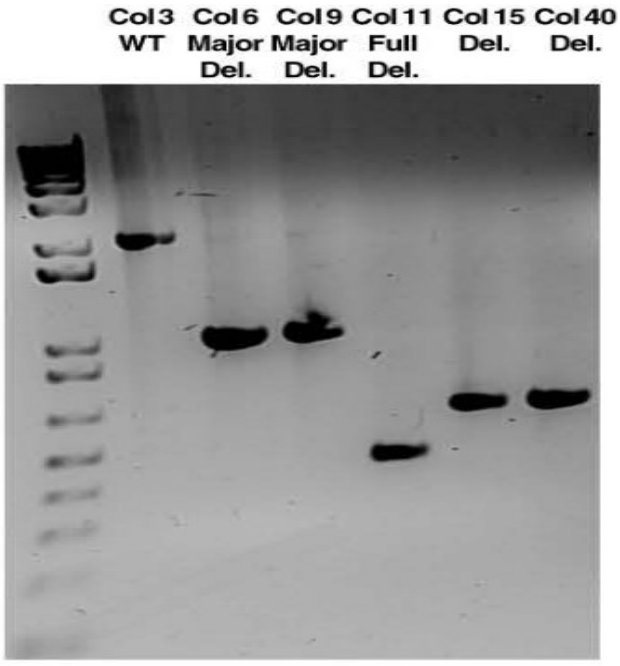


Figure 22: Overview of the TFF1e1 enhancer region and deletion

Overview of the TFF1e1 deletion in MCF7 cell lines. Top panel is GRO-seq tag counts mapped on the UCSC genome browser for the TFF1e1 enhancer. Bottom panel shows the binding of ERα, P300 and MED1 to TFF1e1. Deletions of different sizes were made for the TFF1e1, two of them are shown in red (full deletion) and black (major deletion).



NRIP1e3 PCR



TFF1e1 PCR

Figure 23: PCR genotyping of NRIP1e3 and TFF1e1 deletions

PCR genotyping of NRIP1e3 and TFF1e1 deletions. Top panel shows the NRIP1e3 PCR product, colonies 3,5,8,12,15 are wild type; colonies 1,10,91,179 are homozygous deletions and colony 17 is a heterozygous deletion. Bottom panel shows TFF1e1 PCR product, colony 3 is wild type, colonies 6,9,11,15 and 40 are deletions of various sizes.

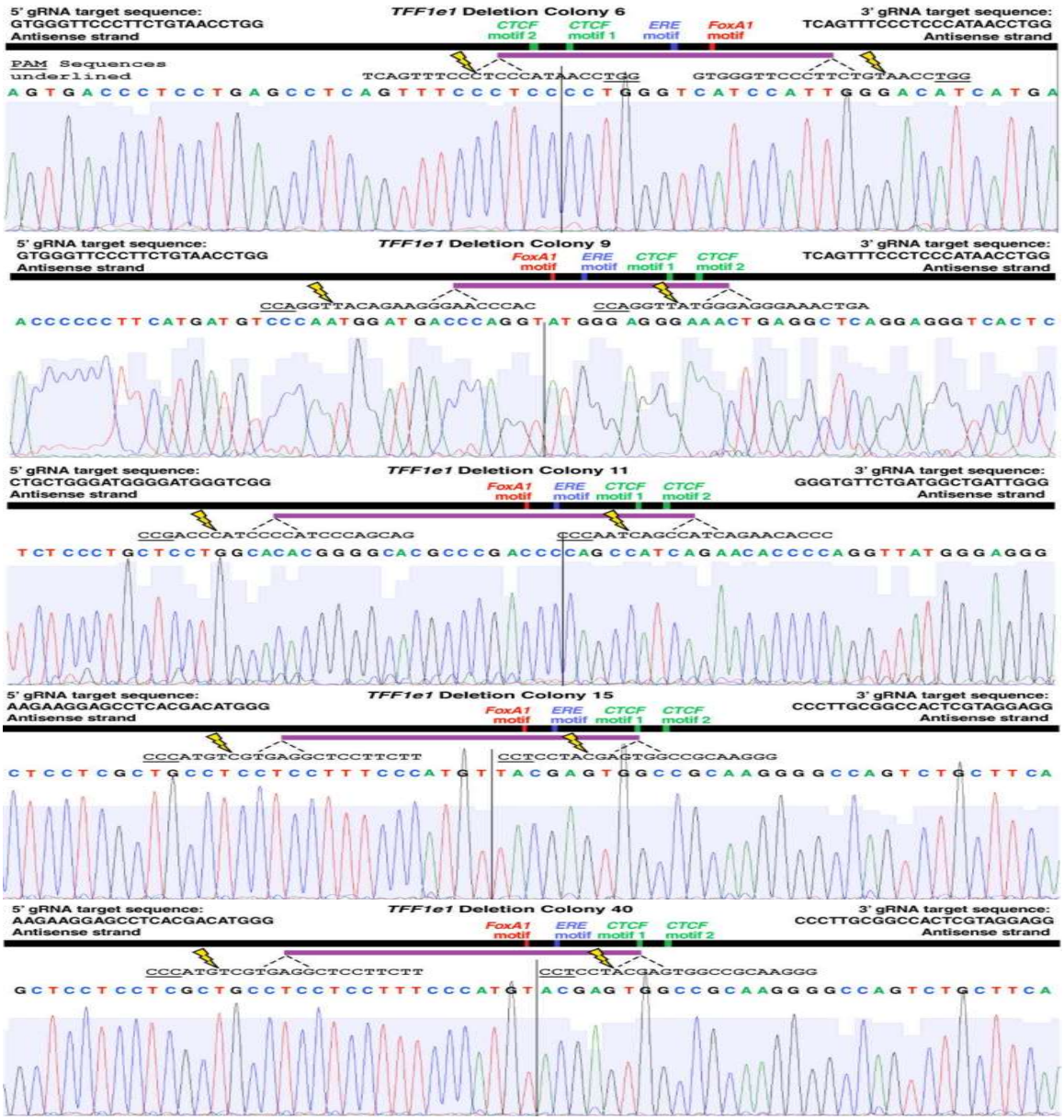


Figure 24: Sanger sequencing of PCR products from TFF1e1 deletions

Diagrams showing the deletions of TFF1e1, black line represents the entire enhancer sequence, purple line represents the deletion site. Guide RNA binding sites are shown as well as the sequences of the 5' and 3' gRNA used.

