Fall 5-22-2015

Transcriptional Regulation by DAX-1 in Pluripotent and Differentiated Cells

Alexandra C. Maramba

University of San Francisco, marambaa@gmail.com

Transcriptional Regulation by DAX-1 in Pluripotent and Differentiated Cells

—————————————————
Thesis title

by
Alexandra Caroline Maramba

—————————————————
Student name

Thesis
Submitted in partial Satisfaction of the Requirements
For the degree of

**Master of Science
In Biology**

In the
College of Arts and Sciences
University of San Francisco
San Francisco, California

Committee in Charge

Approved:_____   8|24|2015
Faculty Advisor   Date

Approved:_____   24 Aug 15
Committee Member   Date

Approved:_____   8/24/15
Committee Member   Date

Approved:_____   8-31-2015
Dean, College of Arts and Sciences   Date

Abstract

DAX-1, an orphan nuclear hormone receptor, acts mainly as a repressor through transcriptional protein complexes. Its unique structure and specific expression raises questions as to what its precise interactions are and how it mediates its repressive function. While it is known to play a role in sexual development and adrenal insufficiency, expression in certain types of cancer suggests additional functions and interactions. Knock in of DAX-1 into a low-DAX-1 expressing cancer cell line has been previously observed to increase apoptosis, while, inversely, down in a high-DAX-1 expressing cancer cell line shows a decrease in apoptosis. Target genes that belong to the TNFα and BCL-2 families have shown changes in expression correlating to the modified levels of DAX-1 in knock-down experiments. Direct regulation of BCL-2, one of the target genes of interest, was investigated further based on mirrored expression changes of DAX-1 in knock-down and knock in experiments through ChIP experiments. These findings emphasize a significant role of DAX-1 in moderating apoptosis in a breast cancer cell line.

In the context of undifferentiated mouse embryonic stem cells, Dax-1 is highly expressed and has been shown to be an important contributor to the pluripotent state. Potential downstream targets of Dax-1 were previously identified based on significant changes in expression when Dax-1 expression was down regulated. Two methods, siRNA and CRISPR-Cas9, were used to decrease Dax-1 expression in the E14 mouse embryonic stem cell line. Direct interactions and other novel stem cell factors were confirmed using analysis of publically available ChiP-

seq data. Ultimately, while Dax-1 is not a master regulator, its transcriptional

control of specific genes that are key in the maintenance of pluripotency is an

important component of stem cell growth and differentiation. Bioinformatic analysis

of ChIP-seq experiments brought to light general patterns as to how Dax-1

contributes to pluripotency, and additional ontologies of Dax-1 target genes for

future studies.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AHC: Adrenal Hypoplasia Congenita

Ago2:   Argonaute-2

AR: Androgen Receptor

BADGE: Bisphenol A diglycidyl ether; 2,2'-[(1-Methylethylidene)bis(4,1-

            phenyleneoxymethylene)]bis-oxirane

BAX: Bcl-2-associated X Protein

BAK: Bcl-2 Homologous Antagonist Killer

BCL-2:   B-cell lymphoma 2

BCL-XL: B-cell lymphoma extra large

BCL-XS: B-cell lymphoma extra small

BH1-4:   Bcl-2 homology domain 1 through 4

Cas: CRISPR Associated Proteins

Caspases: Cysteine-aspartic Proteases; Cysteine-dependent aspartate-directed

            proteases

cDNA: Complimentary DNA

ChIP: Chromatin Immunoprecipitation

ChIP-seq: Direct High-throughput Chromatin Immunopreciptation Sequencing

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

crRNA: CRISPR repeat-spacer RNA

DAX-1:   Dosage Sensitive Sex Reversal Adrenal Hypoplasia Congenita, critical

            region on the X Chromosome, Gene 1

DBD: DNA Binding Doman

DMEM: Dulbecco's Modified Eagle Medium

DNA: Deoxyribonucleic Acid

DSS: Dosage Sensitive Sex Reversal

ER: Estrogen Receptor

ERα: Estrogen Receptor alpha

FBS: Fetal Bovine Serum

FITC: Fluorescein Isothiocyanate

GAPDH: Glyceraldehyde 3-Phosphate Dehydrogenase

GRTH: Gonadotropin-regulated Testicular Helicase

H: The amino acid histidine

HDR: Homology Directed Repair

I: The amino acid isoleucine

IgG: Immunoglobulin G

Indel: Insertion/deletion mutation

LBD: Ligand Binding Doman

L: The amino acid leucine

mESC: Mouse Embryonic Stem Cell

MOMP: Mitochondrial Outer Membrane Permeabilization

mRNA: Messenger RNA

NC: Negative Control

NEAA: Non-essential Amino Acids

NHEJ: Nonhomologous End Joining

NHR: Nuclear Hormone Receptor

NR: Nuclear Receptor

RBS: Phosphate Buffered Saline

PCR: Polymerase Chain Reaction

PI: Propidium Iodide

PS: Phosphatidylserine

PVDF: Polyvinylidene Fluoride

qPCR: Quantitative Polymerase Chain Reaction

RNA: Ribonucleic Acid

RNAi: RNA Interference

RISC: RNA Interference Silencing Complex

RT-PCR: Real Time Polymerase Chain Reaction

SF-1:   Steroidogenic Factor-1

siRNA: Short Interfering RNA

shRNA: Short-hairpin RNA

TBS: Tris Buffered Saline

TNFα: Tumor Necrosis Factor alpha

TRADD: TNFRSF1A-associated via death domain

TSS: Transcription Start Site

UT: Untreated

WT: Wild-type

x: represents any amino acid in the protein sequence

X-ChiP: Cross-linked ChiP

Chapter 1

Nuclear Hormone Receptors and DAX-1

Introduction

Nuclear hormone receptors (NHRs) are a superfamily of eukaryotic transcription factors that are involved in the regulation of specific target genes for various biological processes through the response to hormones and other metabolic ligands [1]. These various proteins share a conserved general domain structure, as illustrated in Figure 1-1A: an N-terminal domain (A/B), a DNA binding domain, or DBD (C), a flexible hinge region (D), a ligand binding domain, or LBD (E), and a C-terminal domain (F). There are seven subfamilies, identified as NR0 through NR6, based on sequence similarity in the conserved DBD and LBD regions of the genes [2]. This sub-categorization helps dictate the gene nomenclature rules.

These proteins generally have an effect on gene transcription upon binding of a ligand [3]. In general, once these transcription factors bind ligand and are activated, they can bind to specific sequences, also known as hormone response elements (HREs), present within the regulator regions of target genes. Typically, NHRs carry out their activity by forming dimers on these response element sequences. The combination of NHRs may either be as homodimers, composed of two identical receptors, or as heterodimers, composed of different receptors [4].

NHRs that act as activators or coactivators contain LxxLL (Leu-x-x-Leu-Leu) motifs in their primary amino acid sequence, which manifest as alpha helices in the protein structure. Also known as NR boxes, the sequence allows for protein-protein interactions, and therefore the formation of transcriptional factor complexes [5, 6].

In contrast, NHRs that generally act as repressors or corepressors have a consensus motif known as a CoRNR box, consisting of an Lxx H/I xxx I/L motif that is manifested as an extended helix, and interacts with similar residues such as the LxxLL motif. The variance in the length of the helical interface allows for a difference in the transcription factor complexes [7].

In humans, there have been 48 different NHRs identified to date. Mice, with 47 known NHRs, share a similar number of this group of proteins encoded in their genome. This conservation of nuclear receptors is reflected in only members of the animal kingdom [8], and is not observed in any other relatives of metazoans [9]. NHRs have been studied in great detail in numerous model systems because of their involvement with key signaling pathways and the potential for targeting in diseases.

Proteins within the NHR family can be further categorized into three groups: steroid hormone receptors, non-steroid hormone receptors, and orphan receptors. Steroid hormone receptors respond to steroids, while non-steroid hormone receptors interact with other ligands such as thyroid hormone, retinoic acid, and vitamin D2. Lastly, the remaining NHRs that have no known ligand interaction are classified as orphan nuclear hormone receptors. [10]

DAX-1 is a unique orphan nuclear hormone receptor that is encoded by the *NR0B1* gene. The protein name stands for dosage sensitive sex reversal, adrenal hypoplasia congenital on the X chromosome, gene 1; for humans, it is written out as DAX-1, while for mice it is noted as Dax-1 [11]. The protein product of the human *NR0B1* gene is 470 amino acids, pieced together from two exons. In contrast to the typical NHR structure, the DAX-1 protein has a repetitive domain consisting of 3.5

alanine/glycine rich repeats that each are about 65 to 70 amino acids long in place of the DNA binding domain, as shown in Figure 1-1B. It shares this unusual domain with a close cousin within the NHR family called SHP (small heterodimer partner), encoded by the *NR0B2* gene [12] [13].

While the ligand-binding domain is conserved, DAX-1 is identified as an orphan due to no known ligand interaction. The lack of the DNA binding domain, no known ligand interaction, and LxxLL sites strongly implies that DAX-1 primarily functions within the cell through interactions with other proteins, such as Wilm's tumor 1 (WT1) and steroidogenic factor 1 (SF-1) [14] rather than direct binding to DNA as a transcriptional factor. And despite having LxxLL sites, which is typical of coactivators, DAX-1 has been shown to act as a corepressor [15]. X-ray crystallographic data has provided some insight as to how DAX-1 mediates its repressive function through binding directly to other NHR proteins. The entire protein has yet to be isolated and crystallized, but the LBD region of DAX-1 has been successfully crystallized in complex with the protein LRH-1, as illustrated in Figure 1-2. The site of interaction with LRH-1 is at an extended surface, therefore occluding binding of other co-activators to LRH-1 [16]. This type of protein-protein interaction is a likely explanation of why DAX-1 is a strong transcriptional repressor.

The original discovery of DAX-1 was through duplications in the gene that caused males to phenotypically appear female, corresponding to the "dosage-sensitive sex reversal" portion of the name [17]. Additional research found that mutations within the same gene led to X-linked adrenal hypoplasia congenital (AHC), an uncommon disorder of adrenal gland development [18]. Outside of these

diseases, DAX-1 has a restricted expression pattern. It is highly expressed in stem cells before differentiation, and is noted to contribute to early embryogenesis [11]. In an adult organism, expression is limited to select steroidogenic organs, especially within the hypothalamus-pituitary-adrenalgonadal axis [19].

More recently, there has been a proposed role of DAX-1 in various cancers, in addition to its role in normal cells. Numerous NHRs have been linked to the pathogenesis and treatment of various cancers [20], and DAX-1 is no exception. Its strong influence on pluripotency, a characteristic seen also in a subset of cancer cells, and unusual structure brought DAX-1 into the spotlight as a candidate for cancer therapy [21]. In this thesis, we study the specific molecular mechanisms and interactions by which Dax-1 functions within mouse ES cells and human DAX-1's influence on processes within normal and breast cancer cells.

A.
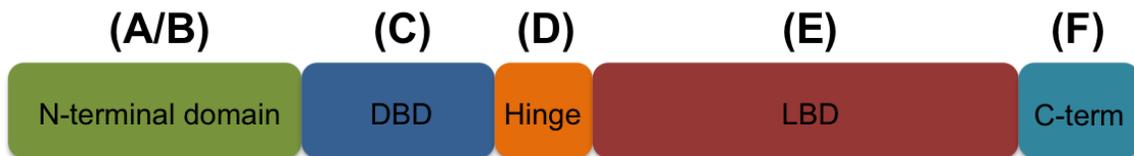
B.

Alanine/glycine rich repeats

Figure 1-1. Functional domain structures of a general nuclear hormone receptor and DAX-1. A. The general nuclear hormone receptor gene regions with the following labeled structures: an N-terminal domain (A/B), a DNA binding domain, or DBD (C), a hinge region (D), a ligand binding domain (E), and a C-terminal domain (F), and B. the DAX-1 protein structure, with the 3.5 alanine/glycine rich repeats on the N-terminal end (R) and a conserved LBD-like domain (E).
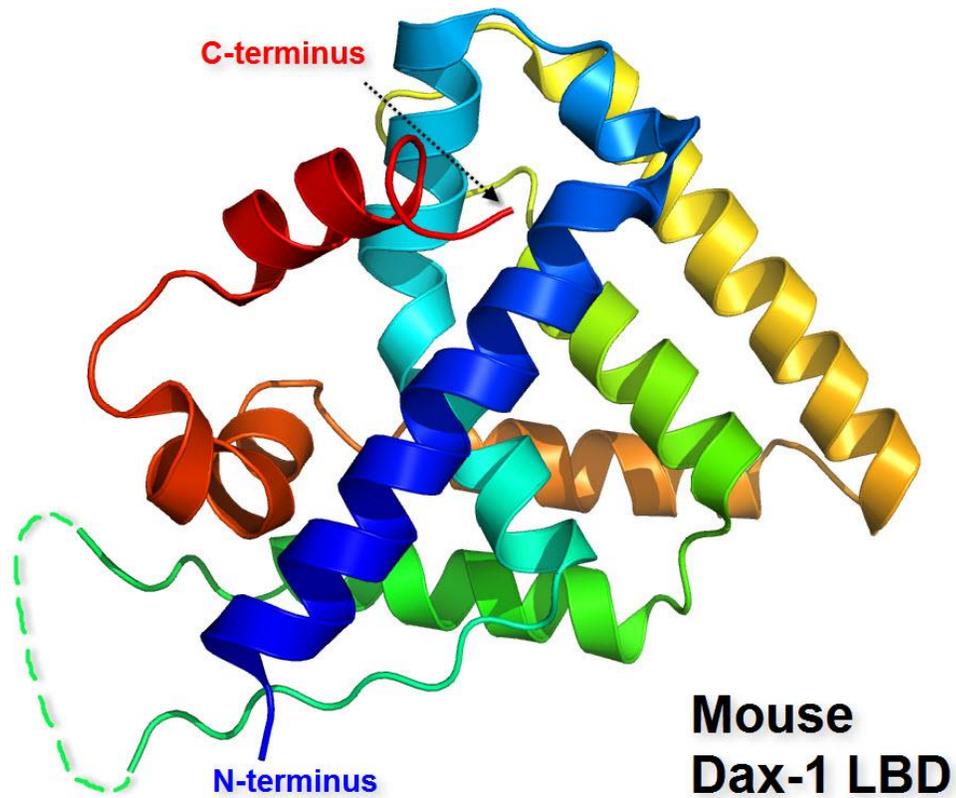
Figure 1-2. X-Ray crystallography illustration of mouse Dax-1 ligand binding domain. Structure is based on the available image of Dax-1 in complex with LRH-1 (Fletterick Group, pdb file 3F5C). The chains are colored from blue to red, in the direction of the N-terminus to C-terminus. Figure generated by P. Foster.

Manipulation of DAX-1 Expression in a Range of Model Systems

Introduction

As previously mentioned, DAX-1 expression is restricted to stem cells and a number of organs within a mature organism. The research detailed in this thesis examines the role of DAX-1 in human breast cancer cells as well as mouse embryonic stem cells. Both these research projects required a means to study and manipulate DAX-1 expression in order to explore the outcome of this change. For these experiments, the human cell lines MCF7 (epithelial breast cancer cells) and MCF10A (mammary epithelial cells) and the mouse embryonic stem cell line E14 were used. MCF7 cells are typically used as a triple positive cell culture model, because these cells express estrogen receptor, progesterone receptor, and HER2 [22]. Transient transfection assays were used to exogenously express DAX-1 in the cell types with little to no DAX-1 expression, while RNAi and CRISPR methods were utilized to reduce or eliminate DAX-1 expression in cell lines with endogenous expression of DAX-1. Once these models had been established, we could explore the outcome of changes in DAX-1 expression in both human cancer cells (Chapter 3) and mouse embryonic stem cells (Chapter 4).

pcDNA expression vector

In order to introduce DAX-1 into the MCF7 cells, an expression plasmid containing the entire human DAX-1 coding sequence was utilized. The expression vector used for these experiments was pcDNA3.1, which contains unique restriction

enzyme target sites in both directions for sequence insertion (Figure 2-1). Other key features of the plasmid include the genes for ampicillin and neomycin resistance, which aid in the selection process after transformation into competent *E. coli*. The entire human DAX-1 coding sequence was amplified by PCR and cloned into the plasmid using the BamHI and XhoI restriction enzyme sites. Confirmation of correct insertion of DAX-1 into pcDNA3.1 was confirmed by DNA sequencing. This expression plasmid was previously generated by C. Tzagarakis-Foster.
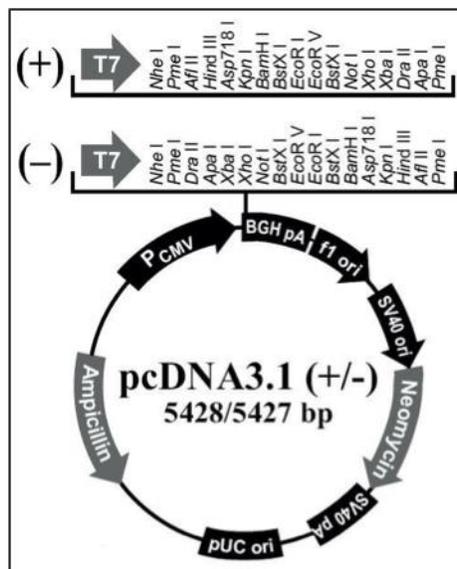
Figure 2-1. pcDNA3.1 vector (Life Technologies) utilized in transient transfection assays.

RNA interference

   RNA interference is a gene silencing technique that utilizes non-coding RNAs to target messenger RNA present in an organism and affect gene expression before translation into protein [23]. Numerous forms of RNA can be utilized, ranging from double stranded RNA to short hairpin RNA, with different structures and stability. While this phenomenon is conserved across most complex organisms, it was not exploited until the discovery of sequence-specific short interfering RNAs (siRNAs). These are 21 nucleotide long sequences, with 19 of those nucleotides specific to the target sequence to form a duplex, and the remaining 2 unpaired at the 3' ends [24].

   The pathway of RNA interference, illustrated in Figure 2-2, starts in the cytoplasm when Dicer, an RNAse III endonuclease, processes siRNA precursors such as dsRNA or shRNAs. Certain kinds of optimized techniques in research skip this processing step and may allow for the transfection of the shortened siRNA from the start. A "mature" siRNA molecule then enters into a multi-protein RNA-induced silencing complex, or RISC. The completed complex pairs the siRNA with the target mRNA at the complementary region, and the protein Argonaut-2 (Ago2) mediates the cleavage of the siRNA-mRNA pair within the complex. Other notable proteins involved in RISC include a transactivating response RNA-binding protein, or TRBP, that helps facilitate the transfer of Dicer-generated fragments to Ago2 for RNA degradation. While there are cleavage independent mechanisms to block translation of the targeted mRNA, the formation of the complex allows for the target to be broken up several times. The resulting smaller RNA fragments are ultimately degraded by exoribonucleases, and the loss of the targeted mRNA in the cell reduces

the amount of protein produced from translation [25]. This technique is widely used

in research for many model organisms, and was applied to both the MCF7 human

breast cancer cells and the E14 mouse embryonic stem cells in experiments to

knock-down endogenous DAX-1 expression.
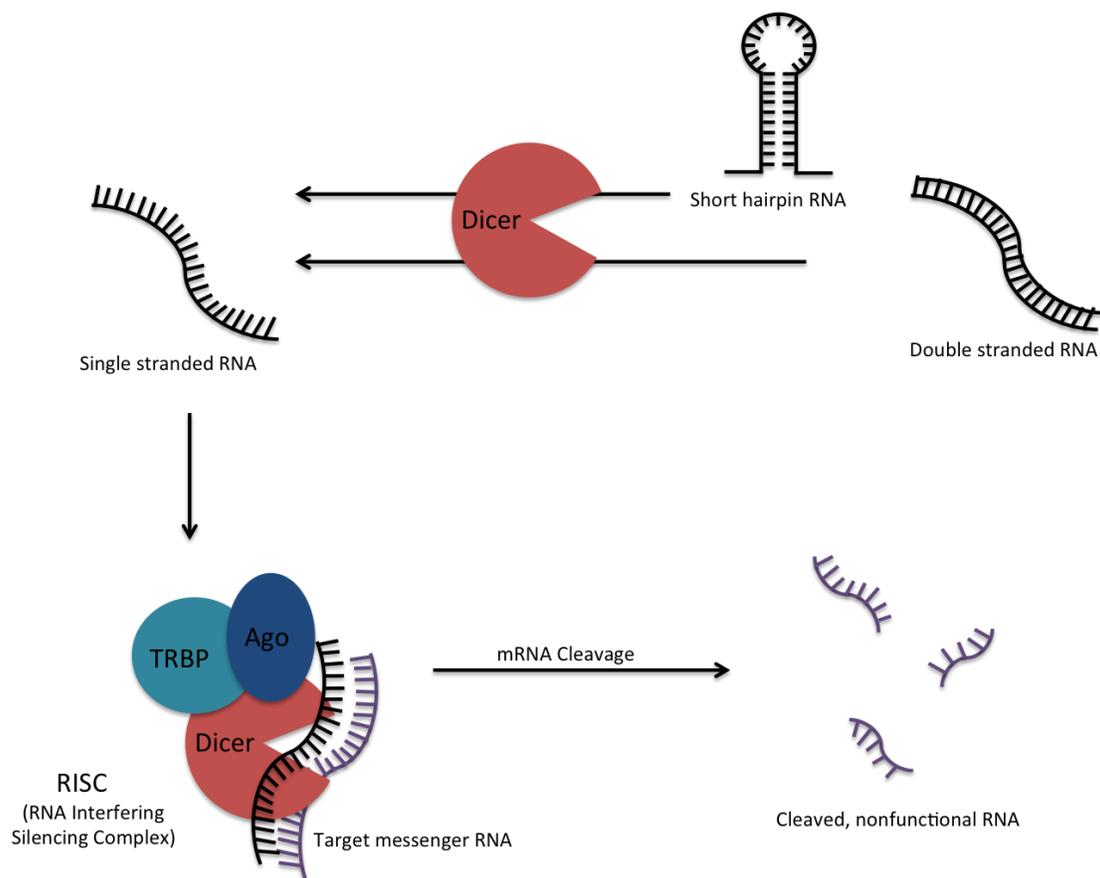
Figure 2-2. RNA interference mechanism. Starting materials range from short-hairpin RNA (shRNA) or double stranded RNA (dsRNA) that is then matured into a simplified form of single stranded RNA, or initially short interfering RNA (siRNA) that is already shortened to a length of twenty to thirty base pairs long. Once a target messenger RNA is found through a section complementary to the siRNA, RISC is formed to capture and degrade the mRNA, resulting in a loss of gene expression at the RNA level.

