Interrogation of the functionality of ERα binding sites with STARR-seq

by

Elif Yapıcı

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Molecular Biology and Genetics

KOÇ
ÜNİVERSİTESİ

August 2022

Interrogation of the functionality of ERα binding sites with STARR-seq

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Elif Yapıcı

and have found that it is complete and satisfactory in all respects
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Assist. Prof. Dr. Gözde Korkmaz (Advisor)

Assoc. Prof. Dr. Nathan Lack (Co-Advisor)

Prof. Dr. Mesut Muyan

Assist. Prof. Dr. İrem Durmaz Şahin

Prof. Dr. Çiler Çelik Özenci

Date: 10.08.2022

To people who make others' life easier rather than being selfish and negative

ABSTRACT

Interrogation of the functionality of ERα Binding Sites with STARR-seq

ELIF YAPICI

Master of Science in Molecular Biology and Genetics

August 10, 2022

Breast cancer (BCa) and Endometrial cancer (EnCa) are the most common cancer types worldwide in the female population. The common point of these cancer types is Estrogen Receptor α (ERα) activity which is a transcriptional factor controlling a variety of genes important for drug response, cell proliferation and survival. Active ERα translocates from cytoplasm to the nucleus where it can bind thousands of regions, ERα binding sites (ERBSs). These regions are known as regulatory elements controlling the genes related to the initiation and progression of BCa and EnCa. Considering the critical role of ERα in both cancer types, a better understanding of how ERα-bound enhancers drive the transcription of target genes is promising to clarify unique pathways and to develop effective treatment strategies.

To characterize the ERBSs in BCa and EnCa, we aimed to assess the enhancer activity of clinical ERBSs in a quantitative manner with a novel massively parallel reporter assay, STARR-seq. This will allow us to generate the first functional "map" of transcriptional activity of ERα in two systems, from these maps ERα regulated enhancers can be classified as inducible ones, constitutively active ones, and inactive ones. As a first step of this collaborative project, clinically relevant ERBSs were identified by ERα ChIP-seq from ERα positive BCa and EnCa patients to create a custom STARR-Seq library. As a model, MCF7 cells (breast cancer) and Ishikawa cells (endometrial cancer) were used to assess the transcriptional activity of candidate ERBSs to assess pathologically important enhancer regions collected from patients.

Within the scope of this thesis, we optimized the experimental conditions to deliver a custom STARR-Seq library to cell lines models, MCF7 and Ishikawa, and after optimization; we prepared samples required for next-generation sequencing and

bioinformatic analysis. The bioinformatics analysis will be performed as the final step. Overall, the final goal is to identify and categorize the activity of clinically detected ERBSs and generate a functional mapping of ERα enhancer activity.

ÖZETÇE

Östrogen Reseptör α bağlanma bölgelerinin STARR-seq ile sorgulanması

ELIF YAPICI

Moleküler Biyoloji ve Genetik, Yüksek Lisans

10 Ağustos 2022

Meme kanseri ve endometriyal kanser, dünya genelinde kadın popülasyonunda en sık görülen kanser türleridir. Bu kanser türlerinin ortak noktası, ilaca yanıt, hücre çoğalması ve hayatta kalması için önemli olan çeşitli genleri kontrol eden bir transkripsiyonel faktör olan Östrojen Reseptör alfa (ERα) aktivitesidir. Aktif ERα, sitoplazmadan çekirdeğe yer değiştirir ve burada genom üzerinde binlerce bölge ile bağ kurabilir. Bu bölgeler, ERα bağlanma bölgeleridir. Bu bölgeler, meme kanseri ve endometriyal kanseri'in başlaması ve ilerlemesi ile ilgili genleri kontrol eden düzenleyici elemanlar olarak bilinir. ERα'nın her iki kanser türündeki kritik rolü göz önüne alındığında, ERα'ya bağlı güçlendiricilerin hedef genlerin transkripsiyonunu nasıl yönlendirdiğinin daha iyi anlaşılması ve etkili tedavi stratejileri geliştirmek için umut vericidir.

ERBS'leri meme kanserinde ve endometriyal kanserde karakterize etmek için, klinik ERBS'lerin arttırıcı aktivitesini yeni bir toplu paralel raportör tahlili olan STARR-seq ile nicel bir şekilde değerlendirmeyi amaçladık. Bu, iki sistemde ERα'nın transkripsiyonel aktivitesinin ilk işlevsel "haritasını" oluşturmamıza izin verecektir, bu haritalardan ERα tarafından düzenlenen arttırıcılar, indüklenebilir olanlar, yapısal olarak aktif olanlar ve aktif olmayanlar olarak sınıflandırılabilir. Bu ortak projenin ilk adımı olarak, klinik olarak ilgili ERBS'ler, özel bir STARR-Seq kitaplığı oluşturmak için ERα pozitif BCa ve EnCa hastalarından ERα ChIP-seq tarafından tanımlandı. Bir model olarak, hastalardan toplanan patolojik olarak önemli güçlendirici bölgeleri değerlendirmek için aday ERBS'lerin transkripsiyonel aktivitesini değerlendirmek için MCF7 hücreleri (meme kanseri) ve Ishikawa hücreleri (endometriyal kanser) kullanıldı.

Bu tez kapsamında, STARR-Seq kütüphanesini MCF7 ve Ishikawa hücrelerine iletmek için deneysel koşulları optimize ettik ve optimizasyondan sonra; yeni nesil dizileme ve biyoinformatik analiz için gerekli örnekleri hazırladık. Biyoinformatik analiz son adım

olarak gerçekleştirilecektir. Genel olarak, nihai hedef, klinik olarak tespit edilen ERBS'lerin aktivitesini tanımlamak ve kategorize etmek ve ERα arttırıcı aktivitesinin fonksiyonel bir haritalamasını oluşturmaktır.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Assist. Prof. Dr. Gözde Korkmaz and Assoc. Prof. Dr. Nathan Lack. At the beginning of this journey, I met with Nathan Lack. I remember the first time we talked about the project. This talk was really exciting for me to learn more about gene regulation mechanisms. Thanks to him I had a chance to observe two different research groups and was involved in scientific discussions with people who are being enthusiastic about doing research. I would like to thank Gözde hoca. She helped me through this journey and answered lots of my questions with all patience. I am grateful to meet with her, observe her work, and discuss any topics when needed. She was always there for me with all calmness and positivity. I started to work with them as a fresh graduate, and at the moment I am moving further thanks to their enthusiasm for doing science so thanks to both of them.

In addition, I would like to thank my thesis defense jury members, Assist. Prof. Dr. İrem Durmaz Şahin, Prof. Dr. Çiler Çelik Özenci and Prof. Dr. Mesut Muyan. Each of their study and journey in the research gives me inspiration separately and being role model in my journey. Also, I would like to special thank to my first supervisor Prof.Dr.Mesut Muyan from METU. His talks to us during my undergraduate studies drive me to imagine more about the story of the cells and continue in this path.

I would like to thank Korkmaz lab members, Hande Ozkan, Elif Kansız, Gamze Turan, Merve Kayıs, Elif Güzar, Ferzan Betül Berber and Dr. Deniz Gülfem Öztürk for their endless support, discussions, friendship and lots of warm and laughful memories.

I would like to thank Lack lab members, Umut Berkay Altıntaş, Betül Ersoy, Tahsin Emirhan Tekoğlu, Doğancan Özturan, Ceren Seref, Tunc Morova, Bengül Gökbayrak, Flora Huang, Ivan Yu, Meriç Dikbas and Kerim Yavuz for their enthusiastic discussions and helpful answers in any topics. Special thanks to Berkay, Emir and Betül for their friendship and warm and enjoyable memories in Istanbul.

My warm thanks for my colleagues at KUTTAM, Merve Kortel, Barış Sergi, Naz Uzunalioğlu, Ece Özmen, Deniz Özen, Mehmet Yunus Çomar, Beyza Dedeoğlu, Canan Bayraktar, Arda Odabaş and Uğur Akcan for their friendship and discussions on any topics when needed. Also, I sincerely thank Merve Kortel to strengthen one of our life mottos “Akışında”.

I would like to thank my flatmates İdil and Buse for their friendship during this journey. Having both of you was like being at home with my family.

Last but not the least, I would like to thank my family, my mother, father, brother and Berna abla for their unconditional love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ERα	Estrogen Receptor alpha
ERBSs	Estrogen Receptor alpha Binding Sites
AR	Androgen Receptor
GR	Glucocorticoid Receptor
PR	Progesterone Receptor
TFs	Transcription Factors
DNA	Deoxyribonucleic acid
STARR-seq	Self-Transcribing Active Regulatory Region Sequencing
ChIP	Chromosome Immunoprecipitation
PCR	Polymerase Chain Reaction
qPCR	Quantitative PCR
CCND1	Cyclin D1
TFF1	Trefoil Factor 1
PEI	Polyethylenimine

Chapter 1:
INTRODUCTION

1.1. DNA Regulatory Elements

DNA is a hereditary molecule of living organisms which is encoded by four different nucleotides Adenine(A), Guanine(G), Thymine(T), and Cytosine(C). Arrangements of these nucleotides with different combinations enable the generation of different functional elements (or units) throughout the genome which are elucidated by different next-generation sequencing techniques and functional investigations [1].

Functional elements meet at distinct compartments in the nucleus to operate and regulate one of the main processes of the cell, transcription. These compartments can be thought of as transcription factories [2]. In these factories information residing in DNA sequences is transferred to RNAs. In other words, the expression of genes is achieved. Transcription factories are composed of different DNA regulatory elements, Transcription Factors (TFs), RNA Polymerase and transcribed RNA strand [3]. DNA regulatory elements possess a significant function in the regulation of gene expression. According to their function, DNA regulatory elements can be listed as promoters, enhancers, repressors/silencers, and insulators[4] . While promoters and enhancers are prominent DNA regulatory elements to drive transcription events positively [5], repressors/silencers and insulators are known as negative regulatory elements [5].

Genome-wide analysis revealed that sequences having enhancer properties are found more than protein-coding genes in a mammalian genome [6,7] and the impact of this regulatory element on diseases and cell identity cannot be underestimated. Abnormality in their functions underlies the causes of different cancers [8] and other disorders such as neurological disorders, diabetes etc. [9,10]. Therefore, a better understanding of enhancers would help us to understand gene regulation mechanisms.

1.1.1.Enhancers

40 years ago, the SV40 DNA sequence, viral DNA, was placed into the recombinant plasmid containing the beta-globin gene and this boosted the gene expression significantly. This was the first observation that the specific DNA sequence able to upregulate the transcription of a gene [11].

Enhancers are distal elements of the transcription start site. The gap between enhancers and promoters can range from 1 kb to 1 Mb and the location might be either downstream or upstream of the target gene [12]. Besides downstream and upstream enhancers, enhancers within the gene body, specifically in introns, are also detected [13]. This DNA regulatory element is mostly known as cis-regulatory, indicating both enhancers and promoters are found in close proximity. However, the enhancer and its target promoter might be present in different chromosomes, too [14–16].

TFs and RNA polymerase are clustered both at the enhancers and promoters which come together to regulate the gene expression of the target gene in a distance-independent manner. The question of how enhancers that are found in long distances meet with their target gene is explained by loop formation mechanisms [17]. With the help of proteins such as Cohesin complex, CCCTC binding factor (CTCF), and YY1[18], DNA strand bends and this bending makes promoter and enhancer contact possible [19]. Enhancer promoter interactions change dynamically during differentiation, within the lifespan, or even during a day. For example, promoter enhancer interaction that is observed during the night could disappear in the afternoon [20–24].

Understanding enhancer-promoter communication is important to elucidate the function of DNA regulatory elements. To understand the interaction between promoter and enhancer, several methods such as Chromosome conformation capture(3C) [25] and single-cell imaging including dCas9-based live-cell imaging [26] have been developed [12]. FISH technique, also, brought important evidence for loop formations[27].

Compared to other DNA regulatory elements, identification of enhancers is difficult because there is no single sign to identify active enhancers [28–31]. For example, epigenetic markers such as histone modifications such as histone H3 acetylated at lysine 27 (H3K27ac) and H3 monomethylated at K4 (H3K4me1) are considered the indicators

of enhancer identity. However, these modifications might be replaced with other histone modifications [32]. Besides epigenetic markers, binding sites of several TFs including nuclear receptors, and co-activators reside in enhancers [33,34]. Regarding that, another sign is the presence of different protein complexes such as p300, CREB-binding protein, and the Mediator complex [35–37]. Another clue that makes sequence to be an enhancer is the accessibility of these sites [38–42]. Another indicator for the identification of enhancers is enhancer RNAs. Like promoters, enhancers also possess motifs to provide a platform for RNA polymerase II [43]. In 2010, bidirectional transcription of enhancer RNAs was detected from enhancer regions[44]. The functionality of enhancer sequence indicates enhancer identity as well [45–47]. As a result, the identification of enhancers is carried out by multiple signs from epigenetic markers to enhancer function of DNA sequence.

To identify these signatures, various methods are available. Firstly, Chromatin immunoprecipitation sequencing (ChIP-seq) is a technique to determine the interaction of histone markers or other proteins with DNA [48–50]. Secondly, the assay for transposase-accessible chromatin using sequencing (ATAC-seq) and DNase-seq techniques are used to detect accessible regions which are nucleosome-depleted in the genome [38–42]. Thirdly, to detect the enhancer RNAs, GRO-seq is used [51]. Lastly, in order to detect enhancer function, a variety of assays both in vivo and in vitro are developed. Integrated reporter assay, massively parallel reporter assays such as STARR-seq [47], Enhancer trap [52], CRISPR based methods [53,54] are the most used ones [12].

In 2013, improvements in sequencing techniques brought a novel notion suggesting super-enhancers and stretch enhancers [55,56]. The main idea of these classes of enhancers indicates that both are clusters of enhancers being a prominent player to build up cell-type-specific genomic networks. For instance, a study in mouse embryonic stem cells showed that a small group of housekeeping genes are regulated by super-enhancers whereas pluripotency genes are under the control of super-enhancers dominantly [56]. Besides control of genes associated with cell identity, analysis in different tissue types and cell lines elicit those super-enhancers control disease-related genes, too [57]. Moreover, super-enhancers differ from typical enhancers in terms of length, content, and density of regulatory proteins (coactivators, TFs.etc) and even the effect on gene expression [56]. While typical enhancers take 100-1000 bp long segments of DNA, super-

enhancers span within 12.5 kb on average [56,58]. The effect of super-enhancers on gene regulation is higher than typical enhancers in the regulation of cell identity [56,58].

In humans, hormonal stimulations are important drivers of establishing cell-specific genomic networks via enhancers. Hormones such as estrogens and androgens, for instance, bind their nuclear receptors [59,60]. As a result of this interaction, the nuclear receptor goes and binds its specific binding sites in the genome [61,62]. In this way, cell-specific hormonal stimulations have a profound impact on the enhancer landscape. This impact is also seen in cancer such as breast cancer and prostate cancer [63,64]. In a recent study, enhancer activities depending on androgen are checked and findings led to the investigation that a specific group of enhancers induced by androgens has a role in the establishment of gene expression programs in prostate cancer [65]. Furthermore, in this study, enhancers are characterized as inducible ones, inactive ones, and constitutively active ones by STARR-seq assay.

1.2. Nuclear Receptors

Cells are exposed to numerous signals from the environment (outside) and from the inside of the cell. Some of them are cell surface receptors and as their name implies, their location is the plasma membrane [66]. They recognize extracellular signals that could not pass through the plasma membrane and uptake of signals lead to activation of various downstream signaling pathways, such as the Epidermal Growth Factor Receptor (EGFR) signaling pathway [67]. On the other hand, other classes of receptors are found in the cytoplasm. Furthermore, these classes of receptors have the ability to translocate to the nucleus and act as transcription factors. Because of their function in the nucleus, they are known as nuclear receptors, or ligand inducible transcription factors [68].

In the 1900s, a better understanding of hormones such as estrogen, progesterone, and testosterone and their roles in the development and reproductive system, and even in clinical diseases drove researchers to unravel their mechanism of action [69]. In 1964, for the first time, the hormone-binding protein was identified by tracking radioactively labeled estradiol, an estrogen hormone, in mature rats [70,71]. Further observations based on the specificity of binding proteins in target cells led to calling these protein receptors [70]. In the 1980s other steroid hormone receptors such as androgen, mineralocorticoid

and glucocorticoid were isolated [69]. Besides steroid hormones such as estrogens and testosterone, thyroid hormones (tyrosine-based hormones) were also followed up by the labeled hormones, and similar hormone-binding proteins were observed for this group of hormones as well [72]. In the following years, the identification of nuclear receptor genes in the human genome and their structural similarities led to classifying them according to their ligands as thyroid hormone and retinoic acid receptors, orphan receptors, and steroid hormone receptors [73].

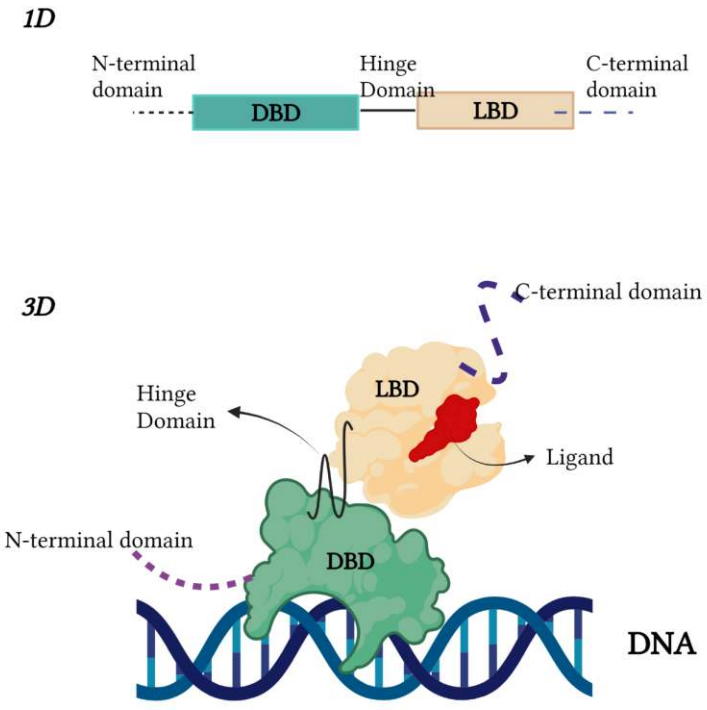


Figure 1 Structural organization of nuclear receptors. 4 domains of the nuclear receptors illustrated. Main domains are DBD (DNA Binding Domain) and LBD (Ligand Binding Domain). Hinge domain is the link between DBD and LBD. N-terminal domain is at the end of DBD whereas C-terminal domain resides in ligand binding domain.

Structural analysis revealed that each nuclear receptor shares structural features [74]. The first universal highly conserved domain of the receptor is DNA binding domain (DBD)[75]. This domain is composed of a zinc finger motif which enables interaction with DNA at the specific *cis*-regulatory element. These specific sites are also called hormone response elements (HRE) [76,77]. The second shared domain is Ligand binding

domain (LBD). This domain recognizes signals such as steroid hormones and this recognition leads to activation of the nuclear receptor and induces conformational change [78–80]. Outside of the DBD and LBD, the N-terminal domain, C-terminal domain, and hinge domain build up the nuclear receptor structure. Hinge domain localized between DBD and LBD. This domain possesses Nuclear Localization Signals (NLSs) that translocate the receptor to the nucleus when it is activated [81,82]. N terminal domain provides a platform to interact with other transcription factors and cofactors that have an impact on gene expression. This domain is, also, named as transactivation domain, or ligand-independent Activation Function 1(AF-1). On the other hand, the C terminal domain resides in the ligand-binding domain and contains a second activation function domain which is named the ligand-dependent Activation Function 2 (AF-2). Owing to cooperation between these two activation domains, the function of nuclear receptors on gene expression is accomplished [83]. Representative structural domains are given in Figure1.

1.3.Estrogen Receptors (ERs)

Estrogen receptors are one of the best studied nuclear receptors in the family of nuclear hormone receptors. Activation of this receptor depends on estrogens which are cholesterol-derived steroid hormones[69]. Having cholesterol-derived steroid structure makes it possible to enter their target cells throughout the cell plasma membrane [69,84]. Naturally produced estrogens in the human body are Estrone(E1), 17 beta-estradiol(E2), and estriol(E3)[85]. Compared to others, E2 is found in circulation mostly [85,86]. Estrogens are known for their role in the development and regulation of the male and female reproductive systems [87–89]. Besides that, they have a profound impact on the regulation of energy metabolism, cardiovascular system, mental health, and even immunity [90–92].

The effect of estrogens is achieved via target receptors which are estrogen receptors [71]. The endoplasmic reticulum membrane, mitochondria, and cytoplasm are locations for different ERs [93]. The ones found in cytoplasm possess the ability to translocate to the nucleus in case of activation. Rather than a single type of estrogen receptor (ER) distinct ERs exists and are known as ERα and ERβ. Like other steroid hormones, both ERα and ERβ are similar in a structure containing 6 different domains; A/B domain(N-terminal

domain), C domain(DNA binding domain), D domain(Hinge region), E domain(Ligand-binding domain), and F domain(C-terminal domain) [94]. Even though the function of shared domains has the same functions in two distinct isoforms, homology between domains shows differences [82,95].

In the human genome, genes encoding these receptors are located in different locations. The ERα gene, known as *ESR1*, is found on chromosome 6, whereas *ESR2* gene expressing ERβ resides on chromosome 14. Moreover, the distribution of ERα and ERβ depends on the cell type. In other words, their expression is tissue-specific [96]. On the other hand, ERβ takes a crucial role in the ovary, prostate, lung, cardiovascular, and central nervous systems [97]. The action of ER depends on interaction with its ligands such as estradiol and this interaction leads to conformational changes which allow dimerization of estrogen receptors. Dimerized ERs translocate to the nucleus where they can regulate target genes. There are various pathways for ERs to regulate genes. The first one is via estrogen response elements (EREs) that reside throughout the genome. Activated ERs directly interact with DNA at these cis-regulatory elements. Further studies elucidated that ERs have the ability to regulate gene expression indirectly, also, via interacting with other DNA-bound transcription factors such as FOXA1 and SRC-1 [98]. Another signaling mechanism for ER is achieved through the G protein-coupled receptors. After post-translational modifications, ER can anchor to the endoplasmic reticulum membrane. At that location ER and G protein-coupled receptors are found in close contact. Thus, activation of ERα led to activation of kinases that would affect several pathways such as ERK/MAPK and PI3K/AKT [82,99,100]. Also, activation of EGFR/IGFR, NFKB, and IL-6/JAK/STAT3 pathways induce activation of ERα by phosphorylation of ERα at a specific residue [101]. In Figure2, signaling pathways for ERs are given.

