Identifying Functional Nuclear Receptor Signaling in Estrogen Receptor-Positive Breast Cancer

A Thesis
Presented to
The Academic Faculty

By

Jonathan Mitchel

In Partial Fulfillment
of the Requirements for the Degree in
Biomedical Engineering in the
College of Engineering, Georgia Institute of Technology

Georgia Institute of Technology
May 2019

Identifying Functional Nuclear Receptor Signaling in Estrogen Receptor-Positive Breast Cancer

Approved by:

Dr John McDonald, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Jeffrey Skolnick
School of Biological Sciences
Georgia Institute of Technology

Date Approved: April 24, 2019

Table of Contents

Abstract

We developed a microarray data analytical method capable of identifying ligand-activated nuclear receptors (NRs) in samples of the MCF7 cell line that are cultured among uncharacterized ligands. Principally, we applied this novel technique to reanalyze microarray data from an MCF7 xenograft experiment to reveal NR signaling that is likely the result of small molecule ligands available in the tumor microenvironment. Furthermore, this method was applied to query NR signaling caused by the small molecules present in fetal bovine serum to identify receptors stimulated by this common culture method. These analyses unveiled the potentially important roles of NR5A1 and NR1H4 in the progression of estrogen receptor-positive breast cancer. Preliminary experiments demonstrate that these receptors may play a role in mediating the transcriptional regulatory action of the estrogen receptor itself, potentially opening novel avenues for the development of more efficacious adjuvant therapies.

Introduction

Nuclear receptors (NRs) are a unique superfamily of proteins that regulate transcription of target genes in response to binding by small molecule ligands. This feature allows NRs to regulate expression programs in response to environmental, metabolic, or hormonally derived small molecules (Sever & Glass, 2013). Therefore, it is not surprising that malfunctional NR signaling can lead to a variety of diseases, including cancer. Due to the presence of ligand binding domains in these receptors, NRs are viewed as easily druggable targets; in fact, NRs already represent 13% of all FDA approved drugs (Whitby et al., 2011). Therefore, by identifying all functional NR activity in a cancer subtype of interest we will be able to repurpose existing NR-targeting drugs to those receptors contributing to any oncogenic transcriptional regulatory programs.

Nuclear Receptors in Breast Cancer:

In breast cancer, 70% of primary tumors are reported to express high levels of the NRs estrogen receptor (ER) and progesterone receptor (PR) (Lumachi, Brunello, Maruzzo, Basso, & Basso, 2013). ER is implicated as one of the primary drivers of this disease, activating pathways involving cell survival, cell cycle, and proliferation when bound by its endogenous ligand, estradiol (E2). Over the last 30 years, treatment with ER modulators such as Tamoxifen or aromatase inhibitors such as Letrozole have been commonly used as adjuvant therapy to prevent tumor relapse (Chang, 2012). These drugs have proven incredibly effective, increasing progression-free survival often upwards of twenty years with relatively minimal side effects. However, nearly 40%-50% of ER/PR-positive breast cancer patients are not initially sensitive to anti-estrogen therapies. Furthermore, many patients develop aquired resistance, eventually

rendering such treatments ineffective (Conzen, 2008). Despite the large amount of research into two receptors, there are only a handful of investigations querying the action of the other 46 NRs. In a review by Garattini et al., the authors propose lists of NRs that may have oncogenic or oncosuppressive function in different breast cancer subtypes (Garattini et al., 2016). In a large-scale CHIP-seq study of NR signaling in MCF7, Kittler et al. was able to identify ligand induced NR-DNA complexes and uncovered vast cross-talk between these receptors (Kittler et al., 2013). These studies, while beneficial to understanding the general functionality of such receptors, are limited in their ability to describe the action of these receptors in vivo, where uncharacterized ligands/signals from the microenvironment likely dictate NR regulatory behavior. The work presented here attempts to address this issue by (a) developing a predictive model of NR signaling from baseline transcript levels in MCF7 cells and (b) applying this model to reveal endogenously activated NRs in vivo.

Nuclear Receptor Regulatory Mechanisms:

To construct such a model, it is important to understand the underlying biological processes that lead from initial small molecule induction of an NR to altered target transcript levels. Within the NR superfamily, there are two distinct subtypes of receptors. The first consists of the steroid-hormone NRs. These generally reside in the cytoplasm bound to heat shock proteins (HSPs) in the absence of ligand. When ligand binds intracellularly, they dissociate from the HSPs, enter the nucleus as homodimers, complex with cofactors, and bind to DNA response elements to regulate transcription. The second type of NRs remain in the nucleus as heterodimers with the retinoid x receptor (RXR). Without ligand present, these complexes are bound to their

response elements and repress transcription of target genes. When ligand binds, they dissociate from the DNA, bind to cofactors, and activate a subset of their target genes.

While one may naively assume that a NR agonist will impact the transcriptional regulation of all target genes having the corresponding response elements, this is almost never the case. Instead, NRs typically regulate a subset of all potential target genes, and this selectivity is governed by several factors. Some of these factors include: the molecular structure of the binding ligand, levels of coactivator and corepressor proteins, target gene methylation status, target gene chromatin structure, growth factor signaling crosstalk, and levels of the endogenous ligand (Long & Campbell, 2015). To further add to the complexity, multiple studies have shown that TF-DNA complexes are often non-functional, leading many to question the practice of identifying target genes by promoter response elements alone (Wu & Lai, 2015; Yang & Wu, 2013). As a result, the nature of NR-signaling results in wide variation in control of gene expression across and within cell types.

Methods and Materials

Model Structure:

To predict the Boolean state of all NRs (ligand-activated or inactive) in a given sample, it was assumed that ligand-activated NRs would have some subset of their target genes differentially expressed between the sample in question and samples of the same cell line where all NR-binding ligands are absent (Figure 1A). To compute differential expression of all target genes for each NR, it was first necessary to construct distributions of NR target gene levels in the unliganded state. To accomplish this, we aquired microarray data for 64 MCF7 samples cultured in charcoal-stripped FBS or no serum with phenol-red free media. The charcoal-stripping process removes any hydrophobic ligands that may enter the cell and activate NRs. In this state it can be assumed that NR signaling is weak for the steroid-hormone binding NRs and repressive for the RXR heterodimer forming NRs that are present (Figure 1B). The raw CEL files for these samples were preprocessed using the frozen RMA (fRMA) method to background correct, normalize, and summarize all probes (McCall, Bolstad, & Irizarry, 2010). The fRMA method was selected because it can process single arrays and because it reduces the impact of batch effects by removing bad probes. After preprocessing, the means and standard deviations for all probe sets across the 64 serum-starved samples were calculated. The distributions for most probe sets were approximately normally distributed as can be seen in a few example distributions below (Figure 2B).