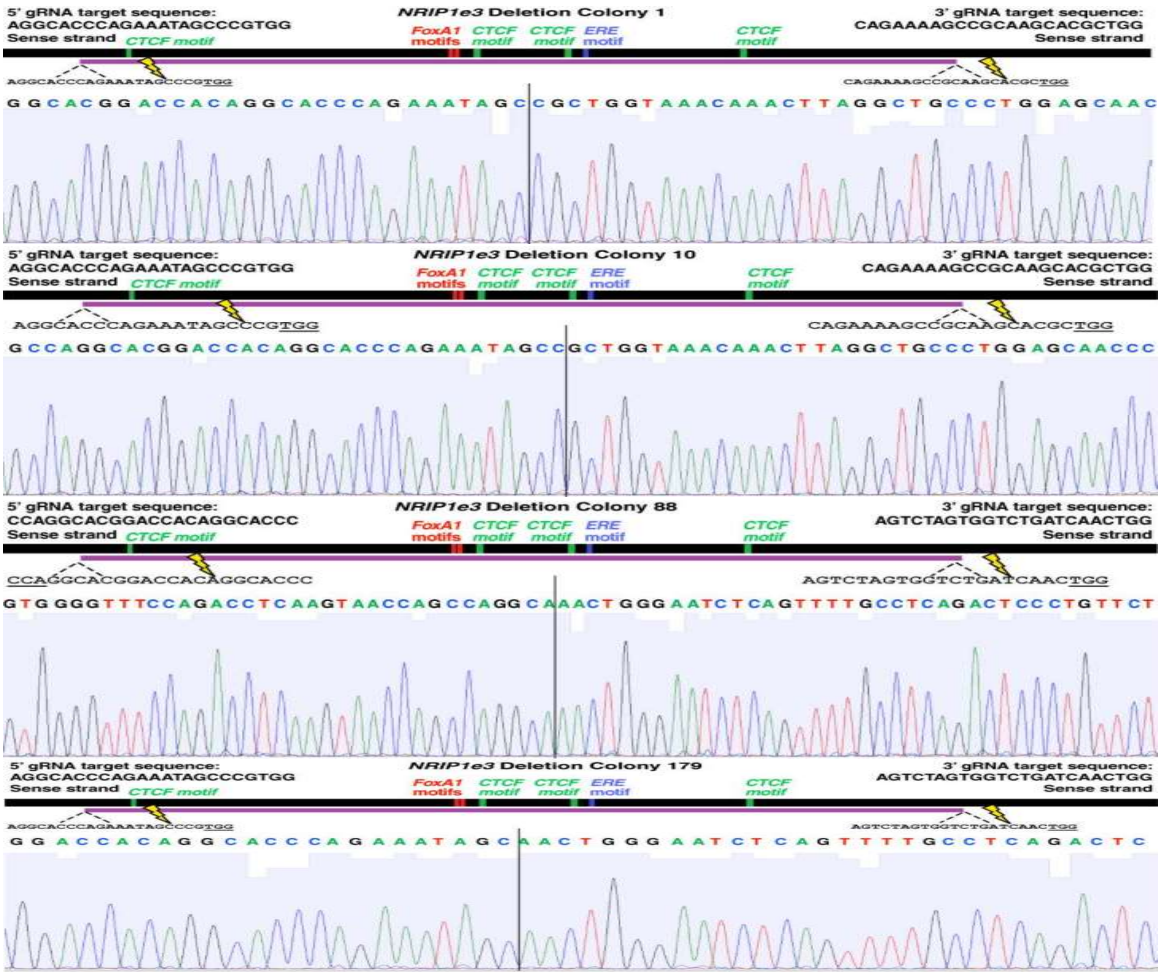


Figure 25: Sanger sequencing of PCR products from NRIP1e3 deletions

Diagrams showing the deletions of NRIP1e3, black line represents the entire enhancer sequence, purple line represents the deletion site. Guide RNA binding sites are shown as well as the sequences of the 5' and 3' gRNA used.

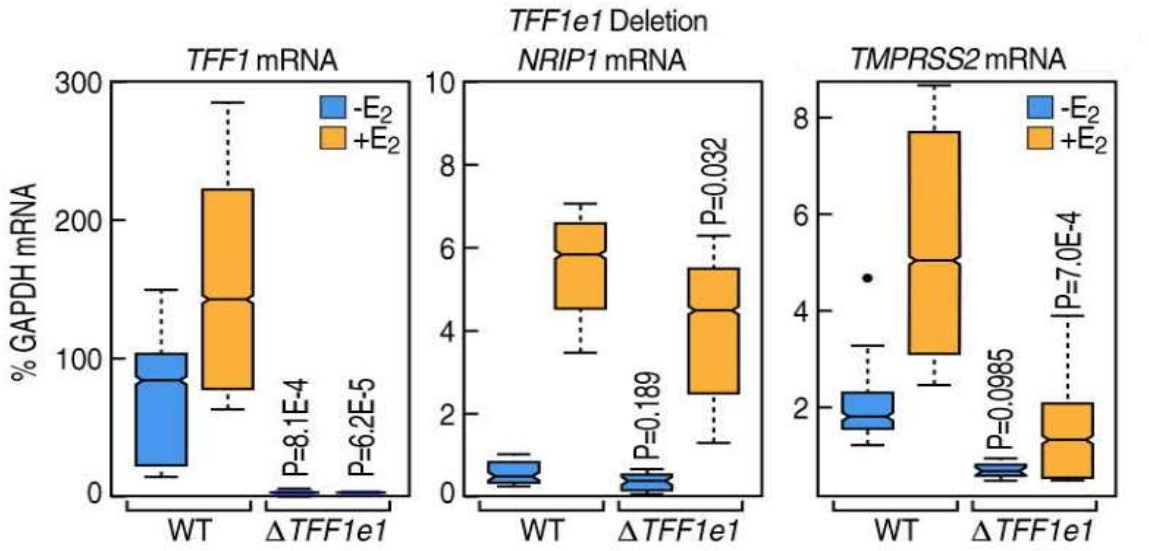


Figure 26: Changes of expression in E_2 regulated genes in TFF1e1 deletion

Expression levels of three highly E_2 induced coding genes on chromosome 21 in wild type and TFF1e1 deletion cell lines. Four independent clonal cell lines of wild type and TFF1e1 deletion cell lines were used, three technical replicates were done for each cell line.

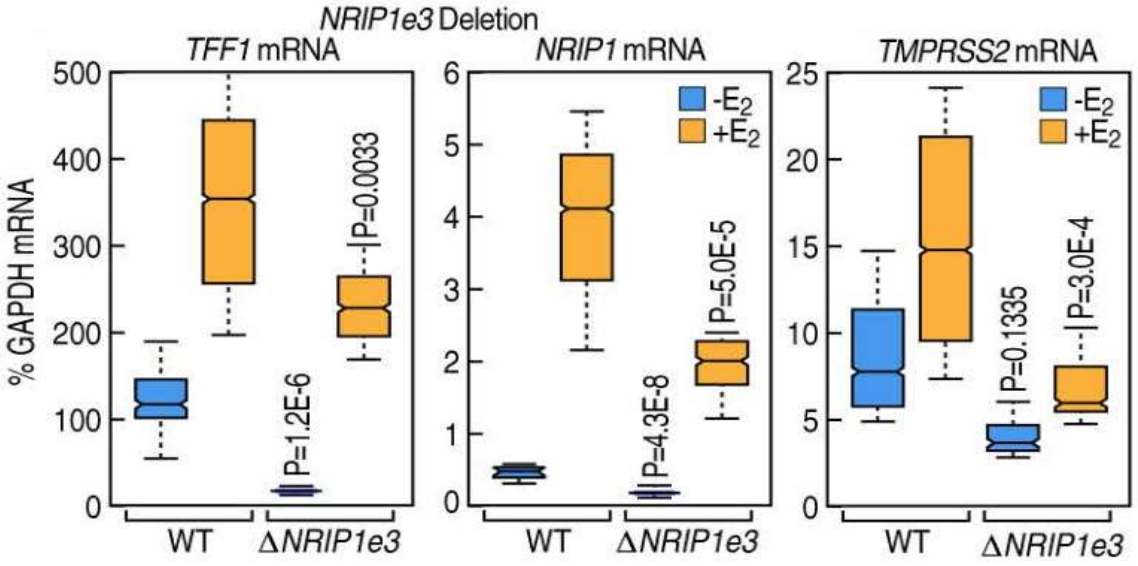


Figure 27: Changes of expression in E_2 regulated gene in NRIP1e3 deletion

Expression levels of three highly E_2 induced coding genes on chromosome 21 in wild type and NRIP1e3 deletion cell lines. Four independent clonal cell lines of wild type and NRIP1e3 deletion cell lines were used, three technical replicates were done for each cell line.