1.4.Estrogen receptor alpha (ERα) signaling in breast and endometrial cancer

The function of ERα possesses a profound impact on cancer progression. Firstly, ERα has a mitogenic activity that affects the proliferation of a cell. As evidence of this ***ERα regulates genes that have a role in the cell cycle and growth*** [54,102]. Furthermore, common mutations in the *ESR1* gene, also, affect the functions of ERα and reinforce the progression of cell growth. Y537S and D538G are the most well-identified mutations,

and these mutations enhance drug resistance [103,104]. Secondly, ***ERα action led to a change in the Extracellular Matrix*** (ECM). Changes in ECM reinforce the invasiveness of tumors[105]. Altogether, the role of ERα in cancer progression led to a focus on ERα as a therapeutic target to treat ER-related cancers such as breast and endometrial cancer. For example, almost 75% of breast cancers (BCas) are known as ERα positive, and hormonal therapies including tamoxifen and fulvestrant specifically target ERα to interfere ERα action pathway in those cancers [106–108]. Also, approximately 90% of type I and 70% of type II endometrial cancer are positive for ERα [109] and the same treatment approach targeting ERα is applied to endometrial cancer [110]. However, even hormonal therapies result in a good prognosis for a certain group of patients, novel treatment options are still needed to provide effective treatment strategies for ERα positive cancer patients. Despite the similar hormonal therapies applied in breast and endometrial cancer, treatment strategies such as tamoxifen in breast cancer caused a 2 to 6-fold increase in the risk of endometrial cancer [111]. Therefore, a better understanding of ERα action is prominent to develop effective treatment strategies for these cancer types.

The first step for the ERα signaling is the activation of ERα which is achieved by the input signal from outside of the cell. Canonical activation of ERα is mediated by its ligand, Estradiol(E2). On the other hand, rather than its own ligand, growth factors (EGF and IGF) or cytokines (TNGa and IL-1B) activate their respective pathways which lead to the phosphorylation of ERα [112]. Phosphorylation on specific residues of ERα result in activation of ERα followed by an increase in gene activation[101]. Since ERα receives the signal from its ligand directly or from other pathways indirectly, activated ERα transfers the signal to other parts of the cell. It can transfer to other signaling pathways such as ERK/MAPK and PI3K/AKT signaling cascades having a role in proliferation and migration[113,114]. In addition to that, signals can be transferred to the nucleus which regulates the expression of ERα responsive genes.

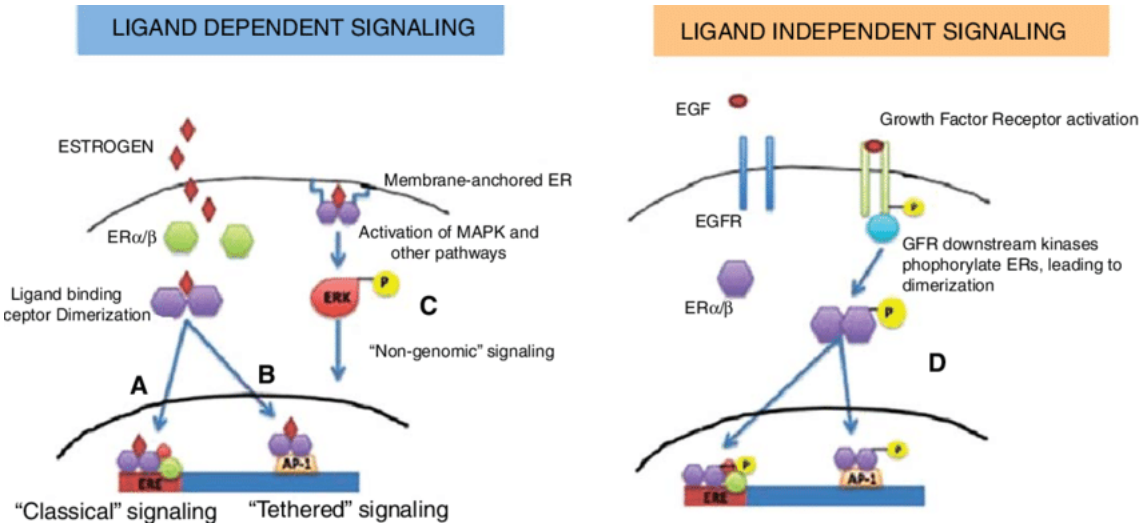


Figure2 Estrogen Receptor Signaling Pathways. Activation of ERα/β can be ligand-dependent and ligand-independent. As a result of ligand-dependence, the genomic signaling pathway gets activated and involved in the regulation of transcription of many genes(A-B-C). Ligand independent signaling pathway gets activated via activation of other cell surface receptors(D)[98].

ERα interacts with *cis*-regulatory elements throughout the genome as a result of estrogen stimulation. A complete set of the genome wide ERα binding sites is known as ERα cistrome [101,115]. ERα cistrome differs between normal mammary epithelial cells and ER+ tumor samples [116]. Also, breast cancer cells gaining endocrine-resistant show changes in ERα cistrome before and after treatment[116]. Thus, a better understanding of ERα-mediated signaling would bring a huge impact to unravel the treatment problems in ERα-related cancers, mainly breast and endometrial cancer. **In this thesis, we will focus on the genomic signaling of ERα.**

1.5.ERα binding sites, ERα mediated gene regulation, ERE elements, mutations, and their effects

E2 stimulation leads to ERα binding throughout the genome. Further identification of the ERα binding sites (ERBSs) revealed that these sites possess a specific motif which is termed estrogen response element (EREs)[117]. 13-19 bp canonical and palindromic sequence which is defined as 5′-GGTCANNNTGACC-3′ provides a platform for ERα dimer. Each monomer interacts with the half-site of the sequence [118,119]. Genome-wide analysis of ERBSs indicates that each ERBS does not contain a full ERE motif

[120]. For instance, whole-genome mapping of the ERα binding site in MCF7, a breast cancer cell line, grouped ERBSs as harboring putative full EREs, half EREs and no ERE ERBSs [120]. In another study, identification of ERBSs in breast cancer cell line (T47D) and endometrial cancer cell line (ECC-1) is carried out and ERBSs are defined as either weak EREs having half ERE motif or strong EREs having full matches with putative ERE motif. In addition to ERBSs with no ERE motif are observed as well [121]. The explanation of ERBSs having half ERE or no ERE is the tethering mechanism. The presence of multiple TF motifs such as FOXA1, GATA3, and ETV4 at ERBSs with a lack of EREs proposes the co-localization of ERα by interacting with other TFs in T47D and ECC-1 cells[121]. Also, further analysis of TFs motifs at ERBSs showed the list of TFs such as FOXA1, GATA4, STAT4, etc. in MCF7 cells[120]. In addition to them, the Sp1 and AP-1, transcription factors, are determined to facilitate indirect ERα binding at specific sites[120,122,123]. Overall, there are two suggested mechanisms for ERα binding selection. One is via direct binding to ERE elements, whereas another mechanism suggests the indirect binding via interacting with other TFs.

ERα cistrome is highly cell type specific. Even though there is about 26,000 palindromic strong ERE across the human genome, only about 9% of them are bound by ERα in differentiated cell[121,124]. Subsets of ERBSs are characterized in breast and endometrial cancer cell lines. Gertz and his colleagues identified shared ERBSs that are ERα bound in both cell line and cell type-specific ERBSs for breast cancer(T47D) and endometrial cancer (ECC-1) cell lines. Further characterization of these sites defined their distinct properties and stated that shared ERBSs are found in closed chromatin regions with strong ERE, whereas cell-type-specific ERBSs with no ERE are located at open sites[121]. Also, like this study, in the same subset of ERBSs, cell type-specific, and overlapping sites are observed for different breast and endometrial cancer cell lines (MCF7 & Ishikawa) [125]. Furthermore, the differences in ERBSs in breast and endometrial cancer were also confirmed in patient tumors. As a result of this study, 15-30% of ERBSs are recorded as shared between breast and endometrial cancer cells[126]. On the other hand, the majority of ERBSs in endometrial and breast cancer cell lines, about 65% of them, are cell type-specific and they are located at accessible sites before E2 administration.[121,126].

Distinct ERBSs profile for breast and endometrial cancer cells is the result of chromatin landscape. One of the factors playing a crucial role to establish a cell-type-specific chromatin landscape is cell-type-specific TFs. A subgroup of TFs, termed Pioneer factors, have the ability to keep chromatin in an open state [127]. In other words, pioneer factors interact with their respective motifs and keep loci accessible for other TFs and co-regulators. Well-known pioneer factors for breast cancer cells are FOXA1[128,129] and GATA3. The motif of these two pioneer factors is found in more than half of ERBSs in breast cancer cells[128,130] . The role of FOXA1 and GATA3 in endometrial cancer cells is sought by many researchers. However, the *de novo* motif analysis on ERBSs showed that endometrial cancer-specific ERBSs do not contain any FOXA1 and GATA3 motifs, whereas over 40% of these sites contain the ETV4 motif. Further analysis showed the differential expression of ETV4, GATA3, and FOXA1 between breast and endometrial cancer cells[121]. In addition to differential ERBSs profiles between cancer types, ERBSs show differences in breast cancer with different clinical outcomes (good and poor) and distant metastases[131].

Besides pioneer factors, the contribution of various co-activators and co-repressors to ER-mediated gene regulation could not be underestimated. The involvement of co-factors depends on the presence of ligand-bound ERα[132]. Ligand binding to ERα makes co-factor binding sites accessible[133]. Co-activators for ERα signaling can be divided into three groups; p160 family members (SRCs or NCOAs)[134], histone acetylases (CBP,p300)[135,136]and p300/CBP associated factor (pCAF)[137]. Recruitment of a number of co-activators at ERα responsive promoters contributes to ERα mediated gene expression significantly[132]. On the other hand, there are co-repressors as another group of co-factors. N-CoR and SMRT are the most known co-repressors in ERα signaling[138]. Co-factor dynamics play a crucial role in ER-mediated gene expression. Tamoxifen, a well-known drug in breast cancer treatment, changes co-factor dynamics to modulate ERα signaling [132]. Treatment of tamoxifen blocks the recruitment of co-activators while inducing the recruitment of co-repressors in breast cancer cells[132]. However, it was known that tamoxifen raises the endometrial cancer risk almost 7-fold, whereas decreases breast cancer recurrence [139]. The opposite action of tamoxifen in breast and endometrial tumors was observed in 1988 as well [140]. Further explanation for the opposite action was done in another study. According to that study, the content of

co-factors in each cell determines the tissue-specific response to tamoxifen-bound ERα[141].

Crosstalk between ERα and other SHRs (Steroid Hormone Receptors) such as PR (Progesterone Receptor), GR (Glucocorticoid Receptor), and AR (Androgen Receptor) have a profound impact on ERα-mediated gene regulation. Firstly, the action of ERa and PR regulating the expression of tumor suppressor genes is detected via separate enhancers[142]. However, the physical interaction of these two SHR is observed by both immunoprecipitation and Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME) experiments as well[142]. This is the indicator for their co-recruitment. Further studies elucidated that ERα and PR were recruited together at the closed chromatin regions. Binding to closed chromatin regions became possible by the chromatin remodeling action of PR. In other words, a subgroup of ERBSs is established by the co-activation of ERα and PR where FOXA1 contributed partially[142]. When co-activation of ERα and PR takes place at specific loci in breast cancer cell lines and in primary tissue, a number of genes associated with estrogen-driven cell proliferation and migration are down-regulated[143]. In line with this finding, co-administration of progesterone and tamoxifen in vivo and ex vivo (primary tissue) decreases tumor growth[142]. Secondly, trans interplay between GR and ERα down-regulates the ERα responsive genes associated with cell proliferation [144]. Interaction of GR at ERα-regulated enhancers having FOXA1 motif recruits the repressor complexes such as N-CoR/SMRT-HDA3 [144]. The presence of GR-induced co-repressor complex removes the ERα-associated Mega Trans complex from ERα-driven enhancers[144]. Besides the antagonist action of GR, the agonist action of GR is also observed. Remodeling activity of GR, like PR, enables to access closed regions in which ERα can interact with other factors such as AP-1[145]. Thirdly, according to many studies both antagonist and agonist action of AR modulators resulted in a reduction of ERα-driven tumor cell proliferation[146–152]. These analyses, also, elucidated the impact of AR on ERα-mediated gene expression in detail. For example, in addition to DHT, a specific ligand of AR, E2 administration induces AR to bind DNA at a specific locus which is different from DHT stimulated binding sites. Furthermore, E2-stimulated AR binding sites contain ERE motifs[147,150]. Additionally, in a recent study, the presence of an AR agonist leads to the replacement of ERα with AR at ERBSs harboring the ERE motif. The binding of

AR at ERBSs blocks the expression of ERα-regulated genes associated with the cell cycle[148]. By the action of AR, ERα cistrome obtained new networks in which ERα co-localized with AR and FOXA1[147]. Overall, the interplay between ERα and other SHRs leads to the reprogramming of ERα cistrome.

Large-scale analysis in multiple cancer types showed that most of the somatic mutations having a role in cancer progression are mapped more in non-coding genomic regions compared to the protein-coding part of the genome[153,154]. A large number of somatic mutations are observed at ERα binding sites. Also, multiple undefined non-coding mutations present in more than one patient at ERBSs are identified[155]. To elucidate the functional significance of mutations at ERBSs, the impact of C to T conversion is checked by the CRISPRi system at a specific locus. As a result, increased expression of multiple distal genes and increased chromatin loop formation, and decreased TF binding are detected[155].

A major number of ERBSs identified in ChIP-PET analysis are located in intragenic regions[120,156]. ERBSs at distal regions are found at least 5 kb apart from 5' or 3' of the closest transcripts[120]. To elucidate whether genes close to ERBSs are ERα responsive or not further analyses are conducted. According to the results, the transcription start site (TSS) of upregulated genes in response to E2 is located within 5 kb apart from ERBS. Furthermore, the list of genes next to the ERBSs contains breast cancer-related ERα target genes[120]. In 2017, more than two ERBSs within 100 kb of TSS was observed for approximately half of E2-induced up-regulated genes in endometrial (Ishikawa) and breast (T47D) cancer cell lines [157] . Next, they focused on 3 different E2-induced genes (*MP17, CISH, FHL2)* having 3 ERBSs within 100 kb of their TSS. The role of each ERBS alone and together with gene regulation is checked by the CRISPRi system. Results highlighted that the collaborative action of multiple ERBSs is essential for E2 responsive gene regulation[157]. In concordance with the collaborative function of multiple ERBSs at a single gene, a number of ERBSs outnumbers the number of E2 responsive genes [121,158]. Findings for collaborative ERBSs are confirmed in a recent study as well[159].

1.6. How to detect and characterize functional ERa binding sites: techniques

Chromatin Immunoprecipitation (ChIP) assay is a technique to investigate protein-DNA interaction[160]. By this method, the target of binding proteins, such as TFs or histones, could be detected. In the first part of this assay, formaldehyde is used to cross-linked DNA-protein interaction and then the whole protein-bound genome is fragmented. Next, using an antibody that is specific to a certain DNA Binding protein (ERα in our case), the DNA-protein complex is precipitated. Later, DNA-protein interaction is reversed, and DNA fragments are purified[160]. The second part of the essay is the analysis of DNA fragments. Methods used in this part bring a different version of ChIP. In the old version of ChIP, a Southern blot was applied to analyze the purified DNA[160]. Improvements in microarray and sequencing technologies lead to expanding the potential of ChIP to analyze the genome-wide binding of a specific protein. Before improvements in sequencing, DNA microarray(chip) technology provides a way to map the binding sites of a specific protein. This version of ChIP is known as ChIP-on-chıp. In this method, purified DNA fragments are hybridized with known DNA fragments that are attached to a solid chip platform. ChIP-on-chıp analysis, firstly, was applied to identify ERBSs on chromosomes 21 and 22 in MCF7. Then a genome-wide analysis of ERα binding sites was conducted again in MCF7. For chromosomes 21 and 22, 57 ERBSs are detected in case of 45 min E2 administration, whereas a whole-genome consists of 3,665 unique ERBSs as a result of performed ChIP-on-chıp experiment upon 45 min E2 administration[161,162]. From previous studies, it is known that 45 min E2 administration results in the maximum ERα binding[132,161]. However, ChIP-on-chip data in this study was reanalyzed by another algorithm and updated to 5782 sites[163]. Another version of ChIP is ChIP paired-end diTags (ChIP-PET). In this method, Tags are ligated to the ends of purified fragments and then by sequencing unknown DNA fragments are identified. By ERα ChIP-PET, 1243 genome-wide ERBSs are discovered. However, 610 of these sites were different from ChIP-on-chıp[120]. In the last version, sequencing technology was combined with ChIP [156]. In this approach, sequenced fragments are aligned to the whole genome according to alignment results, and identification of ERBSs is done. As a result of the application of ERα ChIP-seq, 10205 high-confidence ERBSs are identified[156]. A comparison of ChIP-seq with ChIP-PET and ChIP-on-chip revealed that ChIP-seq encompasses the major amount of identified

ERBSs in the microarray and cloning approach. In addition to shared sites, ChIP-seq provides a larger number of sites including unique ones compared to other techniques[156]. Improvements in sequencing reinforce the usage of ChIP-seq with low cost and high efficiency. Also, compared to other techniques ChIP-seq is conducted in a short period of time compared to others. In further studies, by overcoming technical limitations ChIP-seq is optimized to study ERBSs in clinical samples such as breast cancer, endometrial, and prostate tumors [164].

To further characterize the ERBSs, like all enhancer regions, different features are checked at the chromatin level and sequence level. This feature shows a correlation with ERα binding. At the chromatin level, the presence of RNA polymerase II, and enhancer-associated histone modifications such as H3K27ac are indicators for functional ERBSs [158,159]. The presence of these indicators also can be detected by ChIP-seq. Another indicator for functional ERBSs is chromatin accessibility. To detect chromatin accessibility ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) is conducted. In this method, Tn5 transposase, a protein having the ability to cleave and paste, reaches open sites of the chromatin, and is cut and tagged by a specific adaptor. Next, by sequencing technology accessibility of the regions is detected [165]. For the ERBSs, accessibility is correlated with ERα binding sites [158,159]. Accessible sites are signatures for the possible binding of various transcription factors and co-activators [166]. ERα co-occupy with numerous transcription factors such as pioneers' factors (FOXA1 and GATA3), and co-regulators (CBP/p300 & NCoR) [61]. Therefore, in the case of ERBSs, the presence of these factors would be also an indicator for functional ERBSs. Another indicator of ERBSs functionality is enhancer RNAs. GRO-seq was applied to ERBSs to measure enhancer RNA. Results showed that the presence of enhancer RNAs is highly correlated with the presence of other chromatin signatures such as H3K27ac, H3K4me1, EP300/CREBBP, RNA Pol II, and chromatin accessibility[61].

Even though there is a correlation between enhancer activity and these signs, neither of them gives an idea about enhancer activity nor the strength of the enhancer. Moreover, these signs are not particular to active enhancers, so false positive identifications are likely to be obtained [46,167]. Therefore, in addition to these chromatin features at ERBSs, functional characterization of ERBSs based on their sequence would bring a

unique aspect to identify enhancers. In order to detect enhancer function based on sequence, a variety of assays both in vivo and in vitro are developed. Integrated reporter assay, massively parallel reporter assays such as STARR-seq [47], enhancer trap [52], CRISPR knockout, and CRISPR interference [53,54] are the most used ones [12]. In this project, we will perform STARR-seq to interrogate the functionality of ERBSs so STAR-seq would be explained.

1.6.1. Self-transcribing active regulatory regions sequencing *(STARR-seq)*

STARR-seq (Self-transcribing active regulatory regions sequencing) method was developed as a massively parallel reporter assay (MPRA) to assess enhancer activity of candidate DNA sequences by Stark lab from the Institute of Molecular Pathology at Vienna in 2013 [47]. In the STARR-seq system, they redesigned a pGL3/4 plasmid in which candidate fragments were placed downstream of the core promoter. In this special design, each candidate fragment is transcribed itself depending on its enhancer activity, and enhancer strength is measured via the abundance of the reporter transcript (Figure3). Thus, STARR-seq enables the characterization of enhancers as strong ones and weak ones. Different from available reporter assays such as luciferase, a large number of candidate enhancers can be tested simultaneously since each candidate produced its own barcode [47]. Even though this is a plasmid-based method, tested candidate sequences are correlated with enhancer-associated histone modifications(H3K27ac) and DHS-seq profile in their endogenous location. Through these and further validations, they introduced a high throughput feasible method to elucidate transcriptional regulatory elements according to their strength residing in the sequences [47].

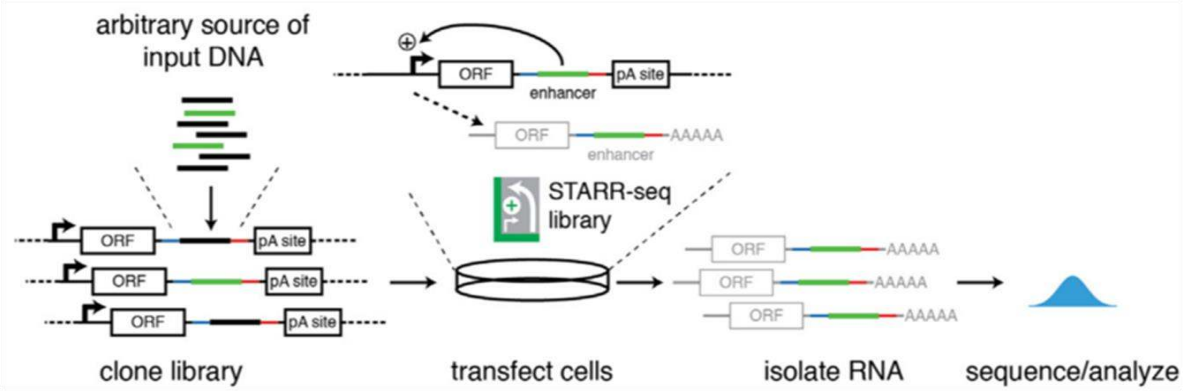


Figure 3.Workflow of STARR-seq. Candidate DNA fragments are cloned into STARR-seq plasmid. Next step is the delivery of the cloned plasmids to cells. Inside the cells each candidate fragments self transcribes itself and produce its own reporter transcripts. Isolation of the reporter transcripts is performed. Last step is sequencing analysis [168].