At this point, differentially expressed probe sets were identified by computing z-scores between the sample of interest and the previously generated distributions. Gene-level expression information was extracted by taking the maximum absolute value z-score among all probe sets for a given gene. Next, gene set enrichment analysis (GSEA) was used to analyze

overrepresentation of NR target gene sets among all probe sets with z-scores above a given threshold. The GSEA was calculated with a hypergeometric test, and a false discovery rate (FDR) correction was applied. NR target gene sets were downloaded from the TTRUST version 2 database (Han et al., 2015). This database was used because its annotations are derived directly from research articles and are not predictions based solely from promoter sequences. Therefore, these annotations are less likely to contain false positives that may dilute enrichment calculations downstream in the analysis. As a note, NR-target gene sets with less than 15 genes were excluded from the analysis as is common practice for GSEA. The microarray platform for all these samples was the Affymetrix GeneChip Human Genome U133 Plus 2.0, and all data was aquired from the Gene Expression Omnibus (GEO). A list of the samples used can be found in the supplemental information.

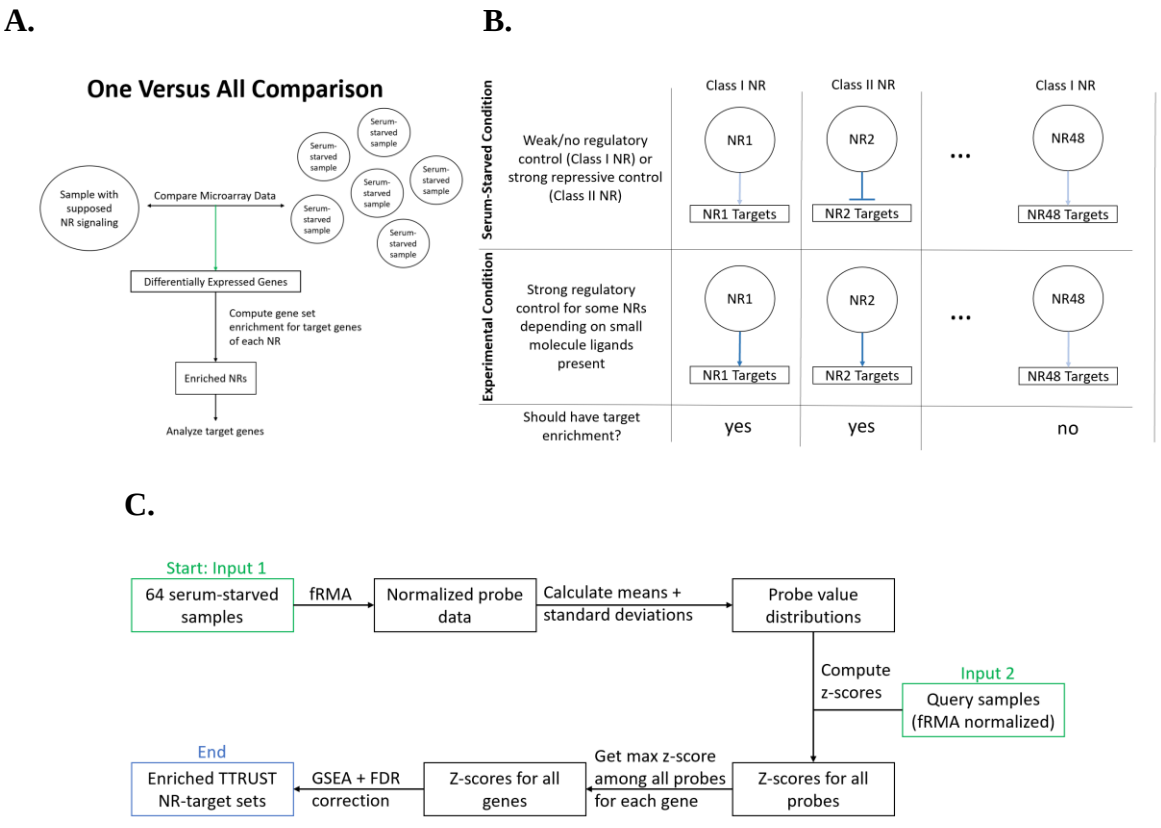

Figure 1. A model for characterizing sample-specific nuclear receptor (NR) signaling. (A) High-level overview of the model design. (B) Logic and assumptions regarding NR regulation in presence/absence of NR binding ligands. Dark blue arrows

indicate strong transcriptional control of target genes, whereas light blue indicates weak regulation. (C) Computational pipeline for this analysis that was applied to MCF7 microarray data.

Model Validation:

To validate the accuracy of the model in identifying ligand-activated NRs, microarray data from 12 experiments in the MCF7 cell line were obtained. These experiments contain samples that were serum-starved with no treatment and samples that were serum-starved followed by treatment with E2. Principal components analysis was used to verify that treatment with E2 caused distinct changes in ER target gene levels and that fRMA reduced batch effects to an acceptable level (Figure 2A). Specifically, principal components were generated from the top 20 probe sets of ER target genes that had the lowest p-values after running Student's t-tests.

After verifying that fRMA sufficiently eliminated batch effects while preserving the anticipated ER target gene expression differences between groups, z-scores were computed for the ER target gene probes as described previously. A visualization of this process can be seen in Figure 2C below for one such experiment. Next, we evaluated whether differentially expression genes identified by z-scores were also differentially expressed in each sample's respective controlled experiment. For the controlled case, an FDR less than 0.05 and a fold change (FC) greater than 1.5 was considered differentially expressed. Finally, the proportion of target gene probe sets above a moving z-score threshold was calculated to ensure that this method of differential expression functions properly.