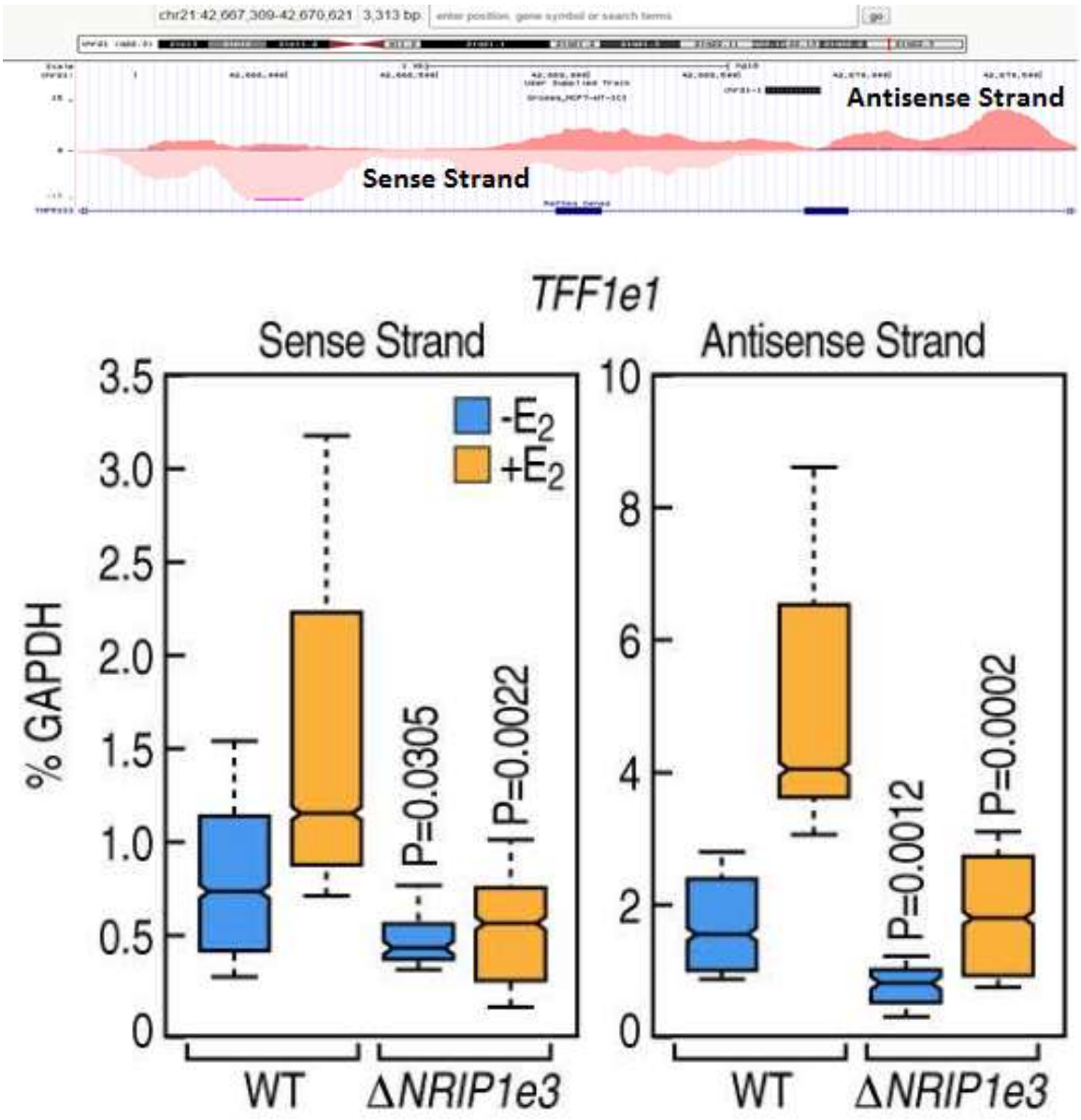


Figure 28: TFF1e1 expression changes in NRIP1e3 deletion

Top panel is UCSC genome browser data of the TFF1e1 locus, the full length of the antisense eRNA strand is not shown, as the full eRNA is over 20kb in length. Bottom panel shows the transcript levels of the TFF1 eRNAs in wild type and NRIP1e3 deletion cell lines. Four independent clonal cell lines of wild type and NRIP1e3 deletion cell lines were used, three technical replicates were done for each cell line.

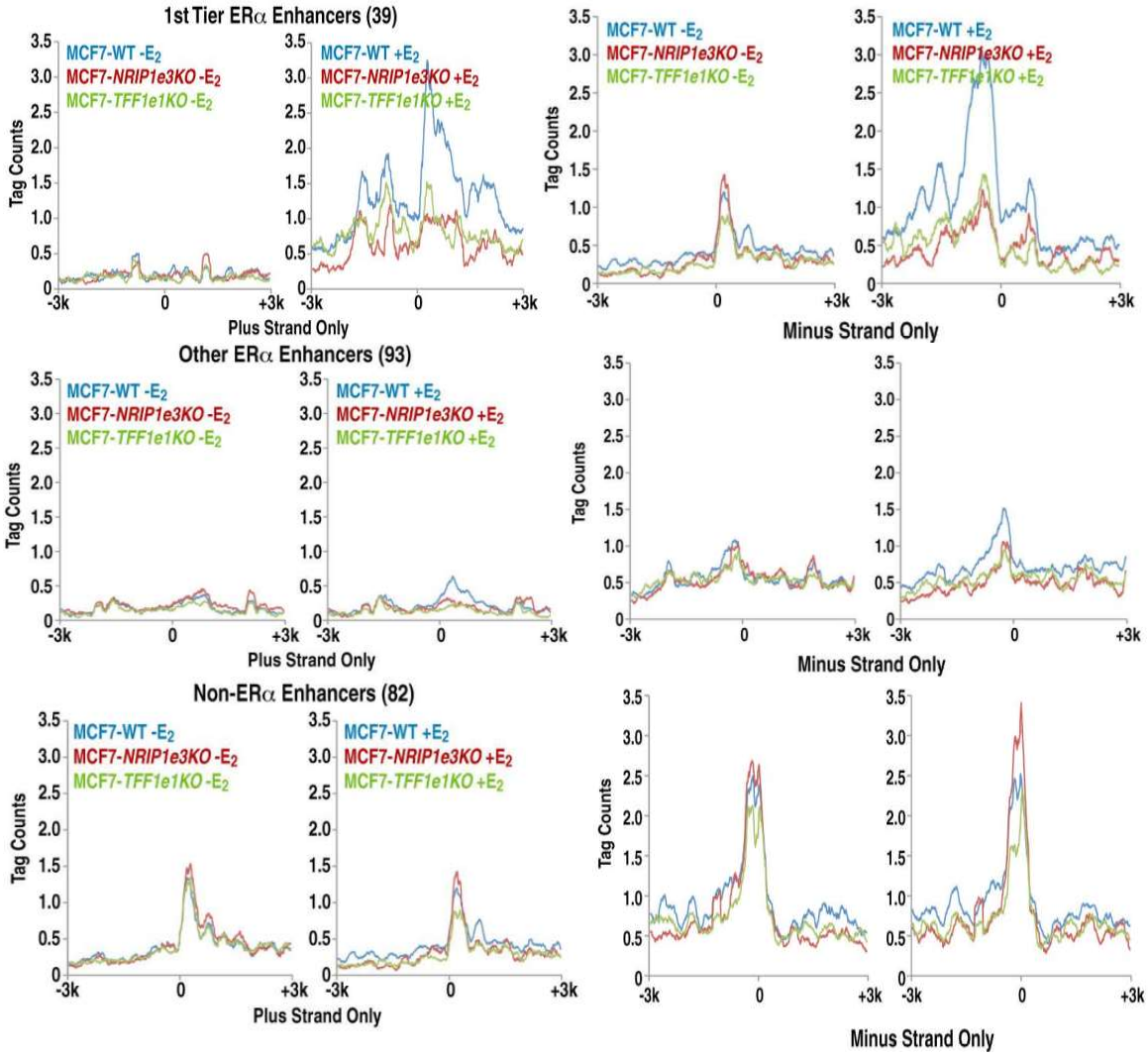


Figure 29: Meta-analysis of eRNA expression in enhancer deletion lines

Meta-analysis of eRNA transcription from first tier ERα enhancers, other ERα enhancers, and non-ERα enhancers on chromosome 21. Transcription from wild type MCF7 cell shown in blue, TFF1e1 deletion cells in green, and NRIP1e3 deletion cells in red.

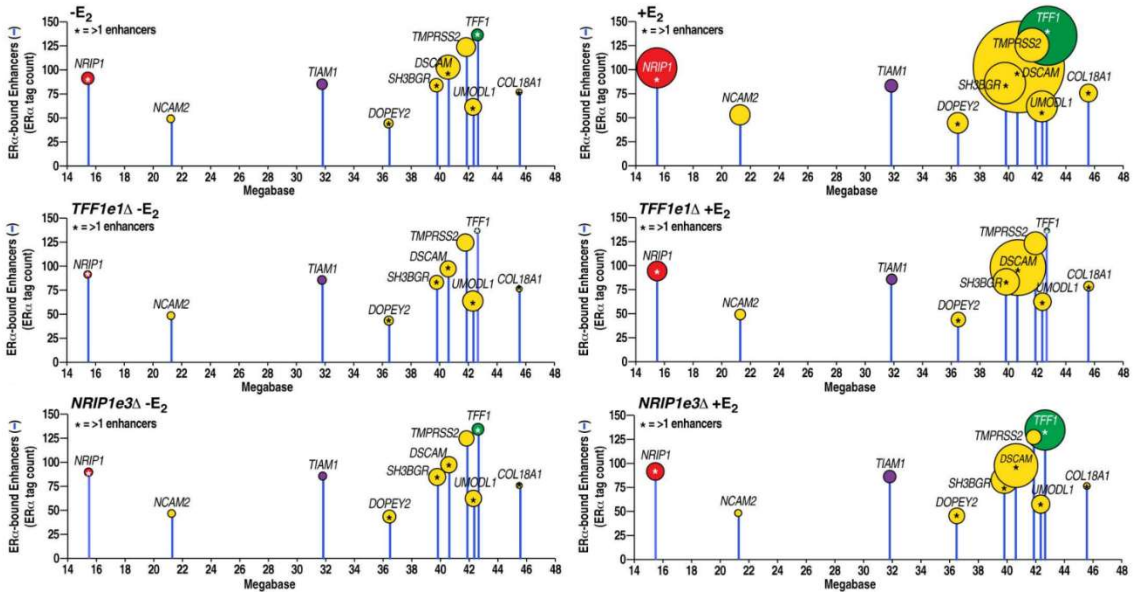


Figure 30: RPKM units from deletion lines on Chr. 21 enhancers

RPKM units from GRO-seq experiments for wild type, TFF1e1 deletion, and NRIP1e3 deletion cell lines represented as the surface area of circles on ten top ERα enhancers.