CRISPR-Cas

A recent technique utilizing an endogenous mechanism called CRISPR-Cas has been propelled as a more permanent modification to protein expression. The name is short for "clustered regularly interspaced short palindromic repeats – CRISPR associated proteins." In contrast to commonly used techniques, such as RNAi that targets a gene at the RNA level in order to reduce its expression, CRISPR affects cell functions through genomic editing [26].

The technique was developed based on microbial immune systems, where RNA-guided endonucleases can interact with specific targets, typically exogenous DNA from viruses or other bacteria, and allows the organism to both eliminate the invading genomic information as well as acquire immunity against similar viruses and plasmids [27]. Optimized plasmids have been developed to contain core elements of the CRISPR-Cas system, selection genes for transformation and cloning, and specific restriction enzymes sites for the insertion of a target guide sequence.

One of the main elements of the most commonly used optimized system is Cas9, part of the Cas protein family. This family, whose name is short for "CRISPR-associated genes," consists of microbial endonucleases than can utilize short RNA sequences as guides to cause damage to genomic DNA [28]. Cas9 is specifically used for research applications because while other systems may require a complex of multiple Cas proteins, Cas9 is large protein that alone is sufficient for generating a mature guide RNA and cleaving the target DNA [29].

The endonuclease can only act on a target DNA by using a guide sequence, often referred to as CRISPR repeat-spacer RNA, or crRNA. This fragment contains

conserved repeat fragments and a variable spacer sequence that is complementary to a target region [30]. While the length of this sequence varies in size depending on the type of Cas protein used, the Cas9 protein is able to create and use a guide with a target sequence as short as 21 nucleotides long. Together with a trans-activating crRNA (tracrRNA), an RNA duplex is formed and matured to form a hybrid, acting as a complete guide for Cas9 [31]. This allows for Cas9 to situate itself at the target location with base-pair matching of the guide RNA and cause DNA damage (Figure 2-3).

The damage caused by the Cas9 endonuclease is a double-stranded break that can be repaired in two ways. Nonhomologous end joining, or NHEJ, is an imprecise type of repair that through immediate ligation to close the break introduces insertion and deletion (or indel) mutations that vary in length [32]. When applied to damage within a gene, this disrupts the translational reading frame; if applied in a noncoding region, this can potentially disrupt a binding site of transcriptional factors. Another type of DNA repair is homology-directed repair, or HDR, which utilizes a template to repair the site of damage. Through the use of the of a donor template, more precise point mutations or insertions can occur at the site of damage [33].

Additional genomic engineering applications of the Cas9-based technique can result in sequence specific insertions, large deletions, genomic rearrangements through inversions or translocations, gene activations, epigenetic modification, or tagging of genomic loci [34]. Applications of this technique include early treatment in haploid cells to increase the frequency of successful gene modification without

selection [35], in human pluripotent stem cells [36], and even in entire living organisms [37]. CRISPR-Cas9 is also applicable to affect multiple gene targets with the inclusion of several guide sequences within the optimized plasmid [38], further expanding the applications while maintaining the simplicity of the system. In the context of this thesis, CRISPR-Cas9 was applied to target the coding sequence of mouse Dax-1 at multiple locations through the design of several guide sequences. The sequences were used individually in CRISPR-Cas9 transfection and cells were selected for proper Dax-1 deletion before further analysis in Chapter 4.
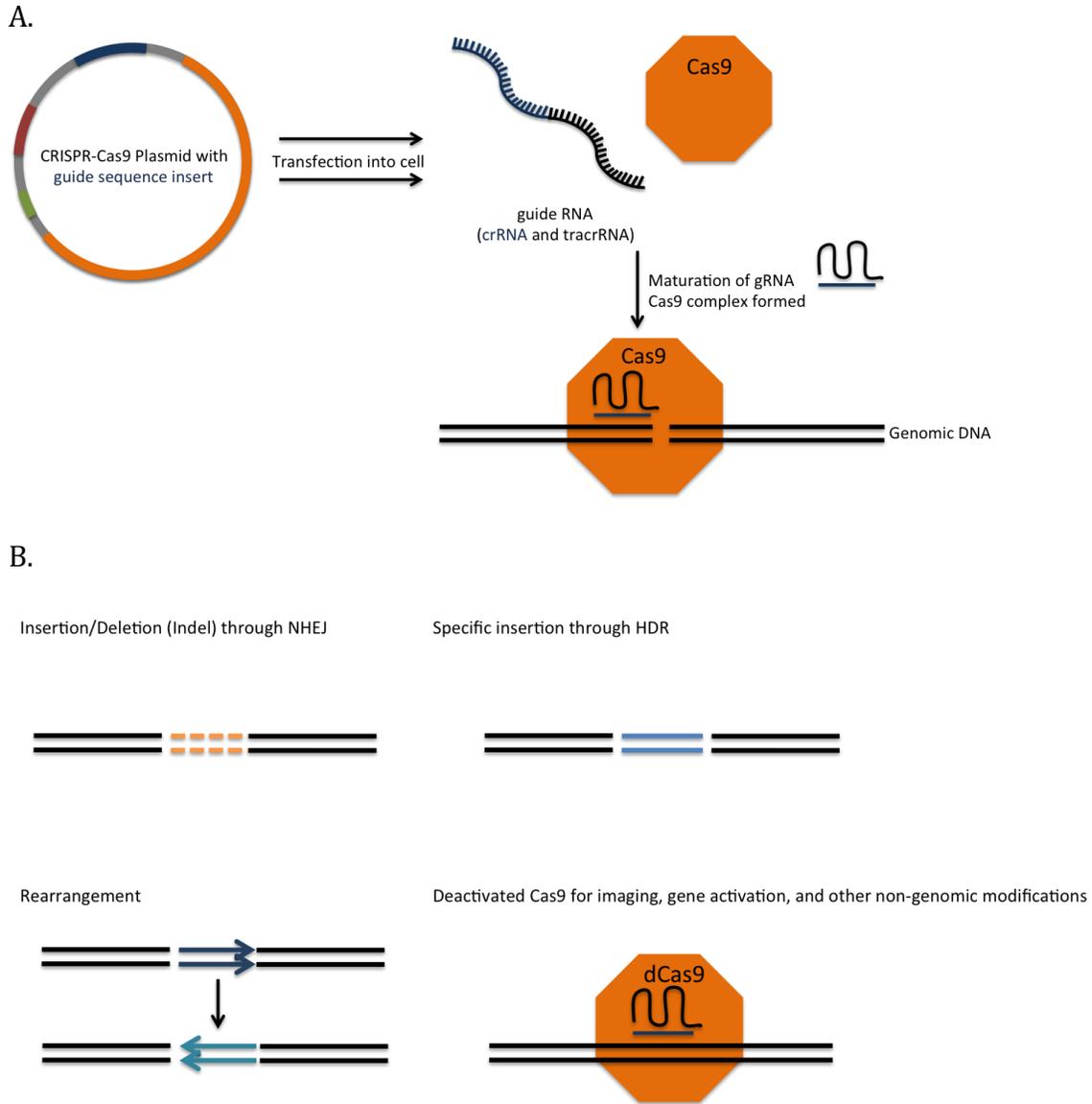
Figure 2-3. CRISPR-Cas9 knock-down mechanism. A. A complete CRISPR-Cas9 plasmid includes a sequence for the Cas9 endonuclease, an inserted guide sequence, and selection genes. The plasmid is transfected into cells or an organism and allows for internal production of the key CRISPR system elements. The ribonucleic-protein complex of Cas9 with a mature guide allows for DNA damage at a target site. This damage can be repaired by either NHEJ or HDR DNA repair mechanisms. B. Based on the type of repair and Cas9 used, there are a number of different outcomes to CRISPR modifications. Without a provided template, the site of damage can be repaired through non-homology end joining and result in random insertions and deletions at the target site. With a template, homology directed repair allows for either the insertion of a provided sequence or an inversion of the targeted section. Deactivation of Cas9 can eliminate the endonuclease aspect and allow for guide direct tagging or modification without damage to the genome.

Methods

Cell Culture

MCF7 epithelial breast cancer cells were obtained from American Type Culture Collection. The cells were cultured and passaged in a 5% carbon dioxide tissue culture incubator at 37°C. Cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) with L-glutamine and without phenol red, supplemented with 10% charcoal-stripped fetal bovine serum (American Type Culture Collection, treated with dextran coated charcoal from Sigma-Aldrich) and 1% penicillin/streptomycin (Gibco).

MCF10A human mammary epithelial cells were obtained from American Type Culture Collection. The cells were cultured in the same incubator conditions as MCF7 cells. Complete MCF10A media used Dulbecco's Modified Eagle Medium (DMEM) with phenol red and additional supplements of 10% fetal bovine serum (American Type Culture Collection), 1% penicillin/streptomycin (Gibco), 100 ng/mL of cholera toxin (Sigma-Aldrich), and mammary epithelial growth medium (MEGM) SingleQuot additives [hydrocortisone (05 μg/mL, human EGF (10ng/mL), insulin (10 μg/mL)] (Lonza Corporation).

The E14 mouse embryonic stem cell line was donated from the lab of Dr. Mohammed El Majdoubi, from Dominican University of California in San Rafael, CA. The feeder-free cells were cultured and passaged in a 5% carbon dioxide tissue culture incubator at 37°C in 0.1% gelatin-coated flasks. The media used for the cultured cells was Dulbecco's Modified Eagle Medium

(DMEM) with phenol red, GlutaMax, and 4.5 g/L D-glucose. For the complete

media, additional supplements of 15% fetal bovine serum (American Type

Culture Collection), 1% penicillin/streptomycin (Gibco), 1% non-essential

amino acids (Gibco), 1,000 U/mL Leukemia Inhibitory Factor (Sigma-

Aldrich), and 0.1mM β-mercaptoethanol (Promega) were added through a

sterile filter. This combination is modified from those outlined by the Cho

Lab [39].

To passage cells, flasks were washed with phosphate buffered saline (PBS)

and treated with Trypsin-EDTA (0.05%) with phenol red (Gibco) at 37°C

until cells lifted from the flask. Cells were then re-suspended and passaged,

ranging from 1:5 to a 1:10 ratio.

DAX1 knock-in into MCF7 cells

A pcDNA3.1-DAX1 construct was provided by Dr. Christina Tzagarakis-

Foster, and transfected into the MCF7 cells using Lafectine RU50 according to

protocol, using 50uL for every μg of plasmid. Treated cells were initially

seeded in "low" media, composed of DMEM without phenol red

supplemented with 1% fetal bovine serum, for up to 24 hours before

treatment in complete media. Cells were treated for 48 hours before

collection. Samples collected from treated cells are referred to herein as

"MCF7+DAX-1".

DAX1 knock-down in MCF10A and E14 cells with siRNA

A set of three synthetic double stranded stealth siRNA oligonucleotides designed to target the human DAX-1 gene were obtained from Life Technologies. The three siRNAs, which targeted human DAX-1 in different regions of the gene, were used for transfection into MCF10A cells, as shown in Table 2-1. Three synthetic *Silencer* pre-designed small interfering RNA (siRNA) oligonucleotides targeting mouse Nr0b1 (Life Technologies) were utilized for Dax-1 knock-down, with sequences and targeting sites listed in Table 2-2. Ambion *Silencer* Negative Control #1 was used as the non-targeting siRNA sample. Six-well plates were seeded with 500,000 MCF10A cells per well in complete media and reverse transfected using Lafectine RU50 (MednaBio). The total concentration of the siRNA for MCF10A cells was 300pmol for each well, while two concentrations of 20pmol and 80pmol of each individual siRNA was used for the mouse embryonic stem cells. Mouse cells were treated a second time with a change of media after 24 hours. Collections for mRNA and protein lysate analysis were carried out after 48 hours.

Statistical Anaylsis

Data from the qPCR analysis are presented with the mean + standard deviation of triplicate experiments, normalized using GAPDH Cq values. Statistical significance is calculated using the two-tailed t-test formula built into Microsoft Excel. A single asterisk (*) represents samples that were

$p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a

triple asterisk (***) represents samples that were $p<0.001$.

Table 2-1. siRNA sequences for MCF10A treatment.

siRNA	Sequence	Target Region
hDAX-1 siRNA 1	CCCAUGACAGAUUCAUCGAACUUAA UUAAGUUCGAUGAAUCUGUCAUGGG	Exon 2 Position 1290
hDAX-1 siRNA 2	CCCAUGACAGAUUCAUCGAACUUAA UUAAGUUCGAUGAAUCUGUCAUGGG	Exon 2 Position 1383
hDAX-1 siRNA 3	CAGUGGCAGGGCAGCAUCCUCUACA UGUAGAGGAUGCUGCCCUGCCACUG	Exon 1 Position 44

Table 2-2. siRNA sequences for mESC treatment. Sequences were not made available with purchase.

siRNA	Target Region
mDax-1 siRNA 1	Exon 1, Position 981
mDax-1 siRNA 2	Exon 1, Position 1187
mDax-1 siRNA 3	Exon 1, Position 987

CRISPR-mediated knock-down in E14 mESC

pSpCas9(BB)-2A-Puro (PX459) plasmid was purchased from Addgene

(plasmid # 48139). The structure of the plasmid along with key features is

illustrated in Figure 2-4. Oligonucleotides for insertion into the CRISPR-Cas9

vector backbone to target mouse DAX1 were designed using the online CRISPR Design web program (http://crispr.mit.edu/). Target areas for each guide are shown in Figures 2-5 and 2-6, with sense and antisense sequences listed in Table 2-3. Annealed oligonucleotides were inserted into the pX459 plasmid using BbsI as the restriction enzyme to cut the insertion site and T4 DNA ligase (New England Biolabs). Proper insertion of the guide sequences into the plasmid backbone was verified through both a BbsI/AgeI double digest, based on the restriction enzyme target sites illustrated in Figure 2-4B, as well as DNA sequencing using a U6 primer (MCLAB, Molecular Cloning Laboratories). Modified plasmids with successful insertions were transformed into competent DH5α *E. coli* (New England Biolabs) and plasmid DNA was extracted using the Plasmid Plus Maxi kit (Qiagen). Complete pX459 with oligonucleotide insertions were transfected into mESCs using Lafectine RU50 1000ng/μL according to the product protocol using reduced media (Dulbecco's Modified Eagle Medium (DMEM) with phenol red; 10% FBS, LIF, β-mercaptoethanol). Cells were treated for 48 hours before selection with 1.0 μg puromycin. Treatment with puromyin was sustained for 48 hours before surviving cells were collected and recovered in normal media. Individual colonies were picked in parallel to grow out individual transfected cells and obtain clonal populations for further analysis [40].

A.
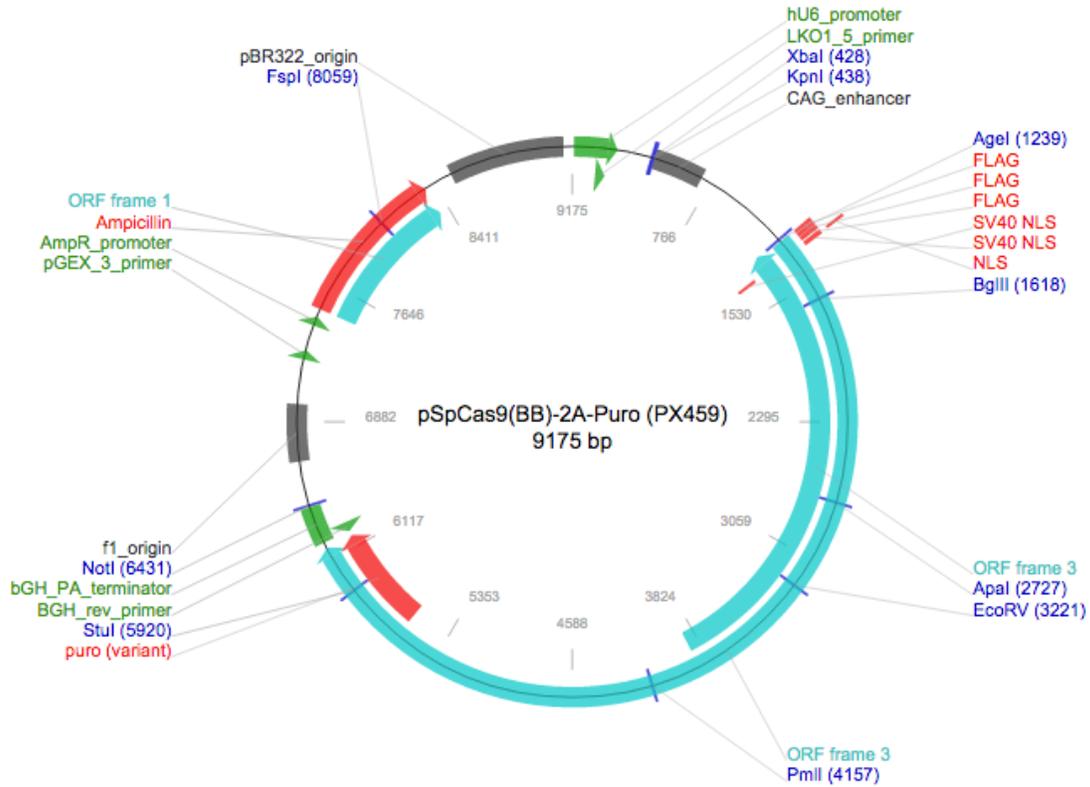
B.
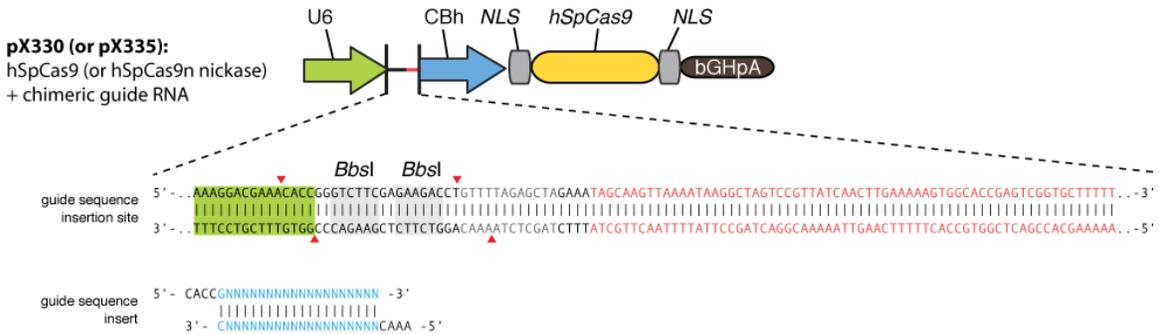
Figure 2-4. Structure summary of the CRISPR-Cas9 optimized plasmid pX459.
A. Overall pX459 structure provided by Addgene. B. Details of BbsI restriction enzyme sites for insertion of a designed guide sequence for CRISPR-Cas9 targeting. Figure provided by Zhang lab.

Table 2-3. Sense and anti-sense sequences for oligonucleotides constructed using the online CRISPR Design web program to target the Dax-1 gene.