The first design of STARR-seq plasmid, derived from pGL3/4 backbone like in all MPRAs, contains minimal core promoters in which candidate sequences are placed downstream of it [169]. This was the site where transcription of reporter transcripts would occur. However, when reporter transcripts were mapped to plasmid besides core promoter-produced transcripts, ORI-derived transcripts were produced. ORI is the origin of replication and is the putative element of each plasmid. ORI is composed of core promoter elements such as DRE, TATA box, etc. and thus ORI could still act as an optional core promoter. This was one of the key problems in the STARR-seq assay since the presence of two core promoters leads to conflict between two core promoters. To overcome this problem, ORI is used as a single-core promoter. As shown in the study, background signal increases in the case of two core promoters, whereas a single-core promoter having only ORI improves STARR-seq signal-like in single candidate luciferase assays. In our experimental system, STARR-seq plasmid harbors ORI as a core promoter which is placed upstream of candidate fragments. Another key problem in STARR-seq arises from the entrance of the plasmid into a cell [169]. Cells respond to the entry of external DNA by the innate immune response which is not specific to plasmid type and delivery method. This response leads to type 1 interferon response via *cGAS, STING, TBK1,* and *IRF7*. The induction of IFN-I genes increases the number of interferon response-related TFs such as *IRF, ISG15,* and *IFI27*. The induction of these TFs has the potential to change the enhancer activities of candidate fragments in the library. To solve

this problem, treatment of TBK1/IKK/PKR inhibitors is suggested. When the induction of interferon-related TFs is inhibited by these inhibitors, STARR-seq peaks (false positives) disappear. In addition to false positives, in the presence of interferon inhibitors, the enhancer activity of a group of candidate fragments increased so IFN signaling may block the bona fide enhancers. In other words, the presence of interferon-related TFs can dominate other enhancer activities and alter the enhancer landscape which gives false-positive and false-negative results. Therefore, interferon response resulting from transfection possesses critical consideration for the STARR-seq method in mammalian cells. However, interferon-stimulated gene induction is not observed for all cell lines [169].

There are various applications of this assay. STARR seq can be applied to identify genome-wide cell type-specific enhancers. Arnold and his colleagues delivered a STARR seq library containing randomly fragmented *Drosophila* genome to different cell types and they obtained quantitative enhancer activity maps having shared and cell type-specific enhancers for each cell type. In addition to that, with the combination of DNA accessibility assays (DHS-seq) endogenously silenced enhancers were detected in *Drosophila* cells [47]. In agreement with these results, endogenously closed enhancers are predicted in HeLa cells, a human cell line, by a similar approach [169].

Rather than genome-wide applications, STARR seq allows assessing enhancer activity of selected genomic regions. For example, candidate fragments could be defined by a combination of DNase-I seq and ChIP-seq results like in the study of Vanhille and his colleagues [170]. Enhancer activity of SNP-containing fragments is pooled, and gene regulatory differences were checked by STARR-seq [171]. In another study, a library was established by cancer risk-associated SNPs located in noncoding regions [8,172]. In this study, regulatory activities of these regions depending on allele-specific changes were assessed and the role of two of the risk-associated variants in the transcriptional regulation of cancer genes was confirmed. Besides constructing a library by a pool of specific candidates, regulatory activity of specific selected genomic loci could be assessed as well. In 2015, Rathert and his colleagues checked the enhancer activity profile of the MYC locus (in a specific cell line having H3K27ac at specific locations). They prepared a STARR-seq library containing a 3.1 Mb region at the MYC locus [173].

A subgroup of regulatory elements on the genome is under the control of extracellular signals such as hormones. Applications of STARR-seq on these regulatory elements provide hormone-responsive enhancer activity maps [174]. In 2014, enhancer activity in response to Ecdysone, a steroid hormone in *Drosophila*, was checked in different cell types with the same library and cell type-specific hormone-responsive activity maps were obtained. This approach was also applied to elucidate the action mechanism of hormone receptor binding sites along the human genome. By the application of STARR-seq functional enhancer activity map of Glucocorticoid receptor (GR) [175] and Androgen receptor (AR) [65] binding sites were obtained. From these maps, constitutively active, inactive, inducible enhancers [65] or induced and repressed regulatory elements [176] in a specific cell can be checked.

Compared to other reporter assays, this unique method enables the assessment of enhancer activity in a high throughput manner, whereas traditional reporter assays test enhancer activity one by one [168]. This method is a plasmid-based method so, unlike integrated reporter assays, the position of the candidate fragment would not be problematic [168,177]. Although enhancer activity is checked ectopically, most of the enhancers characterized by this method are found inaccessible parts of the genome and had typical chromatin features of enhancers [47]. That is validation for their function in an endogenous context. Also, the cell-type specificity of the enhancers can be tested accurately. On the other hand, by this method, closed enhancer regions that cannot be predicted by other chromatin features can be checked [47].

1.7. Aim of this thesis

ERα ChIP-seq analysis was performed by Wilbert Zwart's group at the Netherlands Cancer Institute from breast cancer and endometrial cancer patients. In this analysis, ~20,000 ERBSs were identified. ***To characterize these ERBSs,*** we propose to test the enhancer activity of these clinical ERBSs with a novel massively parallel enhancer assay, STARR-seq. This would allow us to obtain a functionality map of clinical ERBSs for both in a quantitative manner. One of the critical steps in STARR-seq is achieving a library transfection in a high-efficiency and cost-effective way. To do that, we aimed ***to choose and optimize the transfection method for ERα positive cell lines*** which are MCF7, breast cancer cell line, and Ishikawa, endometrial cancer cell line.

Chapter 2:

MATERIALS AND METHODS

2.1. Preparation of positive controls

2.1.1 Optimization of PCR for enhancer regions

MACHEREY-NAGEL Genomic DNA isolation kit was used for the isolation of genomic DNA from MCF7 cells. CCND1 enhancer region (known as "enh588", chr11:69,329,989-69,330,492) and TFF1 enhancer region(chr21:43,796,360-43,796,859) which are known ERα regulatory regions [178,179] were amplified with Polymerase Chain Reaction (PCR) using primers in Table1, and Phusion High-Fidelity DNA Polymerase (NEB Cat#M0530S).

In addition to PCR components, dNTP (Invitrogen, 18427013) and GC buffer (NEB, B0519) were used. The two different sizes of these regions were amplified. For the amplification of 1 kb regions, the PCR reaction was optimized in the extension step for 15 sec to avoid unspecific amplification, and for the amplification of 500 bp, 30 sec extension was applied. The temperatures of PCR reaction cycles were as follows; 98°C for 10 seconds, 65°C for 25 seconds, 72°C for 15 seconds as 25 cycles. For the primers in Table1, the annealing temperatures were optimized. In order to avoid random mutations in enhancer regions, the amount of the Phusion was optimized to 1 µl. PCR is carried out in the Veriti^{TM} 96-Well thermal cycler. PCR product was checked in the agarose gel (0.1%, w/v). Then, the PCR product was purified from the gel with a PCR and gel clean-up kit (Macherey Nagel, Cat#740609).

Table 1. Primers to amplify positive controls (Primers to amplify ERα regulated enhancers (TFF1 and CCND1 enhancers) are given. For each enhancer Forward(F) and Reverse(R) primers are designed. Primers are designed to obtain two different sizes of corresponding enhancer. Underline sequence indicates assembly sites for Gibson Assembly cloning)

Enhancer	Type	Sequence (5'-3')	Amplification Size
TFF1	F	ATTGATCTAGAGCATGCACACTGCAGCTTTCGGAGGGC	500 bp
	R	AGCGGCCGGCCGAATTCGCCTTCTCACACACATCCCCTC	
TFF1	F	ATTGATCTAGAGCATGCAGATGTCCCTCGGCCCCTTTG	1 kb
	R	AGCGGCCGGCCGAATTCGAGTTTCCCTCCCATAACCTGG	
CCND1	F	ATTGATCTAGAGCATGCAATCTTGTGAATGGTTTGGCAC	500 bp
	R	AGCGGCCGGCCGAATTCGTCCATCCTGCTACGTTGAGAATTAC	
CCND1	F	ATTGATCTAGAGCATGCAGCCTAGGAACACACAGCC	1 kb
	R	AGCGGCCGGCCGAATTCGTTTCGCAGCCTTAAGAATGTC	

2.1.2. Restriction enzyme digestion of STARR-seq_ORI vector

STARR-seq_ORI vector (Addgene, Plasmid#99296) was digested by AgeI (New England Biolabs, Cat#R0552S) and SalI (New England Biolabs, Cat#R0138L). The map of the plasmid is given in Figure4. Since these are two incompatible enzymes and there is no common buffer in which both AgeI and SalI exhibit more than 50% activity while avoiding star activity, a sequential double digestion reaction was necessary. For this, NEBCloner online protocol (https://nebcloner.neb.com/#!/protocol/re/sequential-heat/SalI,AgeI) was followed. In the first reaction, 1 µg plasmid digestion with 1 µl AgeI in a total reaction volume of 50 µl was carried out with NEBuffer r1.1 at 37°C for 1 hour, following the heat inactivation step at 65°C for 20 minutes. Then, the salt concentration was adjusted to 100 mM by addition of NaCl for adequate SalI activity. 1 µl of SalI was added and incubated at 37°C for 1 hour. After the heat inactivation step at 65°C for 20

minutes, digestion was completed. Digestion products were run on agarose gel (0.1%, w/v) and purified with a PCR and gel clean-up kit (Macherey Nagel, Cat#740609).

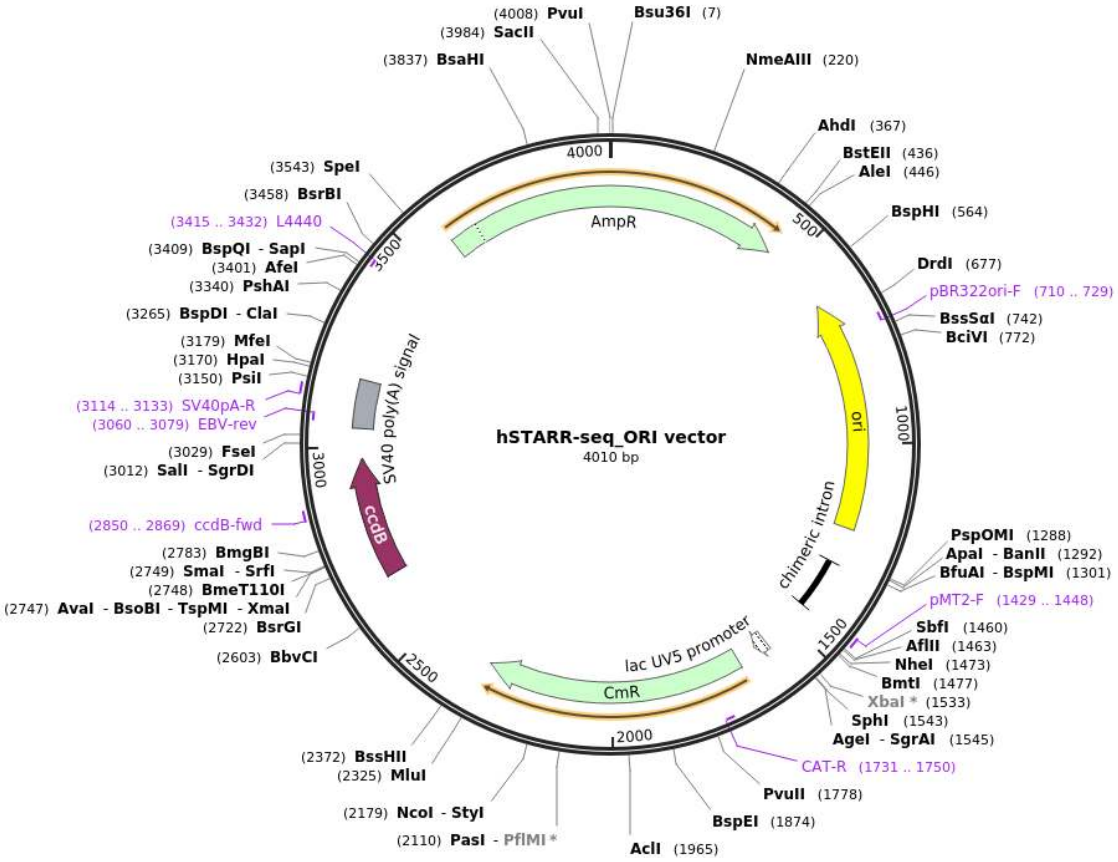


Figure 4 The map of hSTARR-seq plasmid. 4010 bp hSTARR-seq plasmid is used to clone positive controls.

2.1.3. Gibson Assembly Cloning Method

Gibson Assembly Cloning Kit (New Englands BioLabs, Cat#E5510S) was used for cloning enhancer of CCND1 gene and enhancer of TFF1 gene. Gibson Assembly master mix, plasmid backbone, and enhancer region fragments were incubated at 50°C for 1 hour in the thermal cycler. It was also stated that the insert amount should be 2 times more than the vector amount in pmol in the reaction.

Primers in Table1 were designed as burying the overhangs considering the ends of the digested vector in which the PCR product will be cloned. An overview of the Gibson

Assembly cloning method is given in Figure 5. Overhangs at the end of each primer are listed in Table 2. Cut site of AgeI and SalI restriction enzymes took into account while designing overhangs.

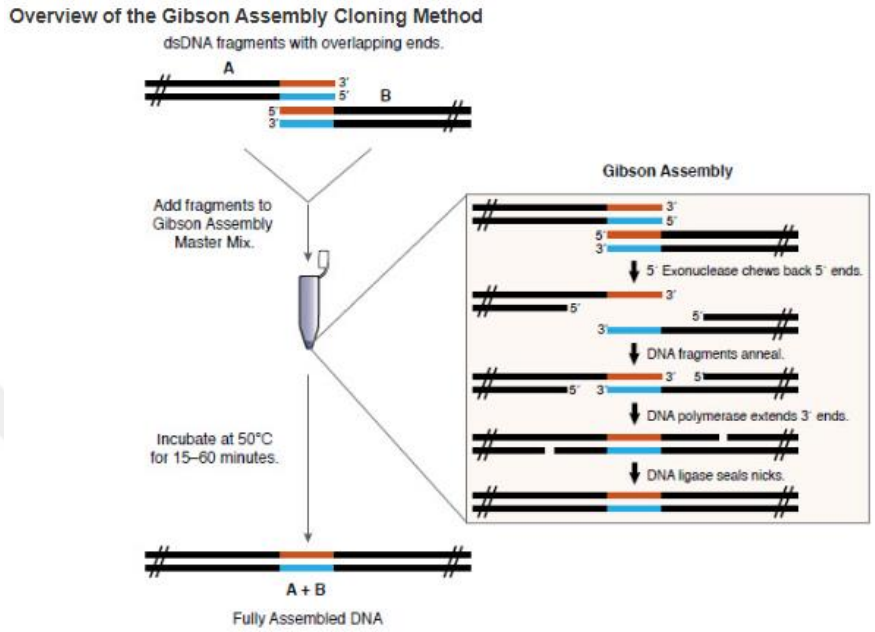


Figure 5 An Overview of the Gibson Assembly Cloning Method. Insert and digested plasmids having overlapping ends at the ends mix with Gibson Assembly Master Mix containing 3 different enzymes (Exonuclease, DNA Polymerase, DNA ligase). Mixture is incubated for 15-60 min at 50°C.

Table 2. **Overhangs end for Gibson Assembly.** These ends designed considering digested plasmid; this sequence is added to PCR primers of each enhancer regions.

Primer	Overhang ends
Forward	ATTGATCTAGAGCATGCA
Reverse	AGCGGCCGGCCGAATTCG

2.1.4. Transformation of vectors and plasmid DNA isolation

Two strains of Escherichia *coli* (E.*coli*) competent cells were used (DB3.1 and Stbl3). After the assembly reaction, the DNA clean-up protocol was followed up. For the 20 µl Gibson reaction, 20 µl isopropanol and 0.4 µl NaCl (5 M) were added. This mixture was

incubated for 15 minutes at room temperature, following the centrifugation at 15000 xg at 4°C for 15 minutes. Pellet was washed twice with ice-cold 1 mL 80% ethanol. After air-drying, the pellet was resuspended in 10 µl ultrapure H_2O.

Heat shock transformation protocol was performed using 50 µl competent cells. For the uncut plasmid DB3.1, the E.*coli* strain was used because of the presence of the ccdB gene in the STARR-seq_ORI plasmid. Gibson product and competent cells were incubated on ice. Then, 45 seconds heat shock at 42°C was applied. Next, incubation on ice for 5 mins was carried out. For the recovery of cells, 950 µl LB broth (Miller) was added and incubated for 1 hour at 37°C before spreading cells onto the LB Agar (Miller) plate including ampicillin (100ug/ml).

After overnight incubation of plates at 37°C, single colonies were picked and added into 5 ml LB Broth medium. Liquid cultures were further incubated overnight at 37°C shaking incubator for further plasmid DNA isolation. Mini- (Macherey-Nagel) and midi-prep (Macherey-Nagel) plasmid DNA isolations were performed according to the manufacturer's instructions.

To confirm colonies 50 ng of cloned plasmids were shipped to Sanger-sequencing. EBV-Reverse Primer was used to sequence plasmids. Sequencing results were checked by alignment tool in SerialCloner.

2.2. Cell Culture

MCF7, a breast cancer cell line, is cultured in Dulbecco's Modified Eagle Medium (DMEM) (Sigma, D6429), supplemented with 10% fetal bovine serum (FBS) (Gibco, Cat#15140148) and 1% penicillin-streptomycin (Gibco, Cat#12140148). Ishikawa, endometrial cancer cell line, is cultured in RPMI-1640 (Gibco, Cat#12633012) supplemented with 10% fetal bovine serum (FBS) (Gibco, Cat#15140148) and 1% penicillin-streptomycin (Gibco, Cat#12140148). During the sub-culturing, 0.05% Trypsin/EDTA (Thermo Fisher, Cat#25300054) was used for detachment of cells. Cells are incubated at 37°C humidified incubator with 5% CO_2 condition. The mixture of 900 µl complete medium (containing 10% FBS) and 100 µl DMSO (10%, v/v) was used to

freeze and store cells at -80°C. Mycoplasma control of the cells was performed each week.

2.3. Optimization of transfection parameters

PEI (Polysciences, Cat#24765) was used as a transfection reagent. The amount and ratio of PEI and plasmid DNA is critical for achieving an efficient transfection. Optimized parameters are given in Table3. DNA and PEI transfection reagents were added into two different tubes, along with the serum-free medium. Then, the PEI-containing solution was transferred to the DNA-containing tube. This mixture was vortexed for 15 seconds and incubated for 15 minutes at room temperature. The medium on top of cells was changed and the transfection mixture was added to the cells drop by drop. As a transfection control, pLenti-CMV-eGFP-Puro plasmid (Addgene, Cat#17448) was transfected. Transfection efficiency was assessed via visualizing fluorescence in the control plate. In Table4, optimization of PEI ratios for two cell lines and time point to detect reporter mRNA expression was given.

Table 3. Optimized parameters for transfection. Small-scale experiments are conducted in 6-well plate, Library transfection is performed in 15 cm plate, transfection parameters DNA amounts and PEI ratios depending on the cell is adjusted as given in the table. While preparing transfection mixture, given amount of serum free medium is used.

Plate	Cell	DNA amount	DNA:PEI ratio - volume of PEI	Serum-free medium
6-well	MCF7	2 µg	1:3 - 6 µl PEI	100 µl
10 cm	MCF7	6 µg	1:3 - 18 µl PEI	200 µl
15 cm	MCF7	10 µg	1:3 - 30 µl PEI	1 mL
6-well	Ishikawa	2 µg	1:2 - 4 µl PEI	100 µl
10 cm	Ishikawa	6 µg	1:2 -12 µl PEI	200 µl
15 cm	Ishikawa	10 µg	1:2 -20 µl PEI	1 mL

Table 4. Parameters to optimize for two cell lines. Different PEI ratios were applied to check transfection efficiency for two MCF7 and Ishikawa cells, Time points are to determine the ideal time to collect reporter transcripts.

Systems to optimize	PEI Ratios	Time Points
MCF7(Breast Cancer Cell Line)	1:3 & 1:6	24 h & 48 h
Ishikawa(Endometrial Cancer Cell Line)	1:2 & 1:3	24 h & 48 h

2.4. RNA Isolation

Trizol Reagent (Invitrogen, Cat#15596018) was used for the RNA isolation step. Recommended protocol in the user guide was followed. RNA concentration was determined using Nanodrop as a unit of ng/µl. Purity was checked by the ratio of Abs_{260}/Abs_{280} and Abs_{260}/Abs_{230}.

2.5. cDNA synthesis

Volume required to have 1000 ng of RNA was calculated for each sample and volume was completed to 9.5 µl. Then, the first 5 µl of dNTP (Invitrogen, Cat#18427013) and 2 µl of random hexamer (Invitrogen, Cat#N8080127) was added and incubated with RNA mixture at 65ºC for 5 minutes. Then, 2 µl of second 5X first strand buffer (Invitrogen, Cat#28025013) and 0.5 µl of Rnasion Ribonuclease Inhibitor (Promega, Cat#N2111) was added. After addition of mastermix to reaction, 1 µl of M-MLV Reverse Transcriptase (Invitrogen, Cat#28025013) was individually added to each tube. The reaction was incubated at a thermal cycler for 1 hour at 37°C at a total volume of 20 µl. Lastly, a heat-inactivation step was carried out at 90°C for 10 minutes. 80 µl nanopure water was added to the reaction mixture for 1:5 dilution and stored at -20°C.

2.6. Real-Time Quantitive Polymerase Chain Reaction

LightCycler® 96 system was used in RT-qPCR experiments. 2X SYBR Green PCR Master Mix (Thermo Fisher, Cat#2309155) and RT-qPCR primers in Table5 were used. To check the expression of reporter transcripts, the forward primer was designed to target the junction site where plasmid backbone and insert join.

Table 5. Primers for qPCR reaction. Primers designed to check expression of reporter transcripts.

Name	Sequence (5' - 3')
Junction_Forward	GAGGCACTGGGCAGGTGTC
qPCR_CCND1_500nt_R	GGTGCCAAACCATTCACAAG
qPCR_CCND1_1kb_R	CTGCTGCTGGGCTGTG
qPCR_TFF1_500nt_R	CCTCCGAAAGCTGCAGTG
qPCR_TFF1_1kb_R	CCTTGGGGATGCTGTGTCTG

2.7. Interrogation of Interferon Response

The delivery of the plasmid to mammalian cells triggers the activation of Type-1 Interferon response as explained in the introduction part[169]. The expression of genes related to this response was checked by the use of primers in Table6 by RT-qPCR.