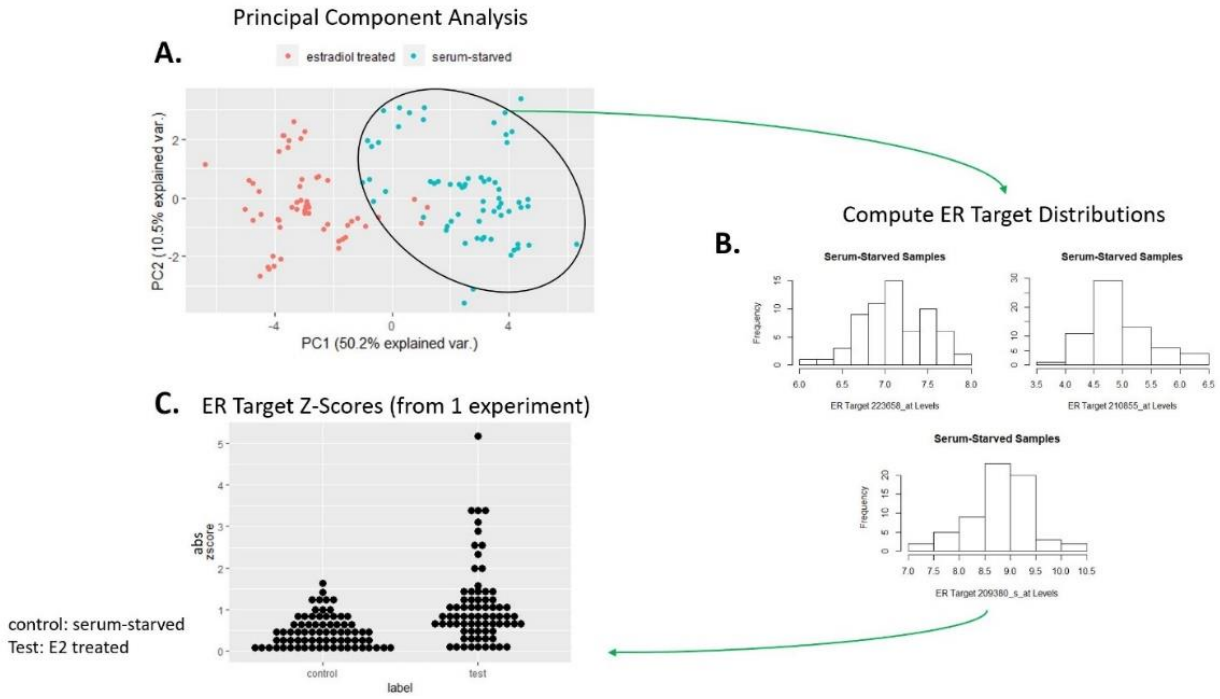

Figure 2. Applying model to estradiol (E2) treated MCF7 samples for validation. (A) PCA of E2 treated MCF7 cells, showing distinct separation between groups based on ER target gene expression values. (B) Distributions calculated for all ER target gene probes (only 3 probes shown here). (C) The resulting differentially expressed genes contain high z-scores, which are present selectively in the E2 treated sample but not in the serum-starved sample.

Next, GSEA for sets of TTRUST listed target genes of all 48 NRs are computed as described above. For the E2 treated samples, precision and recall were calculated to verify that most experiments resulted in selective enrichment of ER target genes and not target genes of other NRs.

Querying NR Signaling in FBS-Cultured and Xenograft Samples:

After validating the predictive power of this method, the pipeline was run to identify active NR signaling in MCF7 mouse xenografts as well as in FBS-cultured samples. A total of 17 MCF7 xenograft samples from a study by Hollingshead et al. were used in this analysis. As was performed in this original study, we removed probe sets from the human microarray chip that also hybridized with RNA from 5 different mouse cell types (Hollingshead et al., 2014).

This ensures the downstream analysis is only of RNA from the MCF7 cells as opposed to stromal cells or fibroblasts that may have been included in the harvested tumors.

In the second case, we analyzed 19 FBS cultured MCF7 samples from a variety of different experiments. FBS is used ubiquitously in many cell culture protocols, and it contains a wide variety of hormones, growth factors, vitamins, lipids, and sugars, many of which are also found in the human body (van der Valk et al., 2010). Therefore, it will be useful to know if which NRs, if any, become active from binding by these small molecules. In a similar procedure as before, z-scores were generated for genes of all NR targets in all samples. Again, using GSEA we identified NR-target gene sets that were overrepresented in the list of genes that had a maximum absolute value z-score greater than 2. Finally, the counts of enriched NRs in all the samples were plotted in a bar graph.

Testing NR1H4 and NR5A1 Agonists:

MCF7 cells were plated in 96 well plates at a seeding density of 10,000 cells per well in serum-free media. The cells remained in serum-free conditions for approximately 24 hours before treatment. Media containing phenol-red was used throughout to mitigate the effects reduced estrogen levels have on cell proliferation. The NR1H4 agonist, GW4064, and the NR5A1 inverse agonist 4-(Heptyloxy)phenol (4HP) were prepared in DMSO. After 48 hours of treatment, cell viability assays were completed using the Tox-8 resazurin based assay, and results are reported as a percentage of .3% DMSO control viability.

Analyzing Effect of GW4064 on MCF7 Cell Morphology:

ImageJ plugins from MorphoLibJ were used to compute a measure of cell elongation using images taken with a brightfield microscope. Steps used for this process include enhancing the image, segmenting by borders, and setting the size opening to remove small artifacts. Circularity was calculated as the measure of cell roundness, which is the inverse of elongation. A Student's t-test was used to evaluate whether there was a difference in the means of the GW4064-treated and untreated groups.

Microarray Analysis:

MCF7 cells were plated in 6 well plates at a seeding density of 400,000 cells per well. As done previously, the cells were grown in serum-free media for approximately 24 hours before treatment. Cells were treated with 9 μM GW4064 for 24 hours and then were harvested for microarray analysis. Three biological replicates were included in both the test and control groups. The Affymetrix GeneChip Human Genome U133 Plus 2.0 was used. The data quality was evaluated using global normalized unscaled standard error (GNUSE) (Supplementary Information) (McCall, Murakami, Lukk, Huber, & Irizarry, 2011). The data were preprocessed using Robust Multi-Array Averaging (RMA) and differentially expressed genes were identified using the Limma package in R (Hobbs et al., 2003). Genes with an FDR corrected p-value less than 0.05 and log2 fold-change greater than 1.5 were considered differentially expressed. GSEA was run on the outputted differentially expressed genes using fast preranked GSEA R package, fgsea (Sergushichev, 2016). MSigDB collections used for this analysis include: H, C2 (subcategories CP:REACTOME, CP:KEGG, CP:BIOCARTA, CP), and C5 (subcategory BP).