Guide Name	Sense	Anti-sense	Target Exon	Color Code
G3	caccGGGTGAGGACCACCCGTGGC	aaacGCCACGGGTGGTCCTCACCC	Exon 1	<mark>green</mark>
G6	caccGCTGAGATTCATCAATAGCG	aaacCGCTATTGATGAATCTCAGC	Exon 2	<mark>pink</mark>

agtttataaagctgtcatagaaatggactttttgccatttaagcatttctttccccaataaaggaaataagttagaggtcaga
gtctaagttaatggcaagagttggaacagagccctaactagctgcctggggttcactgctagctcttctcttccccaggtag
aggcaggaggggtggagtgaagaaggaaaggtggtatgtggtatgctagttccagtgctgagactctcccttggatttcc
agcttctagggagtgtttgccccttgagctttcgaggtcatggccacacacattcaagcacaaaggcgcgtccccctctgc
gcccttgtccaagaggaggaggcggacgcgcttgcgtgcgcattcagtataaataagtcccaagcggcggccactgggc
AGAACGAGCTACAGGAGCCTCAGGCCATGGC<mark>***GGGTGAGGACCACCCGTGGC***</mark>**AGGGCAG**
CATCCTCTACAATCTACTGATGAGCGCGAAGCAGAAGCACGCGTCTCAGGAAGAGCG
AGAGGTGCGCTTGGGGGCTCAGTGCTGGGGTTGCGCCTGCGGTGCTCAGCCCGTCCT
GGGTGGGGAGAGACTGTCCGGCGGGCAAGCCAGGTCCCTCTTGTACCGCTGCTGCTT
TTGTGGGGAGAATCACCCGCGCCAGGGTGGCATCCTCTACTCTATGCTCACCAACGC
CAGGCAGCCAAGCGTGGCGACCCAGGCGCCGAGGGCACGATTCGGAGCACCTTGCTG
GGGCTGCGCCTGCGGCAGCGCAGAGCCCCTGGTGGGCAGAGAGGGGCTGCCGGCTGG
CCAGGCCCCCTCGCTCCTGTACCGCTGCTGCTTCTGCGGAGAAGAGCACCCGAGGCAG
GGCAGCATCTTATACAGCTTGCTCACTAGCGCTCAGCAAACGCACGTGTCTCGGGAA
GCACCCGAGGCACATCGCAGAGGCGAGTGGTGGCAGCTGTCCTACTGTACCCAGAGT
GTGGGTGGCCCAGAGGGGCTGCAGAGCACACAGGCCATGGCGTTCCTGTACCGCAGC
TATGTGTGCGGTGAAGAGCAGCCCCAGCAGATCAGCGTTGCCTCTGGCACGCCCGTG
AGCGCAGACCAAACACCAGCGACCCCGCAAGAGCAGCCGAGGGCTCCCTGGTGGGAC
GCCTCACCTGGTGTGCAGCGTCTGATCACACTCAAGGATCCACAGGTGGTGTGCGAG
GCAGCGTCCGCTGGCCTGTTGAAGACCCTGCGCTTTGTCAAGTACTTGCCCTGCTTCC
AGATCCTGCCCCTAGATCAGCAGCTGGTGCTGGTGCGGAGCTGTTGGGCGCCCCTAC
TCATGCTTGAGTTGGCCCAAGATCACCTGCACTTCGAGATGATGGAGATCCCGGAGA
CCAACACGACGCAGGAAATGCTTACCACCAGGCGGCAGGAGACCGAAGGTCCAGAGC
CTGCAGAGCCCCAGGCCACAGAGCAGCCACAGATGGTGTCCGCGGAGGCTGGGCACT
TGCTCCCAGCTGCTGCGGTCCAGGCCATCAAGAGTTTCTTTTTCAAGTGCTGGAGTC
TGAACATTGACACCAAAGAGTATGCCTATCTGAAAGGGACCGTGCTCTTTAACCCAG
gtaagcattgtcaaccttgtgcactgtctcttcaggtcagaaaagcaccactacagacatttatgagtttttagtaggggttt
tggagcccacttgactgtgactgactctgcaaatctgctgcacacatcaggaataccctgtcaagtggtcaaatgggtgag
cttctgccagaaacttggcactggattggtggtgatagtgtgtgtgtgtggggggggggggggcaggggttggggggag
ggtggggaacgaaccagtaagctttgtacctggctctctattatctgctgctacagaagtttctgtcttcataggacctggct
aactcttc

Figure 2-5. CRISPR targets within the Dax-1 gene (first exon). Sequence of the
first exon of the Dax-1 gene, with CRISPR targets for guide 3, highlighted <mark>green</mark> and
underlined. Intron/non-coding nucleotides are in lowercase, and exon/coding
regions are in uppercase.

tattaagtgtgccagtgtttgtgcacactttagtgtgtccagtgtatgtgcacacatcctttcatccttagaagtgttgcttctg
aaaacttagaacttatttcagtccattattttaggccccagaaaaccctttaatcaagaagctagggttctttttaatcggg
atgggactaaaaaaaatgagtttggcctaaaccataaagatcctgtggtgagctgtttaaaataaagttttctcctttcag
ACCTGCCTGGCCTGCAGTGCGTGAAATACATTGAGGGTCTTCAGTGGAGAACCCAGC
AGATCCTTACTGAGCACATCCGGATGATGCAGAGAGAGTACCAGATCAGATCCGCTG
AACTGAACAGTGCCCTTTTCCT*GCTGAGATTCATCAATAGCG*ATGTCGTCACTGAACT
CTTTTTCAGGCCCATCATTGGTGCAGTCAGCATGGATGATATGATGCTGGAGATGCT
CTGTGCAAAGCTGTGAAGGCATATGTCCACTCAAGTGCATTTTACTATAGATGGAG
AAAGCGGTCGTAGCTGTAGGCAGAAGAGTGCTAAAATTTGTGAAACCAAACTTTCT
TGTATTTTTACATGCATAGTATATTTGTATTCAATTGAAGAAATACTTTAGTTACA
ATGTAAAAATCCCTCTGCTCCATACTGGCTTCTTGTGAAGGAAAGCATTTGCAAACA
AATCACTATTTCTGTATATCTCTAAGAGTGTGGTACTAGGCTAACAAGCTAATTTC
ATAAAAATAACATCTCTTTCCATTAACCCCGCCAAATAAAATTTATAATATTAACTT
TTAATAAAATTTAAGGTACTAACCGTTaaacggactggagagttttcttggggtttacagcttcaaata
ctttcaaatctagctaaaataactgaagatacaatcgtgagtaacttcatggaatttaatgctacagtatgtcactggttaa
aaaaaaaaaaggttattttgcctggcttcactgcttccattaacagtataaaagattgtcttcattcctgcctggaactcaa

Figure 2-6. **CRISPR targets within the Dax-1 gene (second exon)**. Sequence of the second exon of the DAX1 gene, with the target for guide 6 (G6) underlined, in italics, and in pink. Intron/non-coding nucleotides are in lowercase, and exon/coding regions are in uppercase.

Results

MCF7 breast cancer cells were successfully treated to express DAX-1, as verified by both standard PCR run through an agarose gel (Figure 2-7A) and quantitative PCR analysis (Figure 2-7B). Knock-in expression was nearly four times the relative untreated expression when normalized with GAPDH using delta Cq calculations, and was considered to be statistically significant with a p-value less then 0.001. These samples were used for further analysis in Chapter 3.

MCF10A cells experienced a reduction in DAX-1 mRNA expression with siRNA treatment using the three transcripts listed in Table 2-1, as verified by standard PCR in agarose gel (Figure 2-8A) and quantitative PCR (Figure 2-8B). Relative expression was reduced to negligible amounts through delta Cq calculations and was statistically significant with a p-value of less than 0.005. These samples were used for further analysis in Chapter 3.

Mouse embryonic stem cells were treated with both siRNA and CRISPR-Cas9 methods to knock-down endogenous expression of Dax-1. The transient knock-down of Dax-1 expression using siRNA transcripts listed in Table 2-2 was verified through agarose gel electrophoresis of standard PCR samples (Figure 2-9A), quantitative PCR analysis (Figure 2-9B), and western blotting (Figure 2-9C). Using two concentrations of siRNA treatment, the cells treated with 20 pmol and 80 pmol of siRNA each experienced a knock-down to half the normal expression when compared to untreated and Scramble/negative control (NC) cells. Both values were statistically significant when compared to the untreated and negative control, with the 20 pmol sample having a p-value of less than 0.005 and the 80 pmol sample

having a p-value less than 0.001. These samples were used for further analysis in Chapter 4.

Development of a pX459 plasmid complete with an inserted guide for Dax-1 CRISPR-Cas9 knock-down in mESC is shown in Figure 2-10. Based on the design of the insertion site (Figure 2-4B), successful insertion of a guide sequence results in the loss of the BbsI restriction enzyme target site. Comparison of a complete plasmid with insertion and one with no insertion is shown in Figure 2-10A. Loss of the BbsI site results in a single linear product when digested with both BbsI and AgeI. However, lack of an insertion preserves the BbsI site resulting in two fragments, including a 1,000 base pair fragment (detectable in Figure 2-10A).

Initially, the six guides, illustrated in Figures 2-5 and 2-6 and listed in Table 2-3, were screened for proper insertion. Out of the six, guide 3 and 6 passed the double digest screen and showed no 1kb dropout fragment, and were subsequently verified by DNA sequence analysis (Figure2-10A and B). Following transfection with the verified plasmids and puromycin screening, several individual mESC colonies that survived treatment were isolated. Whole flask samples (labeled with the letter A) and isolated colonies (labeled from 1 to 10) were used for further Dax-1 knock-down verification from the two treatments of G3 and G6 CRISPR-Cas9 guides that were selected for. These were tested using standard PCR in agarose gel (Figure 2-11A), quantitative PCR (Figure 2-11B), and western blots (Figure 2-11C). Based on the results, isolates G3-1 and G6-1 were used for further analysis in Chapter 4.
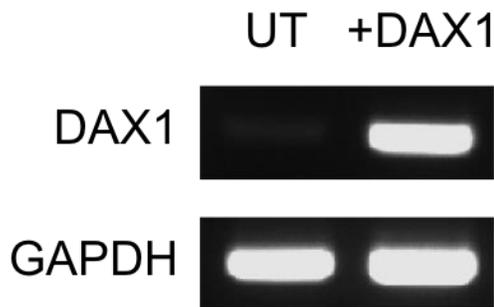
A.   B.
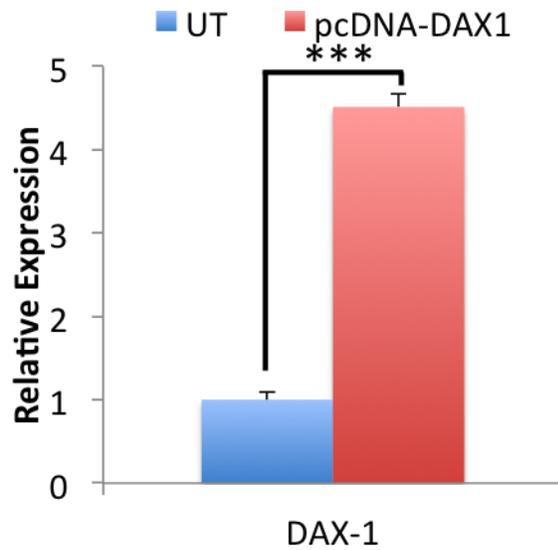
Figure 2-7. Introduction of DAX-1 expression in MCF7 cells by transient transfection with pCDNA DAX-1. A. Agarose gel electrophoresis of PCR samples comparing DAX-1 expression in untreated and transfected samples (GAPDH is shown as a positive control). B. qPCR results comparing the relative expression of DAX-1 mRNA in untreated and transfected cells. Error bars on qPCR results represent standard deviation of the mean.
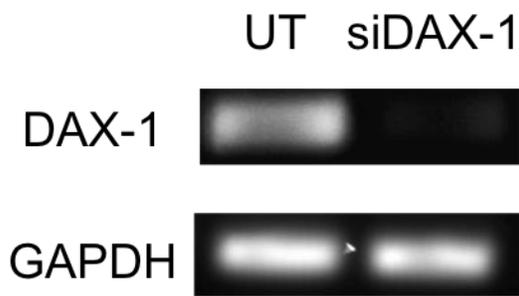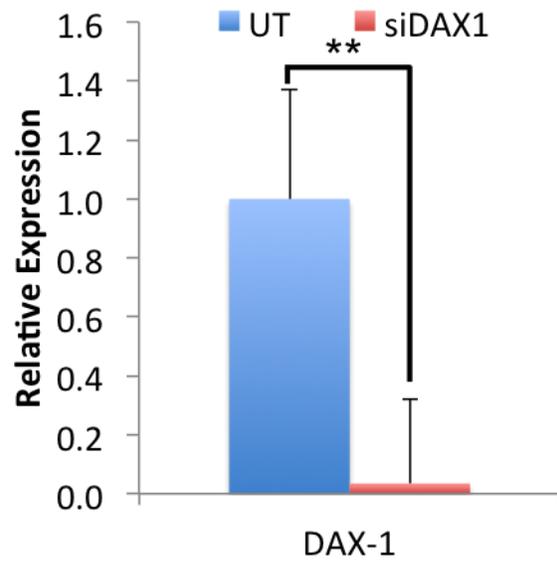
Figure 2-8. siRNA DAX-1 knock-down in MCF10A. A. Agarose gel electrophoresis of PCR samples comparing DAX-1 amounts in untreated and transfected samples (GAPDH is shown as a positive control). B. qPCR results comparing the relative expression of DAX-1 mRNA in untreated and transfected cells. Error bars on qPCR results represent standard deviation of the mean.

A.

B.

C.
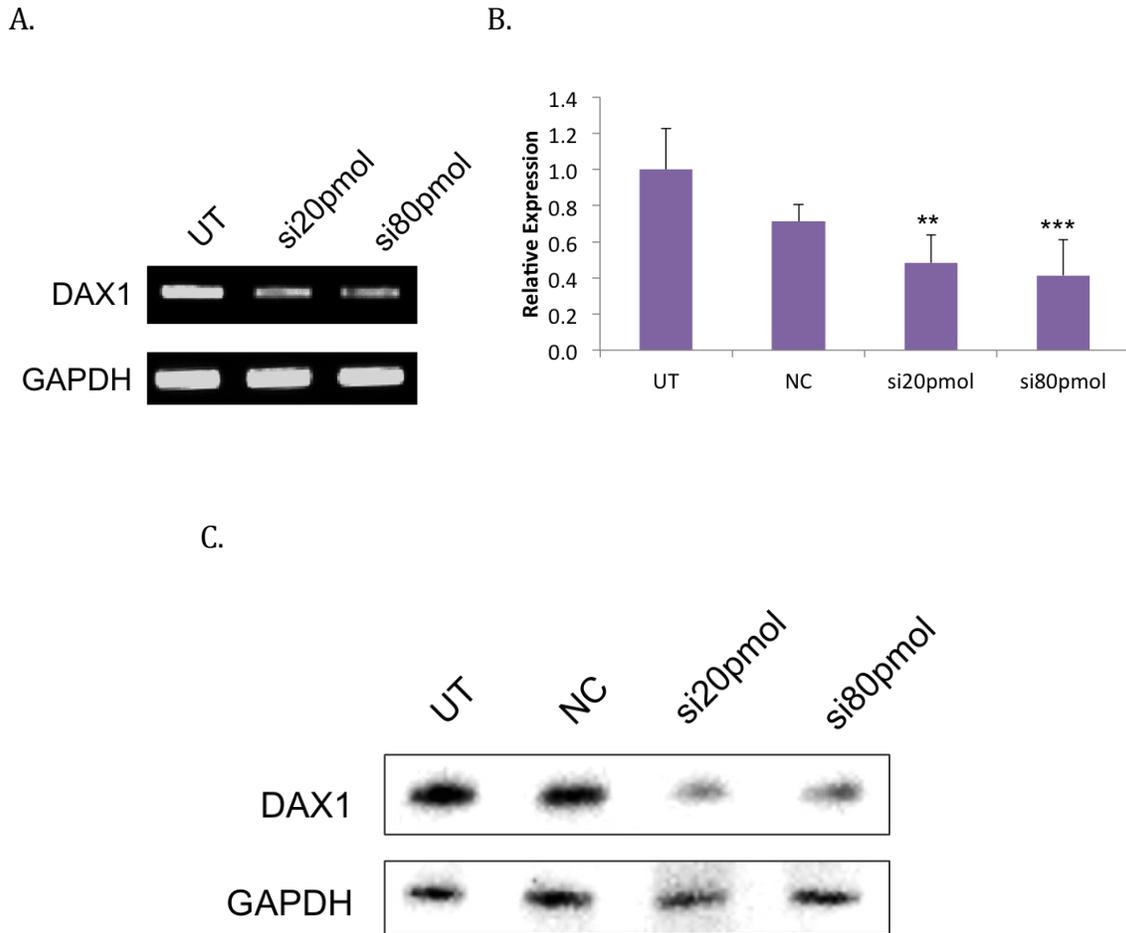
Figure 2-9. siRNA Dax-1 knock-down in mESC. A. Agarose gel electrophoresis of PCR samples comparing Dax-1 amounts in untreated and transfected samples (GAPDH is shown as a positive control). B. qPCR results comparing the relative expression of Dax-1 mRNA in untreated and transfected cells. Error bars on qPCR results represent standard deviation of the mean. C. Western blot of Dax-1 protein expression in untreated and transfected mESC. GAPDH protein expression is included as a loading control.

A.
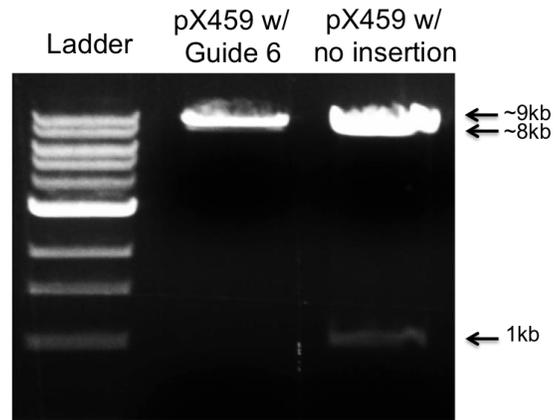
B.
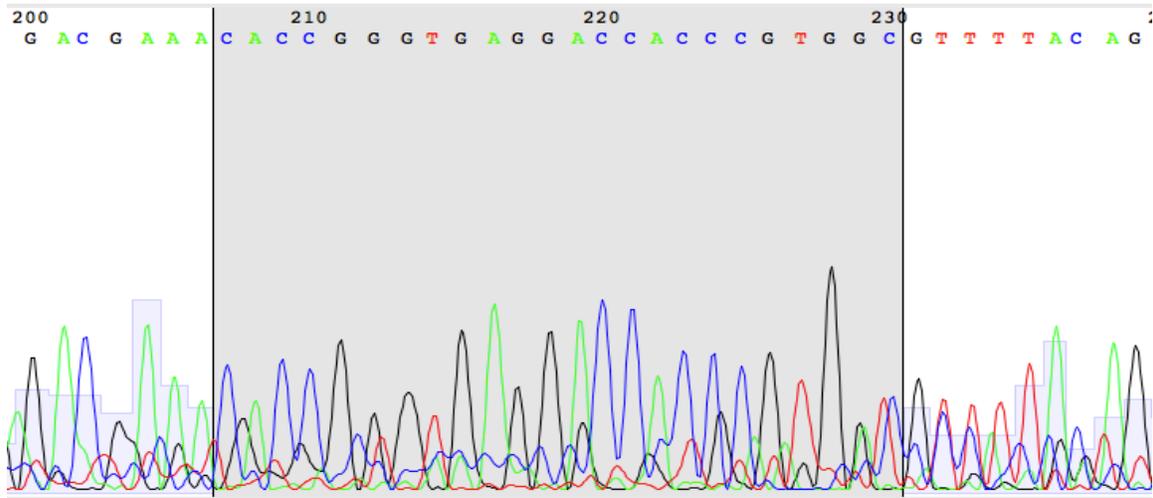
C.
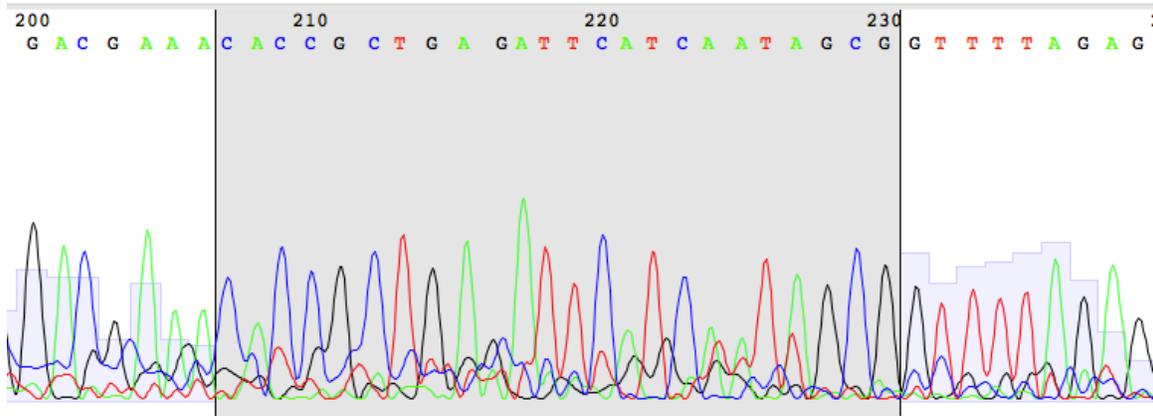
Figure 2-10. Dax-1 specific CRISPR-Cas9 construct design. A. Screen cloned samples for proper insertion through a double digest with restriction enzymes AgeI and BbsI. The guide insertion site has BbsI target sites, so without an insert, a double digest will show a 1kb fragment. With guide sequence insertion, loss of the BbsI target site will lead to a single cut at the AgeI site and result in a linear product around 9kb length. B. Sequencing results verifying guide sequence insertion (highlighted in grey) in a guide 3 sample. C. Sequencing results verifying guide sequence insertion (highlighted in grey) in a guide 6 sample.