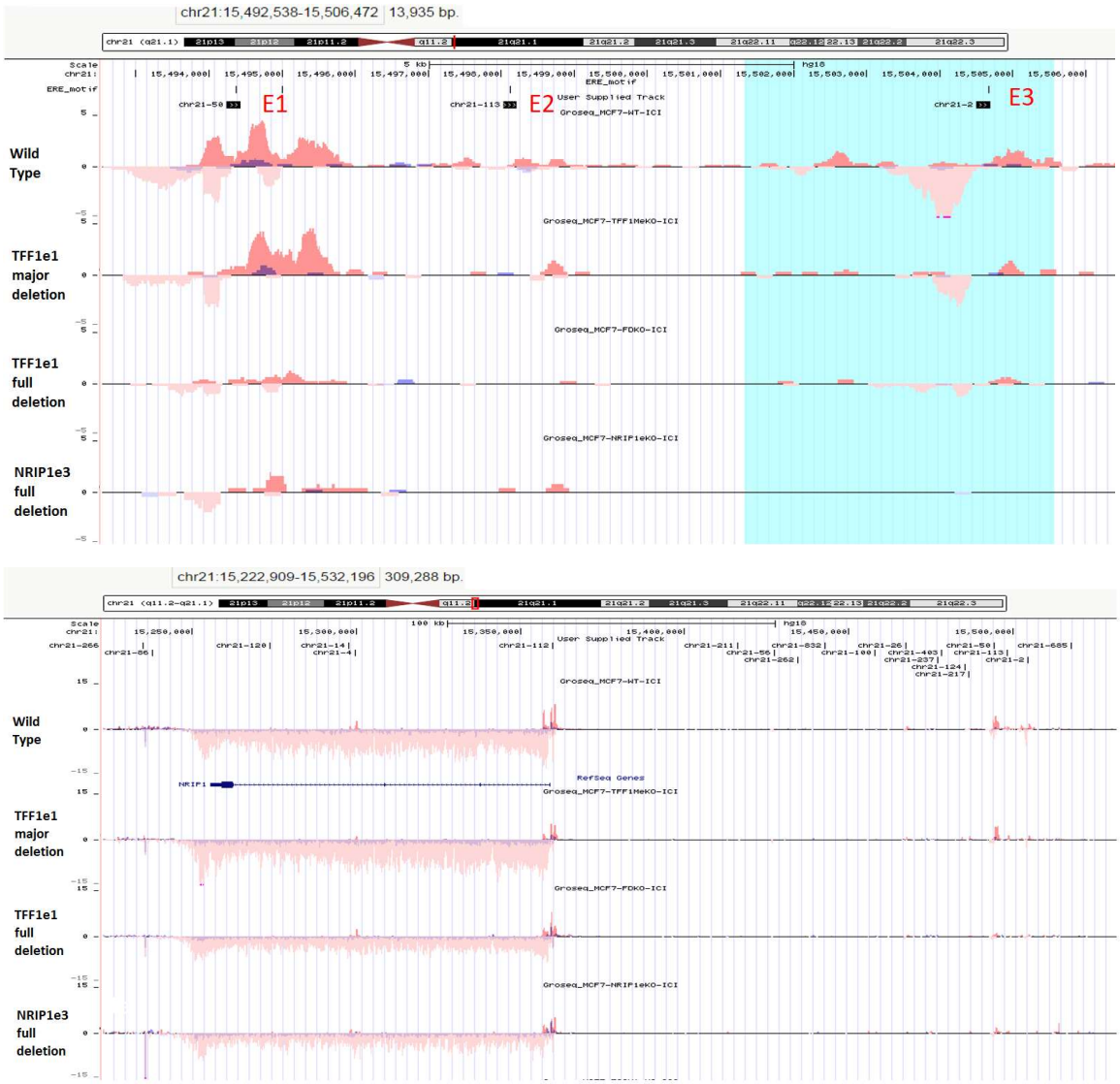


Figure 31: GRO-seq data from the NRIP1 enhancers and gene

GRO-seq data mapped onto the UCSC genome browser from wild type, TFF1e1 major deletion, TFF1e1 full deletion and NRIP1e3 full deletion MCF7 clonal cell lines. Top panel shows transcription of the NRIP1e1-3 superenhancer cluster, bottom panel shows transcription of the NRIP1 mRNA.

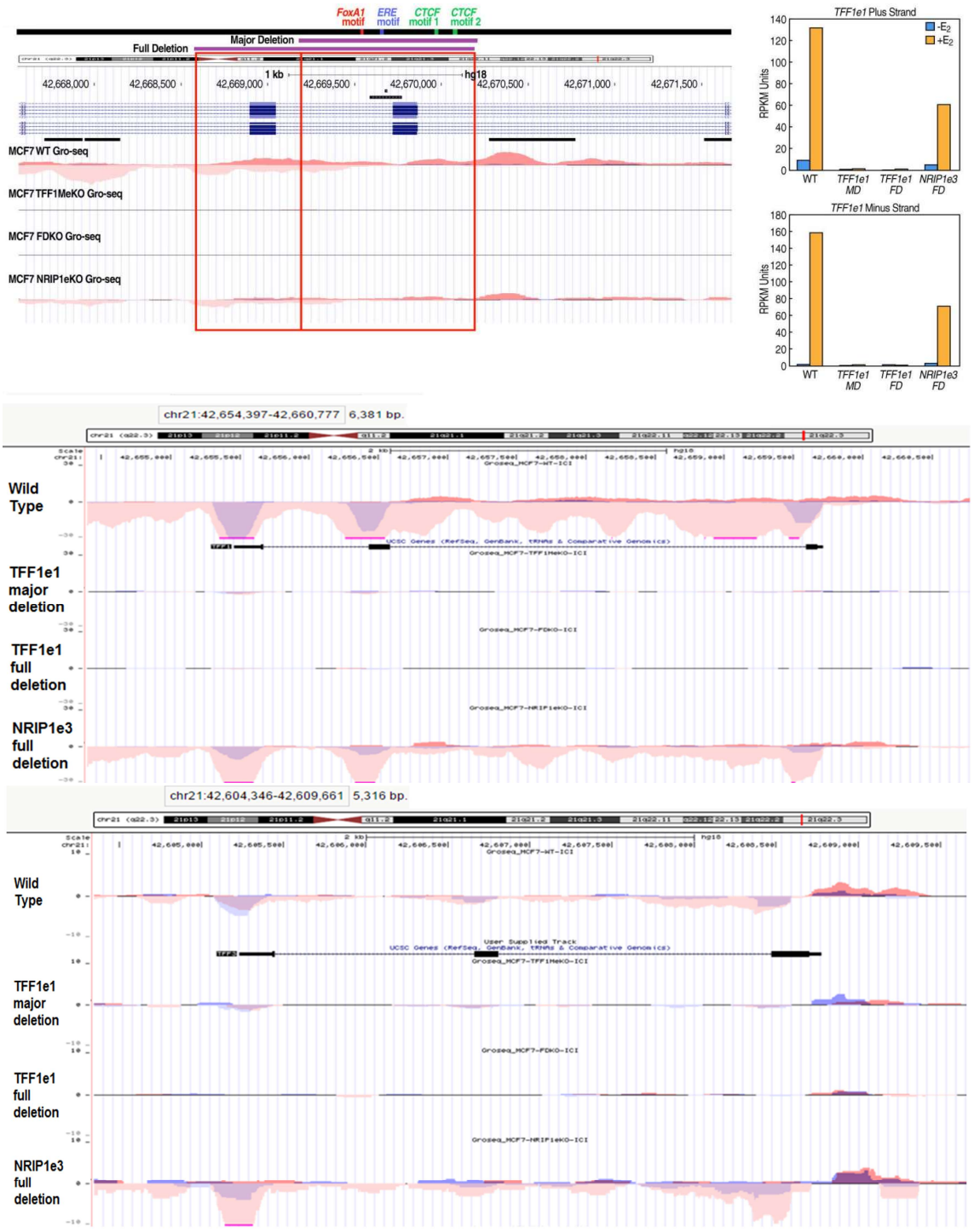


Figure 32: GRO-seq data from the TFF1 enhancer and gene family

GRO-seq data mapped onto the UCSC genome browser from wild type, TFF1e1 major deletion, TFF1e1 full deletion and NRIP1e3 full deletion MCF7 clonal cell lines. Top panel shows transcription of the TFF1e1 eRNA, bottom panels show transcription of the TFF1 and TFF3 mRNA.