The resulting list of pathways was collapsed to highlight pathways containing independent sets of genes.

RT-qPCR:

To assess whether NR5A1 is inducible in the MCF7 cell line, we measured mRNA levels of three NR5A1 target genes in samples treated with or without NR5A1 inverse agonist, 4-(Heptyloxy)phenol, at 12.5 μM concentration. Cells were initially serum-starved for 24 hours; after 24 hours of treatment, cells were harvested for RT-qPCR. The target genes measured include CYP19A1, CYP11B1, and STAR. GAPDH was used as a reference gene. Forward and reverse primers were purchased from Sigma-Aldrich and are MIQE approved. Exact sequences for the forward and reverse primers can be found in the supplemental information. Delta Ct values were calculated as the difference between the Ct for the target and the Ct of GAPDH for each sample. Two biological replicates were generated for each group.

Results

Model Validation:

In E2 treated MCF7 samples, probes that resulted in absolute value z-scores greater than 2 were differentially expressed in their respective controlled experiments, with the median proportion at 100%. This result can be viewed in Figure 3 below.

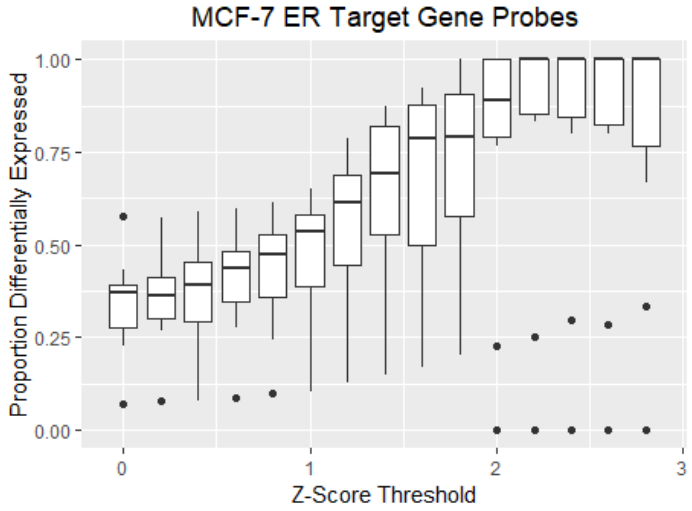

Figure 3. Proportions of differentially expressed probes at varying z-score thresholds for 12 experiments of MCF7 cells with serum-starved control versus serum-starved + estradiol treatment groups.

Enrichment calculations for overrepresentation of NR target gene sets amongst E2 induced genes with absolute value z-score greater than 2 shows that this model has a precision of 93% and a recall of 52% at an FDR cutoff of 0.1 (Figure 4). Androgen receptor (AR) and NR1I2 were excluded from these calculations because previous reports suggest E2 can directly bind these proteins, though likely with lower affinity compared to ER (Mnif et al., 2007; Yeh, Miyamoto, Shima, & Chang, 1998).

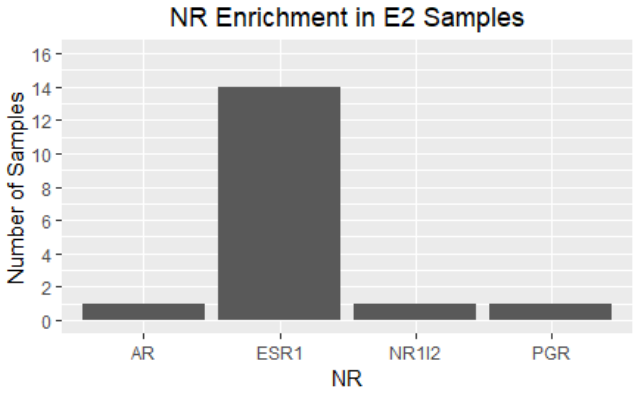

Figure 4. Model enrichment precision and recall using FDR cutoff of 0.1. There were 27 E2-treated samples total (replicates were averaged). Precision: 14/15 = 93%, recall: 14/27 = 52%

NR Signaling in FBS Cultured and Xenograft Samples:

Computing enriched NR target gene sets in MCF7 samples cultured in FBS resulted in multiple samples showing AR and ER signaling with one sample appearing to exhibit NR5A1 signaling (Figure 5A). The xenograft samples displayed a wider variety of NR signaling. Notably, NR1H4 was an unexpected result that showed up in high frequency (Figure 5B).

A. **B.**

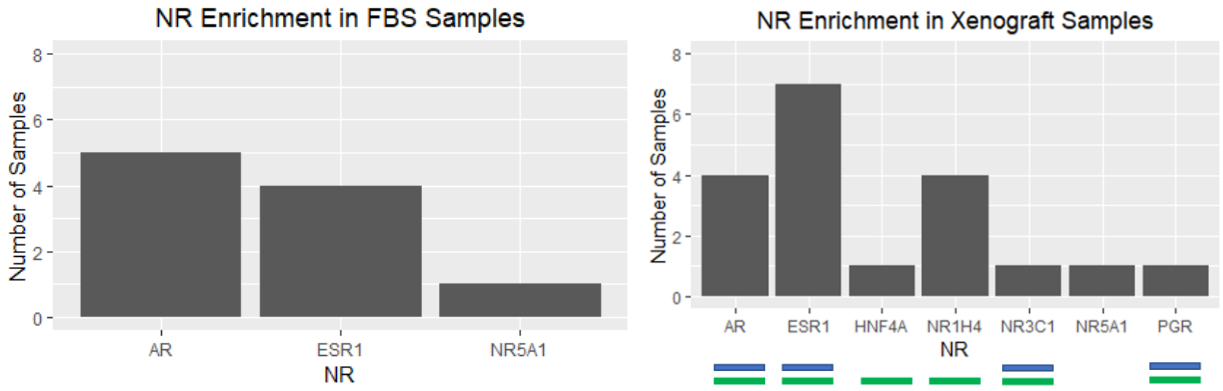

Figure 5. Resulting enriched nuclear receptors in MCF7 samples in (A) FBS cultured condition and (B) xenografts. Cutoff for enrichment was the same as used in model validation with FDR < 0.1. The blue marks in (B) denote previous literature validation from CHIP-seq of each respective NR in MCF7. Green marks denote literature evidence of a previous Western blot.

Next, the target genes that contributed to enrichment of each NR were plotted for each sample (Figure 6).