A.
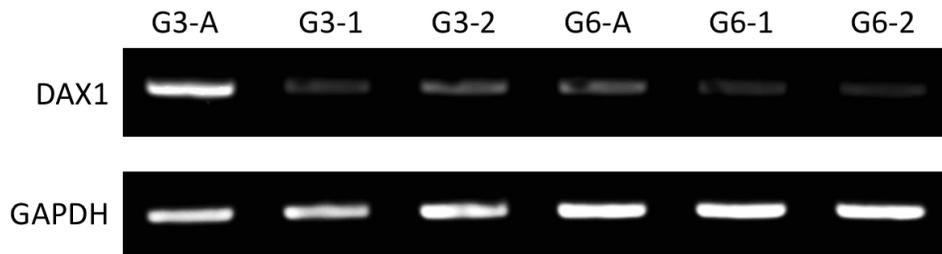
B.
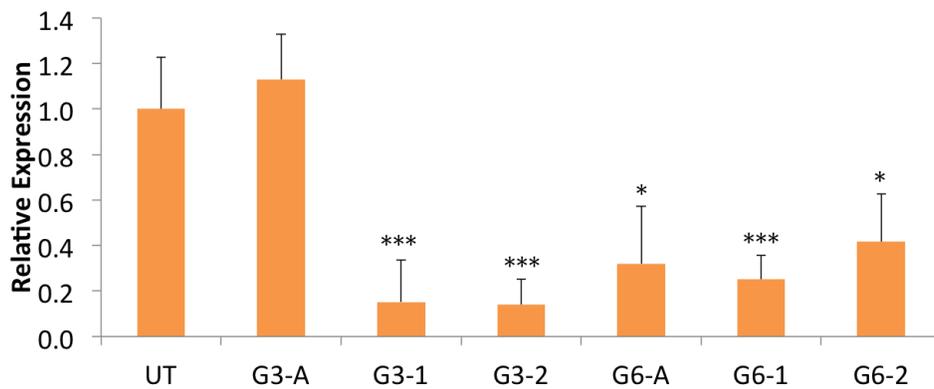
C.
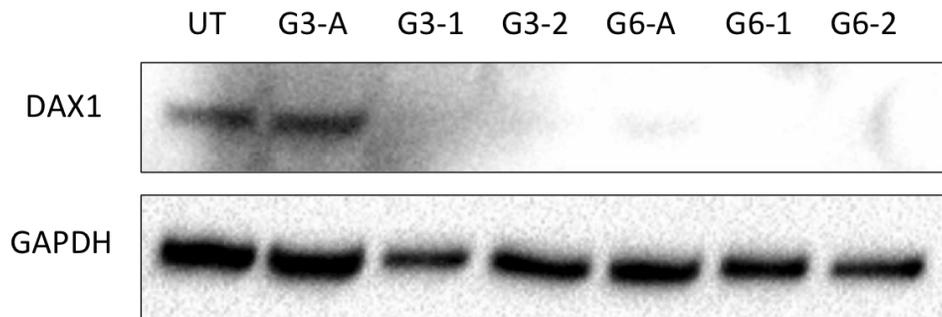
Figure 2-11. CRISPR Dax-1 knock-down in mESC. A. Agarose gel of PCR samples comparing Dax-1 amounts in untreated and treated G3 and G6 samples (guide targets illustrated in Figure 2-5 and 2-6) with GAPDH as a positive control, B. qPCR results comparing the relative expression of Dax-1 mRNA in untreated and G3 and G6 transfected cells. Error bars on qPCR results represent standard deviation of the mean. C. Western blot analysis of Dax-1 protein expression in untreated and G3 and G6 transfected cells. GAPDH is included as a loading control.

Discussion

The successful treatment of the human cell lines MCF7, with the introduction of DAX-1 expression through a pcDNA3.1 vector, and MCF10A , with the knock-down of DAX-1 expression through siRNA, was carried out to see the effect of mirrored DAX-1 modification in the same cellular background yet with different endogenous DAX-1 expression levels. This system allows for subsequent analysis of downstream targets within specific pathways in Chapter 3.

The treatment of E14 mESC with both siRNA and CRISPR-Cas9 allowed for two different methods with the same outcome – knock-down expression of endogenous Dax-1. Unlike the human mammary epithelial cell model, there is no variation in wild-type stem cells and Dax-1 is endogenously expressed in relatively high amounts. Each method uses different mechanisms for knocking down gene expression, and using both will allow any possible off-target effects of an individual method to be disregarded. The results of these treatments to mESC are used for subsequent analysis of candidate Dax-1 target genes in relation to pluripotency in Chapter 4.

Chapter 3

DAX-1 Mediates Expression of Apoptosis-Inducing Genes in Human Breast Cancer

Introduction

Amongst the various human diseases, cancer poses a great threat to the human population. As the life expectancy rises, so does the risk of cells functioning abnormally within the human body. The umbrella term of cancer mistakenly simplifies the complexity of the disease as a whole, which affects nearly every type of cell. It is expected that there will be about 1.5 million new diagnosed cases in the United States in 2015. Estimates for 2015 in the United States alone put breast cancer as the second leading cause of cancer death for women, with about 40,000 expected cancer deaths (Figure 3-1). Research of specific types of cancer is made possible with cell culturing of commercially available cell lines, such as MCF7 [41], that are well characterized and give great insight into the mechanisms and changes that contribute to onset and persistence of cancer.

An early study linking DAX-1 to androgen receptor activity in human breast cancer cells saw a significant decline in cell proliferation with the introduction of exogenous DAX-1 into MCF7 [42]. Additionally, experiments in the Tzagarakis-Foster lab brought to light the connection of DAX-1 to the decline of cell survival. Studies within our lab have confirmed that DAX-1 affects cell growth and survival from the aspect of proliferation through interaction with estrogen receptor, cyclin D, and Ki-67 [43]. Results from cell proliferation assays not only showed a decline in growth rate, but an overall decrease in cell population (Tzagarakis-Foster et al,

unpublished), indicating that cell death is another cause of the decline of survival as well. If DAX-1 was to be involved as a transcription factor with this cellular process, it is likely that it contributes to apoptosis, or programmed cell death.

Results observed in MCF7 cells prompted further investigation in an opposing model system with DAX-1 highly and endogenously expressed. The lung adenocarcinoma cell line A549 was treated with siRNA to knock-down DAX-1 expression, and the treated cells were compared with untreated cells in an apoptosis PCR array (Nguyen and Tzagarakis-Foster, unpublished). Analysis and comparison of the gene expression of the genes (Figure 3-2) provided in the array allowed for the selection of candidate DAX-1 targets that work within known cell death pathways. Candidate genes that had demonstrated a fold-change, either decrease or increase, greater than two-fold were selected for further analysis. The number of genes that showed a change in expression with knock-down further confirmed a strong degree of DAX-1 influence on apoptosis.

Estimated Cancer Deaths in the US in 2015

		Men 312,150	**Women** 277,280		
Lung & bronchus	28%			26%	Lung & bronchus
Prostate	9%			15%	Breast
Colon & rectum	8%			9%	Colon & rectum
Pancreas	7%			7%	Pancreas
Liver & intrahepatic bile duct	5%			5%	Ovary
Leukemia	5%			4%	Leukemia
Esophagus	4%			4%	Uterine corpus
Urinary bladder	4%			3%	Non-Hodgkin lymphoma
Non-Hodgkin lymphoma	4%			3%	Liver & intrahepatic bile duct
Kidney & renal pelvis	3%			2%	Brain & other nervous system
All other sites	24%			23%	All other sites

Figure 3-1. Estimated Cancer Deaths in the US (Cancer Facts and Figures 2015, American Cancer Society).
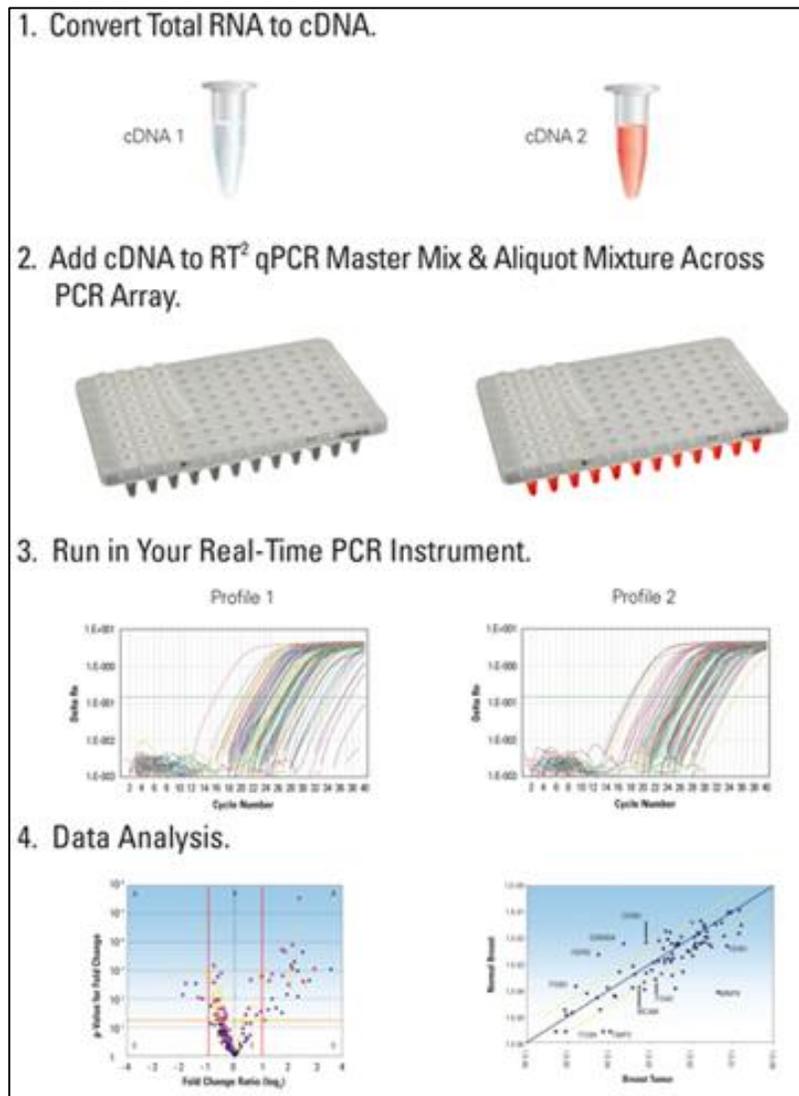
Figure 3-2. Workflow of PCR Array (SABiosciences). Comparison of an untreated and treated sample is done by running individual plates and measuring relative expression of candidate genes, then calculating the fold change difference between samples.

Apoptosis is an efficient way for an organism to maintain tissue homeostasis and to clear old or aberrant cells. [44]. Evading apoptosis is one of the chief hallmarks of cancer [45]. Altering major apoptotic pathways gives cancerous cells an advantage over natural and artificial stressors that typically activate a cell death cascade. The cellular pathways that lead to cell death can be described as either extrinsic or intrinsic. Extrinsic programming is caused by extracellular death ligands interactions with transmembrane protein receptors. Intrinsic programming is triggered by intracellular signals, such as elevated oncogene signaling and DNA damage, or upstream activity from extrinsic signaling pathways. While these pathways differ in their different specific molecular targets, they converge and lead to morphological changes within the cell as it begins to die. Cells going through early apoptosis can be identified through light microscopy by their characteristic shrinkage and visible chromatin condensation. When going through the later stages of apoptosis, the plasma membrane separates from the cytoskeleton causing the cell to start blebbing, organelles get fragmented into apoptotic bodies, and the dead cell is then consumed by macrophages through phagocytosis [44].

The breakdown of internal materials can be categorized into two pathways, distinguished by the use of caspases. The caspase-independent Granzyme A pathway is carried out through single-stranded DNA damage [46]. The more common pathway depends on caspases, the shortened name of cysteine-aspartic proteases, that are generally present as inactive proenzymes in a living, active cell [47]. Among the twelve known caspases, five are clearly involved with apoptosis. The five caspases are subcategorized by their mechanism of action: caspases 8 and 9

are initiators while caspases 3, 6, and 7 are executioners [48]. Once activated, these enzymes cleave proteins at aspartic acid residues, with each one specific depending on the neighboring amino acids [49] [50]. The inactivation and degradation of key proteins leads to the breakdown of internal components, and eventual breakdown of DNA through CAD, or caspase-activated DNase [51].

The results from the knock-down of DAX-1 saw a change in expression of genes that fell within two major apoptotic, caspase-dependent pathways. The first category fell under the TNFα pathway, an extrinsic pathway that influences both cell proliferation and apoptosis. The TNFα ligand interacts with two receptors: TNFR-1, also known as p55, and TNFR-2, also known as p75. TNFR-1 mediates most apoptotic signaling [41], as seen in Figure 3-3, while TNFR-2 is upstream of anti-apoptotic, cell proliferative, and cell survival signals such as NF-κB [52]. Upon interaction of the ligand with a death receptor like TNFR-1, the cytoplasmic "death domain" of the extracellular receptor recruits adapter proteins such as TRADD, FADD, and RIP. This recruitment associates with procaspase-8, resulting in the activation into caspase-8 and leading to downstream apoptotic effects [53].
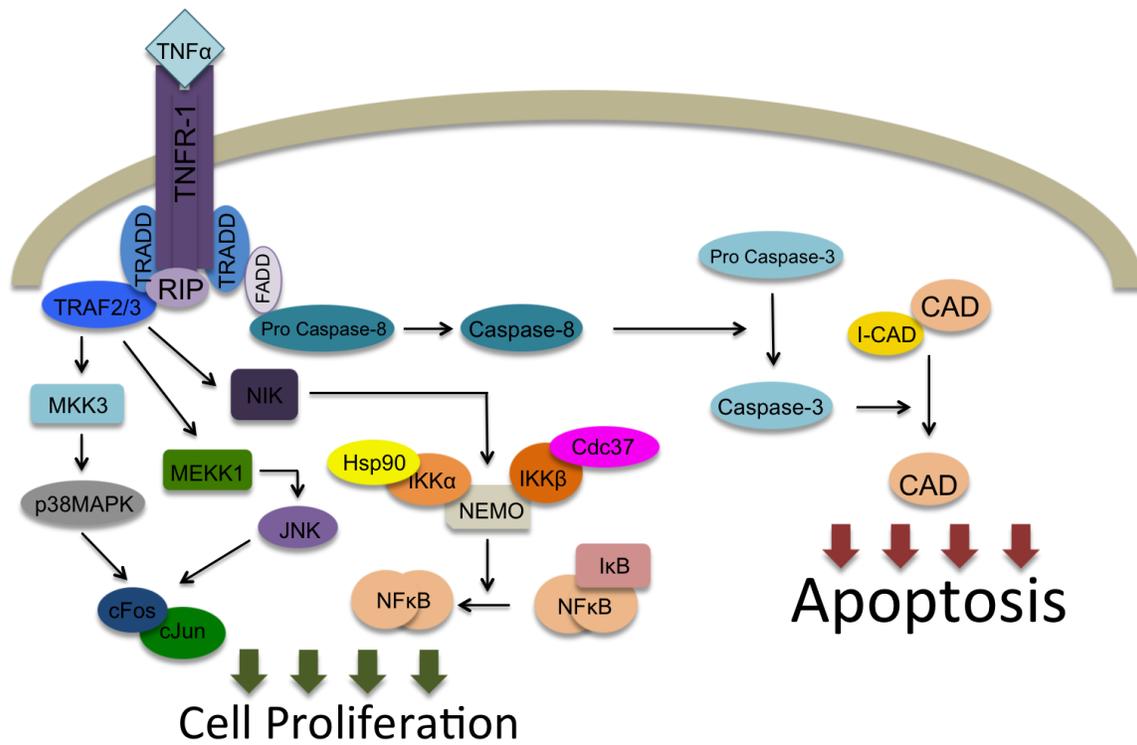
Figure 3-3. TNFα pathway, with the major interactions leading to either cell proliferation or apoptosis through the triggering of the caspase cascade.

The second general category of genes that saw a change in expression in response to reduced DAX-1 was the BCL-2 family. This is a major intrinsic pathway, illustrated in Figure 3-4, that is strongly linked to programmed cell death, and is a major target for therapy due to its sole involvement with apoptosis alone [54]. There are three categories of proteins based on function and structure. Four distinct domains within the BCL-2 family, ranging from BH1 to 4, relate to their interactions with one another [55]. The first, most upstream members of the BCL-2 family are the BH3-only domain proteins. These proteins, such as BNIP3 and BID, are only able to interact with isoforms of BCL-2 [56]. By inhibiting anti-apoptotic members of BCL-2 and activating pro-apoptotic members, BH3-only proteins cause mitochondrial outer membrane permeabilization (MOMP), which releases internal proteins of the mitochondria and leads to the caspase cascade described earlier [57].

As mentioned earlier, while these two categories are distinct, they ultimately converge to result in the outcome of cell death. There is some crosstalk between the two different pathways when deciding on the activation of caspases. Ultimately, the genes were investigated by organizing them based on their function, rather than by protein family. The genes analyzed that inhibit apoptosis when expressed, or anti-apoptotic genes, are listed in Table 3-1, while genes that function to promote apoptosis, or pro-apoptotic genes, are listed in Table 3-2.
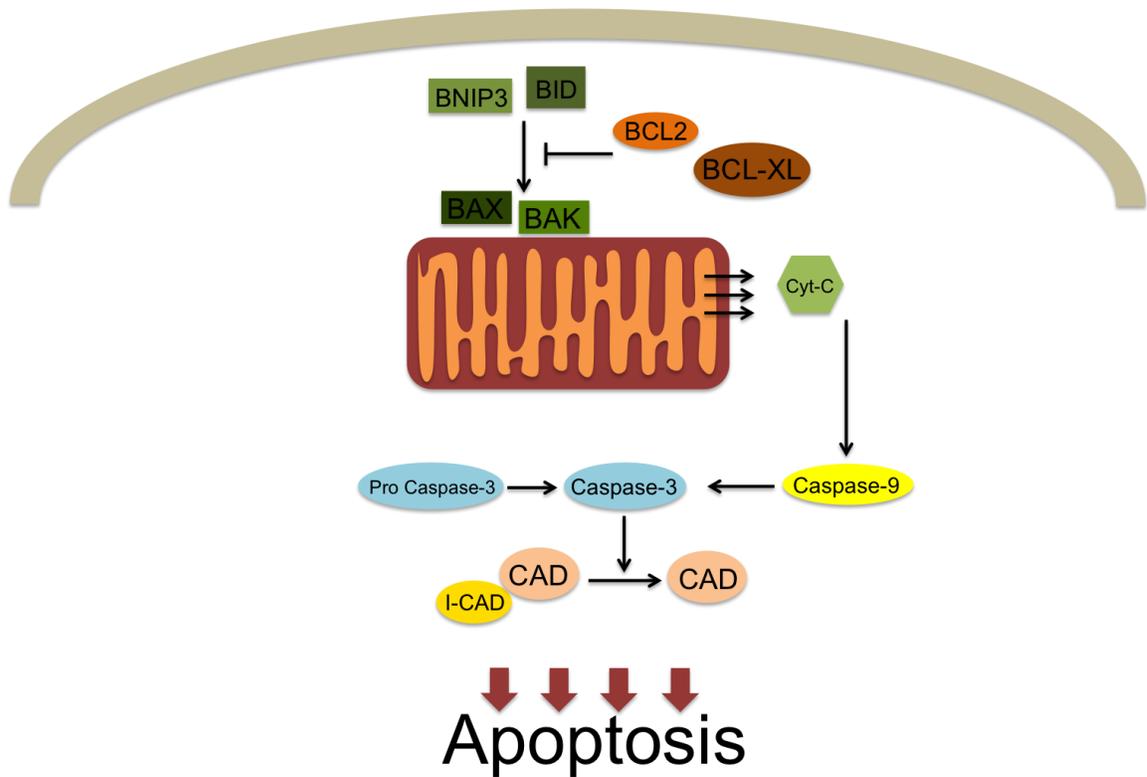
Figure 3-4. BCL-2 family apoptosis pathway. BH3-only proteins like BNIP3 and BID are able to activate BAX and BAK proteins to cause membrane permeabilization of the mitochondria and release internal components like cytochrome c that cause activation of caspases.

Table 3-1. Anti-apoptotic candidate target genes.

Gene Name	Function
TNFα	Encodes a protein that is a death receptor ligand; when it binds to TNFR2, activates both cell proliferation and apoptotic pathways downstream [52]
TRADD	Encodes a death domain containing protein that suppresses TRAF2 recruitment of inhibitor-of-apoptosis proteins (IAPs) [58], but also has anti-apoptotic function because of influence on NFκB [59]
TRAF1	Encodes a protein that associates with TRAF2 to form a complex needed to activate NFκB and activate IAPs [58]
BCL-XL	Large isoform of BCL-2, encodes a transmembrane protein that is present in the mitochondria, prevents the release of cytochrome c [56]
BCL-L2	Isoform of BCL-2 that prevents the release of cytochrome c [56]
BCL2-A1	Isoform of BCL-2 that prevents the release of cytochrome c [56]
BCL-L10	Isoform of BCL-2 that prevents release of cytochrome c [56]

Table 3-2. Pro-apoptotic candidate target genes.

Gene Name	Function
BCL-XS	The shorter isoform of BCL-2 that acts as an apoptotic activator by allowing for the release of cytochrome c [56]
BCL2-L1	One of the smaller isoforms of BCL-2 that allows for the release of cytochrome c [56]
BAX	A BH1 through 3 domain containing protein that associated with the membrane to release cytochrome c [56]
BAD	A BH3-only protein that forms a heterodimer with anti-apoptotic proteins to prevent the inhibition of apoptosis [56]
TRAF3	Activates the immune response and contributes to the lymphotoxin-beta receptor signaling complex, which induces NFκB activation and cell death [58]

Methodology

While the changes in the gene expression in the mentioned pathways were significant enough to suggest some influence, some sort of measurement of the actual occurrence of apoptosis with DAX-1 presence is required to confidently associate DAX-1 with apoptosis. Annexin V and propidium iodide staining can be used to quantify cells undergoing apoptosis through cell flow cytometry. During early apoptosis, the membrane protein phosphatidylserine is flipped to the exterior of the cell, which allows for Annexin V to bind and tag a cell. As a cell progresses through the breakdown of internal components, DNA is more readily accessible and stained by propidium iodide (Figure 3-5A). Cell flow cytometry allows for the comparison of the degree of staining with these two reagents to identify and characterize cells within a population sample (Figure 3-5B).