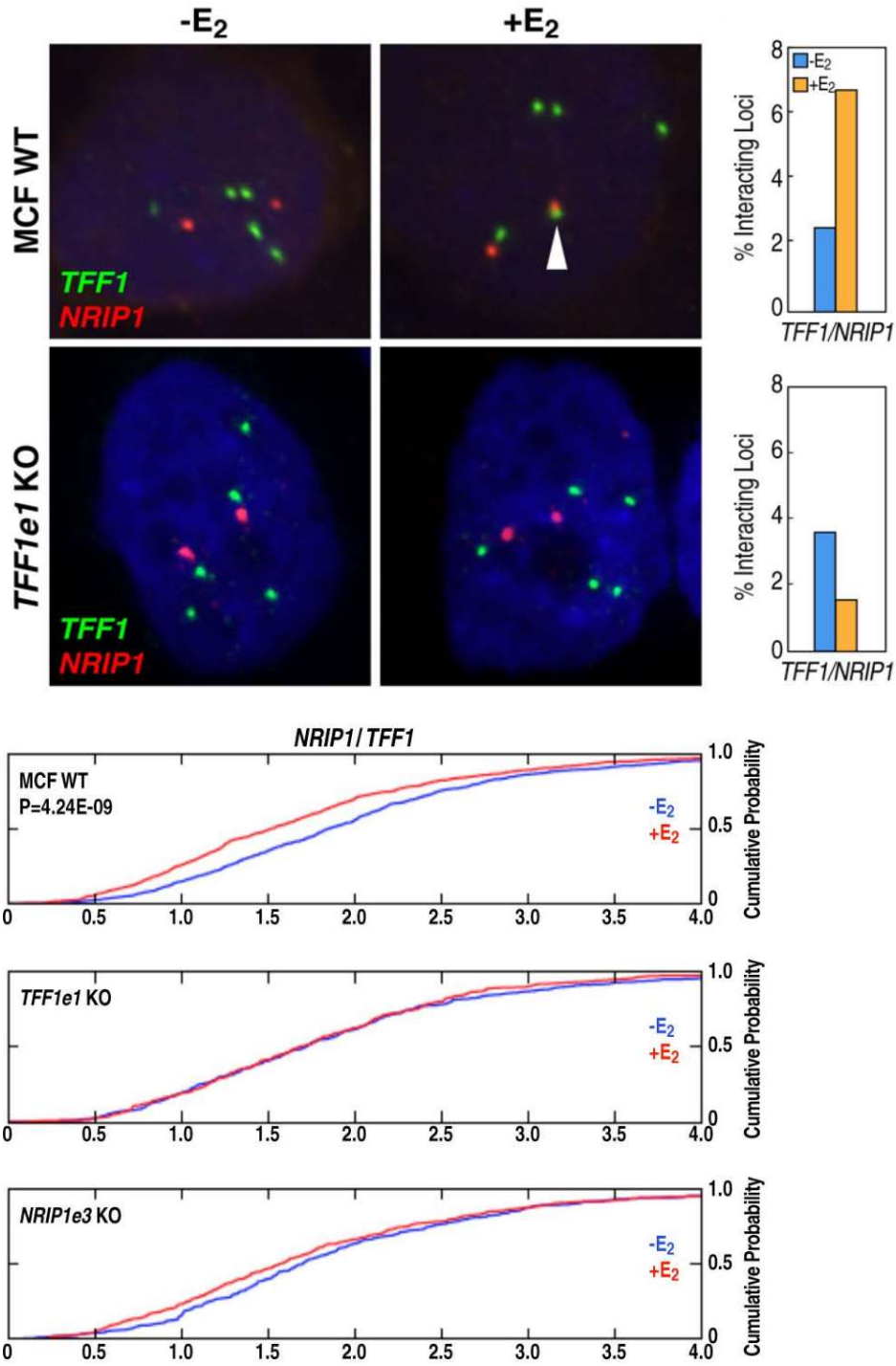


Figure 33: Effect of deletions on E_2 induced TFF1 and NRIP1 proximity

Top panel, deletion of the TFF1e1 locus negates the E_2 mediated increases in interactions between these NRIP1 and TFF1. Bottom panel, the induced proximity of TFF1 and NRIP1 is completely abolished by TFF1e1 deletion and diminished by deletion of NRIP1e3.

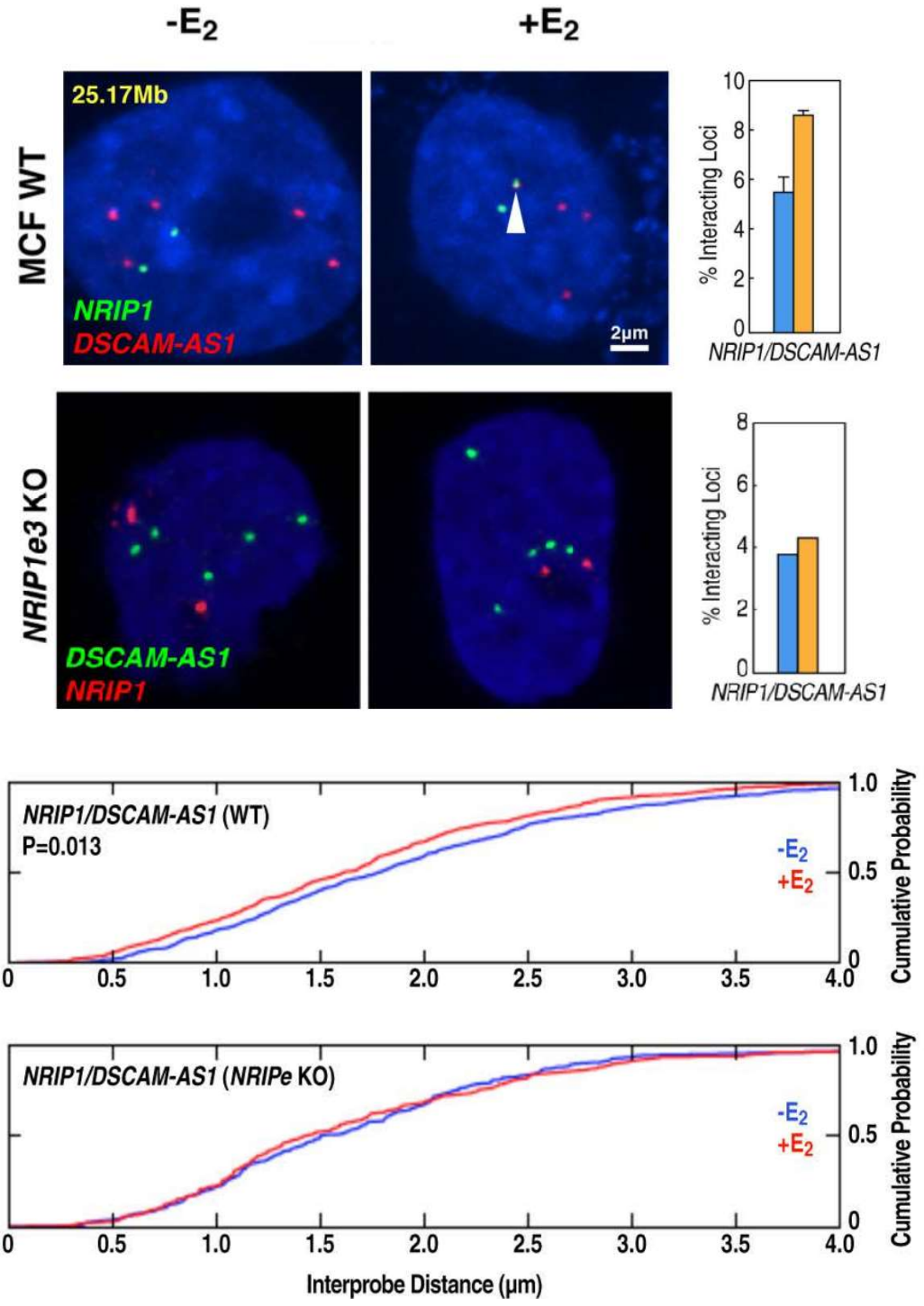


Figure 34: Effects of deletions on E_2 induced NRIP1 and DSCAM proximity

Top panel, deletion of the NRIP1e3 locus negates the E_2 mediated increases in interactions between NRIP1 and DSCAM-AS1 enhancers. Bottom panel, the induced proximity of NRIP1 and DSCAM-AS1 is diminished by deletion of NRIP1e3.