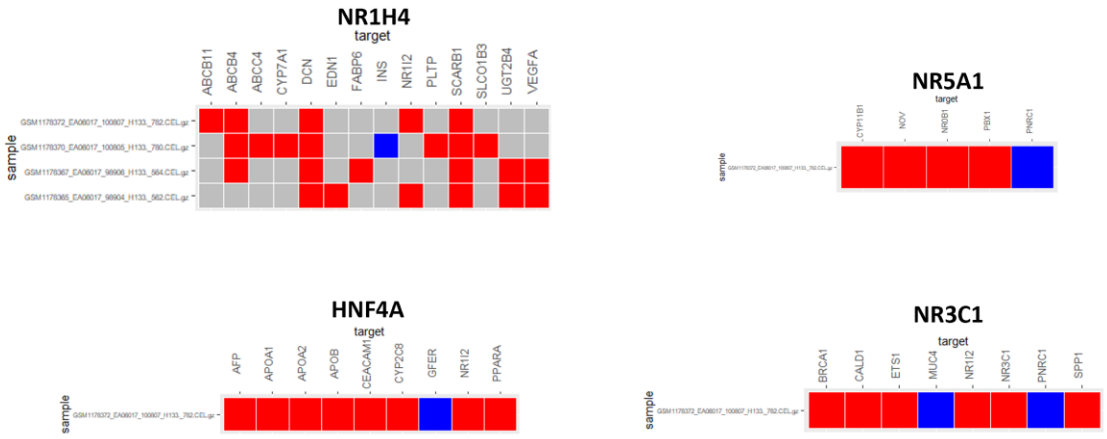

Figure 6. Target genes contributing to enrichment of the respective nuclear receptors in the xenografted MCF7 samples.

Since enrichment was found for NR1H4 target genes in multiple xenograft samples, we wanted to further investigate the role of this receptor. First, the selective NR1H4 agonist, GW4064, was used to treat the MCF7 cells. The effects of this drug on MCF7 cell viability can be viewed below (Figure 7). The IC50 concentration for this compound was found to be approximately 9 µM. It was also observed that culturing the MCF7 cells without FBS does not significantly affect their growth rate.

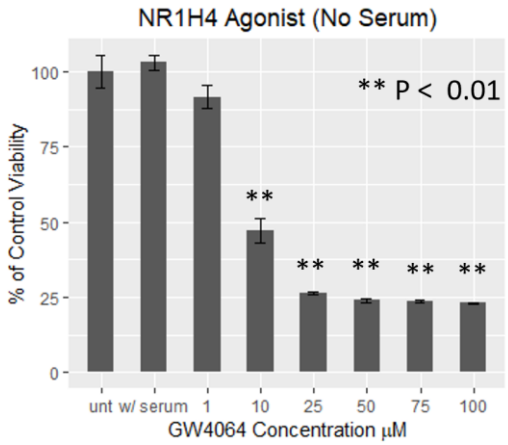

Figure 7. Effect of GW4064 on MCF7 cell viability using Tox-8 assay. IC50 concentration is 9 µM. Student's t-tests were used to determine statistical differences between the treatment groups and the untreated .3% DMSO control.

Next, we evaluated the effect of this NR1H4 agonist on the morphology of the MCF7 cells. As can be seen in Figure 8, treatment with the IC50 concentration of GW4064 causes the MCF7 cells to become less elongated, as is indicated by the increased circularity.

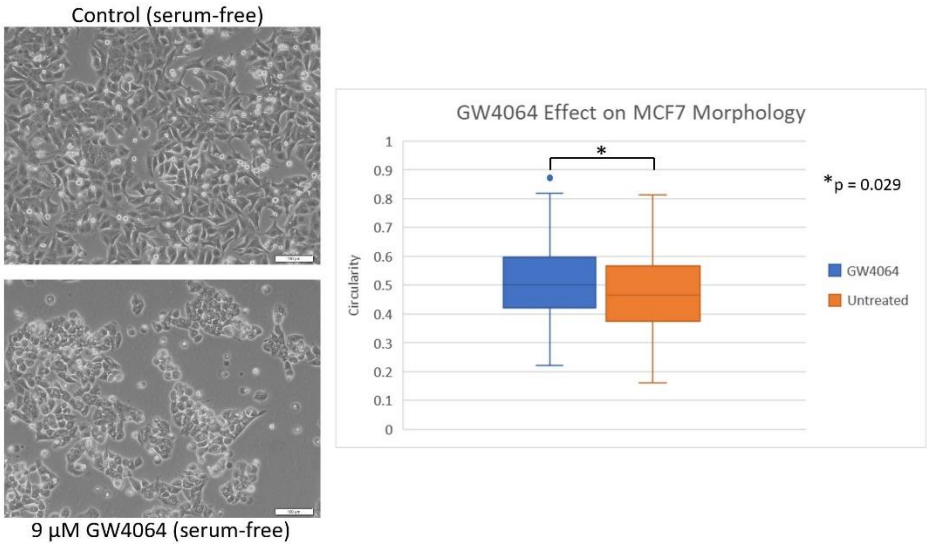

Figure 8. Effect of 9 µM GW4064 on MCF7 cell morphology. Circularity was measured using ImageJ. Student's t-tests were used to determine statistical differences between the treatment groups and the untreated. n^{GW4064} = 101 cells. n^{unt} = 198 cells.

When comparing gene expression levels between GW4064-treated and untreated cells from the microarray experiment, around 1000 genes were found to be differentially expressed (Supplementary Information). The top enriched pathways from running fgsea on the ranked list of differentially expressed genes can be found in Figure 9 below.