If a candidate gene displays a mirrored change in expression with changes in the amount of DAX-1 presence, chromatin immunoprecipitation (ChIP) is an assay that can be used to test for the presence of DAX-1 within the regulatory region of the target gene. One form of this technique, called X-ChIP, crosslinks proteins attached to DNA. With the proteins locked onto the genome, cells are lysed and the chromatin is sheared through sonication into smaller fragments for easier analysis. The target protein, still attached to DNA, is selected using antibodies or any other method specific to the protein. The DNA and proteins are then separated, and the fragments of DNA are used for further analysis (Figure 3-6).
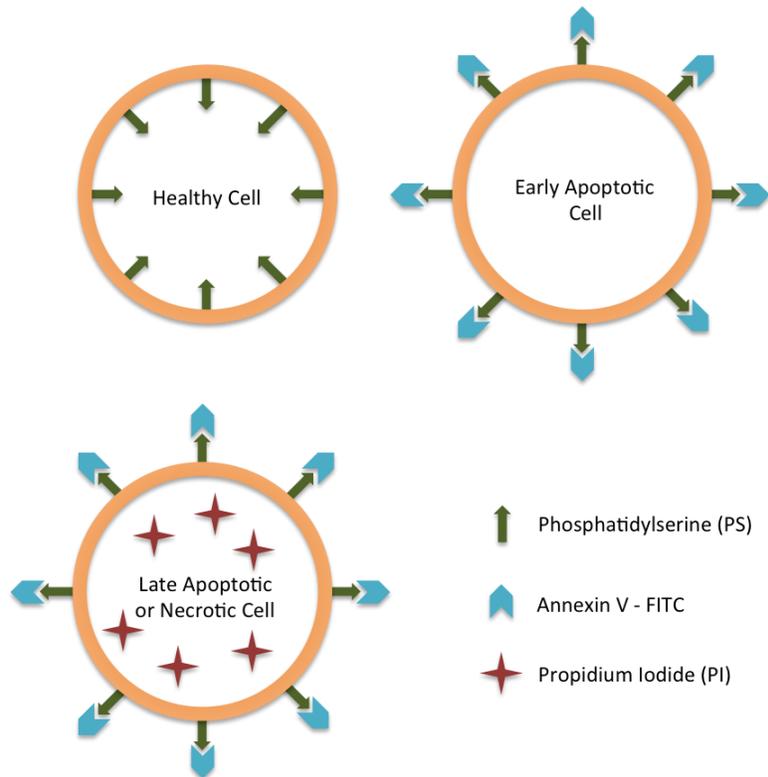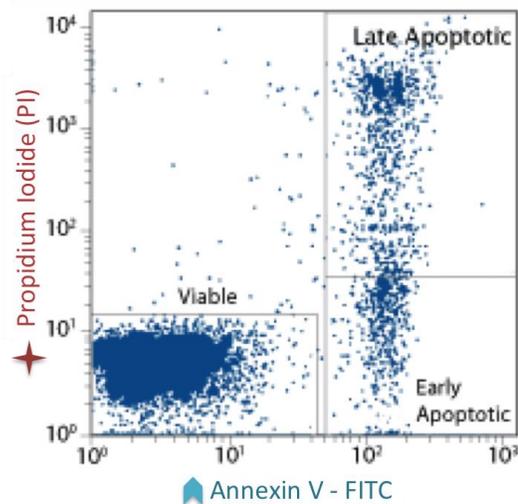
Figure 3-5. Annexin V and PI staining of apoptotic cells. A. The attachment of the dyes Annexin V-FITC and propodium iodide (PI) depend on the accessibility of the membrane protein phosphatidylserine (PS) and chromatin, respectively. Cells that are going through early apoptosis will have PS flipped outwardly, and cells going through late apoptosis or necrotic cells will also have chromatin readily stained with PI due to breakdown of the nucleus. B. Quantification of cells based on the amount of staining of both reagents is carried out by cell flow cytometry and graphed on the computer.
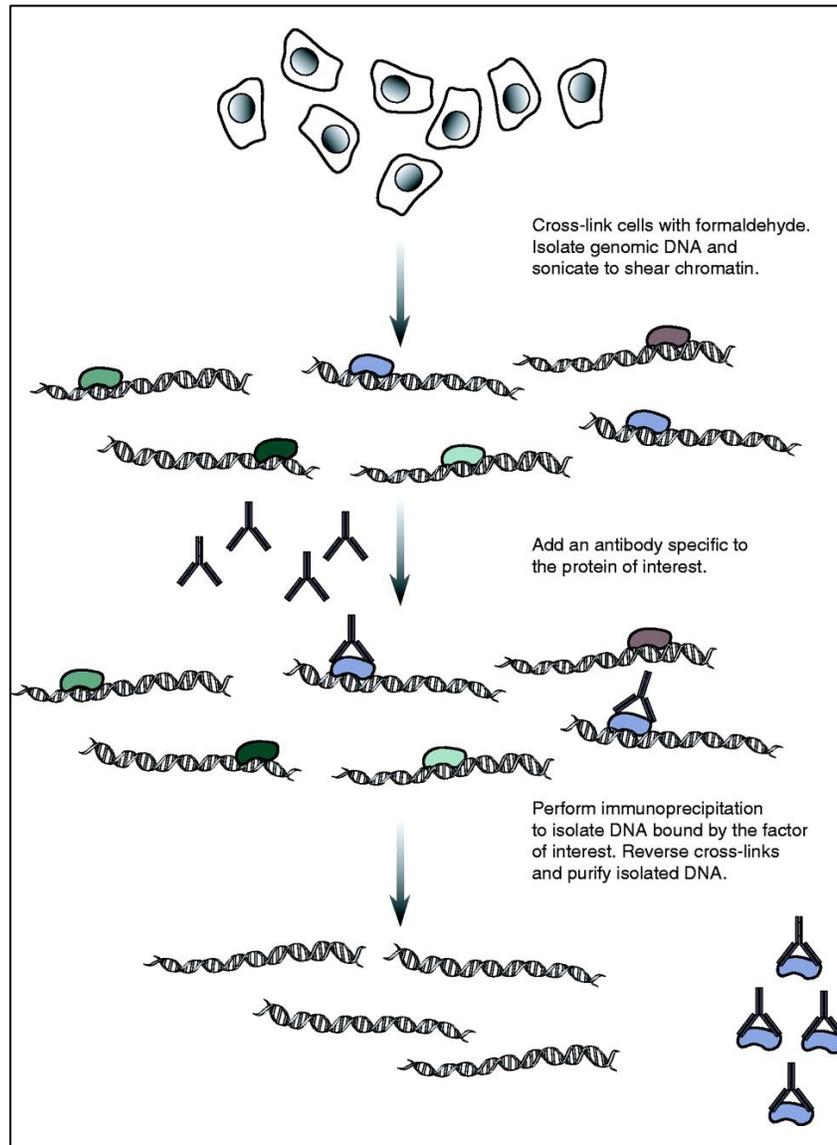
Figure 3-6. Workflow of X-ChIP Assay (AbCam). Untreated or treated cells are treated with formaldehyde in order to leave proteins in contact with genomic DNA to stay attached. Cells are then lysed and the chromatin is sheared to 300 to 500 base pair fragments. Antibodies are added, and samples are immunoprecipitated to isolate the fragments with protein(s) of interest. Proteins are released and the remaining DNA fragments are purified for downstream analysis.

Materials and Methods

RNA and cDNA extraction

MCF7 human breast cancer cells and MCF10A human normal breast cells, untreated and treated as stated in Chapter 2, were lysed and messenger RNA was collected using QIAGEN RNeasy Kit according to the manufacturer's protocol. Nucleic acid concentrations were measured using the GE Nanovue Plus Spectrophotometer. Complementary DNA (cDNA) was then synthesized from 1 μg of collected mRNA using various kits (High Capacity cDNA Reverse Transcription Kit from Applied Biosystems; M-MuLV Reverse Transcriptase from New England Biolabs; QuantiTect Reverse Transcription Kit from QIAGEN) according to the corresponding protocol.

Small Molecule Inhibitor

BADGE (Bisphenol A diglycidyl ether; 2,2'-[(1-Methylethylidene)bis(4,1-phenyleneoxymethylene)]bis-oxirane), a TNFα inhibitor that is known to induce apoptosis in tumor cells, was prepared to a working concentration of 1 mM in 100% ethanol from the stock (Santa Cruz Biotech, cat no. sc-202487). 100 μL was used in each well of a twelve-well plate. Cells were treated for 48 hours before collection and flow cytometry analysis.

Flow Cytometry

MCF7 cells were plated in twelve-well dishes and either untreated, transfected with pcDNA-DAX-1 as stated in Chapter 2, treated with BADGE

alone, or BADGE and pcDNA-DAX-1. After collection, each treatment of cells was left unstained, stained with Annexin V alone, stained with PI alone, or stained with the combination of Annexin V and PI. Cells were quantified using the BD Accuri C6 Cytometer and its provided computer program, with a sample size of 10,000 cells per reading.

Polymerase Chain Reaction (PCR)

Annealing temperatures were optimized using Promega GoTaq Green Master Mix. Primers listed in Table 3-5 and any subsequent primers used were all ordered from Integrated DNA Technologies. Conditions in Table 3-3 were used, with optimized annealing temperatures for specific primer pairs used for step 3. Upon completion, samples were removed from the 4°C hold and stored until further analysis.

Table 3-3. Thermocycler conditions and steps for standard PCR.

Step	Temperature	Time
1	95°C	5 minutes
2	95°C	30 seconds
3	53-60°C	30 seconds
4	72°C	30 seconds
5	Repeat steps 2 to 4 x thirty times	
6	4°C	∞

Real-Time PCR (qPCR)

Quantitative PCR analysis was performed on untreated and treated cDNA
collected from MCF7 and MCF10A cells to analyze the relative gene
expressions of target genes, normalized with GAPDH expression. SYBR green
master mixes (BioRad and QIAGEN) and the respective primer pairs for the
genes listed in Table 3-5 were used for apoptotic gene analysis. Primers with
a source as "construct" were designed using NCBI Primer-BLAST using the
available cDNA sequences online. Each sample was prepared using 10 μL of
SYBR green master mix, 0.25 μL of the forward primer, 0.25 μL of the reverse
primer, 7.5 μL nuclease free water, and 2 μL of cDNA. All samples were run in
triplicate for 40 cycles using the two-step amplification program, outlined in
Table 3-4, with the BioRad CFX96 Touch Real-Time PCR Detection System.
Results were analyzed using the BioRad CFX Manager™ software and
Microsoft Excel.

**Table 3-4. Thermocycler conditions and steps in the two-step amplification
program for qPCR.** Fluorescent dye detection occurred at step 3.

Step	Temperature	Time
1	95°C	5 minutes
2	95°C	30 seconds
3	55-60°C	15 seconds
4	Repeat steps 2 and 3 x thirty-nine times	

Statistical Anaylsis

Data from the qPCR analysis are presented with the mean + standard

deviation of triplicate experiments. Statistical significance is calculated in the

same manner described in Chapter 2.

Table 3-5. Primers used for apoptosis qPCR analysis.

Gene	FWD	REV	Source
TNFα	GGAGAAGGGTGACCGACTCA	CTGCCCAGACTCGGCAA	Construct
TRAF2	TGGCTGGCCGCATACC	TGTAGCCGTACCTGCTGGTGTA	Construct
TNFR1	TACATTGCAGCCTCTGCCTC	AGAGCTTGGACTTCCACCGT	Construct
TNFR2	ACATCAGACGTGGTGGTGCAA	CCAACTGGAAGAGCCAAGTC	[59]
TRADD	AGCTCATAGTGAACCGGCCA	TCCCTCAGTGCTCGACAGC	Construct
BCL-XS	CTGCACCTGACGCCCTTCACC	CACATGACCCCACCGAACTCAAAGA	Construct
BCL-XL	GATCCCCATGGCAGCAGTAAAGCAAG	CCCCATCCCGGAAGAGTTCATTCACT	Construct
BAX	ATGGAGCTGCAGAGGATGATT	TGAAGTTGCCATCAGCAAACA	Construct
BCL2-L10	GCCTTCATTTATCTCTGGACACG	GAAGGTGCTTTCCCTCAGTTCTT	Construct
BCL2A1	CCCGGATGTGGATACCTATAAGGAGA	GTCATCCAGCCAGATTTAGGTTCA	Construct
ERα	GACAGGGAGCTGGTTCACAT	AGGATCTCTAGCCAGGCACA	Construct

Western Blot

Protein lysates were collected from untreated and treated cells grown in six well plates according to protocol using M-PER Mammalian Protein Extraction Reagent (Thermo Scientific). Protein concentrations of collected cell lysates were determined by Bradford Protein Assay using a bovine serum albumin (BSA) dilution standard. Lysates normalized to 10,000 ng total per well were prepared for SDS-PAGE gel electrophoresis using Novex NuPage LDS and Reducing Agent (Thermo Scientific). Samples were electrophoresed through a NuPage 4-12% Bis-Tris precast gel at 200 V for 45 minutes in 1X MOPS running buffer and then transferred onto a PVDF membrane in NuPage Transfer buffer with 20% methanol. The PVDF membrane(s) were blocked in 5-10% Blotto in TBST prior to incubation with appropriate antibodies (Santa Cruz Biotechnology, Inc. and ActiveMotif) as listed on Table 3-6 in a 1:1000 dilution in Blotto. Blots were then treated with secondary mouse or rabbit antibodies based on the host of the primary antibodies diluted 1:2000 in 5% Blotto in TBST. Protein bands were developed using the Clarity ECL Western Blotting substrate kit (BioRad) according to protocol and visualized using a BioRad Gel Doc™ XR+ System with the ImageLab software.

Table 3-6. Monoclonal Antibodies used for Western Blotting.

Protein Target	Host Organism	Cat. No, Company
Beta-Actin	Mouse	sc-82278, Santa Cruz
GAPDH	Rabbit	3683, Cell Signaling
DAX-1	Mouse	39983, ActiveMotif
BCL-2 (XL isoform)	Rabbit	GTX127958, GeneTex
BAX	Rabbit	GTX109683, GeneTex

Chromatin Immunoprecipitation (ChIP)

MCF7+DAX1 cells were collected for treatment using the AbCam 500 ChIP kit according to protocol. Antibodies for the protein targets listed in Table 3-7 were used. Primers used to analyze the promoter regions of two possible targets, broken down into individual regions (Figure 3-7) were designed using NCBI's Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and are shown in Table 3-8.

Figure 3-7. Targeted regions of the A. BCL-2 and B. BAX promoter. Each region has an individual length of about 200 to 300 base pairs.

Table 3-7. Antibodies used for ChIP assay.

Protein Target	Host Organism	Cat. No, Company
Histone H3	Rabbit (polyclonal)	1791, AbCam
DAX-1	Mouse (monoclonal)	39983, ActiveMotif
ERα	Rabbit (polyclonal)	GTX100634, GeneTex

Table 3-8. Primer sets corresponding to promoter regions of BCL-2 and BAX.

Gene	Region	FWD	REV
BCL-2	Region 3	CCCCTCGTCCAAGAATGCAA	CATCTCCCGCATCCCACTC
	Region 4	ACGCCCCATCCAGCCG	CCCGCGCGGTGAAGG
BAX	Region 1	GGGTGGCTCAAGCCTGTAATC	GAGTGCAGTGGCCCAATCAT
	Region 2	GAGCCATGATTGGGCCACTG	CCAGGCAGGACGTTATAGAT

Results

Introduction of DAX-1 increases number of apoptotic cells in population

Cell flow cytometry was used to quantify the number of early and late apoptotic cells. The number of cells that fell in the right hand quadrants in Figure 3-4A were quantified in Figure3-4B. There was a significant increase of apoptotic cells in the MCF7-DAX-1 samples, specifically from 1.3% to 23.2%. Cells treated with BADGE alone saw a similar significant increase from 1.3% to 29.7%. There was no significant difference between singular treatment of BADGE or pcDNA-DAX-1 alone compared to BADGE and pcDNA-DAX-1 (31.3%) in combination.

A.
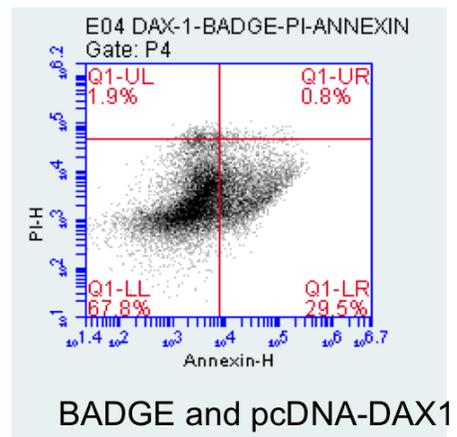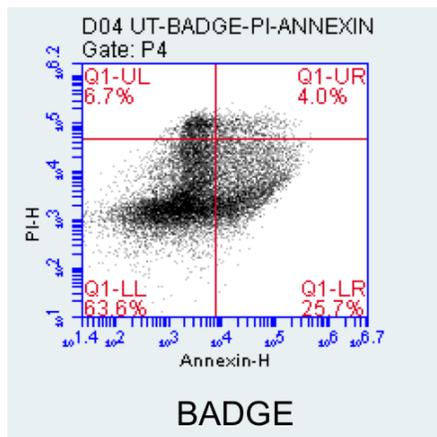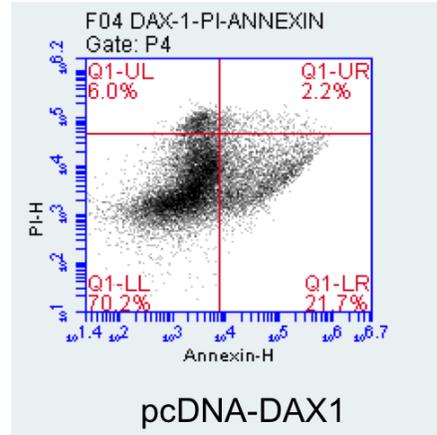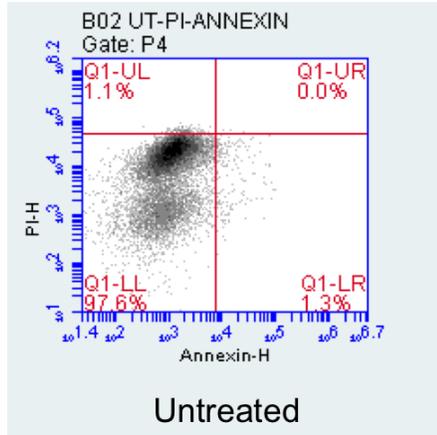
B.
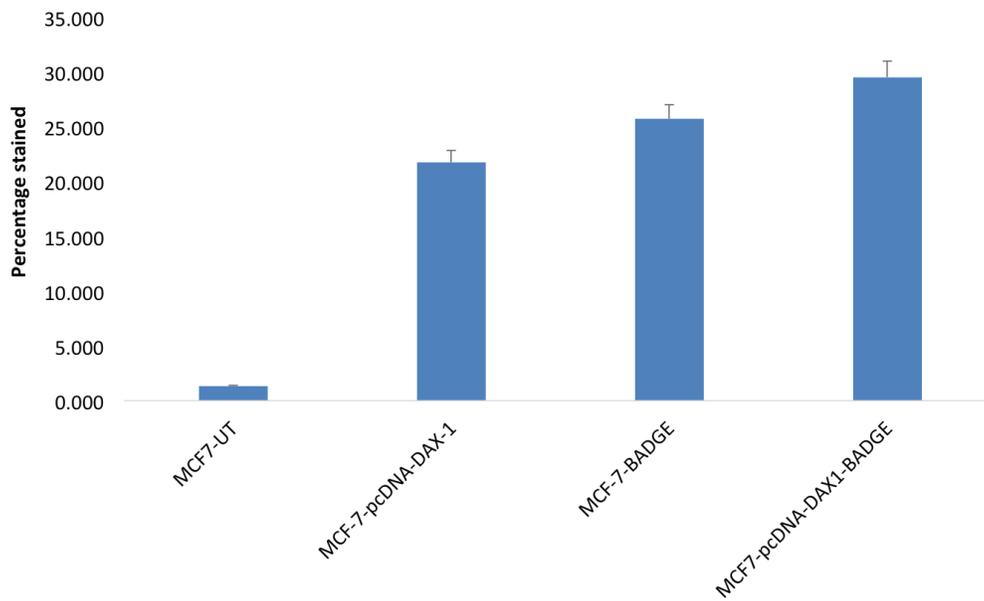
Figure 3-8. Cell flow cytometry of Annevin V and propidium iodide stained MCF7 cells untreated and treated with pcDNA-DAX1. A. Contour diagram of FITC-Annexin V/PI flowcytometry of MCF7 calls after 48 hour collection following transfection with pcDNA-DAX-1, and either no treatment or treatment with BADGE, or untransfected (UT) cells. B. Quantification of the right-hand quadrants from each cell flow cytometry sample to give value to the number of apoptotic cells with or without treatment. Error bars shown represent 5% error of shown value.