In order to revert/block the interferon response, cells were co-treated with C16 inhibitor targeting PKR (Sigma-Aldrich, Cat#I9785-5MG) and BX-795 inhibitor targeting TBK1/IKK (Sigma-Aldrich, Cat#SML0694-5MG). The final concentration of each inhibitor was calculated as 1 nM.

Table 6. RT-qPCR primers used for interferon response genes and housekeeper gene

GENE	TYPE	Sequence
ACTB	F	GTTGTCGACGACGAGCG
	R	GCACAGAGCCTCGCCTT
TUBB	F	TCGATGCCATGTTCATCACT
	R	TAACCATGAGGGAAATCGTG
IFIT1	F	TGCTCCAGACTATCCTTGACCT
	R	TCTCAGAGGAGCCTGGCTAA
ISG15	F	AGCATCTTCACCGTCAGGTC
	R	GCGAACTCATCTTTGCCAGT

IFI27	F	GCCACAACTCCTCCAATCAC
	R	ATCAGCAGTGACCAGTGTGG
OAS3	F	GGTGAGGAGCCTCGAGTAGA
	R	CTGAAGAGCTGGACGGATGT
IRF7	F	AGGGTGACAGGTACGGCTCT
	R	CTCCTGGAGAGGGACAAGAA
MX1	F	GATGATCAAAGGGATGTGGC
	R	AGCTCGGCAACAGACTCTTC

2.8. Hormone Deprivation Experiments

DMEM (low glucose, pyruvate, no glutamine, and no phenol red) (Gibco, Cat#11880028) supplemented with 5% Charcoal Striped Serum (CSS) (Biowest, Cat#S181F) and 1% penicillin-streptomycin (Gibco, Cat#12140148) was used in hormone deprivation experiments. Workflow of the hormone deprivation experiment was given in Figure6. Estradiol (E2) (MedChemExpress, Cat#HY-B0141) treatment was started (final concentration of 10 nM) after 24h of transfection, and cells were collected after 6h for RNA isolation - termination of the experiment at a total of 30h after transfection. For blocking interferon response in cells, BX-795/C16 inhibitor co-treatment was started on Day4, where the medium change takes place.

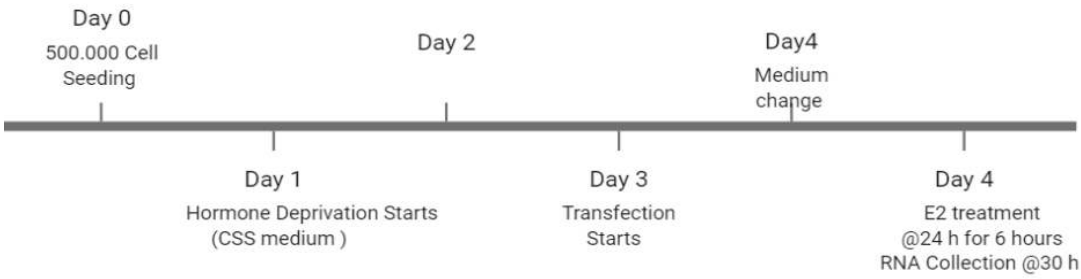


Figure 6 Workflow for Hormone Deprivation Experiments (In Day0 number of cells given for 6 well experiments is seeded. In Day1 CSS (Charcoal Striped Serum) containing phenol free medium is added to cells after washing 3 times to remove remaining from complete media. In Day2 cells incubated within hormone deprived media; In Day3, Transfection starts; In Day4, after 24 h of transfection, Estradiol(E2) treatment starts for 6 hours and at 30 h samples are collected.)

2.9. Library transfection

2.9.1. Library preparation and Content of the library

The ER-STARR-seq library is composed of ERα binding sites (ERBSs). These sites were identified by ChIP-seq analysis in the ER+ breast cancer and endometrial cancer patients in the Netherlands Cancer Institute (NKI) in the group of Wilbert ZWART. The library preparation steps were performed in the Nathan LACK's group by Flora Huang and Tunç Morova. Here we transfected the library to Ishikawa and MCF7 cells (KORKMAZ group).

There are three classes of ERBSs available in the ER-STARR-seq library; breast cancer-specific ERBSs, endometrial cancer-specific ERBSs, and ERBSs which are common in both cancer types (Figure 7).

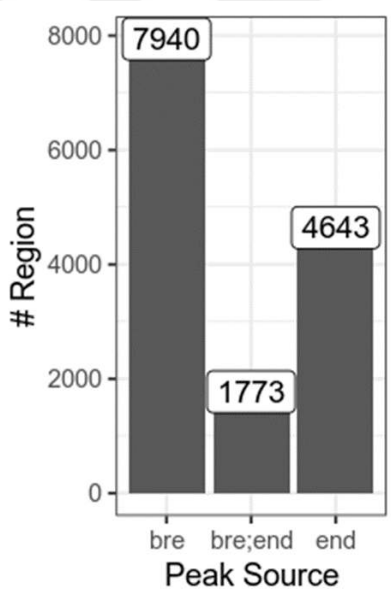


Figure 7 Content of ER-STARR-seq Library Library composed of three subgroups of ERBSs. First group is breast cancer specific ERBSs whereas second one is endometrial cancer specific ERBSs. The last group contains the ERBSs that common in both cancer types.

2.9.2. Coverage calculations

After transfection, coverage had been calculated to get an insight into the delivery success of the ER-STARR-Seq library to MCF7 and Ishikawa. Considering the transfection efficiency and cell death we started the experiment with 10M cells as an initial cell number. To calculate the coverage, the final number of cells was counted at the end of each experiment for each condition (vehicle and E2 treated). The formula was given in Figure8.

$$\text{Formula to calculate Coverage:} \quad \frac{\text{Final cell number}}{\text{x Transfection Efficiency}}$$

Figure 8. The Formula for coverage calculation. Final number of cells multiplied by transfection efficiency to calculate coverage.

Transfection experiments were performed with ER-STARR-Seq library and positive control plasmid (pLenti-CMV-eGFP-Puro, Addgene, Cat#17448) for both MCF7 and Ishikawa cells. Before the collection of the RNA samples, transfection efficiency was calculated via observing the transfection positive control. To calculate the transfection efficiency of the ER-STARR-Seq library, flow cytometry analysis was performed coupled with fluorescence microscopy imaging with eGFP transfected positive control. After GFP fluorescence images taken by fluorescent microscope, cells were trypsinized and collected as a pellet. Non-transfected control cells were used for compensation and to mask the autofluorescence of cells. Forward and side scatter height dimensions were used to map viable cell populations. For each cell, acquired settings were adjusted. The gain was adjusted for Ishikawa cells as FSC:12, SSC:25, FITC:1, and for MCF7 as FSC:8, SSC:28, FITC:1.

Chapter 3:
RESULTS

Experimental setup for transfection of ER-STARR-seq library was optimized. In the first step, ERα-regulated positive control plasmids were constructed. Then, transfection experiments were performed with PEI reagent. Parameters for the transfection with PEI reagent and STARR-seq system (insert size & time point) were optimized for both cell lines. Interferon response was considered for both cells and parameters for inhibitor treatments were optimized to prevent interferon response. Lastly, library transfection was carried out according to the optimized transfection protocol.

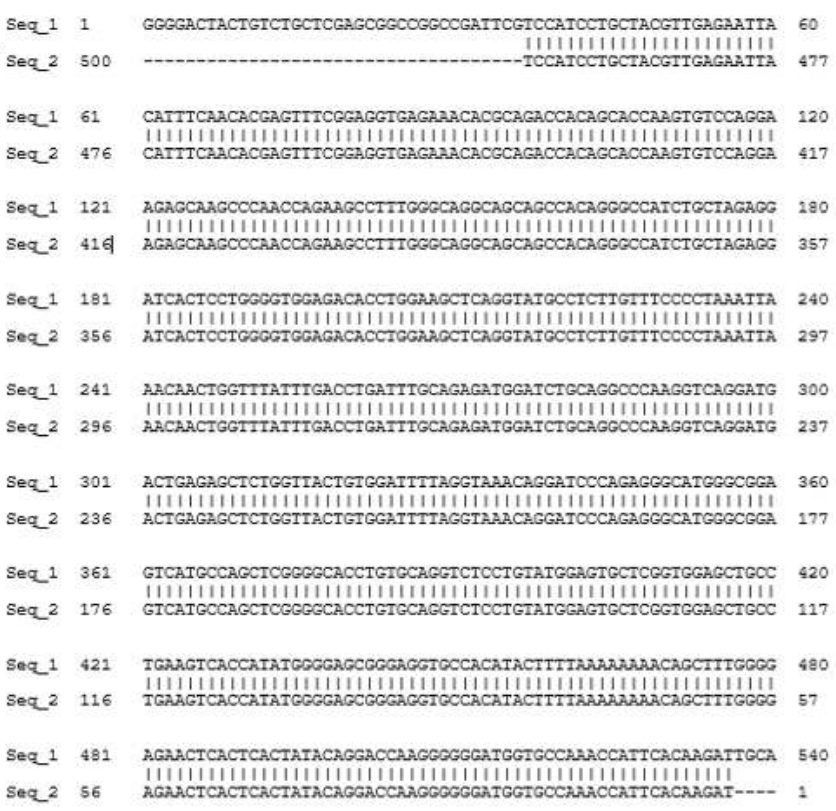
3.1. Confirmation of positive control plasmids

ERα regulated regions (CCND1 enhancer and TFF1 enhancer) were selected to clone into STARR-seq plasmid. These constructs would be used in optimization experiments to ensure that whether the self-transcription from the plasmid works in both cell lines. After constructs were ready, transformation was performed to make multiple copies of the cloned plasmids. During the replication of the cloned plasmid errors can be arise in the DNA sequences because of the replication machinery of bacteria [181]. In addition to transformation, during PCR step errors might be added to sequence because of the function of DNA polymerase as well [182]. Since even a single nucleotide change can affect the enhancer activity of the sequence, after cloning step, isolated plasmids from different colonies sent to Sanger sequencing to confirm the sequence of cloned plasmid.

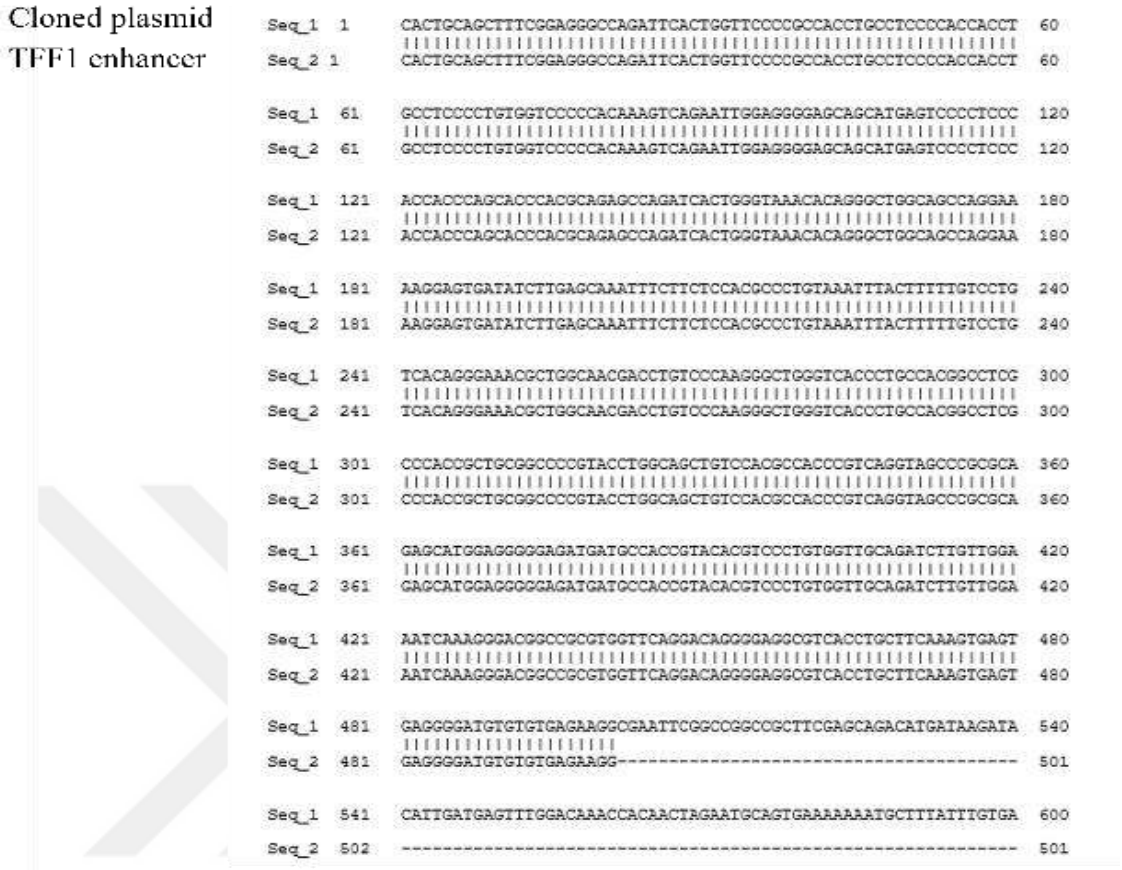
Alignment results of DNA sequences coming from sequencing to enhancer region that is cloned to plasmid was given in Figure9.

A

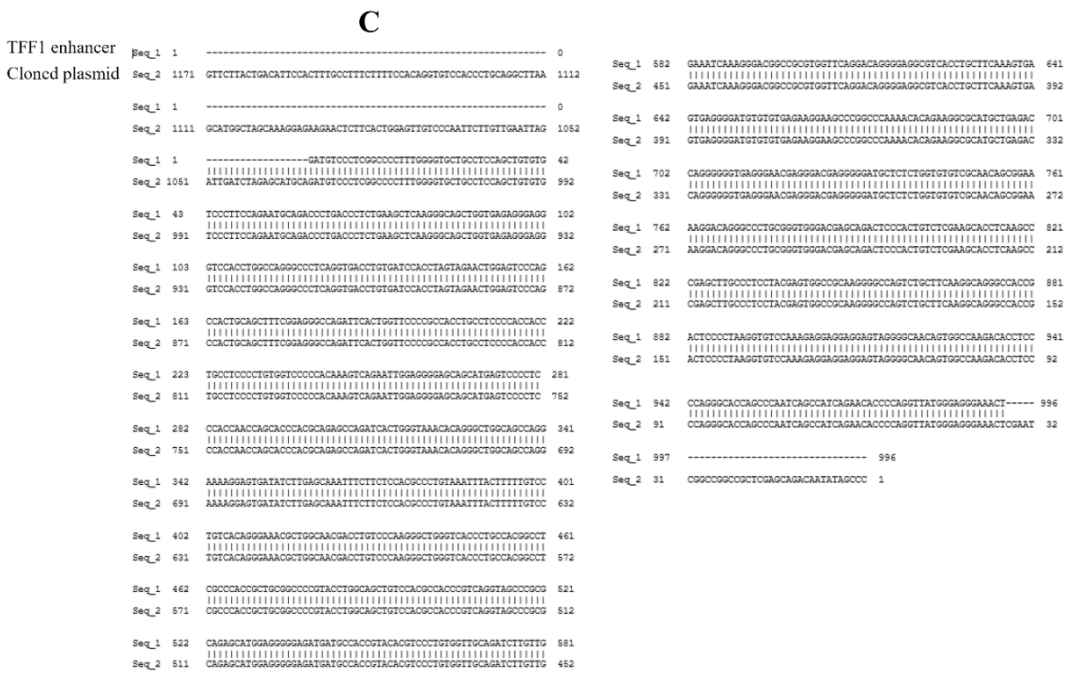
Cloned plasmid
CCND1 enhancer



B


Cloned plasmid
TFF1 enhancer

C


TFF1 enhancer
Cloned plasmid

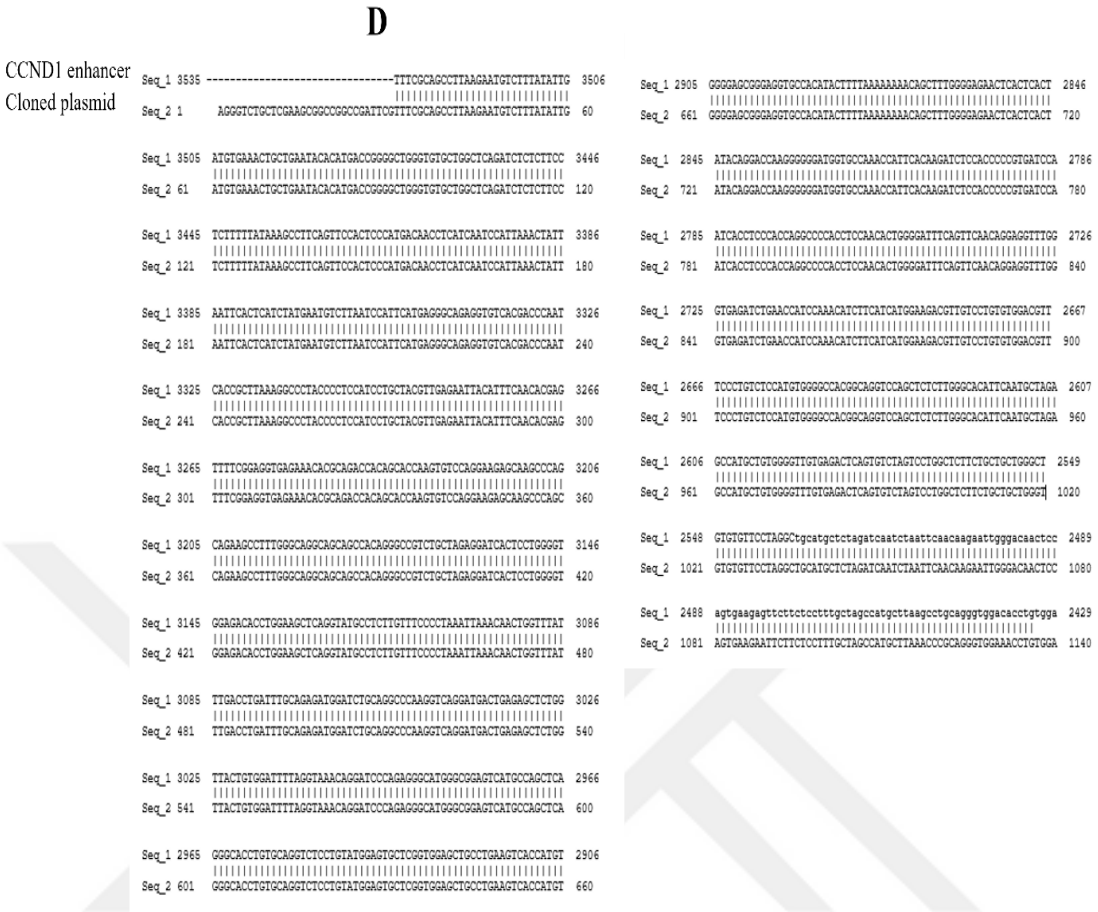


Figure 9. Alignment results of positive control plasmids. Cloned plasmids sequenced to confirm DNA sequence. To check DNA sequence Alignment was performed with alignment tool of Serial Cloner. Sequenced plasmid aligned to enhancer region cloned into plasmid. Figure9A shows the alignment results for CCND1 enhancer 500 bp cloned plasmid, Figure9B shows the alignment results for TFF1 enhancer 500 bp cloned plasmid. Figure9C and Figure9D are the results of 1 kb enhancer region cloned plasmids, respectively, CCND1 enhancer and TFF1 enhancer.

3.2. Optimization of Transfection parameters for MCF7 and Ishikawa

In this part, cloned plasmids were delivered to MCF7 and Ishikawa cells. Parameters for the delivery method and for STARR-seq system were optimized in these experiments.

Transfection is a method to deliver foreign nucleic acid constructs across the membrane of the cell. There are different methods such as biological-based methods, physical methods, and chemical-based methods such as cationic polymers-based reagents [180]. Firstly, our aim was to decide on a highly efficient and cost-effective method to introduce

the STARR-seq library to Ishikawa and MCF7 cell lines. Therefore, we focused on cationic polymer-based transfection methods. We tried three chemical transfection reagents such as Fugene, Lipofectamine, and PEI. Examinations on MCF7 and Ishikawa cell lines showed that PEI(Polyehylenimine) provided the promising transfection efficiency for these cell lines. Therefore, we continued to optimize transfection parameters with PEI reagent for MCF7 and Ishikawa cells.

Optimization of transfection parameters was performed in terms of two different parameters as the time point for sample collection to examine exogenous gene expression and the ideal PEI ratio for efficient transfection. Firstly, the sample collection time is important after STARR-seq library transfection to detect the reporter transcripts. Expression from the plasmid was checked at two different time points: 24 h and 48 h after transfection. According to expressions from each plasmid, at 24 h reporter transcripts showed the highest expression. An expression pattern between 24 h to 48 h was noted as the reliable interval to collect reporter transcripts for both cell lines. In the next step, different PEI ratios were optimized. Because of the toxic effect of PEI in Ishikawa cells, the 1:3 PEI ratio decreased transfection efficiency, but it was observed that the 1:2 PEI ratio was proper for transfection. Therefore, the 1:2 PEI ratio was noted as an ideal ratio for this cell line. For MCF7, 1:3 and 1:6 PEI ratios were examined and the 1:3 PEI ratio gave higher expression for our reporter genes compared to the 1:6 PEI ratio. For these sets of experiments, three biological replicates had been completed and the combined result is represented in Figure10. As a note, our STARR-seq fragments are 500 bp and because our system depends on a special STARR-seq plasmid and this plasmid-based system is highly sensitive to the three-dimensional structure of STARR-seq plasmid, we checked the effect of fragment length in order to be sure about our system works with 500 bp fragments. The expression of the 500 bp insert containing the CCND1 enhancer(enh588) is significantly higher than the expression of the 1 kb insert containing the CCND1 enhancer in both cell lines. Therefore, the ideal length of the insert for the STARR-seq plasmid is observed as 500 bp in the STARR-seq plasmid library.