Pathway	Gene ranks	NES	pval	padj
GO_MITOTIC_NUCLEAR_DIVISION		1.99	2.3e-04	2.8e-02
GO_CELL_CYCLE_G2_M_PHASE_TRANSITION		1.94	2.2e-04	2.8e-02
GO_RESPONSE_TO_TOPOLOGICALLY_INCORRECT_PROTEIN		1.87	2.2e-04	2.8e-02
GO_TRANSLATIONAL_TERMINATION		1.83	2.2e-04	2.8e-02
HALLMARK_APOPTOSIS		1.81	2.2e-04	2.8e-02
GO_PROTEIN_FOLDING		1.78	2.2e-04	2.8e-02
HALLMARK_E2F_TARGETS		1.76	2.2e-04	2.8e-02
REACTOME_CELL_CYCLE_MITOTIC		1.75	2.3e-04	2.8e-02
HALLMARK_SPERMATOGENESIS		1.74	2.2e-04	2.8e-02
HALLMARK_MTORC1_SIGNALING		1.72	2.2e-04	2.8e-02
HALLMARK_G2M_CHECKPOINT		1.71	2.2e-04	2.8e-02
GO_POSITIVE_REGULATION_OF_PROTEIN_COMPLEX_ASSEMBLY		1.66	2.2e-04	2.8e-02
GO_PROTEIN_LOCALIZATION_TO_ORGANELLE		1.56	2.4e-04	2.8e-02
GO_PROTEASOMAL_PROTEIN_CATABOLIC_PROCESS		1.55	2.3e-04	2.8e-02
GO_AMIDE_BIOSYNTHETIC_PROCESS		1.45	2.4e-04	2.8e-02
GO_GLYCOPROTEIN_METABOLIC_PROCESS		-1.72	1.8e-04	2.8e-02
NABA_CORE_MATRISOME		-1.78	1.8e-04	2.8e-02
REACTOME_SIGNALING_BY_NOTCH		-1.80	1.8e-04	2.8e-02
GO_LIPID_PHOSPHORYLATION		-1.89	1.8e-04	2.8e-02
GO_MULTICELLULAR_ORGANISM_METABOLIC_PROCESS		-2.05	1.8e-04	2.8e-02
HALLMARK_ESTROGEN_RESPONSE_EARLY		-2.19	1.8e-04	2.8e-02

Figure 9. Top enriched pathways from results of microarray experiment. NES is the normalized enrichment score. The adjusted p-values are computed using an FDR correction.

In addition to NR1H4, we decided to further investigate NR5A1 since it was an unexpected result identified in one FBS sample and one xenograft sample. To assess the inducibility of NR5A1 in the MCF7 cell line, the effect of a selective NR5A1 inverse agonist, 4HP, was evaluated on cell proliferation. As seen in the figure below, 4HP reduced cell viability with an IC50 concentration between 10 µM and 100 µM (Figure 10).

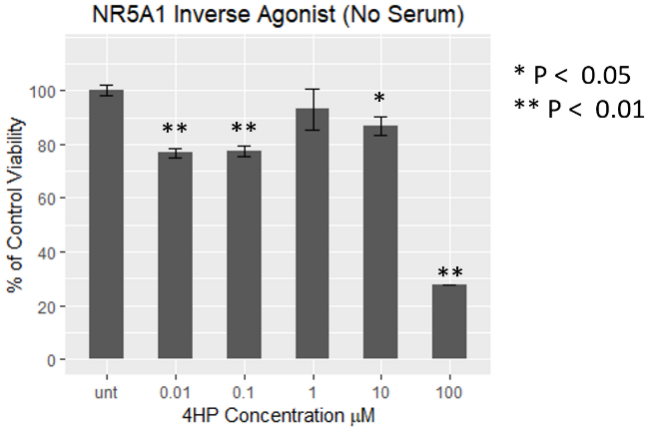

Figure 10. Effect of 4-(Heptyloxy)phenol on MCF7 cell viability using Tox-8 assay. IC50 concentration is approximately 40 µM. Student's t-tests were used to determine statistical differences between the treatment groups and the untreated control.

To gather more evidence that NR5A1 is present and can be pharmaceutically modulated, RT-qPCR was used to measure changes in levels of the NR5A1 target genes CYP19A1, CYP11B1, and STAR. Both STAR and CYP19A1 had increased ΔCt values in the 4HP-treated samples. (Figure 11). This indicates a reduction in expression of these two genes as a result of treatment.

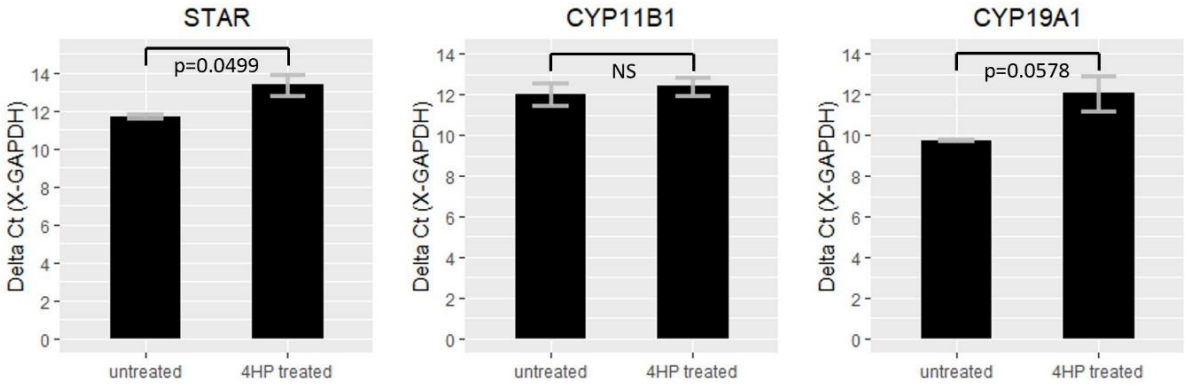

Figure 10. RT-qPCR experiment showing the effect of 12.5 µM 4-(Heptyloxy)phenol (4HP) on expression of NR5A1 target genes in the MCF7 cell line. The relative mRNA level was calculated using the ΔΔCt method, with GAPDH as a reference gene. Student's t-tests were used to determine statistical differences between the ΔCt values of the 4HP treated samples and the untreated samples.

Discussion

Model validation using MCF7 cells treated with or without E2 demonstrates (a) that probes with absolute value z-scores above 2 are mostly differentially expressed and (b) that using this threshold, the model predicts NR signaling with high precision and reasonable recall. Such high precision allowed us to confidently query NR signaling in FBS cultured cells and in mouse xenografts.