Changes in candidate genes within MCF10A DAX-1 knock-down cells

MCF10A cells that underwent siRNA treatment in order to transiently knock-down DAX-1 expression, treated as stated in Chapter 2, were lysed for collection of mRNA. Complementary DNA synthesized from the collected mRNA of treated and untreated cells were then analyzed by qPCR to compare gene expression changes. Among the anti-apoptotic candidate genes that were analyzed (Figure 3-9), BCL-L2 (A), BCL-XL (B), TNFα (D), and TRADD (G) increased in expression, while BCL2-A1 (C), TRAF1 (E), and BCL-L10 (F) decreased in expression. For the pro-apoptotic genes that were analyzed (Figure 3-10 A-E), BAX, BAD, BCL-XS, BCL-L1, and TRAF3 all showed varied degrees of significant decreased expression.
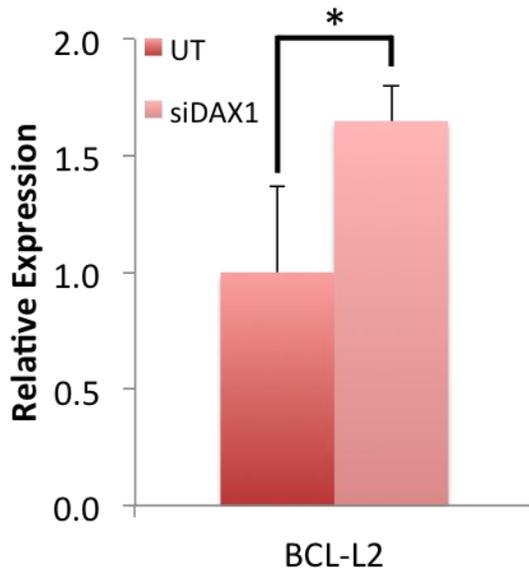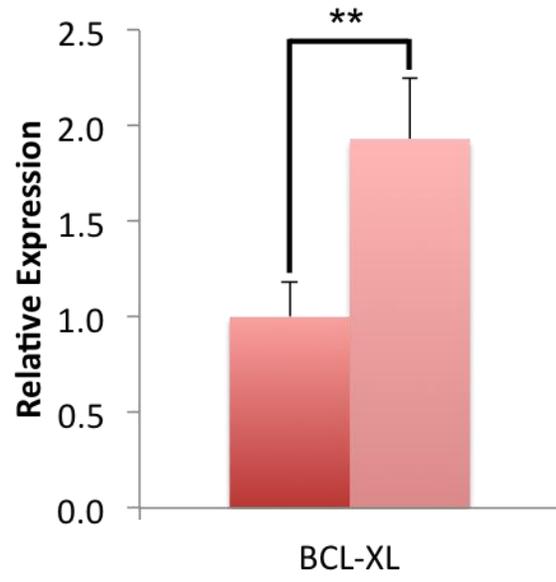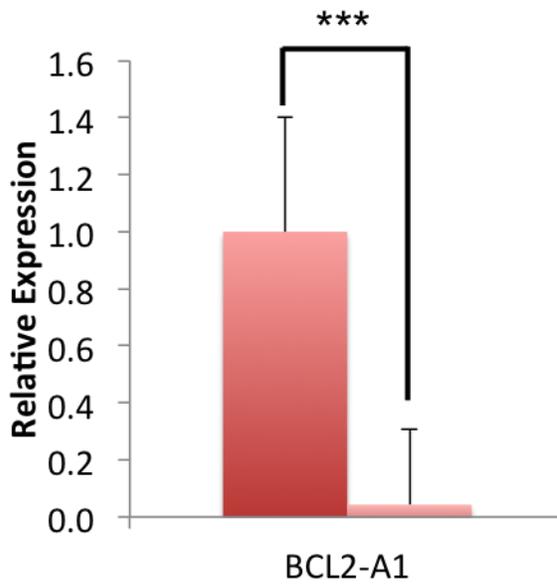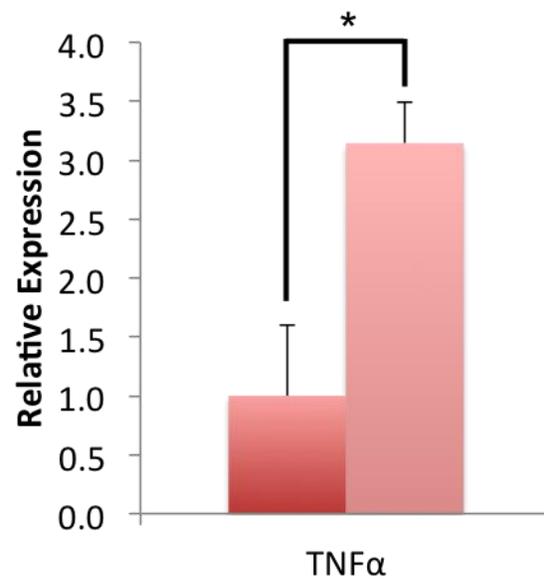
E.
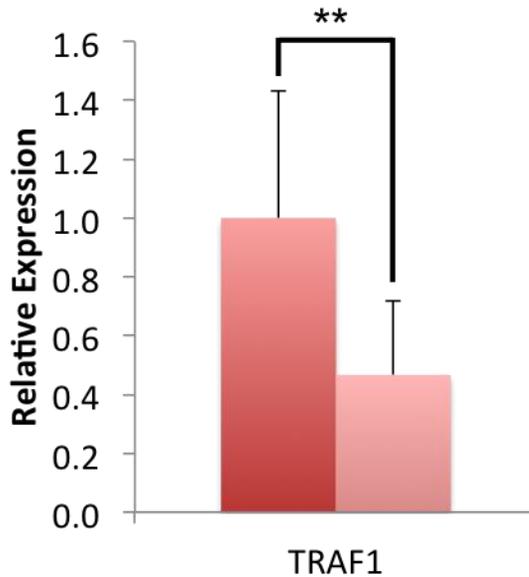
F.
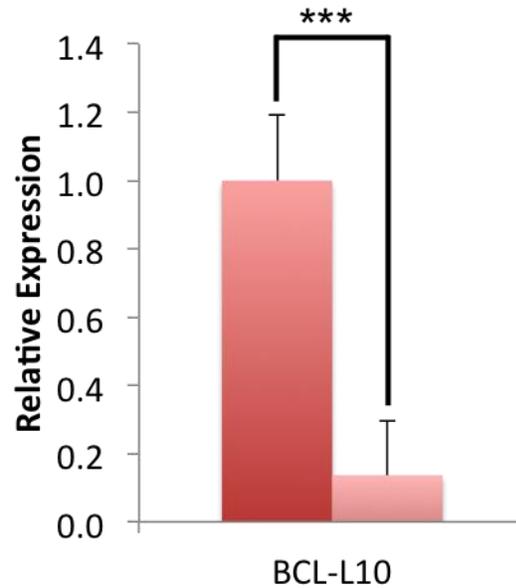
G.
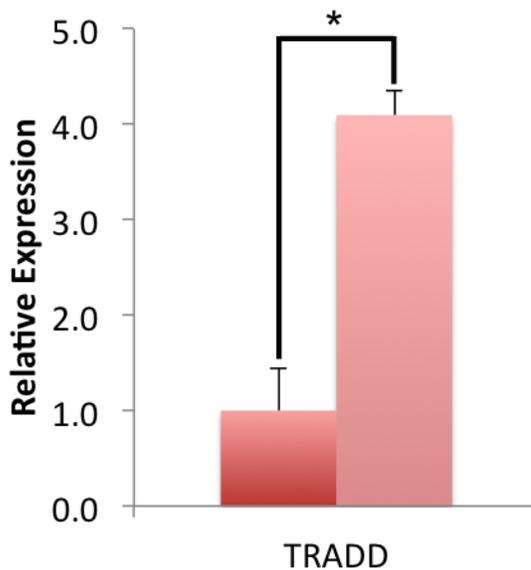
Figure 3-9. qPCR results of anti-apoptotic candidate genes tested in MCF10A knock-down total mRNA samples. Relative expression values were normalized using the expression of GAPDH to compare treated siRNA knock-down samples to untreated samples. Results shown correspond to candidate genes A. BCL-L2, B. BCL-XL, C. BCL2-A1, D. TNFα, E. TRAF1, F. BCL-L10, and G. TRADD. A single asterisk (*) represents samples that were $p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a triple asterisk (***) represents samples that were $p<0.001$. Error bars on qPCR results represent standard deviation of the mean.

A.
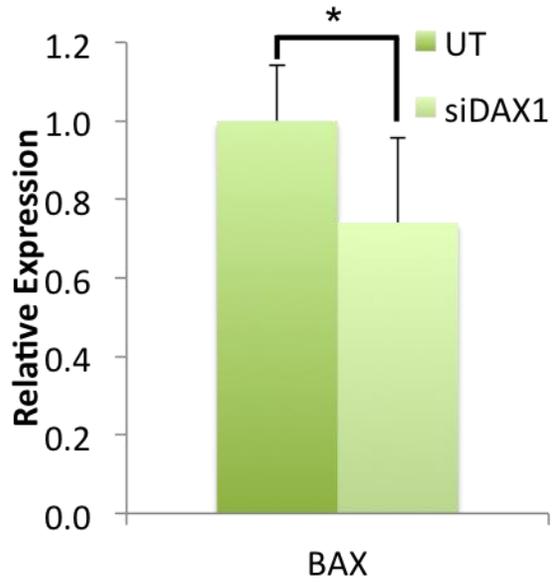
B.
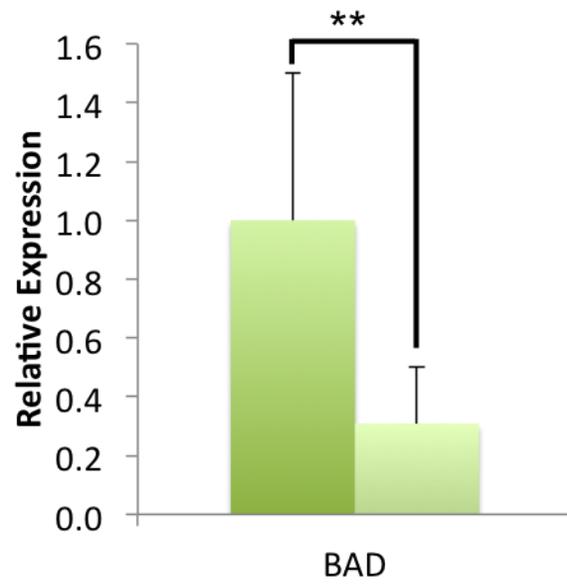
C.
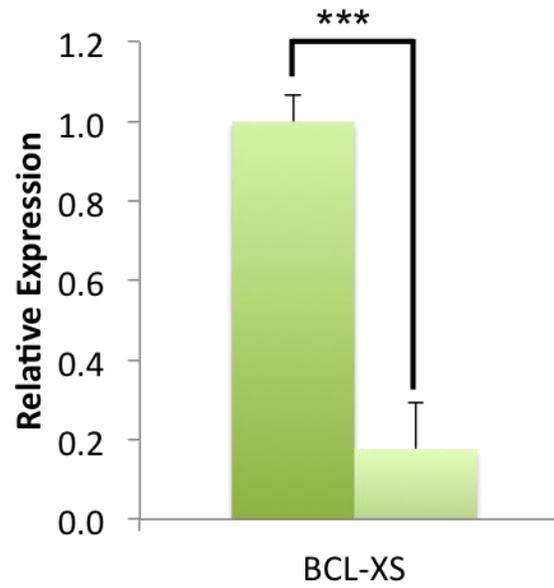
D.
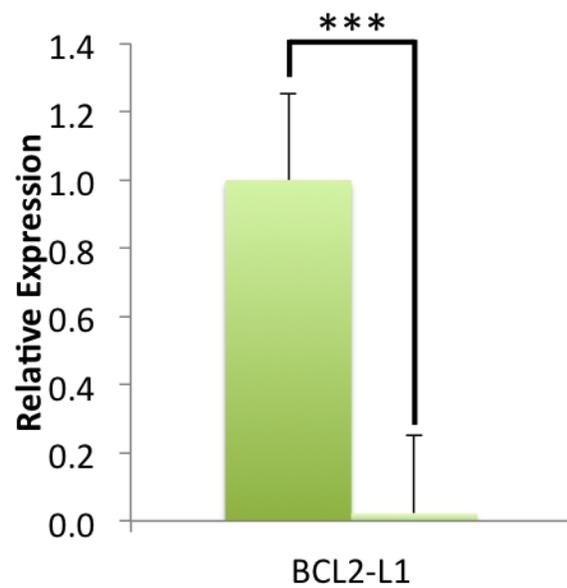
E.
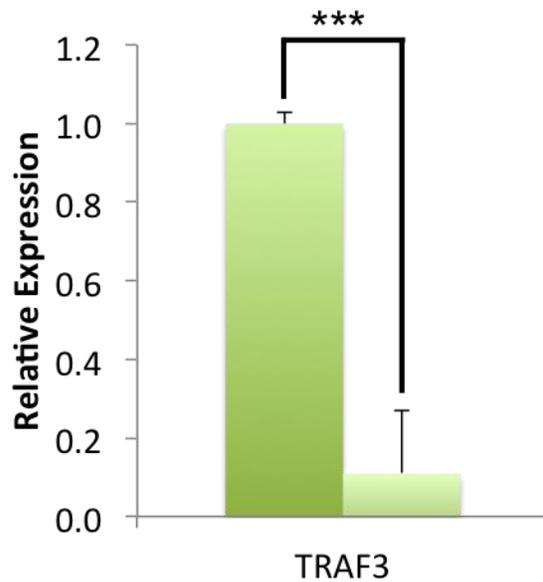
Figure 3-10. qPCR results of pro-apoptotic candidate genes tested in MCF10A knock-down total mRNA samples. Results shown correspond to the candidate genes A. BAX, B. BAD, C. BCL-XS, D. BCL-L1, and E. TRAF3. A single asterisk (*) represents samples that were $p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a triple asterisk (***) represents samples that were $p<0.001$. Error bars on qPCR results represent standard deviation of the mean.

Changes in candidate genes within MCF7 pcDNA-DAX-1 cells

The cDNA synthesized from collected mRNA of MCF7 cells, treated as described in Chapter 2 to express DAX-1, was analyzed by qPCR to compare gene expression changes of specific candidate genes. The gene expression of anti-apoptotic gene candidates BCL-XL, TRAF1, and TRADD (Figure 3-11 A, C, D) showed a significant decrease in expression with DAX-1 treatment , while BCL-L2 (Figure 3-11B) saw slight decrease but was not statistically significant. Of the pro-apoptotic genes analyzed (Figure 3-12), BAX (A) saw a significant increase in expression, while BCL-XS (C) and TNFR1 (F) saw a significant decrease. The genes BAD (B), BCL2-L1 (D), and TRAF3 (E) also saw a decrease in expression, but were not statistically significant.
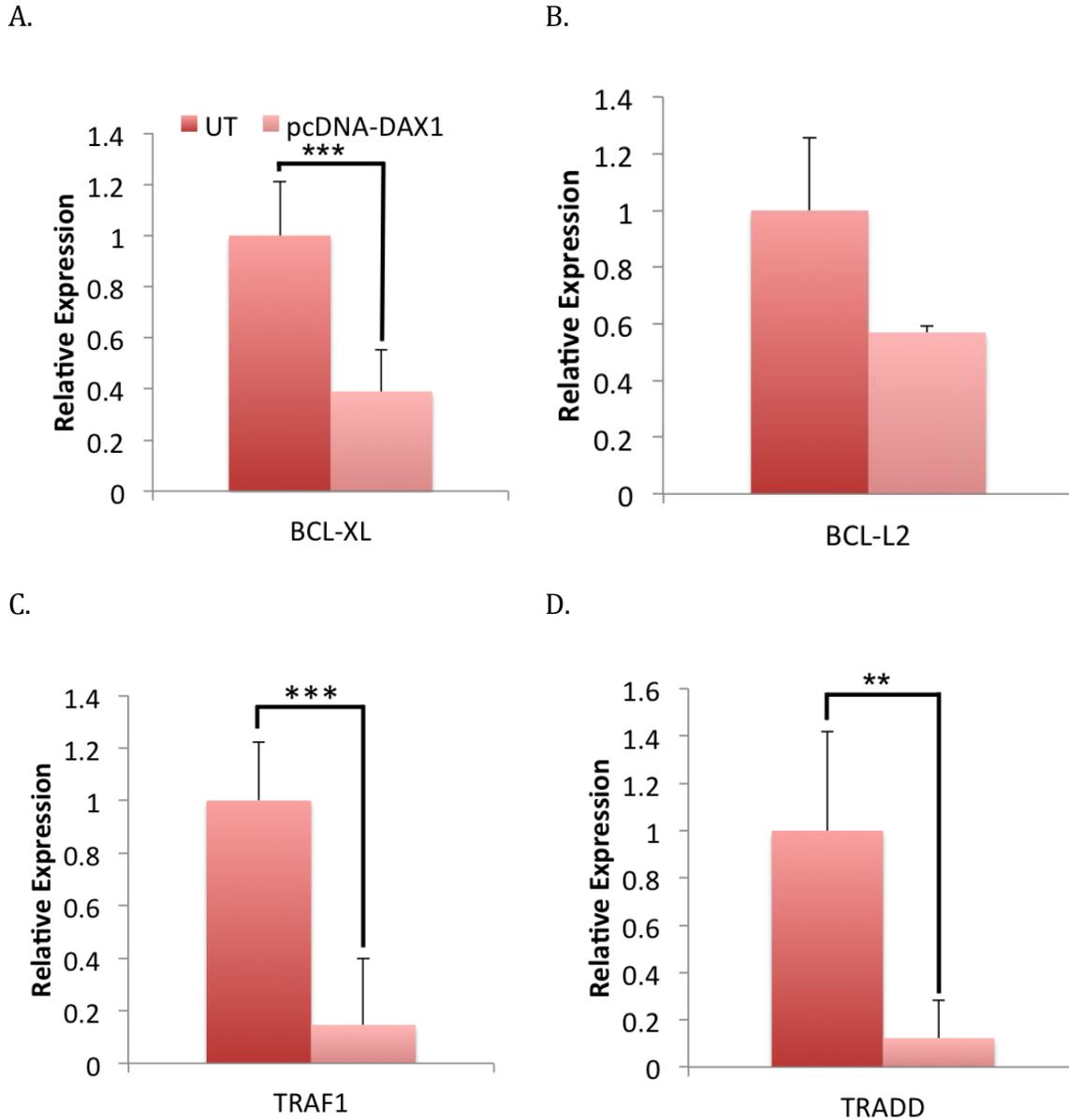
Figure 3-11. qPCR results of anti-apoptotic candidate genes tested in MCF7 knock-down total mRNA samples. Results shown correspond to the A. BCL-XL, B. BCL-L2, C. TRAF1, and D. TRADD genes. A single asterisk (*) represents samples that were $p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a triple asterisk (***) represents samples that were $p<0.001$. Error bars on qPCR results represent standard deviation of the mean.
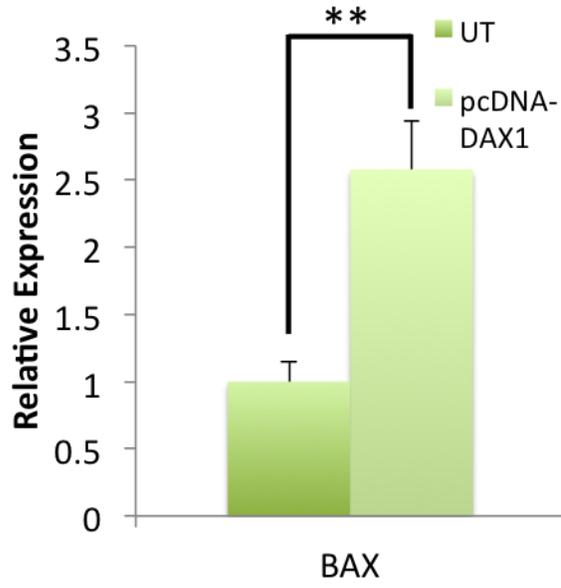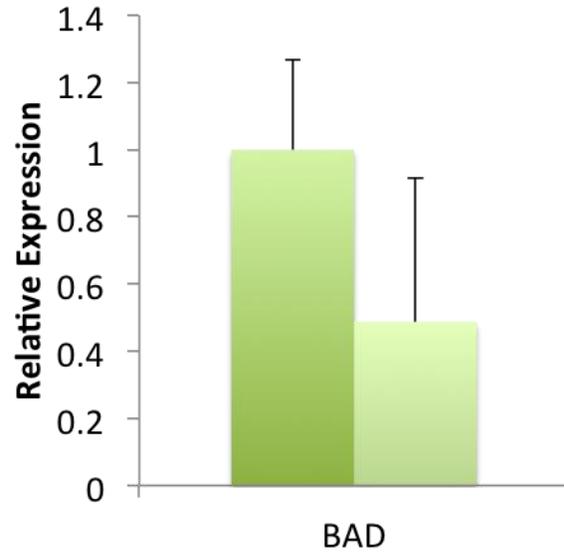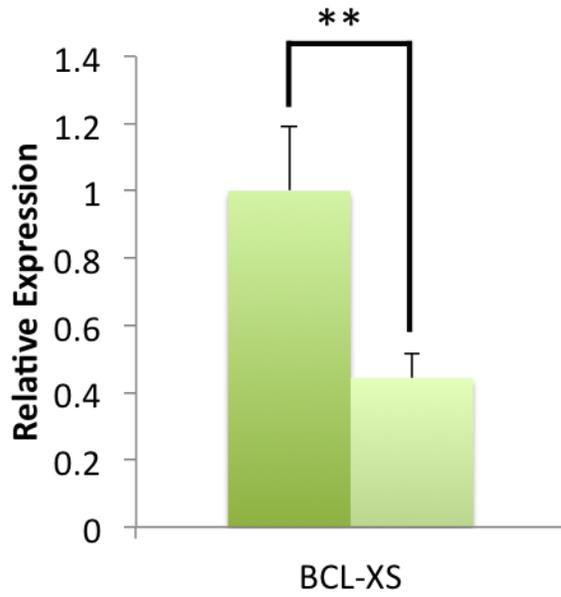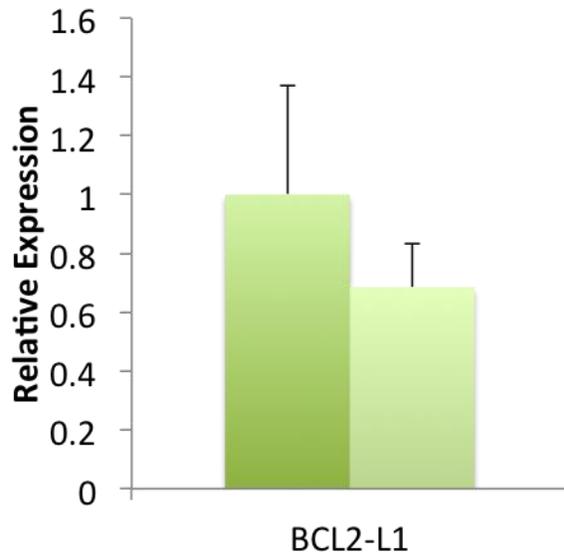
E.
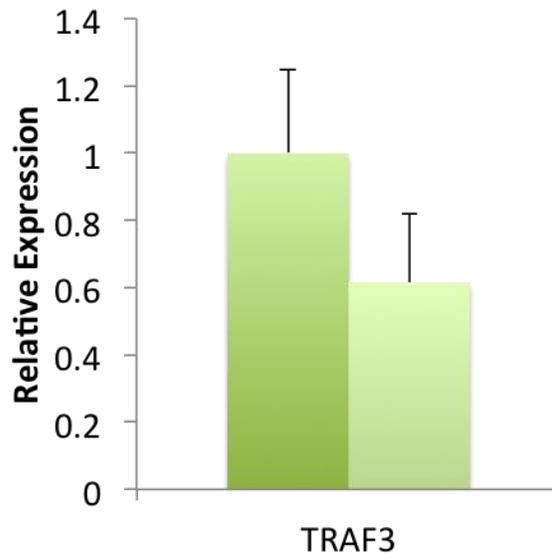
F.
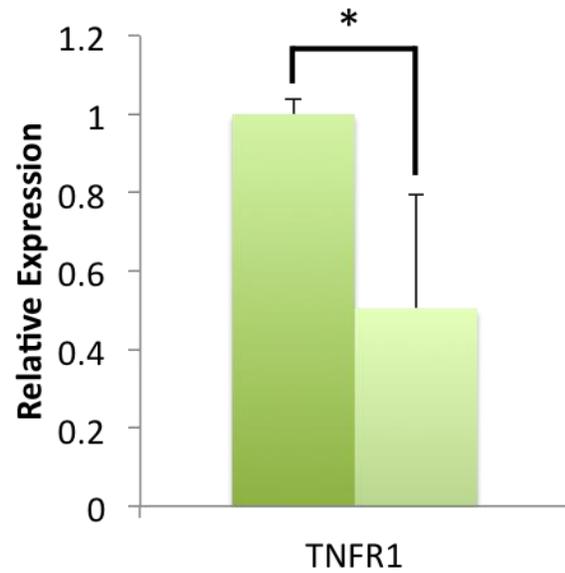
Figure 3-12. qPCR results of pro-apoptotic candidate genes tested in MCF7 knock-down total mRNA samples for candidate genes A. BAX, B. BAD, C. BCL-XS, D. BCL2-L1, E. TRAF3, and F. TNFR1. A single asterisk (*) represents samples that were $p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a triple asterisk (***) represents samples that were $p<0.001$. Error bars on qPCR results represent standard deviation of the mean.