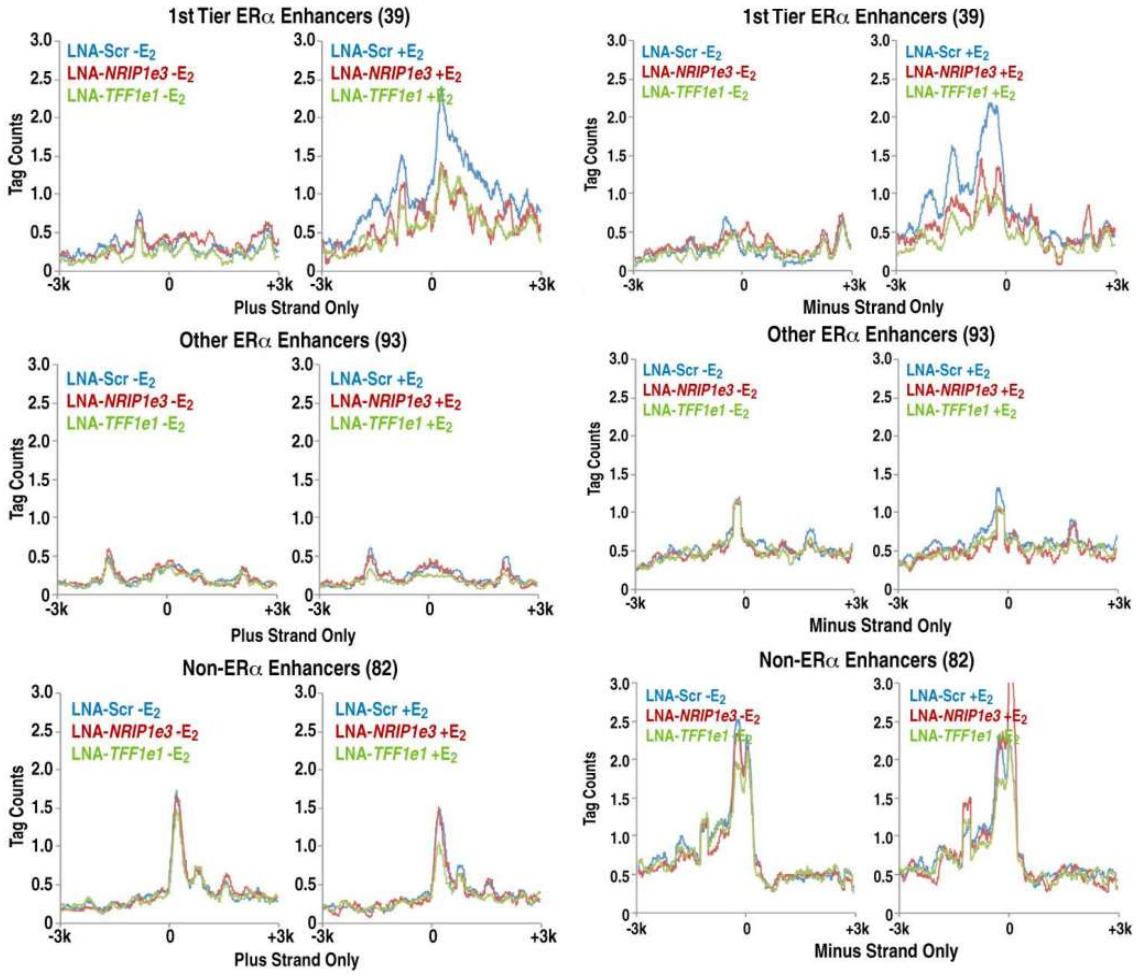


Figure 35: Meta-analysis of GRO-seq in LNA treated cells

Meta-analysis of GRO-seq experiment in which LNA-Scramble, LNA-TFF1e1 and LNA-NRIP1e3 were transfected into MCF7 cells 24 hours before adding E_2 ligand for 60 minutes. Transcription of enhancer RNAs at three sets of enhancers were compared: 1st tier ERα enhancers, other ERα enhancers and non-ERα enhancers.

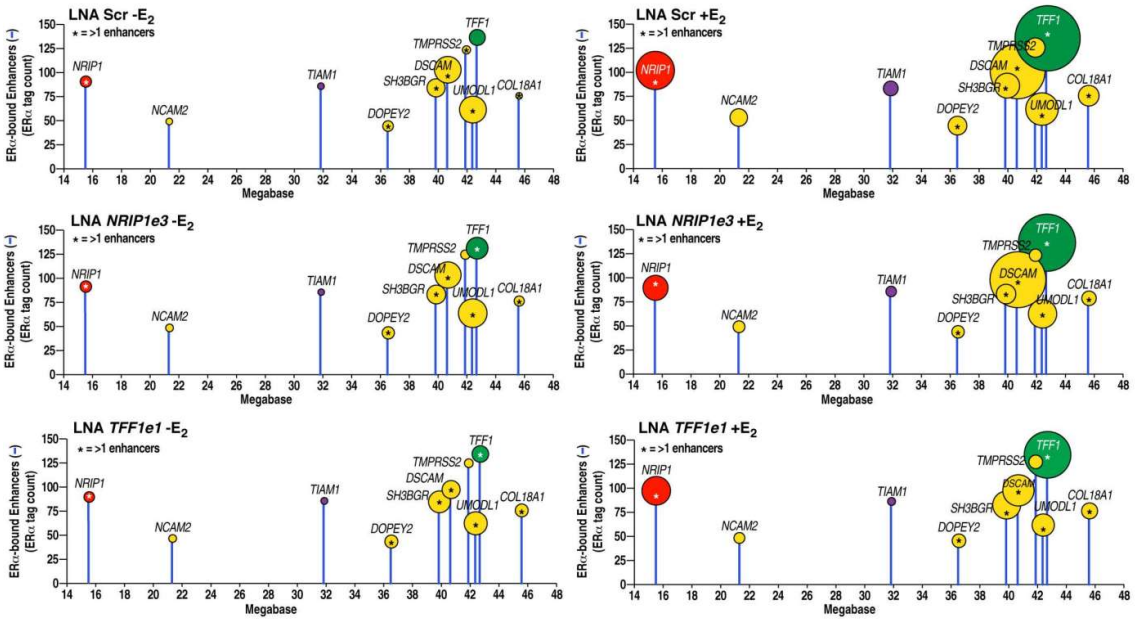


Figure 36: RPKM units from LNA experiments on Chr. 21 enhancers

RPKM units from GRO-seq experiments for LNA scramble, LNA TFF1e1, and LNA NRIP1e3 deletion cell lines represented as the surface area of circles on ten top ERα enhancers.