While one could generally conduct an experiment comparing several FBS cultured samples to several serum-starved samples, this would likely not capture the vast heterogeneity in small molecule composition that varies considerably between FBS batches. Over the 19 FBS cultured samples, our results demonstrated enriched signaling for ER and AR, confirming that levels of estrogens and androgens in FBS are often high enough to saturate binding sites on these receptors. One particularly interesting result from this analysis was the observed enrichment of NR5A1, also referred to as Steroidogenic Factor 1 (SF-1). No previous literature has shown the expression levels of this receptor in MCF7, and at least one RT-qPCR experiment shows there may be low mRNA expression of this gene (D Clyne, Speed, Zhou, & R Simpson, 2002). To test the hypothesis that NR5A1 is present and inducible in this cell line, we treated the MCF7 cells with a selective NR5A1 inverse agonist and measured changes in the levels of several well established NR5A1 target genes using RT-qPCR. In previous reports, STAR was found to be robustly affected by changes in NR5A1 (Del Tredici et al., 2008; Utsunomiya et al., 2008). Here, we also saw a significant change in STAR expression. One other NR5A1 target gene of interest is CYP19A1, which encodes the aromatase enzyme. Aromatase catalyzes androgens to estrogens and is the target of aromatase inhibitors used to treat post-menopausal breast cancer patients. Here we found CYP19A1 to be potentially affected by 4HP with a resulting p-value of 0.0578.

This result merits further investigation into NR5A1 as a potential target for breast cancer and as a potentially confounding factor that may affect anti-estrogen therapies.

When applying our computational method to MCF7 mouse xenograft samples generated in a previous study, we identified several NRs as potentially ligand-bound in these samples. This result was likely due to the presence of endogenous small molecules present in the tumor microenvironment. Through this analysis, NR1H4, also referred to as the Farnesoid X Receptor (FXR), was identified as activated in 4 separate samples. Bile acids are the endogenous ligands for FXR and have been reported to be at high levels in breast cysts and serum of newly diagnosed breast cancer patients (Costarelli & Sanders, 2002). Therefore, bile acids were likely also present in the xenograft tumor microenvironment, causing the resulting enrichment of FXR target genes. While there has been contradictory evidence regarding the oncogenic or anti-oncogenic activity of FXR, treatment with FXR synthetic agonist, GW4064 has previously shown antiproliferative and apoptotic effects in MCF7 cells (Garattini et al., 2016). Several mechanisms were proposed for these observed anti-cancer effects, but the complete transcriptional response to selectively targeting FXR has yet be investigated (Swales et al., 2006). In our experiments, we verified the antiproliferative effect of this drug and conducted a microarray experiment to further probe the genome wide transcriptomic response to activating FXR. The results of this analysis show that this drug may impact the cancer cell's mitotic function, apoptotic function, extracellular matrix structure, and estrogen signaling. The impacted estrogen signaling was a particularly interesting result because it strengthens prior evidence that FXR directly interacts with ER (Journe et al., 2008). We also observed a measurable difference in cell morphology for MCF7 cells treated with or without GW4064. Specifically, this treatment decreased elongation of the cells, making for a more epithelial-like morphology. Such a change

may affect the cancer's metastatic potential. These results add to the existing body of evidence that FXR may indeed serve as a useful druggable target in breast cancer, and it gives us insight into the underlying mechanisms by which it may act.

The predictive computational method developed here has several benefits that extend beyond the scope of this study. First, it allows for the querying NR signaling in MCF7 samples that do not have a corresponding serum-starved control. Without this stringency, we can conduct meta-analyses on dozens of experiments that were conducted for other purposes besides NR signaling. This method can also help identify active NR signaling present in control samples that could potentially impact the outcome of a given treatment (i.e. a treatment that may impact hormone levels).

Despite its success, our computational model is limited in two ways. First, some NR gene sets do not meet the GSEA requirement for having at least 15 target genes, and therefore we cannot evaluate potential signaling for all NRs. In future work, we would reevaluate the model using one or more alternate transcription factor target gene databases so that more NRs can be queried. A second limitation of this method is that it requires a sizeable number of serum-starved samples for a given cell type to serve as a model of NR signaling in the unliganded condition. We used MCF7 because it had the largest number of these serum-starved samples in the public database. In future work, it will be useful to culture a variety of cell types in this manner and measure global gene expression levels. Furthermore, this method could eventually be extended to include serum-starved patient-derived cell lines and cell lines grown in 3D culture. This would lend even more insight into patient-specific NR signaling processes and would help establish a novel approach to developing personalized hormonal therapies for cancer. Finally, it would be

beneficial and relatively easy to develop a similar approach using RNA-Seq, since this is becoming the standard for high-throughput gene expression measurement.

Acknowledgements

I would like to thank Minati Satpathy for her guidance in experimental techniques as well as Lilya Matyunina for assisting with the microarray experiment. I would also like to thank John McDonald and Jeffrey Skolnick for their advice and mentorship throughout the project.

Supplemental Information

A list of all differentially expressed genes from the GW4064-treated MCF7 microarray experiment can be found here:

https://docs.google.com/spreadsheets/d/1o_3RoJ35J0eiIww53FP5qQwgQe_2P1UTeZNo-9EfXqU/edit?usp=sharing

Or access through attached file: DEG_list.xlsx

A list of all GEO samples used in this study can be found here:

https://docs.google.com/spreadsheets/d/1goAYcmG4jJXCWZX0Y-Eud7Ca-oUnfqDx57FR188Y9i0/edit?usp=sharing

Or access through attached file: GEO_list.xlsx

RT-qPCR Primer Sequences:

FH1_CYP11B1: ATCTTCCACTACACCATAGAAG

RH1_CYP11B1: GTGGATTTGAACATGACCTC

FH1_CYP19A1: GGTGAGAGAGACATAAAGATTG

RH1_CYP19A1: TTCAGGATAATGTTTGTCCC

FH1_STAR: GACAAAGTGATGAGTAAAGTGG

RH1_STAR: CAGCTCGTGAGTAATGAATG

Microarray Quality Control Results with GNUSE (all arrays pass):

	Array 1	Array 2	Array 3
Control Samples	0.9481537	0.9456906	0.9375703
Test Samples	0.9797399	0.9725803	0.9697121

References

Chang, M. (2012). Tamoxifen resistance in breast cancer. *Biomolecules & Therapeutics, 20*(3), 256-267. doi:10.4062/biomolther.2012.20.3.256

Conzen, S. D. (2008). Minireview: Nuclear receptors and breast cancer. *Molecular Endocrinology, 22*(10), 2215-2228. doi:10.1210/me.2007-0421

Costarelli, V., & Sanders, T. (2002). *Plasma bile acids and risk of breast cancer* (Vol. 156).

D Clyne, C., Speed, C., Zhou, J., & R Simpson, E. (2002). *Liver receptor homologue-1 (lrh-1) regulates expression of aromatase in preadipocytes* (Vol. 277).