Selected genes further analyzed for expression changes reflected in protein expression

The genes BAX (Figure 3-10A and Figure 3-12A), the large BCL-2 isoform BCL-XL (Figure 3-9B and Figure 3-11A), and TRADD (Figure 3-9G and Figure 3-11D) were observed to have significant mirrored expression with the opposing DAX-1 expression models, as listed in Table 3-9. BAX, a pro-apoptotic protein, and BCL-XL, an anti-apoptotic isoform of the BCL-2 protein, were selected for further analysis of DAX-1 transcriptional regulation. The changes in gene expression were reflected in the amount of protein produced based on DAX-1 level changes, as seen in Figure 3-13. BAX saw an increase when DAX-1 was present and a decrease when DAX-1 expression was reduced, while BCL-XL saw a decrease with DAX-1 introduction and increase with DAX-1 knock-down.

Table 3-9. Comparison of candidate genes in MCF10A and MCF7 cells. Relative increase in expression is indicated by an up arrow (↑), a decrease is indicated by a down arrow (↓), and degree of significance is noted in parentheses. Significance is either noted with a number of asterisks or labeled (n.s.) for not significant. A single asterisk (*) represents samples that were $p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a triple asterisk (***) represents samples that were $p<0.001$. Not all treated samples were tested for each candidate genes.

Type	Gene	MCF10A (siDAX-1)	MCF7 (pcDNA-DAX-1)
Anti-apoptotic	BCL-L2	↑ (*)	↓ (n.s.)
	BCL-XL	↑ (**)	↓ (***)
	BCL2-A1	↓ (***)	
	TNFα	↑ (*)	
	TRAF1	↓ (**)	↓ (***)
	BCL-L10	↓ (***)	
	TRADD	↑ (*)	↓ (**)
Pro-apoptotic	BAX	↓ (*)	↑ (**)
	BAD	↓ (**)	↓ (n.s.)
	BCL-XS	↓ (***)	↓ (**)
	BCL-L1	↓ (***)	↓ (n.s.)
	TRAF3	↓ (***)	↓ (n.s.)
	TNFR1		↓ (*)
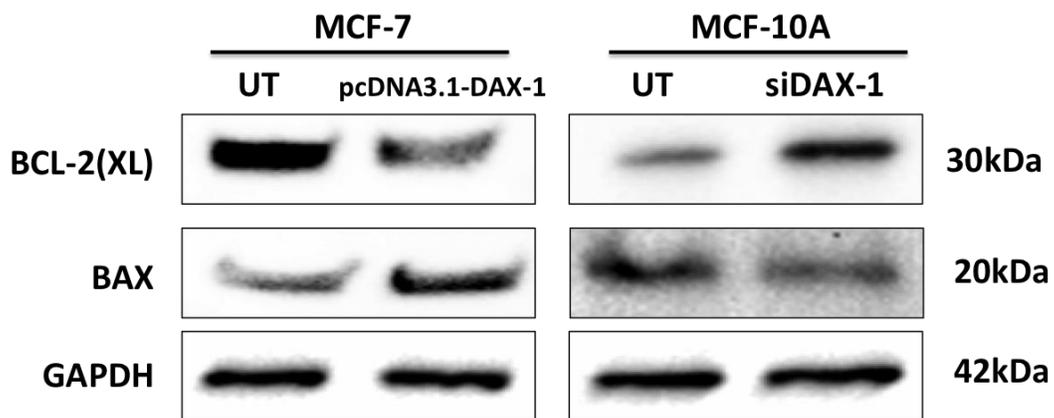
Figure 3-13. Western blot of candidate DAX-1 targets BCL-XL and BAX. BCL-XL, the anti-apoptotic isoform of BCL-2, saw a decrease of protein expression with the introduction of DAX-1, and an increase with the knock-down of DAX-1 expression. BAX, a pro-apoptotic protein in the BCL-2 family, was observed to have an increase in expression with the introduction of DAX-1, and a decrease in expression with the knock-down of DAX-1 expression. GAPDH was used as a control.

ChIP analysis of BCL-2 and BAX promoter regions

The promoter regions of BCL-2 and BAX were broken up into five regions, with each region having a length of about 200-300 base pairs (Figure 3-7). Among the regions tested, the presence of DAX-1 was detected in region 4 for BCL-2, that corresponds to a section that spans a range of 442 to 225 base pairs upstream of the BCL-2 transcription start site. This is shown in both a detectable band in agarose gel electrophoresis of PCR samples using primers for region 4 (Figure 3-14A) and a significant qPCR value when compared to both negative and positive controls (Figure 3-14B). Results from region 3 were used in comparison as a section with no detectable DAX-1 binding. Additionally, ERα is known to bind to the BCL2 promoter based on a response to estradiol [60], and was similarly detected with significant qPCR values.

DAX-1 was also detected in Region 2 of the BAX promoter region (results not shown), which corresponds to a distance of 761 to 484 base pairs upstream of the transcription start site.
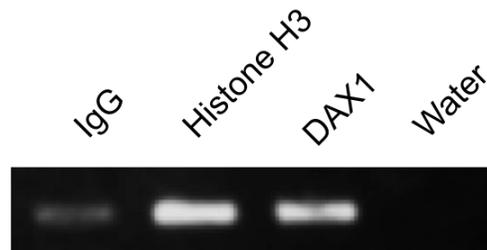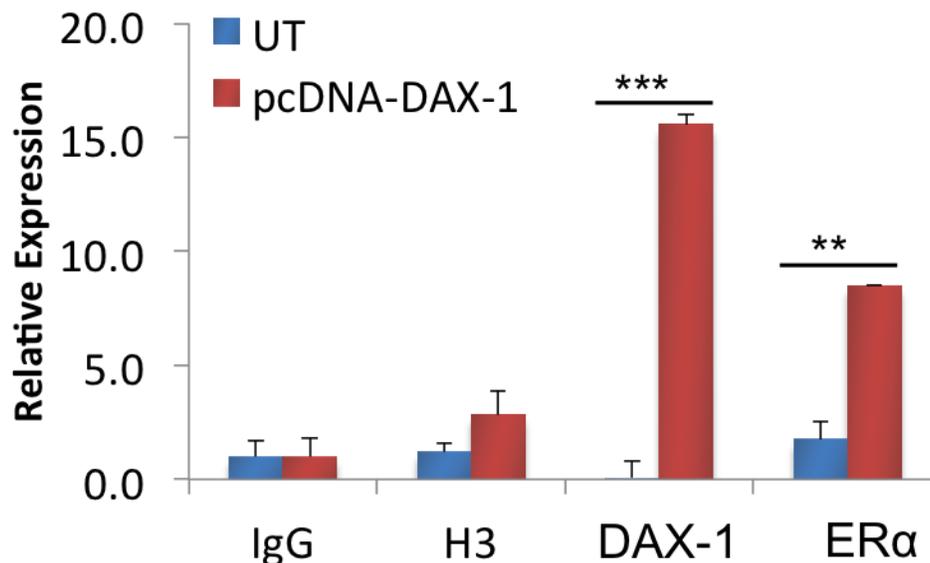
Figure 3-14. Chromatin immunoprecipitation (ChIP) experiments targeting DAX-1 binding to the BCL-2 promoter. A. Agarose gel electrophoresis of PCR products using primers targeting region 4 of the BCL-2 promoter and B. qPCR results comparing the transcript amounts from ChIP experiments using IgG (as a negative control), Histone H3 (as a positive control), DAX-1, and ERα antibodies. A single asterisk (*) represents samples that were $p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a triple asterisk (***) represents samples that were $p<0.001$. Error bars on qPCR results represent standard deviation of the mean.

Discussion

After verifying the success of DAX-1 expression modification in Chapter 2, the influence of DAX-1 on apoptosis was tested through cell flow cytometry of cells stained with Annexin V and PI (Figure 3-8). The increase in the number of apoptotic cells in the entire population was comparable to a known apoptotic inducer BADGE, confirming the association of DAX-1 with programmed cell death overall. There was no additive effect of the combination of DAX-1 knock-in with BADGE treatment. The treatment with both DAX-1 and BADGE appears to affect the same pathway, rather than through different and alternate pathways. The BADGE reagent affects apoptosis through TNFα very far upstream, therefore the DAX-1 target(s) would likely lie downstream of the affected members of the pathway.

The MCF7 and MCF10A cells, treated as described in Chapter 2, served as the opposing DAX-1 expression models. Genes that clearly showed mirrored expression changes with the opposing models would be selected for further analysis, as these changes strongly suggest a direct interaction as caused by varied DAX-1 expression, rather than by the changes in the balance of apoptotic pathways. Among the analyzed candidate genes, three showed significant mirrored changes: the large BCL-2 isoform BCL-XL (Figure 3-9B and Figure 3-11A), TRADD (Figure 3-9G and Figure 3-11D), and BAX (Figure 3-10A and Figure 3-12A). TRADD was ruled out for further analysis because some of its regulation occurs post-transcriptionally, like by gonadotropin-regulated testicular helicase (GRTH) interacting with TRADD mRNA [61]. Meanwhile, research into BCL-2 and its family members has focused to a large degree on transcriptional regulation due to key recognition elements [62], with

some more recent research into post-transcriptional regulation [63]. For these reasons, BCL-2 (interchangeably referred to as BCL-XL because of the large, anti-apoptotic isoform) and BAX were selected for further analysis. The changes in gene expression were reflected in the changes of the protein level (Figure 3-13). This not only confirms that DAX-1 affects the protein level at the gene level, but also reaffirms the idea that with DAX-1 apoptosis is promoted, while removing DAX-1 shifts away from an apoptotic profile.

The ChIP assays showed the presence of DAX-1 far upstream from the BAX gene transcription start site, and relatively closer to the BCL-2 transcription start site (Figure 3-14A). DAX-1 is unable to bind to the promoter directly on its own, therefore it is likely DAX-1 binds to another transcription factor already occupying the promoter, much like a "piggyback" fashion. Available ChIP assays also showed that ERα is detected in the same region on the BCL-2 promoter, and this was confirmed with our own experiments (Figure 3-14B). Given the strong relation of DAX-1 and ER in breast cancer [64], the suggestion of an interaction between these two proteins is highly likely. This can be further confirmed if DAX-1 would have less of an effect on apoptosis in an ER negative breast cancer line. Additionally, this means of interacting with apoptosis through ER interaction may also be reflected in other key biological processes important to cancer cells. These results reveal a fraction of an answer to the overall question of DAX-1's positive contributions to breast cancer, but strongly suggest that DAX-1 regulates the expression BCL-2 with a transcription factor partner that, in turn, affects the apoptotic aspect of cell survival.

Chapter 4

Dax-1 Knock-down Effects on Pluripotent Gene Network in Mouse Embryonic Stem Cells

Introduction

Stem cells are characterized by two defining properties: self-renewal and pluripotency [65]. Self-renewal is the continued replication of a cell lineage, allowing for preservation of population of cells, while pluripotency is the capacity of a cell to differentiate into a variety of cell types, more notably into all three germ layers: the ectoderm, mesoderm, or endoderm [66]. There are distinct pluripotent stem cell types, distinguished by when the cells are harvested from an embryo. As illustrated in figure 4-1, embryonic stem cells (ESCs) are harvested from the inner cell mass (ICM) of a blastocyst after embryo day 5, while the outer cells can be cultured into trophoblast stem cell (TSC) lines. As the blastocyst matures and the ICM begins to divide into distinct layers by the fifth and sixth embryonic day, embryonal carcinoma cells (ECCs) can be harvested from the embryo [67].

Embryonic stem cells are the most commonly used cells in mouse and human stem cell research. *In vitro* culturing of mouse ESC was successfully developed to maintain pluripotency of cells in tissue culture with the presence of leukemia induction factor (LIF) [68]. The embryonic stem cells typically require the presence of feeder cells that served as the source of LIF, but continued harvesting of mouse ESC has allowed for specific lines that can be maintained without feeder cells [69], which are adherent growth-arrested cells that are incapacitated and have the sole purpose of releasing key nutrients required for the maintenance of the stem cell

state [70]. These cell lines can be implanted in a developing embryo and still differentiate into all three types of germ cells, allowing for genetic modification of entire organisms.

Established ES cell lines are maintained by complex protein networks [71, 72], with distinct key transcription factors such as Oct3/4 [73], Sox2 [74], and Nanog [75] that strongly drive the maintenance of pluripotency, and additional factors such as Stat3 [76] and *c-Myc* [77] that contribute to self-renewal. These connections were made through various knock-down studies, both of individual and a combination of candidate factors [78, 79]. Through these studies, no one single gene was the sole factor for maintaining the pluripotent state. Likewise, since the maintenance of a pluripotent state has a number of factors centered in the complex interactions, it was found that reversing differentiated adult mouse fibroblasts needed a combination of some of these key factors (OCT3/4, Sox2, c-Myc, and Klf4) to return stem cell-like state [80]. Untangling the network of transcription factors that keeps pluripotency in check therefore opens up a range of therapeutic possibilities in the medical world.

When Dax-1 was knocked down in mouse stem cells, it was found that the loss of Dax-1 alone was not sufficient to change the phenotype of the population of cells, just like the other key factors. However, there were interactions with many key pluripotency genes that brought Dax-1 into the spotlight as an important contributor to the pluripotent state [81].
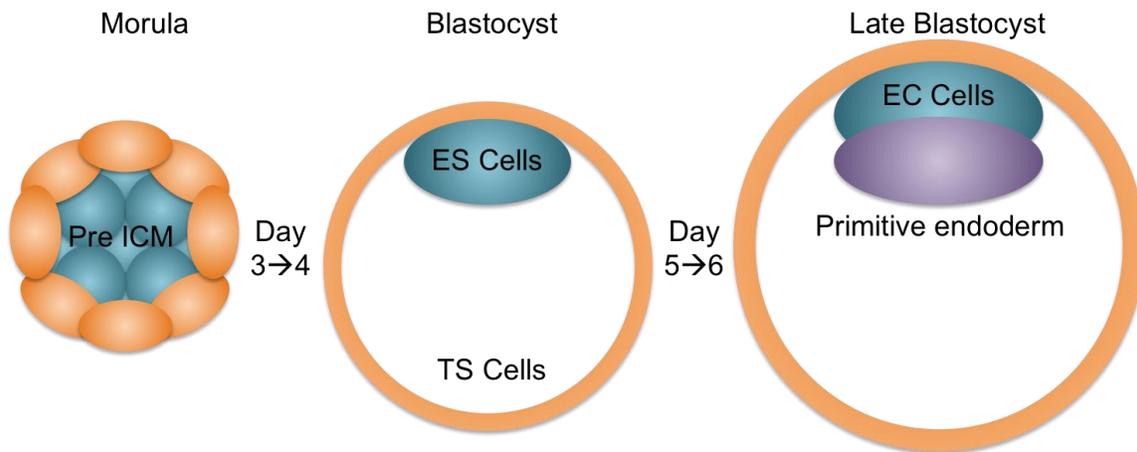
Figure 4-1. Stem cell origins in the mammalian embryo. Pre-ICM (inner cell mass cells) transition to ICM cells during blastulation at embryonic day 3-4. After this occurrence, ES cells can be harvested from the organism and established *in vitro*.

Target Genes

Research carried out previously in the Tzagarakis-Foster lab identified a number of notable genes that Dax-1 appears to interact with, based on changes in gene expression between wild-type mESCs and Dax-1 knocked-down mESCs [82]. Candidate genes were selected based on observations of a significant fold change in expression through tests in a pluripotency PCR array. The known roles and contributions to overall cell activity of a subset of the selected genes are elaborated, and give hints to the mechanisms by which Dax-1 is contributing to pluripotency, as shown in Table 4-1. Furthermore, additional genes of interest were added to the pool of genes examined based on known interactions with Dax-1 in contexts outside of stem cells.

Table 4-1. Dax-1 candidate target genes in mESC and their role in mediating pluripotency and differentiation.

Target Gene	Description
Nanog	Key player in maintenance of pluripotency and self-renewal [83]
Pou5f1; OCT3/4	Key player in maintenance of pluripotency and self-renewal [73, 84]
GATA6	Drives differentiation and organogenesis [85]
SOX2	Known to interact with Oct3/4; together with *SOX3*, encodes the protein SOXB1. Key play Key player in maintenance of pluripotency and self-renewal [86]
SOX9	Known to interact with DAX1 in male sexual development through the expression of SRY from the Y chromosome [87]
SOX17	Drives differentiation and cell fate determination [86]
T	*Brachyury* is part of the family of T-box transcription factors; contributes to the formation and organization of the mesoderm layer during embryogenesis; also refered to as T-Brachyury [88]
Lefty2	Part of the TGFbeta family, which as a whole is involved in cell growth, cell differentiation, apoptosis, homeostasis; required for the left-right asymmetry determination of organogenesis [89]
AMH	Anti-Mullerian hormone, also known as *MIS* (Mullerian inhibiting substance), is part of the TGFbeta family as well. It is known primarily for its role in male sexual development through inhibition of Mullerian ducts, allowing for proper development of Wolffian ducts. [90]

Materials and Methods

RNA and cDNA extraction

mESC were treated using siRNA and CRISPR experiments as described in Chapter 2. The mRNA of these cells were extracted and isolated as described in Chapter 3.