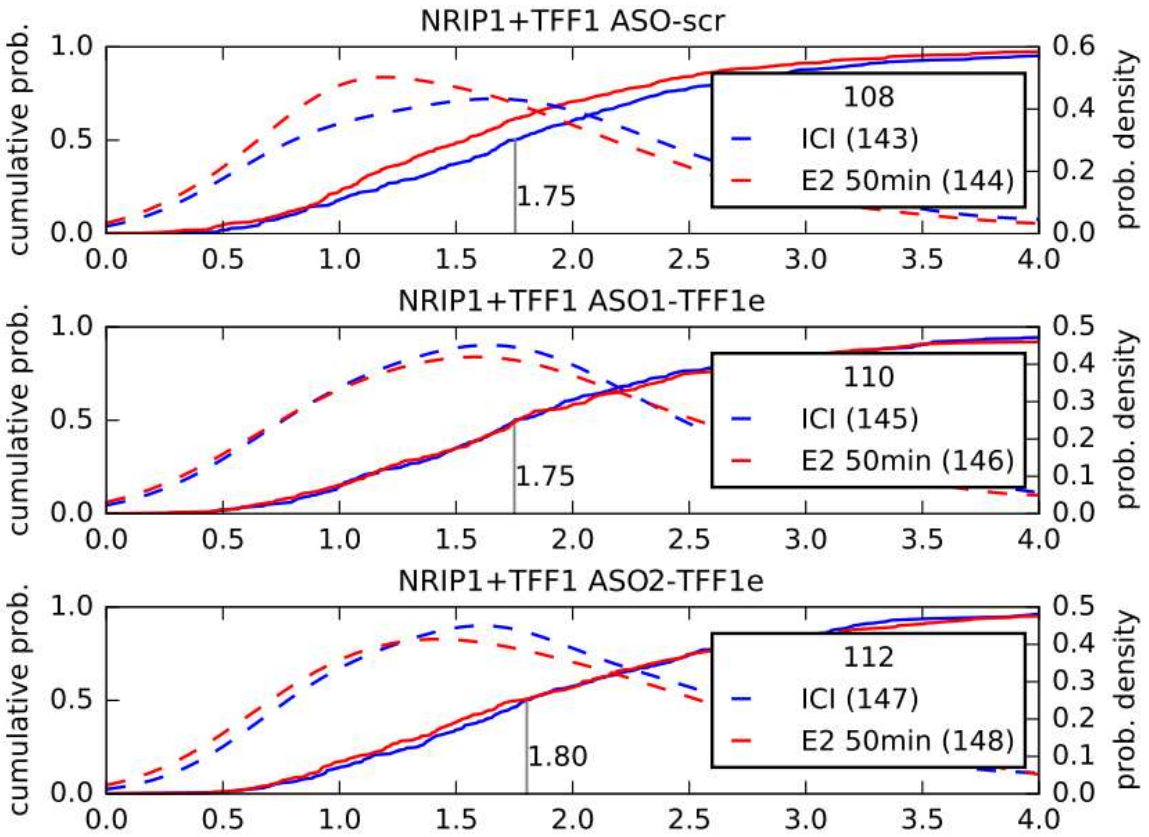


Figure 37: Cumulative distribution of distances in ASO treated cells

Cumulative probability distribution (solid line) and probability distribution (dashed line) of the spatial distances between the NRIP1 and TFF1 loci in MCF7 cells treated with ASO-Scrambled, ASO1-TFF1eRNA, and ASO2-TFF1eRNA.

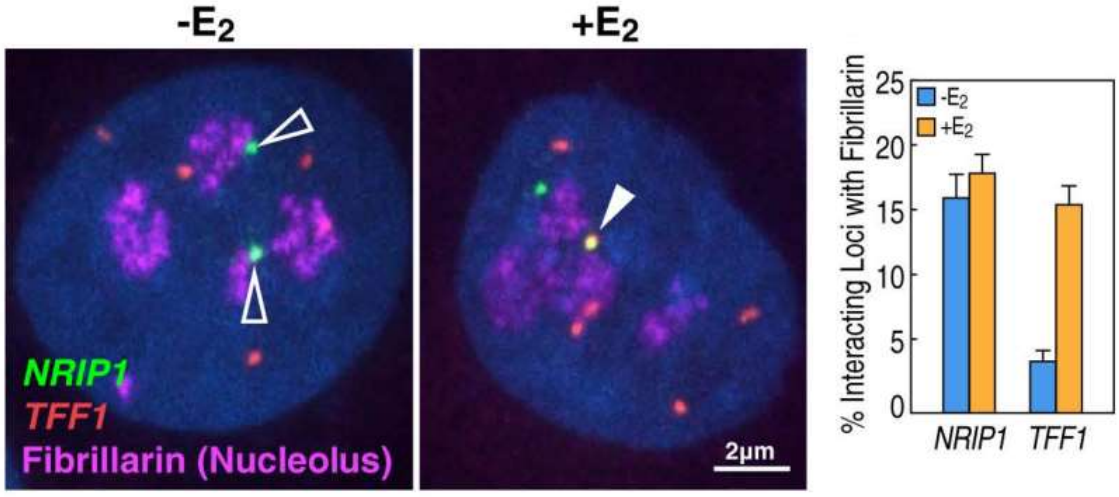


Figure 38: Immuno-FISH data of TFF1 and NRIP1 with fibrillarin

Immuno-FISH data regarding colocalization of the NRIP1 and TFF1 loci with fibrillarin protein, a component of the dense fibrillar component (DFC) of the nucleolus. The interaction of NRIP1 and fibrillarin is E_2 independent, while the interaction of TFF1 with fibrillarin increases more than three-fold after addition of ligand.

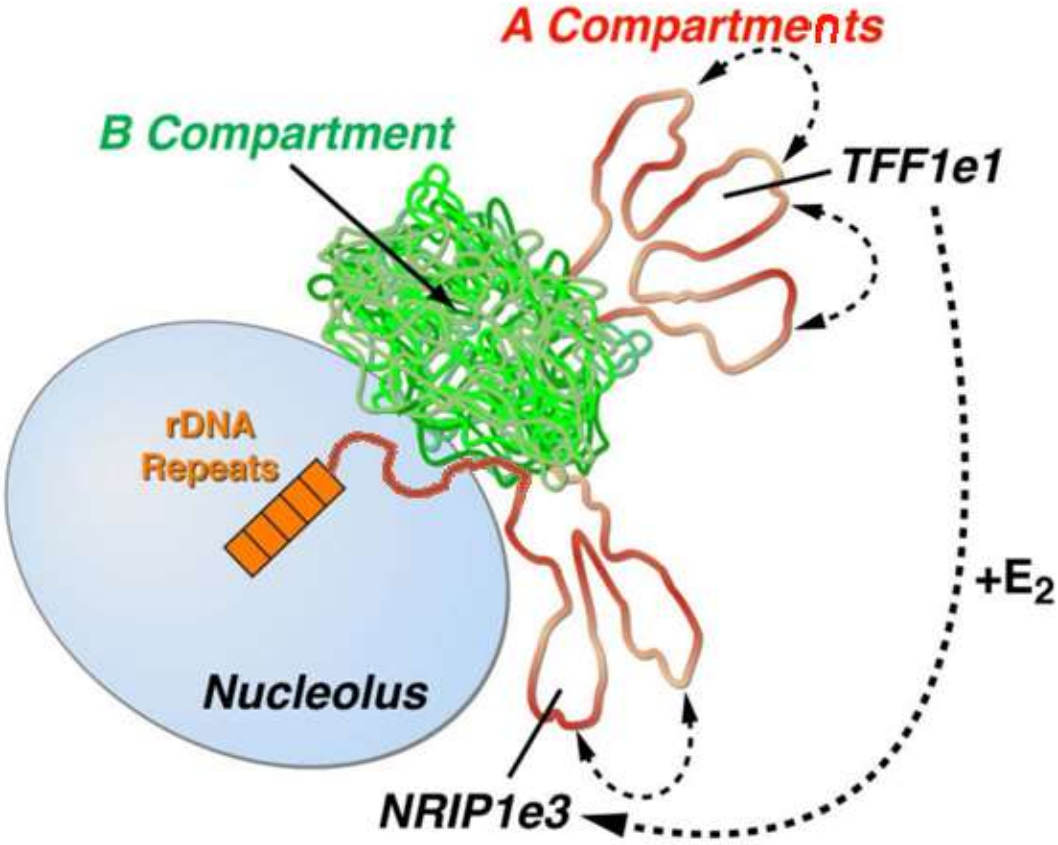


Figure 39: General model of TFF1e1 towards the nucleolus with E_2 ligand

General model of the movement of the TFF1 locus towards the nucleolus after treatment of chromosome 21 with E_2 ligand.