Del Tredici, A. L., Andersen, C. B., Currier, E. A., Ohrmund, S. R., Fairbain, L. C., Lund, B. W., . . . Piu, F. (2008). Identification of the first synthetic steroidogenic factor 1 inverse agonists: Pharmacological modulation of steroidogenic enzymes. *Molecular Pharmacology, 73*(3), 900. doi:10.1124/mol.107.040089

Garattini, E., Bolis, M., Gianni, M., Paroni, G., Fratelli, M., & Terao, M. (2016). Lipid-sensors, enigmatic-orphan and orphan nuclear receptors as therapeutic targets in breast-cancer. *Oncotarget, 7*(27), 42661-42682. doi:10.18632/oncotarget.7410

Han, H., Shim, H., Shin, D., Shim, J. E., Ko, Y., Shin, J., . . . Lee, I. (2015). TRRUST: a reference database of human transcriptional regulatory interactions. *Scientific Reports, 5*, 11432-11432. doi:10.1038/srep11432

Hobbs, B., Collin, F., Antonellis, K. J., Irizarry, R. A., Speed, T. P., Scherf, U., & Beazer-Barclay, Y. D. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics, 4*(2), 249-264. doi:10.1093/biostatistics/4.2.249

Hollingshead, M. G., Stockwin, L. H., Alcoser, S. Y., Newton, D. L., Orsburn, B. C., Bonomi, C. A., . . . Collins, J. (2014). Gene expression profiling of 49 human tumor xenografts from in vitro culture through multiple in vivo passages--strategies for data mining in support of therapeutic studies. *BMC Genomics, 15*(1), 393-393. doi:10.1186/1471-2164-15-393

Journe, F., Laurent, G., Chaboteaux, C., Nonclercq, D., Durbecq, V., Larsimont, D., & Body, J.-J. (2008). Farnesol, a mevalonate pathway intermediate, stimulates MCF-7 breast cancer cell growth through farnesoid-X-receptor-mediated estrogen receptor activation. *Breast Cancer Research and Treatment, 107*(1), 49-61. doi:10.1007/s10549-007-9535-6

Kittler, R., Zhou, J., Hua, S., Ma, L., Liu, Y., Pendleton, E., . . . White, Kevin P. (2013). A comprehensive nuclear receptor network for breast cancer cells. *Cell reports, 3*(2), 538-551. doi:https://doi.org/10.1016/j.celrep.2013.01.004

Long, M. D., & Campbell, M. J. (2015). Pan-cancer analyses of the nuclear receptor superfamily. *Nuclear receptor research, 2*, 101182. doi:10.11131/2015/101182

Lumachi, F., Brunello, A., Maruzzo, M., Basso, U., & Basso, S. M. M. (2013). Treatment of estrogen receptor-positive breast cancer. *Current Medicinal Chemistry, 20*(5), 596-604. doi:http://dx.doi.org/10.2174/092986713804999303

McCall, M. N., Bolstad, B. M., & Irizarry, R. A. (2010). Frozen robust multiarray analysis (fRMA). *Biostatistics (Oxford, England), 11*(2), 242-253. doi:10.1093/biostatistics/kxp059

McCall, M. N., Murakami, P. N., Lukk, M., Huber, W., & Irizarry, R. A. (2011). Assessing affymetrix GeneChip microarray quality. *BMC bioinformatics, 12*, 137-137. doi:10.1186/1471-2105-12-137

Mnif, W., Pascussi, J.-M., Pillon, A., Escande, A., Bartegi, A., Nicolas, J.-C., . . . Balaguer, P. (2007). Estrogens and antiestrogens activate hPXR. *Toxicology Letters, 170*(1), 19-29. doi:https://doi.org/10.1016/j.toxlet.2006.11.016

Sergushichev, A. A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*, 060012. doi:10.1101/060012

Sever, R., & Glass, C. K. (2013). Signaling by nuclear receptors. *Cold Spring Harbor Perspectives in Biology, 5*(3), a016709. doi:10.1101/cshperspect.a016709

Swales, K. E., Korbonits, M., Carpenter, R., Walsh, D. T., Warner, T. D., & Bishop-Bailey, D. (2006). The farnesoid x receptor is expressed in breast cancer and regulates apoptosis and aromatase expression. *Cancer research, 66*(20), 10120.

Utsunomiya, H., Cheng, Y.-H., Lin, Z., Reierstad, S., Yin, P., Attar, E., . . . Bulun, S. E. (2008). Upstream stimulatory factor-2 regulates steroidogenic factor-1 expression in endometriosis. *Molecular endocrinology (Baltimore, Md.), 22*(4), 904-914. doi:10.1210/me.2006-0302

van der Valk, J., Brunner, D., De Smet, K., Fex Svenningsen, Å., Honegger, P., Knudsen, L. E., . . . Gstraunthaler, G. (2010). Optimization of chemically defined cell culture media – Replacing fetal bovine serum in mammalian in vitro methods. *Toxicology in Vitro, 24*(4), 1053-1063. doi:https://doi.org/10.1016/j.tiv.2010.03.016

Whitby, R. J., Stec, J., Blind, R. D., Dixon, S., Leesnitzer, L. M., Orband-Miller, L. A., . . . Ingraham, H. A. (2011). Small molecule agonists of the orphan nuclear receptors steroidogenic factor-1 (sf-1, nr5a1) and liver receptor homologue-1 (lrh-1, nr5a2). *Journal of medicinal chemistry, 54*(7), 2266-2281. doi:10.1021/jm1014296

Wu, W.-S., & Lai, F.-J. (2015). Functional redundancy of transcription factors explains why most binding targets of a transcription factor are not affected when the transcription factor is knocked out. *BMC Systems Biology, 9*(Suppl 6), S2-S2. doi:10.1186/1752-0509-9-S6-S2

Yang, T.-H., & Wu, W.-S. (2013). Inferring functional transcription factor-gene binding pairs by integrating transcription factor binding data with transcription factor knockout data. *BMC Systems Biology, 7*(Suppl 6), S13-S13. doi:10.1186/1752-0509-7-S6-S13

Yeh, S., Miyamoto, H., Shima, H., & Chang, C. (1998). From estrogen to androgen receptor: a new pathway for sex hormones in prostate. *Proceedings of the National Academy of Sciences of the United States of America, 95*(10), 5527-5532.