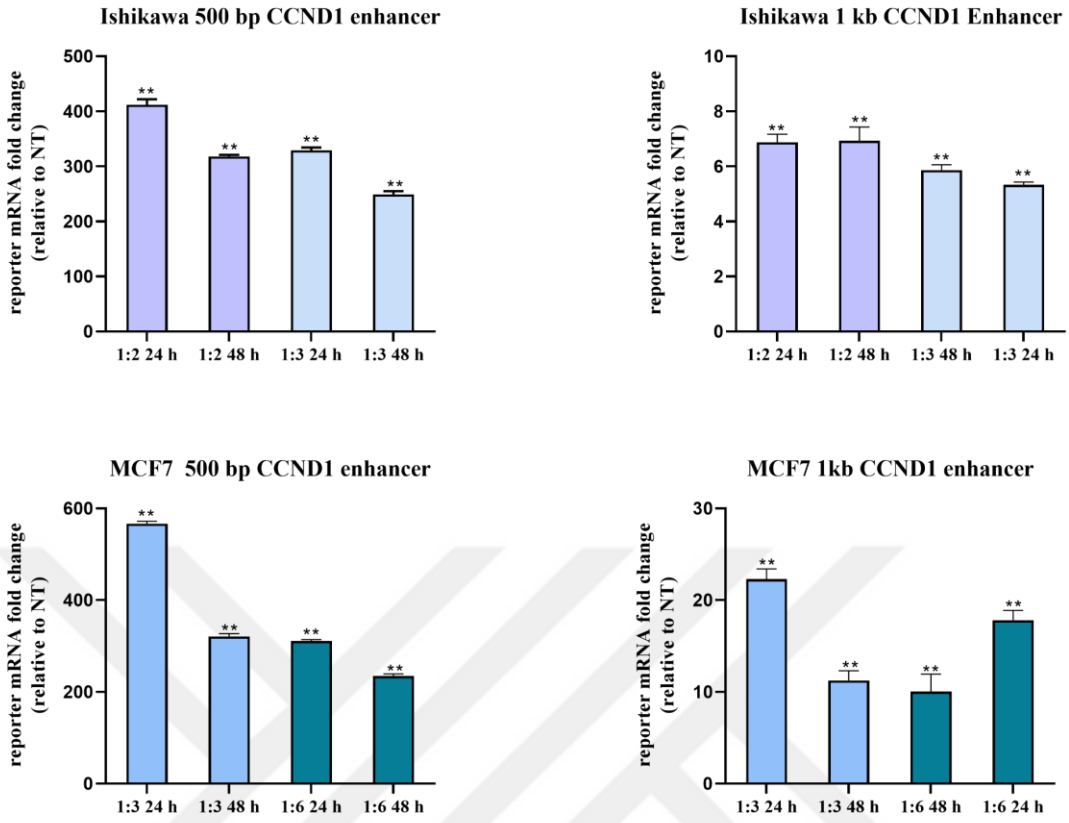


Figure 10. Optimization of Transfection Parameters as PEI ratio and time point for Ishikawa (A) and MCF7 (B) cells and examination of the effect of fragment length for the STARR-Seq system. RT-PCR results indicated that fragment length is important and 500 bp fragments were more suitable for the STARR-Seq system. In addition, PEI ratio was determined as 1:2 for Ishikawa cells and 1:3 for MCF7 cells. In both cases the 24 h-48 h interval is suitable to collect samples. (GAPDH is used as a housekeeping gene for Ishikawa, TBP is used as a housekeeping gene for MCF7 cells). Two tailed student's t-test relative to NT (Non-Transfected control) was performed. *P<0.05, **P<0.01,n.s., non-significant.

3.3. Interferon Response of Ishikawa and MCF7

Interferon response is a crucial consideration in the STARR-seq method. The delivery of plasmid inside to the cell can trigger the immune system-related type-I interferon response. The activation of this pathway results in differential expression of TFs in the whole cell landscape and alters the enhancer activity which affects the expression of transcript coming from STARR-Seq plasmid. Therefore, in this section, type-1 interferon response for each cell line was checked after the transfection of the STARR-Seq plasmid with CCND1 reporter fragment.

Type-1 interferon response was examined by the expression of genes such as OAS3, IRF7, and IFI27 which were activated upon induction of this response. For the Ishikawa, interferon response was not observed since the expression of related genes (5 out of 6 genes) were not significantly changed. However, for MCF7, interferon response was reported by 3 important genes as OAS3, IRF7, and IFIT1 with fold changes of around 30-fold as shown in Figure 11. Regarding these results, inhibitors treatment as suggested in the paper[169], was planned for MCF7 cells.

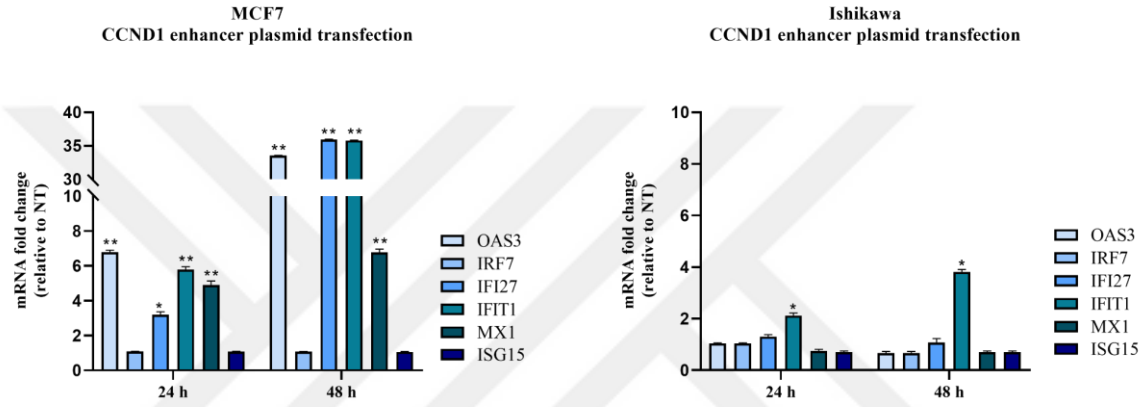


Figure 11. Examination of immune system-related type-I interferon response in MCF7 and Ishikawa cells. In Ishikawa cells, interferon response was not observed by interrogating the expression pattern of 6 different interferon response-related genes. Conversely, in MCF7 cells there was a certain interferon response that was observed with 4 genes that were upregulated approximately 30-fold after plasmid transfection. Two-tailed student's t-test relative to NT (Non-Transfected control) was performed. *P<0.05, **P<0.01,n.s., non-significant.

3.4. Hormone Deprivation Experiments

ERα responsive regulatory elements which potentially act as enhancers are regulated by hormones specifically Estradiol and until now it has been shown that a complete medium containing phenol, affects the ERα responsive regulatory elements with its estrogenic effects. Therefore, the certain function of these regulatory elements also should be investigated with a phenol-free hormone-deprived medium. In this section, it was tried to optimize the experimental setup required to reflect the response of ERα regulatory elements only coming from Estradiol.

Firstly, we aimed to check transfection efficiency in the hormone-deprived medium. To do that, pLenti-CMV-eGFP-Puro plasmid that involves the eGFP reporter gene was transfected in the hormone-deprived medium with optimized PEI ratios by following the workflow given in the Figure 6. As shown in Figure12, the efficiency of transfection was noted as approximately %50 for both cell lines examined via microscope. which was evaluated as sufficient for our library transfection experiment. Therefore, experiments after now had been performed with hormone deprived medium to prevent the effects of estradiol independent responses of ERα -regulated elements.

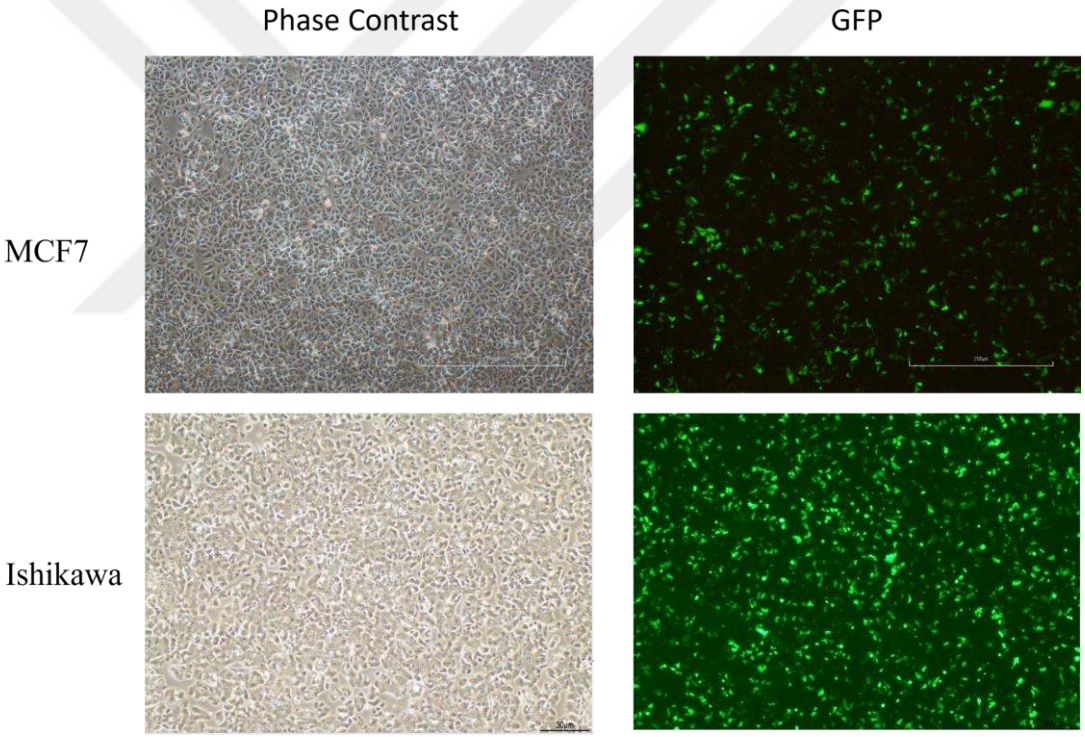


Figure 12. MCF7 and Ishikawa Cell lines Transfected with pLenti-CMV-eGFP-Puro in the phenol-free hormone-deprived medium. The transfection efficiency was evaluated in hormone deprived medium for MCF7 and Ishikawa cell lines. For both of them, transfection efficiency was decreased in hormone deprived medium compared to the complete medium; however, efficiency in this condition was sufficient to continue for STARR-Seq experiments.

3.4.1 Without inhibitor

In these experiments, the workflow in Figure 6 (given in the material method) was followed up. CCND1 enhancer (enh588) containing STARR-Seq plasmid and TFF1 enhancer containing STARR-Seq plasmids were transfected individually to MCF7 and Ishikawa cells. For each condition, effects of Estradiol (E2) without interferon response inhibitors were investigated after hormone deprivation was carried out. In MCF7 cells, 4-fold increase in the expression of CCND1(enh588) plasmid was observed in response to Estradiol, whereas approximately 2-fold change was observed for TFF1 plasmid. In Ishikawa cells, CCND1(enh588) expression was increased approximately 10-fold after treatment with E2. For TFF1, fold change was observed around 60-fold. As shown in Figure13, the enhancer strength of the corresponding regulatory element differs depending on cell type and is directly regulated by Estradiol. Despite the differential activity, the system can be defined as responsive for both cell lines.

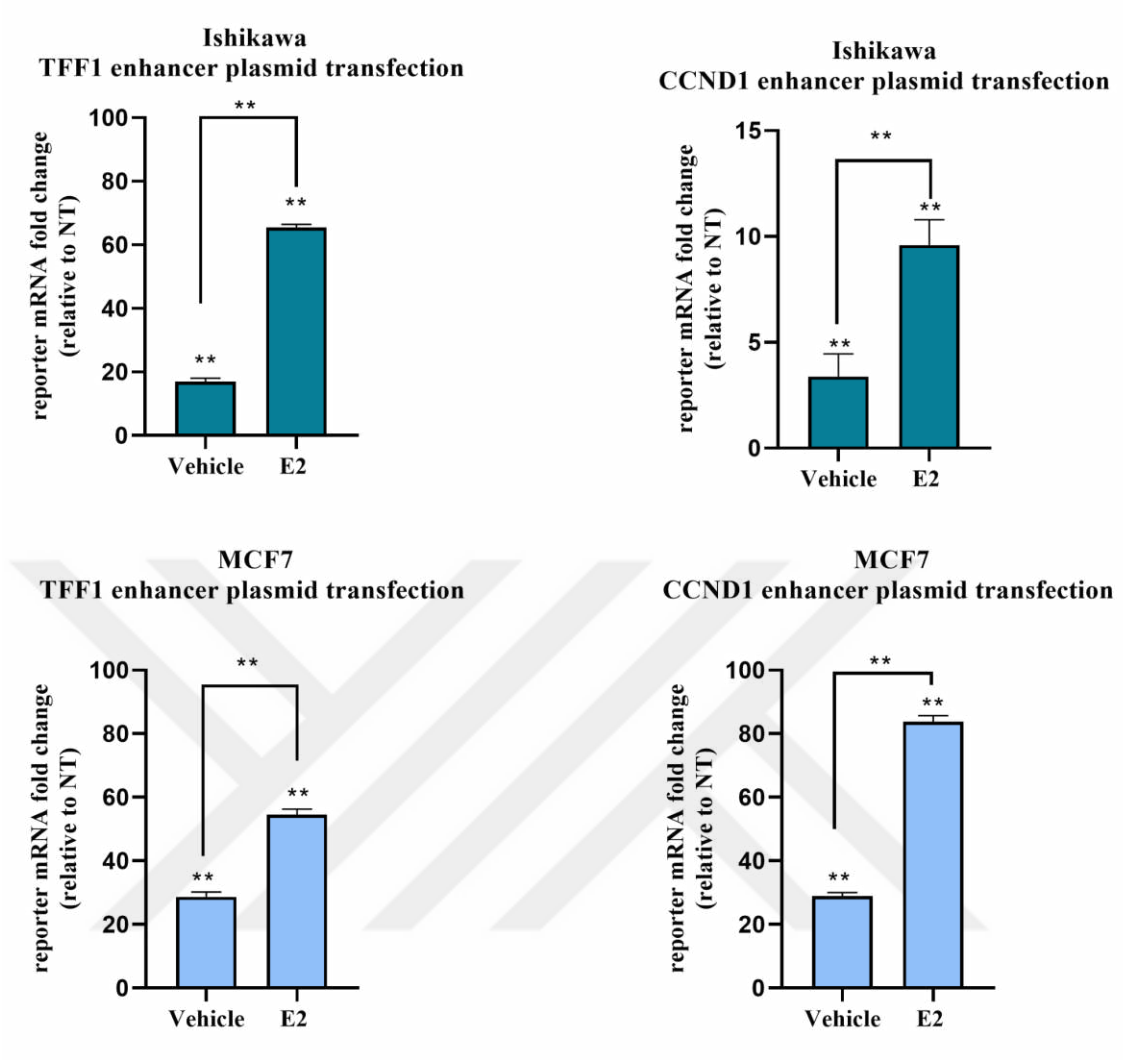


Figure 13. Differential expression of reporter fragments in STARR-seq plasmid after E2 treatment. For both cell lines, reporter fragment expression had been increased significantly after E2 induction. Even though there were efficiency differences between cell lines and reporter fragments, these results clearly indicated that the STARR-Seq system is responsive to E2. (Beta-actin is used as a housekeeping gene, E2: Estradiol) Two-tailed student's t-test relative to NT(Non-Transfected control) and also relative to the vehicle was performed. *P<0.05, **P<0.01,n.s., non-significant.

3.4.2. With Inhibitors

In this section, we examined the transfection efficiency and aimed to investigate the effects of Estradiol (E2) with interferon response inhibitors in hormone deprived medium. As shown in Figure11, MCF7 is highly sensitive to foreign plasmid delivery and activates interferon response. Even though there was no response in Ishikawa cells as shown in

Figure11, transfection experiments and subsequent investigations were performed in both MCF7 and Ishikawa cells to be sure about consistency between each cell line group and to compare them at the end of the STARR-Seq screening.

Therefore, transfection efficiency was investigated in the condition as hormone deprived medium with interferon response inhibitors. In these optimization experiments, two parameters were optimized as the time point those inhibitors were applied and the amount of the inhibitors. There were two-time points to start inhibitor treatment: Day3 which starts with transfection and Day4 when discarding the PEI-containing transfection media. As shown in Figure14, transfection efficiency decreased significantly when transfection and inhibitor treatment started at the same time point in Day3 as observed in both cell lines. Thus, the ideal time point to start inhibitor treatment was determined as Day4 when discarding PEI-containing transfection media.

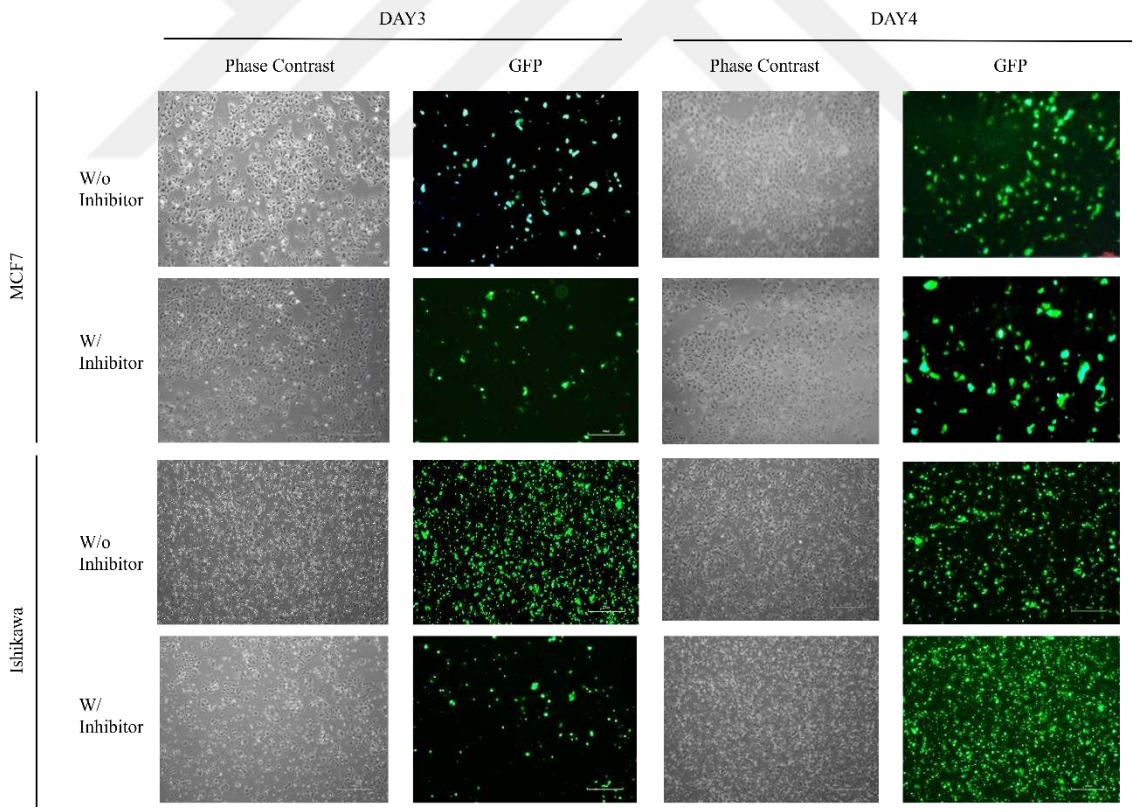


Figure 14. The evaluation of the optimum time point for interferon response inhibitor treatment. For MCF7 and Ishikawa cell lines, transfection of eGFP containing reporter plasmid was performed and 1 nM interferon response inhibitors were applied in indicated time points. In both MCF7 and Ishikawa, Day4 was selected as the most proper time to apply inhibitors.

After then, we investigated the activity of inhibitors by applying 1 nM from each inhibitor for each cell line. In the case, treatment with 1 nM inhibitors and concomitant E2 induction, interferon response was evaluated as shown in Figure 6. In MCF7, 1 nM inhibitors effectively abolished the interferon response related to unspecific induction of ERα regulated enhancers. As shown in Figure15, a combination of inhibitors and subsequent induction with E2 decreased the recorded response of CCND1 enhancer (Enh588). In Ishikawa cells, there was no effect of interferon response inhibitors because these cells were already unresponsive to plasmid delivery to activate interferon response. In addition, we evaluated the expression of interferon response regulated genes and we checked the inhibition could be provided with 1 nM inhibitors. As shown in Figure15, 1 nM inhibitors could effectively prevent the interferon response in MCF7 determined by the expression of genes that were upregulated in the case of no inhibitor condition.