Table 1: ERα tag counts on the 39 strongest ERα enhancers on Chromosome 21

ERα tag counts at the 39 strongest ERα enhancers on Chromosome 21 in MCF7

ER alpha ChIP Intensity +E2	Location on Chromosome 21 (hg18)
113.3	chr21:15,297,346-15,300,000
47.3	chr21:15,467,000-15,468,000
87	chr21:15,492,537-15,506,472
73.5	chr21:15,880,000-15,885,000
32.8	chr21:17,421,622-17,423,896
46.3	chr21:21,290,833-21,293,318
34.2	chr21:29,147,500-29,148,000
33.2	chr21:29,671,888-29,692,013
34.6	chr21:29,802,364-29,805,068
84.7	chr21:31,822,983-31,824,535
106.2	chr21:35,021,838-35,030,350
44	chr21:36,246,238-36,253,287
40.7	chr21:36,481,523-36,489,192
60.4	chr21:36,528,259-36,540,820
35.1	chr21:36,906,262-36,908,602
55.2	chr21:37,561,516-37,563,415
34.6	chr21:37,855,372-37,864,281
77.2	chr21:38,061,873-38,063,441
38.4	chr21:38,853,888-38,856,854
73.5	chr21:39,200,453-39,209,600
80	chr21:39,816,573-39,823,358
94.1	chr21:40,610,583-40,620,226
47.3	chr21:40,631,659-40,635,901
25.3	chr21:40,675,604-40,680,490
123.5	chr21:41,875,116-41,901,438
30	chr21:41,968,479-41,980,178
57.1	chr21:42,352,377-42,354,013
29.5	chr21:42,378,621-42,401,771
132.4	chr21:42,659,014-42,671,342
37.2	chr21:42,683,630-42,684,807
28.1	chr21:43,299,536-43,300,842
54.8	chr21:44,439,209-44,443,076
22.9	chr21:44,484,735-44,485,604
43.5	chr21:45,316,957-45,319,292
73	chr21:45,570,825-45,575,781
60.3	chr21:45,629,805-45,632,829
44.5	chr21:46,112,019-46,113,768
101.1	chr21:46,536,035-46,540,004

Table 2: ChIP-seq tag counts for: AP2γ, FOXA1, GATA3, P300 and MED1

ChIP-seq tag counts for transcription factors: AP2γ, FOXA1, GATA3, P300 and MED1 binding on ten top ERα enhancers on chromosome 21, E2 treatment is 100nM for 60 minutes.

				ChIP Seq Tag Counts									
PeakID	Chr	Start (bp)	End (bp)	**AP2γ ICI**	**AP2γ E2**	FOXA1 ICI	FOXA1 E2	**GATA3 ICI**	**GATA3 E2**	P300 ICI	P300 E2	**MED1 ICI**	**MED1 E2**
NRIP1e1-3	chr21	1.5E+07	1.6E+07	**14.19**	**61.18**	8.66	26.27	**8.69**	**20.26**	14.02	55.88	**8.9**	**23.43**
NCAM2e1	chr21	2.1E+07	2.1E+07	**4.39**	**6.8**	2.89	3.72	**3.72**	**5.67**	7.87	11.67	**7.04**	**11.71**
TIAM1e1	chr21	3.2E+07	3.2E+07	**7.77**	**18.7**	9.3	23.52	**3.72**	**9.32**	4.1	26.52	**3.34**	**10.21**
DOPEY2e1-2	chr21	3.6E+07	3.6E+07	**20.27**	**74.78**	38.16	63.91	**5.38**	**16.2**	34.89	64.72	**24.85**	**32.73**
SH3BGRe1	chr21	4E+07	4E+07	**24.32**	**83.71**	16.35	23.14	**17.79**	**26.33**	41.38	61.89	**25.22**	**36.64**
DSCAM-AS1e1-2	chr21	4.1E+07	4.1E+07	**10.47**	**53.54**	13.46	26.27	**10.76**	**22.68**	22.91	68.26	**30.03**	**46.56**
TMPRSS2e1	chr21	4.2E+07	4.2E+07	**21.96**	**129.17**	31.74	58.81	**7.03**	**4.46**	43.78	141.46	**24.47**	**82.29**
UMODL1e1	chr21	4.2E+07	4.2E+07	**17.9**	**36.54**	30.46	51.16	**12.83**	**27.55**	45.15	34.3	**23.36**	**17.72**
TFF1e1	chr21	4.3E+07	4.3E+07	**35.47**	**107.5**	22.13	73.12	**14.9**	**24.71**	105.34	202.99	**44.12**	**157.98**
COL18A1e1	chr21	4.6E+07	4.6E+07	**7.43**	**28.89**	5.45	8.43	**7.03**	**3.65**	18.81	31.12	**10.01**	**19.82**

Table 3: MSD between pairs of ERα enhancer probes on Chromosome 21

The median spatial distance between pairs of ERα enhancer probes on Chromosome 21.

	Median distance in µm		MSD Percent Change	Volume of a sphere with radius = median distance in μm^3		Volume Percent Change
Enhancer Pair	ICI	E2		ICI	E2	
NRIP1/NCAM2	1.02	0.94	-7.84%	4.45	3.48	-21.73%
NRIP1/TIAM1	1.25	1.25	0.00%	8.18	8.18	0.00%
NRIP1/DOPEY2	1.34	1.44	7.46%	10.08	12.51	24.10%
NRIP1/DSCR3	1.68	1.41	-16.07%	19.86	11.74	-40.88%
NRIP1/DSCAM-AS1	1.81	1.64	-9.39%	24.84	18.48	-25.61%
NRIP1/TFF1	1.86	1.52	-18.28%	26.95	14.71	-45.43%
NRIP1/COL18A1	2	2	0.00%	33.51	33.51	0.00%
DSCR3/TIAM1	1.27	1.12	-11.81%	8.58	5.88	-31.41%
DSCR3/COL18A1	1.41	1.31	-7.09%	11.74	9.42	-19.80%

Table 4: GRO-seq RPKM from ten top enhancers in WT and mutant lines

GRO-seq RPKM values from ten top ERα enhancers on chromosome 21 in wild type, TFF1e1 major deletion, TFF1e1 full deletion, and NRIP1e3 full deletion MCF7 cell lines.

	RPKM Units from GRO-seq							
	WT-ICI	**WT-E2**	TFF1eMD-ICI	TFF1eMD-E2	**TFF1eFD-ICI**	**TFF1eFD-E2**	NRIP1eFD-ICI	NRIP1eFD-E2
NRIP1e1-3	**12.62**	**137.68**	8.11	80.91	**4.32**	**32.75**	5.63	24.18
NCAM2e1	**4.73**	**31.46**	4.29	6.81	**4.39**	**9.73**	5.03	4.90
TIAM1e1	**8.21**	**13.33**	3.51	8.80	**5.58**	**9.69**	5.21	6.52
DOPEY2e1	**7.46**	**35.51**	8.63	21.11	**6.63**	**18.13**	14.93	21.26
SH3BGRe1	**14.98**	**146.06**	18.64	44.82	**14.75**	**57.00**	25.64	79.32
DSCAM-AS1e1-2	**51.15**	**697.00**	56.14	381.74	**23.84**	**258.86**	22.92	161.47
TMPRSS2e1	**29.93**	**96.56**	16.92	28.67	**25.44**	**40.46**	19.87	20.24
UMODL1e1	**22.86**	**78.83**	18.12	24.01	**35.69**	**27.00**	21.71	27.78
TFF1e1	**12.44**	**292.30**	2.45	4.17	**2.57**	**2.34**	10.90	132.48
COL18A1e1	**3.21**	**27.70**	2.81	3.89	**3.14**	**8.18**	2.93	3.75

Table 5: GRO-seq RPKM from ten top enhancers in LNA treated cells

GRO-seq RPKM values from ten top ERα enhancers on chromosome 21 in LNA Scramble, LNA TFF1e1, and LNA NRIP1e3 full deletion MCF7 cell lines.

	RPKM Units from GRO-seq					
	LNA SCR ICI	**LNA SCR E2**	LNA NRIP1e3 ICI	LNA NRIP1e3 E2	**LNA TFF1e1 ICI**	**LNA TFF1e1 E2**
NRIP1e1-3	**6.19**	**56.10**	6.19	23.43	**5.17**	**27.47**
NCAM2e1	**1.69**	**4.41**	4.23	9.25	**1.93**	**4.99**
TIAM1e1	**1.35**	**11.16**	1.58	8.63	**1.78**	**1.37**
DOPEY2e1	**11.08**	**27.67**	14.42	15.03	**14.93**	**17.63**
SH3BGRe1	**9.13**	**26.72**	15.35	23.92	**20.39**	**41.93**
DSCAM-AS1e1-2	**34.06**	**137.19**	37.53	76.64	**20.39**	**53.08**
TMPRSS2e1	**2.68**	**12.38**	4.11	4.41	**3.39**	**7.94**
UMODL1e1	**29.04**	**60.13**	39.40	41.64	**22.78**	**28.05**
TFF1e1	**8.63**	**182.28**	17.75	139.10	**12.30**	**108.38**
COL18A1e1	**1.73**	**21.71**	2.33	7.31	**6.92**	**8.69**