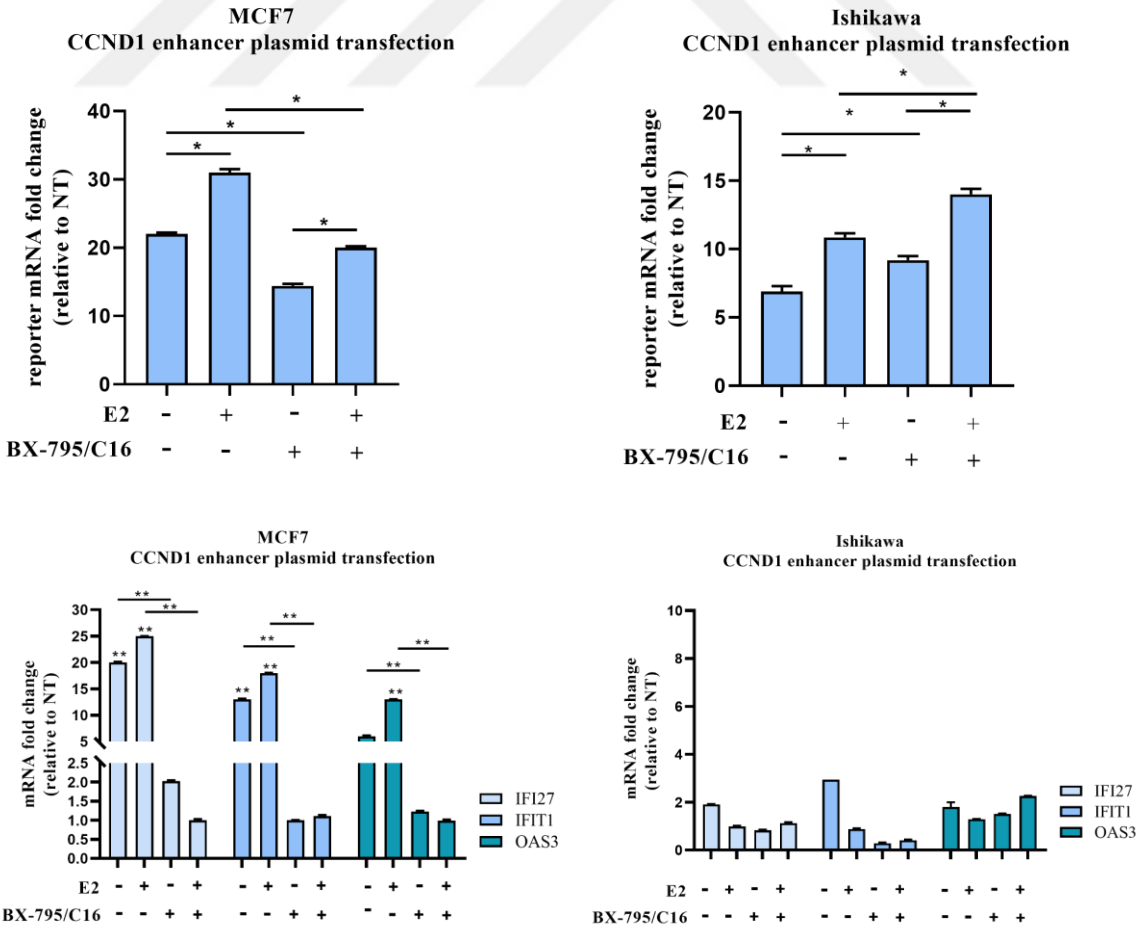
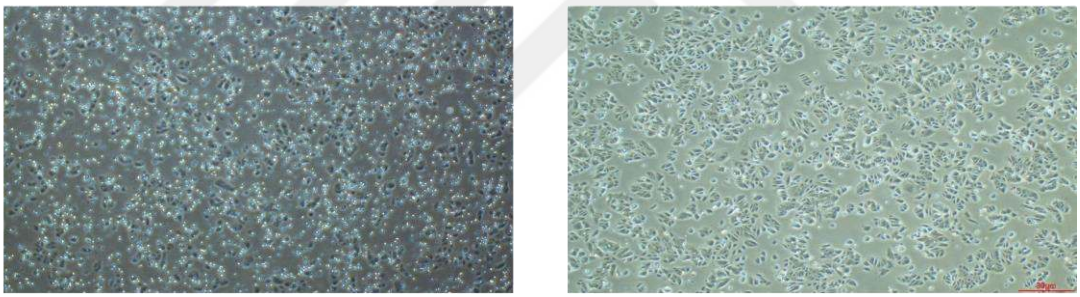


Figure 15. Optimization of the amount of interferon response inhibitors. In these experiments, 1 nM inhibitors were applied while removing PEI-containing transfection medium as at Day4. In MCF7, interferon response inhibitors apparently prevent the effect of interferon response by diminishing the E2 induction response which only represents the specific responses. In Ishikawa cells, the inhibitors had no effect because these cells did not show interferon response. Two tailed student's t-test relative to NT (Non-Transfected control) and also relative to vehicle and E2 condition to check significance of inhibitor effect was performed. *P<0.05, **P<0.01,n.s., non-significant.

Even though 1 nM inhibitors provided an efficient obstruction for the interferon response in MCF7, the toxicity was observed in Ishikawa cells realized by excessive cell death. The undetached death cells were observed under the microscope as indicated in Figure 16. In addition, cells were counted by trypan blue staining and the percentage of death cells and live cells were calculated. After that point, we were ensured the toxic effects of 1 nM inhibitors that we could not proceed with indicated concentration.

a.



b.

		Inhibitors(-)		*Inhibitors(+)*	
	NT	VEH	E2	VEH	E2
Death Cells (%)	%6	%21	%19	%45	%48
Live Cells(%)	%94	%79	%81	%55	%52

Figure 16. Evaluation of toxicity of 1 nM interferon response inhibitors for Ishikawa cells. In microscope imaging, it was clearly seen that 1 nM inhibitors caused an excessive cell death and cells appeared highly unhealthy to proceed experiments with these concentrations. In addition, we would like to be sure about cell death observed under microscope by certainly occurring in our experimental groups by counting cells with trypan blue. As indicated, death cell percentage was quite high in the condition with inhibitors compared with no inhibitor applied condition.

Because of the excess toxic effect of interferon response inhibitors for Ishikawa cells, the concentration of interferon response inhibitors was determined to be decreased 0.5 nM as

by half. However, as shown in Figure17, 0.5 nM interferon response inhibitors did not prevent interferon response in MCF7 cells effectively. In these experiments, besides interferon response, the expression of reporter transcripts is checked by RT-PCR. For MCF7 reporter transcripts coming from enhancer 588 showed changes in presence of inhibitors. In the 1 nm inhibitor treatment, expression coming from plasmid is reduced 10-fold compared to in the absence of inhibitor conditions. On the other hand, for 0.5 nm treatment showed 10-fold increase compared to its control, in the absence of inhibitors. For Ishikawa, 588 plasmid expression showed similar expression both in the presence of inhibitor and in the absence of inhibitor.

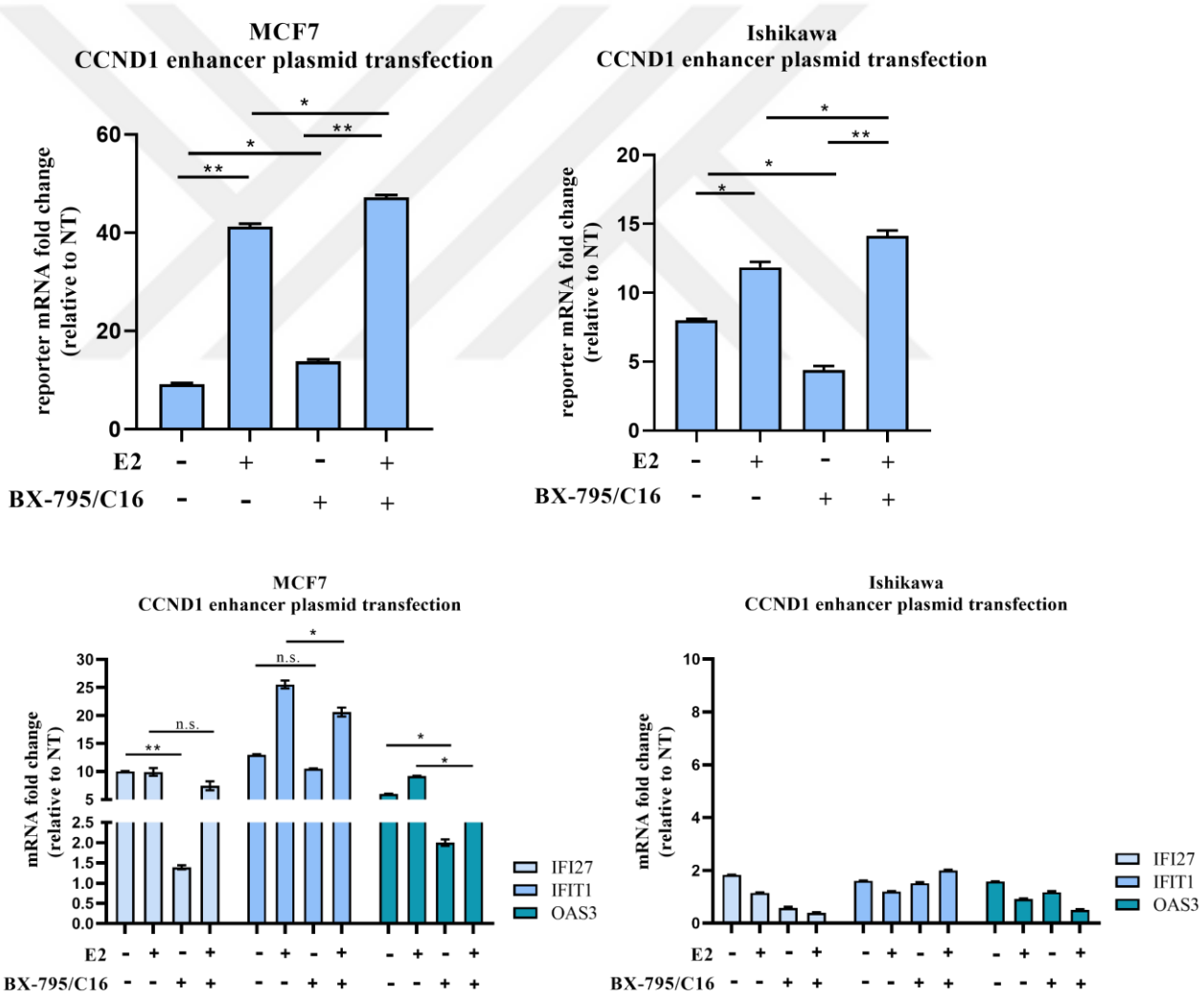


Figure 17. Evaluation of 0.5 nM interferon response inhibitors for MCF7 and Ishikawa cells. In these experiments, 0.5 nM inhibitors were applied while removing the PEI-containing transfection medium as at Day4. In MCF7, interferon response inhibitors could not prevent the effect of interferon response effectively as in the case of 1 nM inhibitors. Unfortunately, there was a clear induction of the CCND1 enhancer element

that was recorded unspecifically coming from interferon response. In Ishikawa cells, inhibitors had no effect on both CCND1 enhancer activity and gene expression of interferon response-related genes.

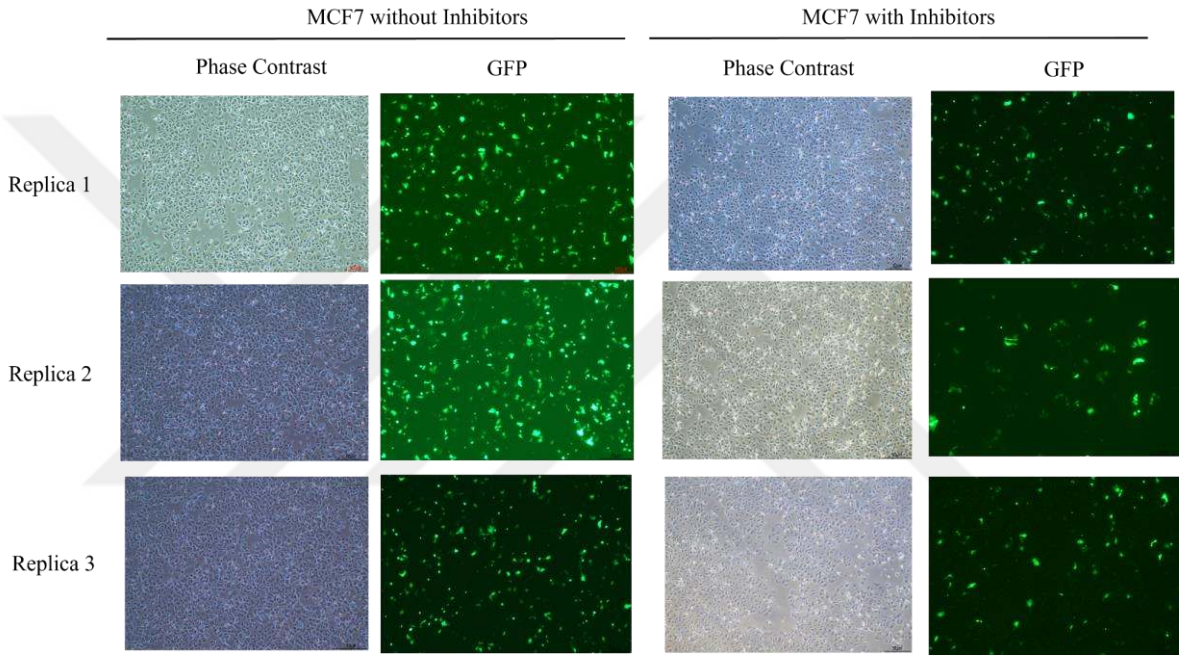
As a result, 1 nM interferon responsive inhibitors could effectively diminish the unspecific induction of ERα enhancers for MCF7 cells while 0.5 nM was not sufficient to inhibit. For Ishikawa cells, 1 nM inhibitors showed high toxicity to perform transfection-based screening with inhibitors. Therefore, we decided to proceed our experiments for MCF7 with and without 1 nM inhibitors while no treatment applied to Ishikawa cells during screening. Because Ishikawa cells did not show an interferon response, we concluded that for this cell line, screening could be proceeded without inhibitors. At the end of these experiments, there will be three groups to evaluate the result coming from screening. For MCF7, with and without inhibitor conditions will be performed while only without inhibitor conditions will be evaluated for Ishikawa cells to compare results between two different cell lines.

3.5. Coverage of ER-STARR-seq library systems

In the functional screening experiments, the final coverage represents the successful delivery of any content that included the fragments-of-interest to the target cells. To evaluate the phenotypic changes after delivering experimental constructs, there should be an expected coverage which enables to be sure about the average number of cells that involve plasmids-of-interest in a library after transfection. In our STARR-Seq screening, we deliver the library of different ERα binding site fragments characterized by ChIP-Seq analysis which were cloned into a special STARR-Seq plasmid as STARR-seq_ORI vector (Addgene, Plasmid#99296). After each transfection experiment, we evaluated the coverage to report the successful delivery of our library to the target cells as MCF7 and Ishikawa as explained in the material and method section.

To do that, the final number of cells was calculated at the end of each replica for each condition. Then transfection efficiency was multiplied by the final number of cells to find the transfected number of cells. Coverage was determined and recorded for subsequent evaluations. For all sets of experiments, transfection efficiency was calculated via Flow Cytometry Analysis which eGFP expression used as a control. In Figure18, the

transfection efficiency was evaluated for MCF7 (with and without interferon response inhibitors) and Ishikawa cells (without interferon response inhibitors) by evaluating the GFP expression as positive control. As can be seen in these results, the transfection efficiency was estimated as around 20% for MCF7 cells in w/o inhibitor conditions while the efficiency regarding coverage was lower in w/ inhibitor conditions. In Ishikawa cells, the transfection efficiency was estimated at around 30% by checking the GFP expression.



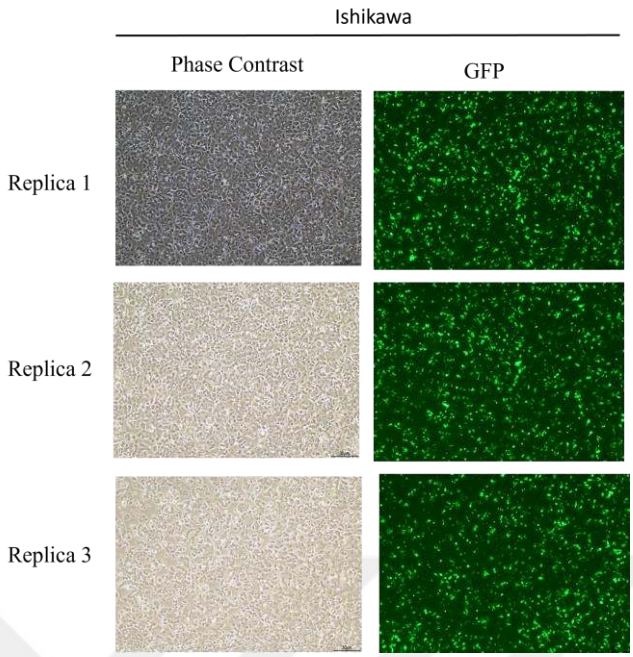
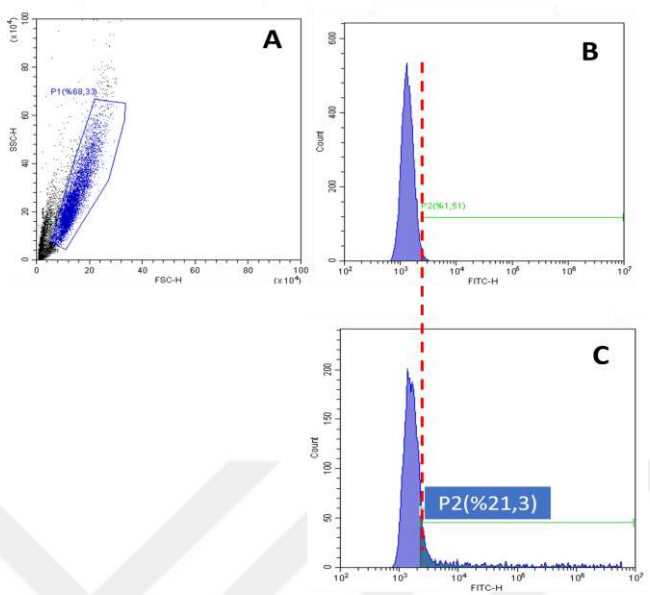


Figure 18. Microscope analysis to estimate transfection efficiency for MCF7 and Ishikawa cells. In MCF7 cells, the efficiency was at around 20% but as can be seen interferon response inhibitors decrease the efficiency. In Ishikawa cells, the transfection efficiency was higher compared with MCF7 at around 30%.

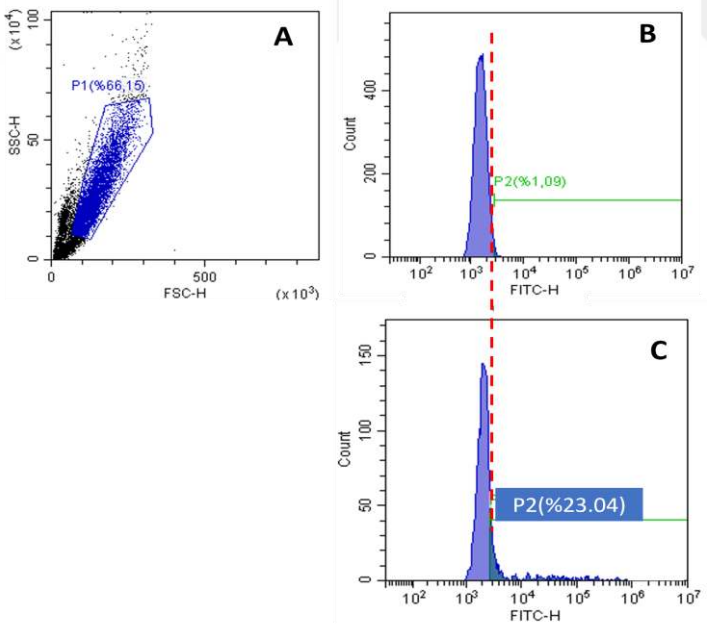
For exact calculations of coverage, flow cytometry analysis had been performed to evaluate transfection efficiency quantitatively by evaluating the eGFP expression that was used as positive control as shown in Figure 19-21. For Ishikawa cells only w/O inhibitor condition was performed, because of toxicity and related excess cell death. Transfection efficiency was recorded as 35%, 24%, and 32% in three biological replicates. For MCF7 cells recorded transfection efficiency in w/O inhibitor condition was 21%, 23%, and 25%, while in w/ inhibitor condition was 11%, 12%, and 14% in three biological replicates for each condition.

MCF7 w/o inhibitors



Replica 1

MCF7 w/o inhibitors



Replica 2

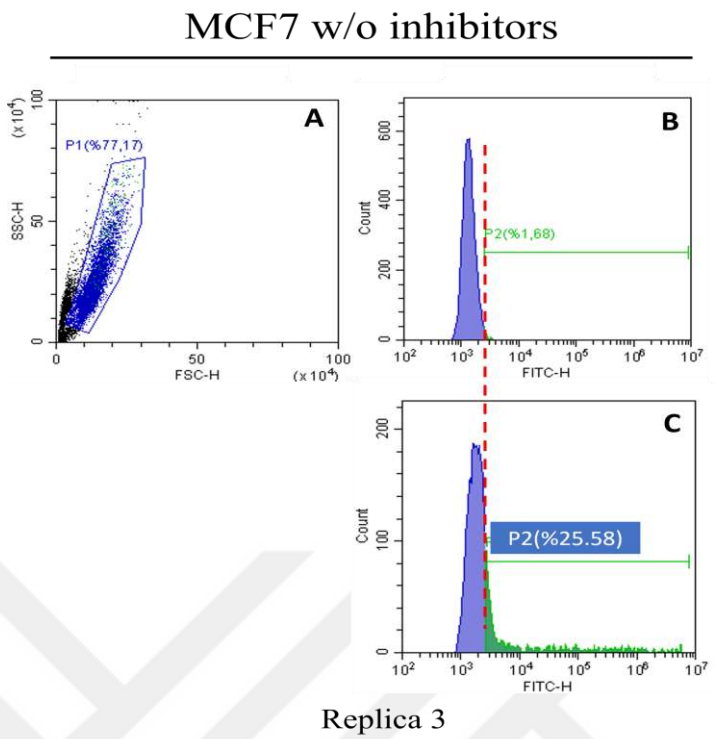
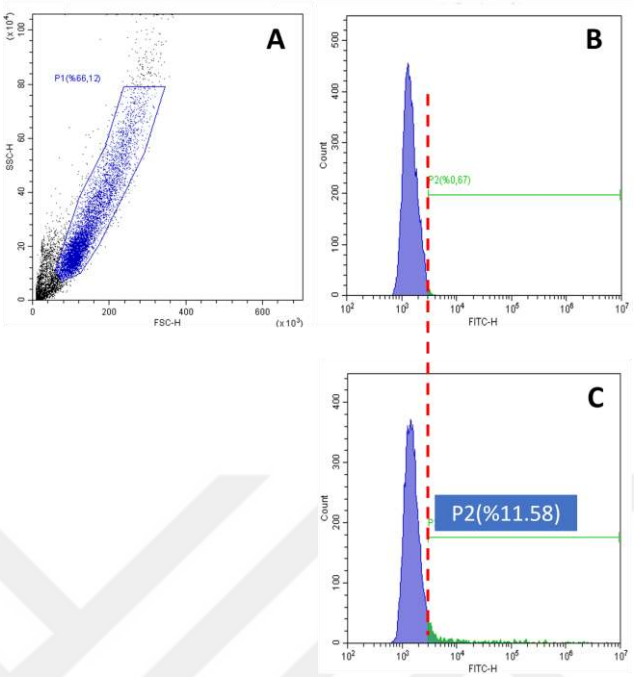


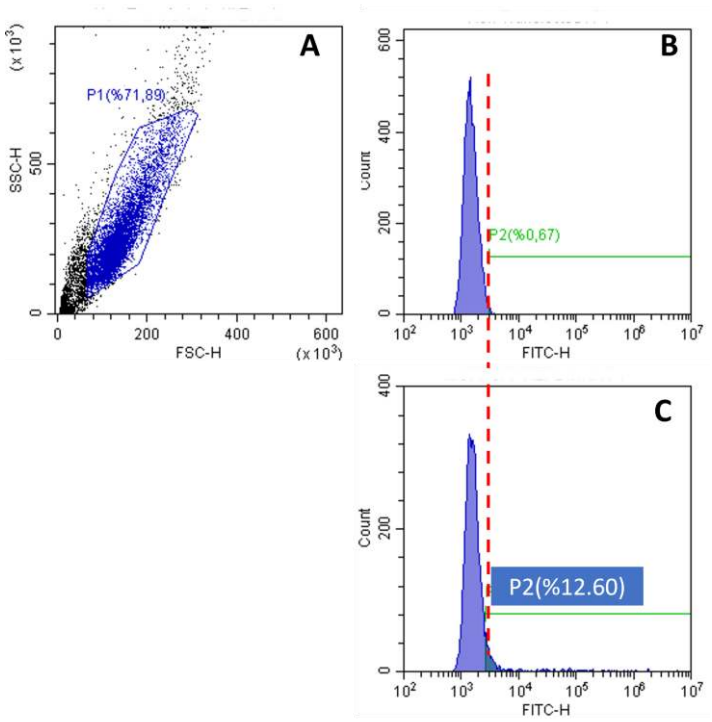
Figure 19. Flow cytometry analysis of MCF7(w/O inhibitor condition). GFP expression that was used as positive control was measured in the FITC channel of flow cytometry to determine transfection efficiency quantitatively and in three biological replicates transfection efficiency of GFP control was accepted as representative result.

MCF7 with Inhibitors



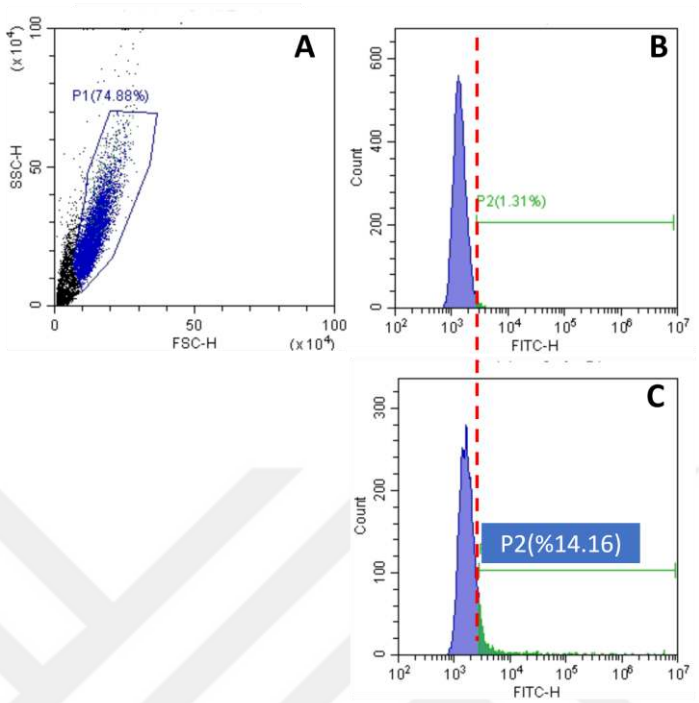
Replica 1

MCF7 with Inhibitors



Replica 2

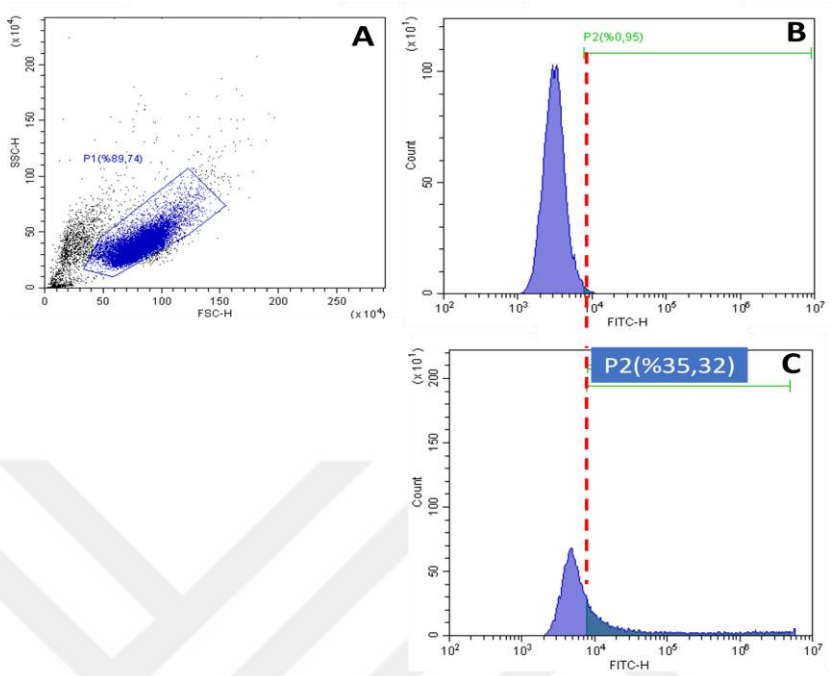
MCF7 with Inhibitors



Replica 3

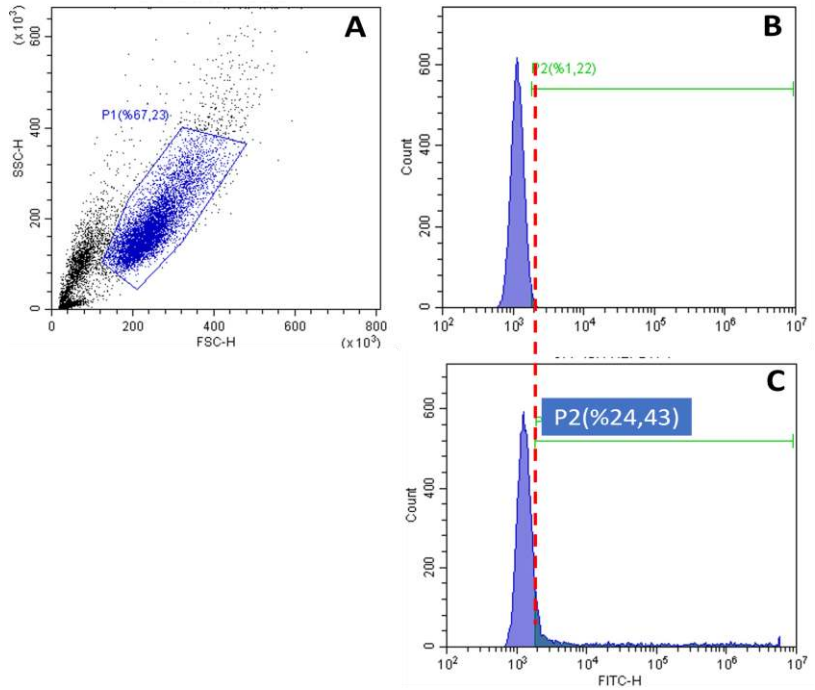
Figure 20 Flow cytometry analysis of MCF7(w/inhibitor condition). GFP expression that was used as positive control was measured in the FITC channel of flow cytometry to determine transfection efficiency quantitatively and in three biological replicates transfection efficiency of GFP control was accepted as representative result.

Ishikawa w/o inhibitors



Replica 1

Ishikawa w/o inhibitors



Replica 2

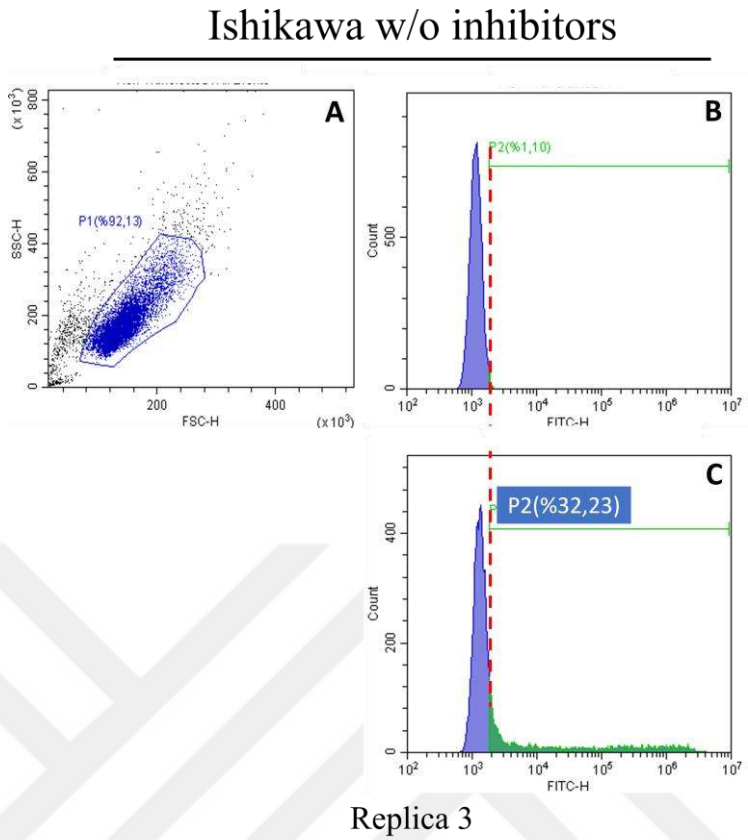


Figure 21 Flow cytometry analysis of Ishikawa (w/O inhibitor condition). GFP expression that was used as positive control was measured in the FITC channel of flow cytometry to determine transfection efficiency quantitatively and in three biological replicates transfection efficiency of GFP control was accepted as representative result.

For Ishikawa cells, the transfection efficiency of our plasmid library consisting of ERα binding site fragments was approximately %35, %24 and %32 in three biological replicates. Based on the recorded transfection efficiencies for these three replicates, coverages were calculated with formula given in Figure 5 at around 50x-60x. For MCF7, our STARR-Seq library transfection efficiency, according to flow cytometry analysis, was recorded as %21, %23 and %25 in three replicates. Coverage was calculated as approximately 27X. On the other hand, transfection efficiency in MCF7 w/ interferon response inhibitors were recorded at around 10% and coverage was calculated approximately 15X for three replicates as indicated in Figure 22.

Cell Line	Replica	Transfection Efficiency	Condition	# of Final Cells	Coverage
Ishikawa	1	%35	E2	171.000.000	~55X
			Vehicle	152.500.000	
	2	%24	E2	215.720.000	~50X
			Vehicle	203.280.000	
	3	%32	E2	200.210.000	~60X
			Vehicle	197.120.000	

Cell Line	Replica	Transfection Efficiency	Condition	# of Final Cells	Coverage
MCF7 (w/o inhibitors)	1	%21	E2	145.350.000	~28X
			Vehicle	135.660.000	
	2	%23	E2	126.900.000	~27X
			Vehicle	120.120.000	
	3	%25	E2	113.980.000	~27X
			Vehicle	112.099.999	

Cell Line	Replica	Transfection Efficiency	Condition	# of Final Cells	Coverage
MCF7 (w/inhibitors)	1	%11	E2	155.010.000	~17X
			Vehicle	162.750.000	
	2	%12	E2	125.820.000	~15X
			Vehicle	130.760.000	
	3	%14	E2	115.240.000	~16X
			Vehicle	120.280.000	

Figure 22 Quantitative transfection efficiency evaluation with Flow Cytometry analysis for Ishikawa and MCF7 cells and subsequent coverage calculation. For each cell line, three replicates were completed and final cell numbers before terminating the experiment were noted to calculate overall coverage.

Chapter 4:

CONCLUSION AND FUTURE DIRECTIONS

STARR-seq-based functional screening is a high-throughput plasmid-based technology that enables the assessment of enhancer activity of candidate DNA fragments both in a quantitative and simultaneous manner. There are several types of application areas of this system and one of them is mapping hormone-responsive enhancer activity as it was demonstrated in the pioneer study focusing on Glucocorticoid Receptors (GR) binding sites. They successfully clarified the activity of each candidate enhancer [175]. Moreover, another study was performed to investigate AR-regulated elements [65]. It could effectively map and characterize the activity of the AR-regulated enhancer elements. Thus, we aimed to characterize the ER-regulated enhancers by creating ***the first functional map of ERα enhancer activity by the STARR-Seq system***.

Working in collaboration with Zwart Lab (Netherlands Cancer Institute), ERα ChIP-seq analysis was conducted on breast cancer and endometrial cancer patients and clinical ERα binding sites had been identified. This captured ERBSs cloned into STARR-seq plasmid and the library containing ERα-regulated enhancer elements was generated (Lack Lab, Vancouver Prostate Centre). Here, we optimized the transfection and treatment conditions to introduce the library to two ERα positive cell lines: MCF7, breast cancer cell line, and Ishikawa, endometrial cancer cell line. We chose PEI reagent, a cationic polymer as the most cost-effective transfection methodology. After optimization of transfection parameters, the length of the insert, and the time point to detect reporter transcripts in response to Estradiol treatment, library transfection was conducted for two systems.

Delivery of the foreign nucleic acids to cells is known as transfection. For the ideal transfection, besides minimal effects on normal physiology and cell toxicity, the considerations can be noted as high transfection efficiency, reproducibility, easy use, and cost-friendly[180]. There are different methods for mammalian cells to achieve ideal transfection from chemical ones to physical ones. Having minimal effect on cell

physiology and being cost-friendly, we continued with the chemical-based methods even though there were transfection efficiency problems. After several tries with cationic lipid carriers (Lipofectamine, FuGENE) and cationic polymer (PEI), the most efficient transfection with low toxicity is detected with PEI reagent in both cell lines. Thus, we continued with the PEI transfection. However, as I will explain in upcoming parts, our coverages were highly low compared to other systems used for mammalian cell transfection such as electroporation.

For Ishikawa and MCF7 cell lines STARR-Seq screening was performed with a library consisting of clinically relevant ERα binding sites. As explained in the results section, we performed STARR-Seq library transfection by using two different ERα hormone receptor-positive cell lines as our models. Transfection was performed for Ishikawa cells without inhibitors for interferon response-related pathways while for Ishikawa cells we had to conduct the experiments with these inhibitors because MCF7 is highly responsive to delivery of foreign DNA by activating the Type-I interferon response. The consideration of this response in mammalian cells is false positive and false negative results as shown in the STARK's paper[169]. As authors showed, enhancer activity of a group of candidate fragments is affected in the presence of type-1 interferon response-related TFs. Therefore, to prevent nonspecific production of reporter transcripts inhibitors to downregulate interferon response was applied in MCF7 cells.

By considering all these factors including transfection efficiency and interferon response, we optimized the systems and transfected the ERα STARR-seq library to Ishikawa and MCF7 cells. To sum up, for Ishikawa cells coverage was approximately 55X. For MCF7 without inhibitors about 27X coverage was noted whereas for with inhibitors around 15X coverages.

In this thesis, optimization of the transfection system and transfection of the custom STARR-Seq library to two ERα positive cell models were performed. The next steps including reporter transcript isolation and sequencing analysis will be conducted in Lack Lab at Vancouver. However, the results could not be ready to discuss in this thesis. In the last part of this chapter, I will explain the expected outcomes and future directions.

After bioinformatics analysis on samples, we will obtain an ERα enhancer activity map that interrogates the transcriptional activity of clinically relevant ERBSs in two different cancer cell lines as MCF7 and Ishikawa. In these maps, we can categorize the ERα regulated regions in terms of their enhancer activity as inducible ones, constitutively active ones, and inactive ones. By comparing the enhancer activity maps belonging to two different cell line models, the differential transcriptional activity that might be related with the recruitment of tissue-specific co-activators and TFs will be investigated. Notably, since it is a plasmid-based system, it is not affected by chromatin accessibility. Therefore, in this system we can also assess the enhancer activity of closed regions.

In summary, ***the first functional map of ERα enhancer activity by the STARR-Seq system*** in two ERα positive cancer cell line models as MCF7 and Ishikawa will be generated to elucidate the potential mammalian gene regulation mechanisms depending on ERα. In further studies, the identification of key enhancers might be used to study disease-causing mutations and the characterization of tissue-specific co-regulators of ERα can be performed. Thus, a better understanding of ERα enhancer activity will bring an insight into the hormonal therapy response and to develop new potential treatment approaches.

BIBLIOGRAPHY

1. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A. 2014;111: 6131–6138.

2. Carter DRF, Eskiw C, Cook PR. Transcription factories. Biochem Soc Trans. 2008;36: 585–589.

3. Edelman LB, Fraser P. Transcription factories: genetic programming in three dimensions. Curr Opin Genet Dev. 2012;22: 110–114.

4. Chatterjee S, Ahituv N. Gene Regulatory Elements, Major Drivers of Human Disease. Annu Rev Genomics Hum Genet. 2017;18: 45–63.

5. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet. 2020;21: 71–87.

6. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004;306: 636–640.

7. de Souza N. The ENCODE project. Nature methods. 2012. p. 1046.

8. Liu S, Liu Y, Zhang Q, Wu J, Liang J, Yu S, et al. Systematic identification of regulatory variants associated with cancer risk. Genome Biol. 2017;18: 194.

9. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman İ, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet. 2014;46: 136–143.

10. Perenthaler E, Yousefi S, Niggl E, Barakat TS. Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development. Front Cell Neurosci. 2019;13: 352.

11. de Villiers J, Schaffner W. A small segment of polyoma virus DNA enhances the expression of a cloned beta-globin gene over a distance of 1400 base pairs. Nucleic Acids Res. 1981;9: 6251–6264.

12. Field A, Adelman K. Evaluating Enhancer Function and Transcription. Annu Rev Biochem. 2020;89: 213–234.

13. Gillies SD, Morrison SL, Oi VT, Tonegawa S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell. 1983;33: 717–728.

14. Geyer PK, Green MM, Corces VG. Tissue-specific transcriptional enhancers may

act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. EMBO J. 1990;9: 2247–2256.

15. Sasaki-Iwaoka H, Maruyama K, Endoh H, Komori T, Kato S, Kawashima H. A trans-acting enhancer modulates estrogen-mediated transcription of reporter genes in osteoblasts. J Bone Miner Res. 1999;14: 248–255.

16. Gribaudo G. Transcriptional regulation of eukaryotic gene expression by cis-acting DNA sequences. G Batteriol Virol Immunol. 1986;79: 3–10.

17. Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nat Rev Genet. 2018;19: 789–800.

18. Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, et al. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell. 2017;171: 1573–1588.e28.

19. Ray-Jones H, Spivakov M. Transcriptional enhancers and their communication with gene promoters. Cell Mol Life Sci. 2021;78: 6453–6485.

20. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet. 2019;20: 437–455.

21. Mermet J, Yeung J, Hurni C, Mauvoisin D, Gustafson K, Jouffe C, et al. Clock-dependent chromatin topology modulates circadian transcription and behavior. Genes Dev. 2018;32: 347–358.

22. Corbin V, Maniatis T. The role of specific enhancer-promoter interactions in the Drosophila Adh promoter switch. Genes Dev. 1989;3: 2191–2120.

23. Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development. 2005;132: 797–803.

24. Marcotte R, Chen JM, Huard S, Wang E. c-Myc creates an activation loop by transcriptionally repressing its own functional inhibitor, hMad4, in young fibroblasts, a loop lost in replicatively senescent fibroblasts. J Cell Biochem. 2005;96: 1071–1085.

25. Miele A, Gheldof N, Tabuchi TM, Dostie J, Dekker J. Mapping chromatin interactions by chromosome conformation capture. Curr Protoc Mol Biol. 2006;Chapter 21: Unit 21.11.

26. Duan J, Lu G, Hong Y, Hu Q, Mai X, Guo J, et al. Live imaging and tracking of genome regions in CRISPR/dCas9 knock-in mice. Genome Biol. 2018;19: 192.

27. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. Nat Rev Genet. 2013;14: 288–295.

28. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009;457: 854–858.

29. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107: 21931–21936.

30. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39: 311–318.

31. Dorschner MO, Hawrylycz M, Humbert R, Wallace JC, Shafer A, Kawamoto J, et al. High-throughput localization of functional elements by quantitative chromatin profiling. Nat Methods. 2004;1: 219–225.

32. Zhang T, Zhang Z, Dong Q, Xiong J, Zhu B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. Genome Biol. 2020;21: 45.

33. Lenardo M, Pierce JW, Baltimore D. Protein-binding sites in Ig gene enhancers determine transcriptional activity and inducibility. Science. 1987;236: 1573–1577.

34. Rao S, Ahmad K, Ramachandran S. Cooperative binding between distant transcription factors is a hallmark of active enhancers. Mol Cell. 2021;81: 1651–1665.e4.

35. Merika M, Williams AJ, Chen G, Collins T, Thanos D. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Mol Cell. 1998;1: 277–287.

36. Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, et al. A high resolution protein interaction map of the yeast Mediator complex. Nucleic Acids Res. 2004;32: 5379–5391.

37. Fan X, Chou DM, Struhl K. Activator-specific recruitment of Mediator in vivo. Nat Struct Mol Biol. 2006;13: 117–120.

38. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008;132: 311–322.

39. Zhang W, Zhang T, Wu Y, Jiang J. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. Plant Cell. 2012;24: 2719–2731.

40. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489: 75–82.

41. Galang G, Mandla R, Ruan H, Jung C, Sinha T, Stone NR, et al. ATAC-Seq Reveals an Enhancer That Regulates Sinoatrial Node Development and Function. Circ Res. 2020;127: 1502–1518.

42. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. Mol

Metab. 2016;5: 233–244.

43. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. 2010;8: e1000384.

44. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010;465: 182–187.

45. Dickel DE, Zhu Y, Nord AS, Wylie JN, Akiyama JA, Afzal V, et al. Function-based identification of mammalian enhancers using site-specific integration. Nat Methods. 2014;11: 566–571.

46. Kwasnieski JC, Fiore C, Chaudhari HG, Cohen BA. High-throughput functional testing of ENCODE segmentation predictions. Genome Res. 2014;24: 1595–1602.

47. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 2013;339: 1074–1077.

48. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S, et al. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods. 2008;5: 829–834.

49. Ramirez-Prado JS, Latrasse D, Benhamed M. Histone Modification ChIP-seq on Plantlets. Bio Protoc. 2021;11: e4211.

50. Nakato R, Sakata T. Methods for ChIP-seq analysis: A practical workflow and advanced applications. Methods. 2021;187: 44–53.

51. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008;322: 1845–1848.

52. O'Kane CJ, Gehring WJ. Detection in situ of genomic regulatory elements in Drosophila. Proc Natl Acad Sci U S A. 1987;84: 9123–9127.

53. Diao Y, Li B, Meng Z, Jung I, Lee AY, Dixon J, et al. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. Genome Res. 2016;26: 397–405.

54. Korkmaz G, Lopes R, Ugalde AP, Nevedomskaya E, Han R, Myacheva K, et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat Biotechnol. 2016;34: 192–198.

55. Parker SCJ, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A. 2013;110: 17921–17926.

56. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity

genes. Cell. 2013;153: 307–319.

57. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155: 934–947.

58. Jia Y, Chng W-J, Zhou J. Super-enhancers: critical roles and therapeutic targets in hematologic malignancies. J Hematol Oncol. 2019;12: 77.

59. Davey RA, Grossmann M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. Clin Biochem Rev. 2016;37: 3–15.

60. Fuentes N, Silveyra P. Estrogen receptor signaling mechanisms. Adv Protein Chem Struct Biol. 2019;116: 135–170.

61. Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. Enhancer transcripts mark active estrogen receptor binding sites. Genome Res. 2013;23: 1210–1223.

62. Kulik M, Bothe M, Kibar G, Fuchs A, Schöne S, Prekovic S, et al. Androgen and glucocorticoid receptor direct distinct transcriptional programs by receptor-specific and shared DNA binding sites. Nucleic Acids Res. 2021;49: 3856–3875.

63. Zhou Y, Bolton EC, Jones JO. Androgens and androgen receptor signaling in prostate tumorigenesis. J Mol Endocrinol. 2015;54: R15–29.

64. Yip C-H, Rhodes A. Estrogen and progesterone receptors in breast cancer. Future Oncol. 2014;10: 2293–2301.

65. Huang C-CF, Lingadahalli S, Morova T, Ozturan D, Hu E, Yu IPL, et al. Functional mapping of androgen receptor enhancer activity. Genome Biol. 2021;22: 149.

66. Yokoyama S, Higashida H. [Classification of cell-surface receptors based on molecular structures and functions]. Nihon Rinsho. 1998;56: 1745–1749.

67. Wee P, Wang Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. Cancers . 2017;9. doi:10.3390/cancers9050052

68. Mazaira GI, Zgajnar NR, Lotufo CM, Daneri-Becerra C, Sivils JC, Soto OB, et al. Nuclear Receptors: A Historical Perspective. Methods Mol Biol. 2019;1966: 1–5.

69. Evans RM. The steroid and thyroid hormone receptor superfamily. Science. 1988;240: 889–895.

70. O'Malley BW, Means AR. Female steroid hormones and target cell nuclei. Science. 1974;183: 610–620.

71. Walter P, Green S, Greene G, Krust A, Bornert JM, Jeltsch JM, et al. Cloning of the human estrogen receptor cDNA. Proc Natl Acad Sci U S A. 1985;82: 7889–7893.

72. MacLeod KM, Baxter JD. Chromatin receptors for thyroid hormones. Interactions

of the solubilized proteins with DNA. J Biol Chem. 1976;251: 7380–7387.

73. Laudet V, Hänni C, Coll J, Catzeflis F, Stéhelin D. Evolution of the nuclear receptor gene superfamily. EMBO J. 1992;11: 1003–1013.

74. Mazaira GI, Zgajnar NR, Lotufo CM, Daneri-Becerra C, Sivils JC, Soto OB, et al. The Nuclear Receptor Field: A Historical Overview and Future Challenges. Nucl Receptor Res. 2018;5. doi:10.11131/2018/101320

75. Wu W, LoVerde PT. Identification and evolution of nuclear receptors in Platyhelminths. PLoS One. 2021;16: e0250750.

76. Ahlquist JA, Franklyn JA, Sheppard MC, Ramsden DB. Thyroid hormone response elements in pituitary genes. Biochim Biophys Acta. 1989;1008: 251–254.

77. Becker PB, Gloss B, Schmid W, Strähle U, Schütz G. In vivo protein-DNA interactions in a glucocorticoid response element require the presence of the hormone. Nature. 1986;324: 686–688.

78. Gim HJ, Park J, Jung ME, Houk KN. Conformational dynamics of androgen receptors bound to agonists and antagonists. Sci Rep. 2021;11: 15887.

79. Billas IML, Moras D. Ligand-binding pocket of the ecdysone receptor. Vitam Horm. 2005;73: 101–129.

80. Egea PF, Klaholz BP, Moras D. Ligand-protein interactions in nuclear receptors of hormones. FEBS Lett. 2000;476: 62–67.

81. Haelens A, Tanner T, Denayer S, Callewaert L, Claessens F. The hinge region regulates DNA binding, nuclear translocation, and transactivation of the androgen receptor. Cancer Res. 2007;67: 4514–4523.

82. Yaşar P, Ayaz G, User SD, Güpür G, Muyan M. Molecular mechanism of estrogen-estrogen receptor signaling. Reprod Med Biol. 2017;16: 4–20.

83. Wärnmark A, Treuter E, Wright APH, Gustafsson J-A. Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation. Mol Endocrinol. 2003;17: 1901–1909.

84. Lundberg B. The solubilisation of some steroids by phosphatidylcholine and phosphatidylcholine-cholesterol vesicles. Chem Phys Lipids. 1979;24: 257–263.

85. Gruber CJ, Tschugguel W, Schneeberger C, Huber JC. Production and actions of estrogens. N Engl J Med. 2002;346: 340–352.

86. Nelson LR, Bulun SE. Estrogen production and action. J Am Acad Dermatol. 2001;45: S116–24.

87. Shoham Z, Schachter M. Estrogen biosynthesis--regulation, action, remote effects, and value of monitoring in ovarian stimulation cycles. Fertil Steril. 1996;65: 687–701.

88. Deroo BJ, Hewitt SC, Collins JB, Grissom SF, Hamilton KJ, Korach KS. Profile of estrogen-responsive genes in an estrogen-specific mammary gland outgrowth model. Mol Reprod Dev. 2009;76: 733–750.

89. Zheng D, Wang X, Antonson P, Gustafsson J-Å, Li Z. Genomics of sex hormone receptor signaling in hepatic sexual dimorphism. Mol Cell Endocrinol. 2018;471: 33–41.

90. Westberg L, Eriksson E. Sex steroid-related candidate genes in psychiatric disorders. J Psychiatry Neurosci. 2008;33: 319–330.

91. Knowlton AA, Lee AR. Estrogen and the cardiovascular system. Pharmacol Ther. 2012;135: 54–70.

92. Moulton VR. Sex Hormones in Acquired Immunity and Autoimmune Disease. Front Immunol. 2018;9: 2279.

93. Govind AP, Thampan RV. Membrane associated estrogen receptors and related proteins: localization at the plasma membrane and the endoplasmic reticulum. Mol Cell Biochem. 2003;253: 233–240.

94. Huang W, Peng Y, Kiselar J, Zhao X, Albaqami A, Mendez D, et al. Multidomain architecture of estrogen receptor reveals interfacial cross-talk between its DNA-binding and ligand-binding domains. Nat Commun. 2018;9: 3520.

95. Hewitt SC, Korach KS. Estrogen receptors: structure, mechanisms and function. Rev Endocr Metab Disord. 2002;3: 193–200.

96. Paterni I, Granchi C, Katzenellenbogen JA, Minutolo F. Estrogen receptors alpha (ERα) and beta (ERβ): subtype-selective ligands and clinical potential. Steroids. 2014;90: 13–29.

97. Lin HK, Oltenacu PA, Van Vleck LD, Erb HN, Smith RD. Heritabilities of and genetic correlations among six health problems in Holstein cows. J Dairy Sci. 1989;72: 180–186.

98. Björnström L, Sjöberg M. Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. Mol Endocrinol. 2005;19: 833–842.

99. Song X, Wei Z, Shaikh ZA. Requirement of ERα and basal activities of EGFR and Src kinase in Cd-induced activation of MAPK/ERK pathway in human breast cancer MCF-7 cells. Toxicol Appl Pharmacol. 2015;287: 26–34.

100. Iorns E, Lord CJ, Ashworth A. Parallel RNAi and compound screens identify the PDK1 pathway as a target for tamoxifen sensitization. Biochem J. 2009;417: 361–370.

101. Mayayo-Peralta I, Prekovic S, Zwart W. Estrogen Receptor on the move: Cistromic plasticity and its implications in breast cancer. Mol Aspects Med. 2021;78: 100939.

102. Lopes R, Korkmaz G, Revilla SA, van Vliet R, Nagel R, Custers L, et al. CUEDC1 is a primary target of ERα essential for the growth of breast cancer cells. Cancer Lett. 2018;436: 87–95.

103. Fanning SW, Mayne CG, Dharmarajan V, Carlson KE, Martin TA, Novick SJ, et al. Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. Elife. 2016;5. doi:10.7554/eLife.12792

104. Jia S, Miedel MT, Ngo M, Hessenius R, Chen N, Wang P, et al. Clinically Observed Estrogen Receptor Alpha Mutations within the Ligand-Binding Domain Confer Distinguishable Phenotypes. Oncology. 2018;94: 176–189.

105. Piperigkou Z, Karamanos NK. Estrogen receptor-mediated targeting of the extracellular matrix network in cancer. Semin Cancer Biol. 2020;62: 116–124.

106. DeFriend DJ, Howell A, Nicholson RI, Anderson E, Dowsett M, Mansel RE, et al. Investigation of a new pure antiestrogen (ICI 182780) in women with primary breast cancer. Cancer Res. 1994;54: 408–414.

107. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst. 1998;90: 1371–1388.

108. Harvey JM, Clark GM, Osborne CK, Allred DC. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol. 1999;17: 1474–1481.

109. Shen F, Gao Y, Ding J, Chen Q. Is the positivity of estrogen receptor or progesterone receptor different between type 1 and type 2 endometrial cancer? Oncotarget. 2017;8: 506–511.

110. Fiorica JV, Brunetto VL, Hanjani P, Lentz SS, Mannel R, Andersen W, et al. Phase II trial of alternating courses of megestrol acetate and tamoxifen in advanced endometrial carcinoma: a Gynecologic Oncology Group study. Gynecol Oncol. 2004;92: 10–14.

111. LE Donne M, Alibrandi A, Ciancimino L, Azzerboni A, Chiofalo B, Triolo O. Endometrial pathology in breast cancer patients: Effect of different treatments on ultrasonographic, hysteroscopic and histological findings. Oncol Lett. 2013;5: 1305–1310.

112. Held JM, Britton DJ, Scott GK, Lee EL, Schilling B, Baldwin MA, et al. Ligand binding promotes CDK-dependent phosphorylation of ER-alpha on hinge serine 294 but inhibits ligand-independent phosphorylation of serine 305. Mol Cancer Res. 2012;10: 1120–1132.

113. Razandi M, Pedram A, Park ST, Levin ER. Proximal events in signaling by

plasma membrane estrogen receptors. J Biol Chem. 2003;278: 2701–2712.

114. Sanchez AM, Flamini MI, Baldacci C, Goglia L, Genazzani AR, Simoncini T. Estrogen receptor-alpha promotes breast cancer cell motility and invasion via focal adhesion kinase and N-WASP. Mol Endocrinol. 2010;24: 2114–2125.

115. Wu XS, Vakoc CR. A pliable ERα cistrome evades therapy. Nature cell biology. 2020. pp. 619–620.

116. Chi D, Singhal H, Li L, Xiao T, Liu W, Pun M, et al. Estrogen receptor signaling is reprogrammed during breast tumorigenesis. Proc Natl Acad Sci U S A. 2019;116: 11437–11443.

117. Kumar V, Chambon P. The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. Cell. 1988;55: 145–156.

118. Klein-Hitpass L, Schorpp M, Wagner U, Ryffel GU. An estrogen-responsive element derived from the 5' flanking region of the Xenopus vitellogenin A2 gene functions in transfected human cells. Cell. 1986;46: 1053–1061.

119. Walker P, Germond JE, Brown-Luedi M, Givel F, Wahli W. Sequence homologies in the region preceding the transcription initiation site of the liver estrogen-responsive vitellogenin and apo-VLDLII genes. Nucleic Acids Res. 1984;12: 8611–8626.

120. Lin C-Y, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, et al. Whole-genome cartography of estrogen receptor alpha binding sites. PLoS Genet. 2007;3: e87.

121. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. Mol Cell. 2013;52: 25–36.

122. Kushner PJ, Agard DA, Greene GL, Scanlan TS, Shiau AK, Uht RM, et al. Estrogen receptor pathways to AP-1. J Steroid Biochem Mol Biol. 2000;74: 311–317.

123. Porter W, Saville B, Hoivik D, Safe S. Functional synergy between the transcription factor Sp1 and the estrogen receptor. Mol Endocrinol. 1997;11: 1569–1580.

124. Joseph R, Orlov YL, Huss M, Sun W, Kong SL, Ukil L, et al. Integrative model of genomic factors for determining binding site selection by estrogen receptor-α. Mol Syst Biol. 2010;6: 456.

125. Bojcsuk D, Nagy G, Bálint BL. Alternatively Constructed Estrogen Receptor Alpha-Driven Super-Enhancers Result in Similar Gene Expression in Breast and Endometrial Cell Lines. Int J Mol Sci. 2020;21. doi:10.3390/ijms21051630

126. Droog M, Nevedomskaya E, Kim Y, Severson T, Flach KD, Opdam M, et al. Comparative Cistromics Reveals Genomic Cross-talk between FOXA1 and ERα in

Tamoxifen-Associated Endometrial Carcinomas. Cancer Res. 2016;76: 3773–3784.

127. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. Genes Dev. 2014;28: 2679–2692.

128. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat Genet. 2011;43: 27–33.

129. Nakshatri H, Badve S. FOXA1 as a therapeutic target for breast cancer. Expert Opin Ther Targets. 2007;11: 507–514.

130. Theodorou V, Stark R, Menon S, Carroll JS. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. Genome Res. 2013;23: 12–22.

131. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012;481: 389–393.

132. Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. Cell. 2000;103: 843–852.

133. Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engström O, et al. Molecular basis of agonism and antagonism in the oestrogen receptor. Nature. 1997;389: 753–758.

134. Anzick SL, Kononen J, Walker RL, Azorsa DO, Tanner MM, Guan XY, et al. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science. 1997;277: 965–968.

135. Chakravarti D, LaMorte VJ, Nelson MC, Nakajima T, Schulman IG, Juguilon H, et al. Role of CBP/P300 in nuclear receptor signalling. Nature. 1996;383: 99–103.

136. Hanstein B, Eckner R, DiRenzo J, Halachmi S, Liu H, Searcy B, et al. p300 is a component of an estrogen receptor coactivator complex. Proc Natl Acad Sci U S A. 1996;93: 11540–11545.

137. Blanco JC, Minucci S, Lu J, Yang XJ, Walker KK, Chen H, et al. The histone acetylase PCAF is a nuclear receptor coactivator. Genes Dev. 1998;12: 1638–1651.

138. Varlakhanova N, Snyder C, Jose S, Hahm JB, Privalsky ML. Estrogen receptors recruit SMRT and N-CoR corepressors through newly recognized contacts between the corepressor N terminus and the receptor DNA binding domain. Mol Cell Biol. 2010;30: 1434–1445.

139. Fisher B, Costantino JP, Redmond CK, Fisher ER, Wickerham DL, Cronin WM. Endometrial cancer in tamoxifen-treated breast cancer patients: findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. J Natl Cancer Inst. 1994;86: 527–537.

140. Gottardis MM, Robinson SP, Satyaswaroop PG, Jordan VC. Contrasting actions of tamoxifen on endometrial and breast tumor growth in the athymic mouse. Cancer Res. 1988;48: 812–815.

141. Shang Y, Brown M. Molecular determinants for the tissue specificity of SERMs. Science. 2002;295: 2465–2468.

142. Mohammed H, Russell IA, Stark R, Rueda OM, Hickey TE, Tarulli GA, et al. Progesterone receptor modulates ERα action in breast cancer. Nature. 2015;523: 313–317.

143. Singhal H, Greene ME, Tarulli G, Zarnke AL, Bourgo RJ, Laine M, et al. Genomic agonism and phenotypic antagonism between estrogen and progesterone receptors in breast cancer. Sci Adv. 2016;2: e1501924.

144. Yang F, Ma Q, Liu Z, Li W, Tan Y, Jin C, et al. Glucocorticoid Receptor:MegaTrans Switching Mediates the Repression of an ERα-Regulated Transcriptional Program. Mol Cell. 2017;66: 321–331.e6.

145. Miranda TB, Voss TC, Sung M-H, Baek S, John S, Hawkins M, et al. Reprogramming the chromatin landscape: interplay of the estrogen and glucocorticoid receptors at the genomic level. Cancer Res. 2013;73: 5130–5139.

146. Cochrane DR, Bernales S, Jacobsen BM, Cittelly DM, Howe EN, D'Amato NC, et al. Role of the androgen receptor in breast cancer and preclinical analysis of enzalutamide. Breast Cancer Res. 2014;16: R7.

147. D'Amato NC, Gordon MA, Babbs B, Spoelstra NS, Carson Butterfield KT, Torkko KC, et al. Cooperative Dynamics of AR and ER Activity in Breast Cancer. Mol Cancer Res. 2016;14: 1054–1067.

148. Hickey TE, Selth LA, Chia KM, Laven-Law G, Milioli HH, Roden D, et al. The androgen receptor is a tumor suppressor in estrogen receptor-positive breast cancer. Nat Med. 2021;27: 310–320.

149. Need EF, Selth LA, Harris TJ, Birrell SN, Tilley WD, Buchanan G. Research resource: interplay between the genomic and transcriptional networks of androgen receptor and estrogen receptor α in luminal breast cancer cells. Mol Endocrinol. 2012;26: 1941–1952.

150. Peters AA, Buchanan G, Ricciardelli C, Bianco-Miotto T, Centenera MM, Harris JM, et al. Androgen receptor inhibits estrogen receptor-alpha activity and is prognostic in breast cancer. Cancer Res. 2009;69: 6131–6140.

151. Ponnusamy S, Asemota S, Schwartzberg LS, Guestini F, McNamara KM, Pierobon M, et al. Androgen Receptor Is a Non-canonical Inhibitor of Wild-Type and Mutant Estrogen Receptors in Hormone Receptor-Positive Breast Cancers. iScience. 2019;21: 341–358.

152. Yu Z, He S, Wang D, Patel HK, Miller CP, Brown JL, et al. Selective Androgen

Receptor Modulator RAD140 Inhibits the Growth of Androgen/Estrogen Receptor-Positive Breast Cancer Models with a Distinct Mechanism of Action. Clin Cancer Res. 2017;23: 7608–7620.

153. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534: 47–54.

154. Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W. Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet. 2014;46: 1160–1165.

155. Yang J, Wei X, Tufan T, Kuscu C, Unlu H, Farooq S, et al. Recurrent mutations at estrogen receptor binding sites alter chromatin topology and distal gene expression in breast cancer. Genome Biol. 2018;19: 190.

156. Welboren W-J, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FC, Span PN, et al. ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. EMBO J. 2009;28: 1418–1428.

157. Carleton JB, Berrett KC, Gertz J. Multiplex Enhancer Interference Reveals Collaborative Control of Gene Regulation by Estrogen Receptor α-Bound Enhancers. Cell Syst. 2017;5: 333–344.e5.

158. Zhang Y, Chan HL, Garcia-Martinez L, Karl DL, Weich N, Slingerland JM, et al. Estrogen induces dynamic ERα and RING1B recruitment to control gene and enhancer activities in luminal breast cancer. Sci Adv. 2020;6: eaaz7249.

159. Carleton JB, Ginley-Hidinger M, Berrett KC, Layer RM, Quinlan AR, Gertz J. Regulatory sharing between estrogen receptor α bound enhancers. Nucleic Acids Res. 2020;48: 6597–6610.

160. Orlando V, Strutt H, Paro R. Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods. 1997;11: 205–214.

161. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell. 2005;122: 33–43.

162. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, et al. Genome-wide analysis of estrogen receptor binding sites. Nat Genet. 2006;38: 1289–1297.

163. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell. 2008;132: 958–970.

164. Singh AA, Schuurman K, Nevedomskaya E, Stelloo S, Linder S, Droog M, et al. Optimized ChIP-seq method facilitates transcription factor profiling in human tumors. Life Sci Alliance. 2019;2: e201800115.

165. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of

native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10: 1213–1218.

166. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet. 2019;20: 207–220.

167. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15: 272–286.

168. Muerdter F, Boryń ŁM, Arnold CD. STARR-seq - principles and applications. Genomics. 2015;106: 145–150.

169. Muerdter F, Boryń ŁM, Woodfin AR, Neumayr C, Rath M, Zabidi MA, et al. Resolving systematic errors in widely used enhancer activity assays in human cells. Nat Methods. 2018;15: 141–149.

170. Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LTM, Fernandez N, et al. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. Nat Commun. 2015;6: 6905.

171. Vockley CM, Guo C, Majoros WH, Nodzenski M, Scholtens DM, Hayes MG, et al. Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. Genome Res. 2015;25: 1206–1214.

172. Zhang P, Xia J-H, Zhu J, Gao P, Tian Y-J, Du M, et al. High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing. Nat Commun. 2018;9: 2022.

173. Rathert P, Roth M, Neumann T, Muerdter F, Roe J-S, Muhar M, et al. Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. Nature. 2015;525: 543–547.

174. Shlyueva D, Stelzer C, Gerlach D, Yáñez-Cuna JO, Rath M, Boryń ŁM, et al. Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. Mol Cell. 2014;54: 180–192.

175. Vockley CM, D'Ippolito AM, McDowell IC, Majoros WH, Safi A, Song L, et al. Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. Cell. 2016;166: 1269–1281.e19.

176. Johnson GD, Barrera A, McDowell IC, D'Ippolito AM, Majoros WH, Vockley CM, et al. Human genome-wide measurement of drug-responsive regulatory activity. Nat Commun. 2018;9: 5317.

177. Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, et al. Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell. 2013;154: 914–927.

178. Pan YF, Wansa KDSA, Liu MH, Zhao B, Hong SZ, Tan PY, et al. Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved

distal response elements. J Biol Chem. 2008;283: 32977–32988.

179. Eeckhoute J, Carroll JS, Geistlinger TR, Torres-Arzayus MI, Brown M. A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. Genes Dev. 2006;20: 2513–2526.

180. Kim TK, Eberwine JH. Mammalian cell transfection: the present and the future. Anal Bioanal Chem. 2010;397: 3173–3178.

181. Woo AC, Faure L, Dapa T, Matic I. Heterogeneity of spontaneous DNA replication errors in single isogenic Escherichia coli cells. Sci Adv. 2018;4: eaat1608

182. Filges S, Yamada E, Ståhlberg A, Godfrey TE. Impact of Polymerase Fidelity on Background Error Rates in Next-Generation Sequencing with Unique Molecular Identifiers/Barcodes. Sci Rep. 2019;9: 3503