Target Genes

Based on the information produced by previous studies [82] and additional genes of interest, the following table lists the forward and reverse primers used for the isolation of gene fragments and measurement of overall gene expression by quantitative PCR.

Table 4-2. Pluripotency candidate gene primers.

Gene	FWD	REV	Source
Dax-1	TCCTGTACCGCAGCTATGTG	TGCAAGTGCAGGTGATCTTG	[82, 91]
GAPDH (1)	ACAGCCGCATCTTCTTGTGCA	GGCCTTGACTGTGCCGTTGAA	[82]
GAPDH (2)	CCATCACCATCTTCCAGGAGCG	AGAGATGATGACCCTTTTGGC	Construct
Nanog	TTCTTGCTTACAAGGGTCTGC	AGAGGAAGGGCGAGGAGA	[91]
GATA6	ATGCTGTCCGGCCTGCCCTA	GTTCACGCACTCGCGGCTCT	[82]
T-Brachyury	CTGCGCTTCAAGGAGCTAAC	CCAGGCCTGACACATTTACC	[81]
Oct3/4	CTCCCGAGGAGTCCCAGGACAT	GATGGTGGTCTGGCTGAACACCT	[81]
SOX2	ACAAGAGAATTGGGAGGGGT	AAAGCGTTAATTTGGATGGG	mouseprimerdepot
SOX9	AGGAAGCTGGCAGACCAGTA	TCCACGAAGGGTCTCTTCTC	mouseprimerdepot
SOX17	TGGAACCTCCAGTAAGCCAG	TCAGATGTCTGGAGGTGCTG	mouseprimerdepot
AMH	GGGAGACTGGAGAACAGCAG	GGTGGAGGCTCTTGGAACTT	mouseprimerdepot
Lefty2	CATGAAGTCCCTGTGGCTTT	TGCAGTAGACTGCTCAGGACC	mouseprimerdepot

Polymerase Chain Reaction (PCR)

Annealing temperatures were optimized using GoTaq Green Master Mix (Promega), Taq 2X Master Mix (New England Biolabs), and FailSafe PCR System (Epicentre) kits. Reactions were run with the same protocol listed in Table 3-3.

Real-Time PCR (qPCR) and Statistical Analysis

Reagents and equipment used for RT-PCR experiments and statistical analyses were carried out as described in Chapter 2.

Results

Gene expression changes with siRNA Dax-1 knock-down

After treating the mouse embryonic stem cells with siRNA transcripts in order to modify high Dax-1 expression, as described in Chapter 2, collected cDNA was used for qPCR analysis of candidate target pluripotency genes (Figure 4-2). Among the genes tested, Nanog and lefty2 saw a significant increase in relative expression after siDax-1 treatment. Expression of the genes SOX2, SOX17, GATA6, and T-Brachyury saw a significant decrease (Figure 4-2 B, C, F, and G), while OCT3/4 and SOX9 (Figure 4-2 H and I) saw a non-significant decrease in expression. The expression of AMH (Figure 4-2D) did not change. These changes are also summarized in part of Table 4-4.

E.

Lefty2

F.

GATA6

G.

T-Brachyury

H.
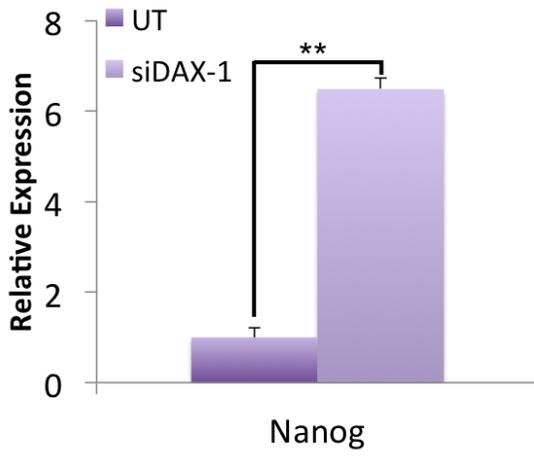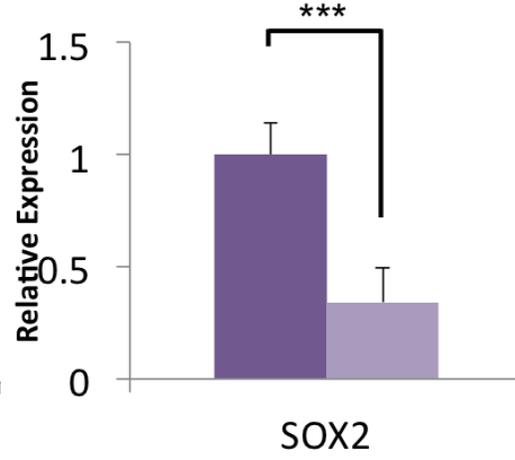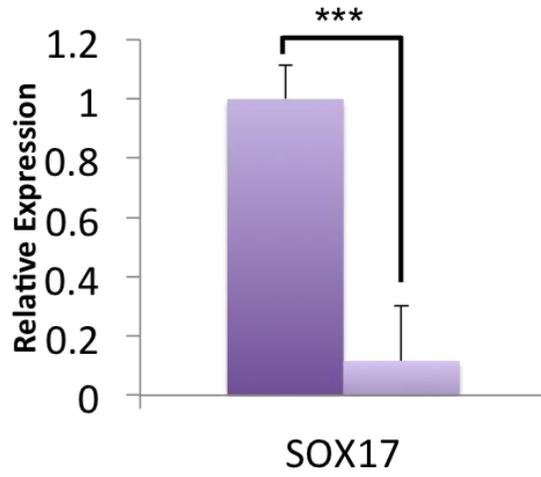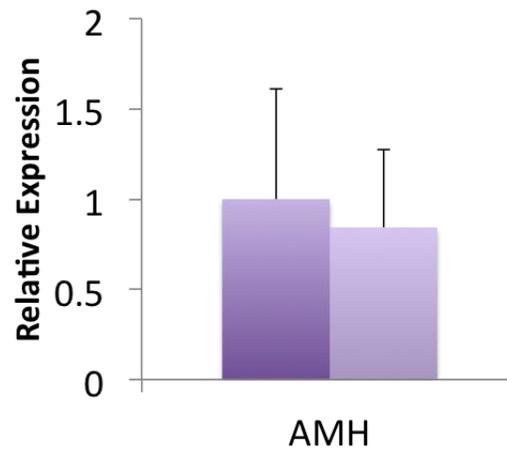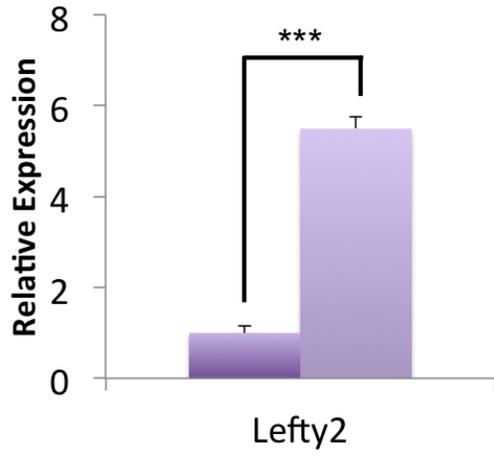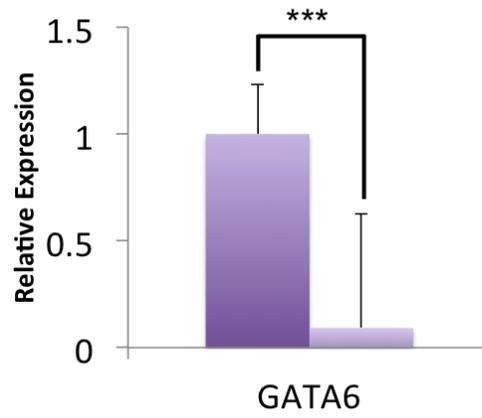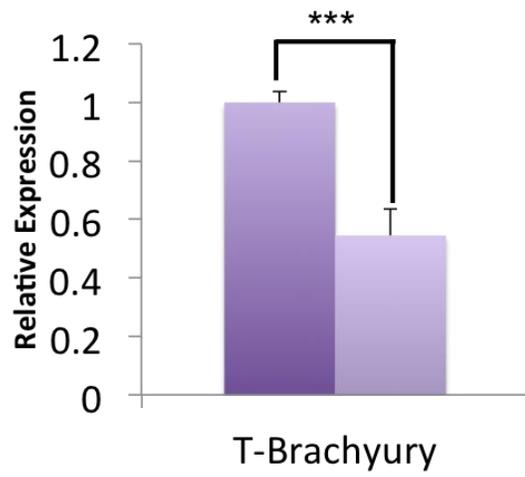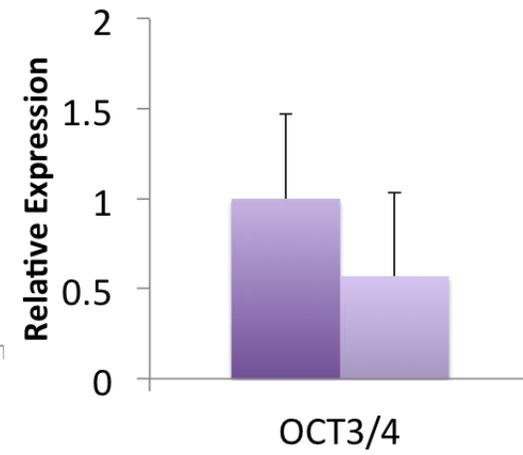
OCT3/4

I.
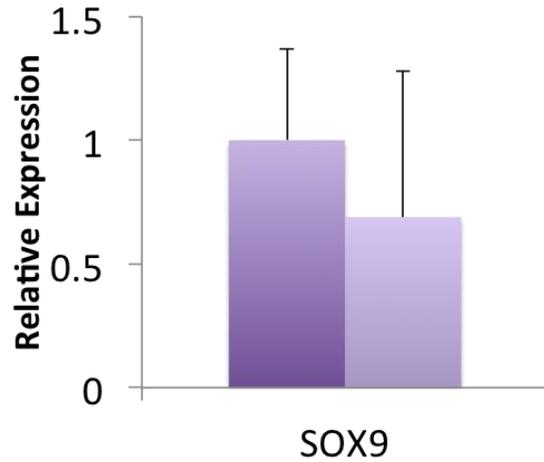
Figure 4-2. qPCR results of siRNA Dax-1 knock-down mESC. Graphs correspond to the candidate genes: A. Nanog, B. SOX2, C. SOX17, D. AMH, E. Lefty2, F. GATA6, G. T-Brachyury, H. OCT3/4, and I. SOX9. A single asterisk (*) represents samples that were $p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a triple asterisk (***) represents samples that were $p<0.001$. Error bars on qPCR results represent standard deviation of the mean.

Gene expression changes with CRISPR-Cas9 treatment

Following the CRISPR-Cas9 treatment described in Chapter 2, stable cells that were treated with CRISPR guide 3 that targeted an early section of exon 1 and showed Dax-1 knockout were harvested and used for analysis of the candidate genes (Figure 4-3). Nanog (A), Lefty2 (E), and T-Brachyury (G) saw a significant increase in expression with the disruption of Dax-1 expression. SOX2 (B), SOX17 (C), GATA6 (F), OCT3/4 (H), and SOX9 (I) saw significant decreases in expression. Lastly, AMH (D) showed no change in expression when comparing untreated to treated samples. These results are also summarized in Table 4-3.

A.

B.

C.

D.

E.

Lefty2

F.

GATA6

G.

T-Brachyury

H.
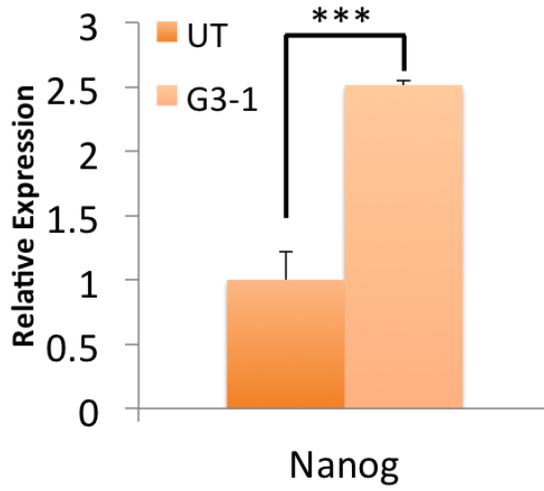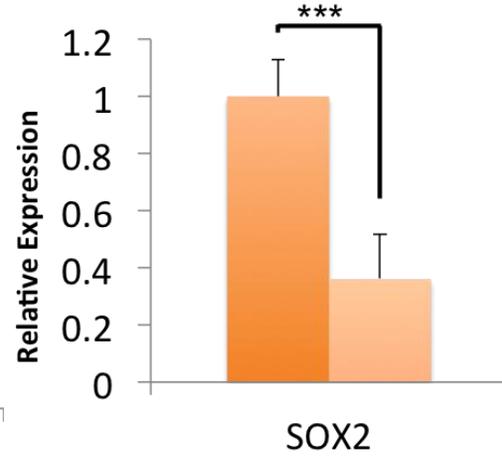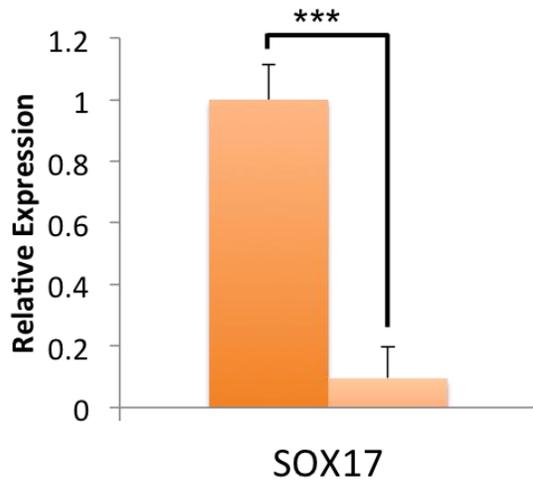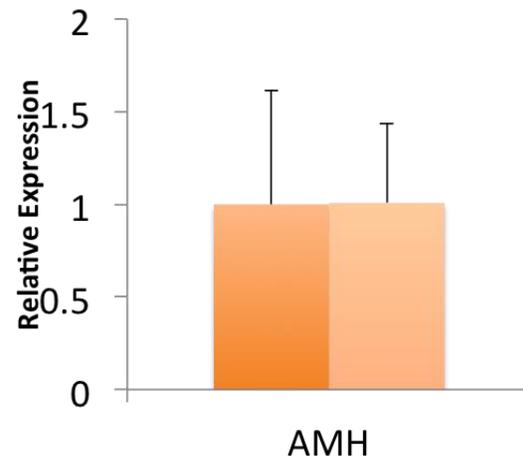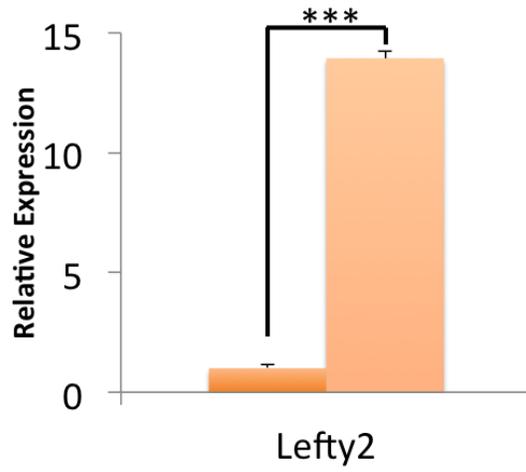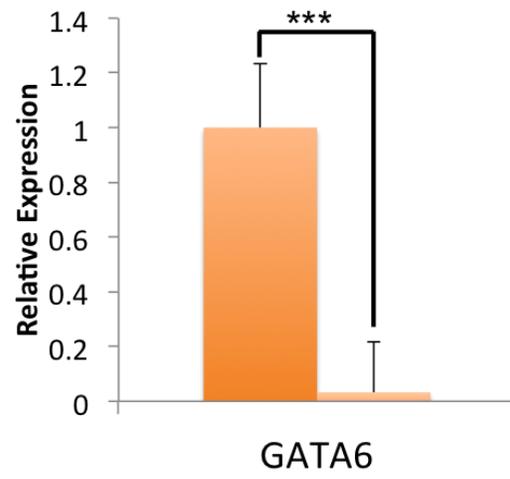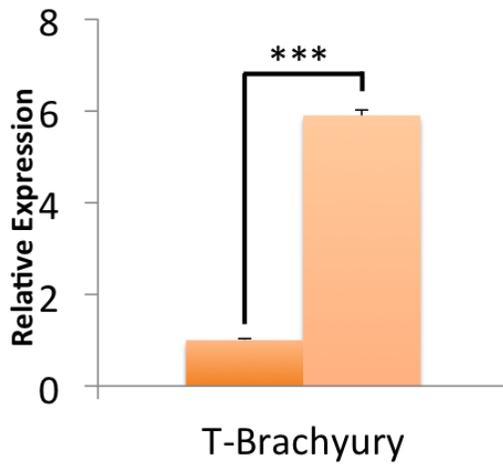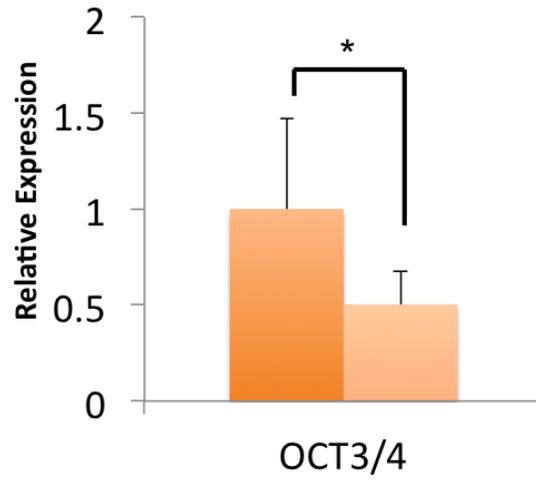
OCT3/4

I.
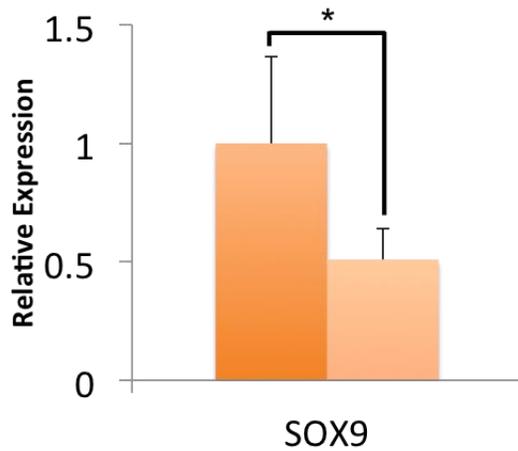
Figure 4-3. qPCR results of CRISPR Dax-1 knock-down mESC. The cells used were treated with guide 3, targeting exon 1 of Nr0b1. Graphs correspond to the candidate genes: A. Nanog, B. SOX2, C. SOX17, D. AMH, E. Lefty2, F. GATA6, G. T-Brachyury, H. OCT3/4, and I. SOX9. A single asterisk (*) represents samples that were $p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a triple asterisk (***) represents samples that were $p<0.001$. Error bars on qPCR results represent standard deviation of the mean.

Table 4-3. Comparison of siRNA and CRISPR Dax-1 candidate gene results.
Relative increase in expression is indicated by an up arrow (↑), a decrease is indicated by a down arrow (↓), no change indicated by an x-symbol (), and degree of significance is noted in parentheses. Significance is either noted with a number of asterisks or labeled (n.s.) for not significant. A single asterisk (*) represents samples that were $p<0.05$, a double asterisk (**) represents samples that were $p<0.005$, and a triple asterisk (***) represents samples that were $p<0.001$.

Gene	siRNA	CRISPR-Cas9
Nanog	↑ (**)	↑ (***)
SOX2	↓ (***)	↓ (***)
SOX17	↓ (***)	↓ (***)
AMH	(n.s.)	(n.s.)
Lefty2	↑ (***)	↑ (***)
GATA6	↓ (***)	↓ (***)
T-Brachyury	↓ (***)	↑ (***)
OCT3/4	↓ (n.s.)	↓ (*)
SOX9	↓ (n.s.)	↓ (*)

Discussion

Using the mESC samples that were treated as outlined in Chapter 2, selected candidate genes were examined for changes in gene expression. In order to further validate the expression changes, two different methods were used to knock-down Dax-1 expression. The comparison using the cDNA of both methods to observed gene expression changes are shown in Table 4-3.

Overall, both methods were successfully able to knock-down Dax-1 expression, both transiently through siRNA and permanently through CRISPR-Cas9. It is very clear that Dax-1 is not a master regulator, since expression of key markers such as Nanog surprisingly increased, while others such as OCT3/4 experienced only a slight decrease in expression. However, changes in particular genes involved with specific cell fates suggest that Dax-1 is not a key regulator because it appears to only be involved with certain pathways towards differentiation. For example, the knockdown of Dax-1 appears to have strong effect on the expression of Lefty2, a gene involved in proper asymmetric development of cells that differentiate to be part of the mesoderm [89]. Another gene that was especially affected with the long-term removal of Dax-1 was T-Brachyury, which is required for mesoderm formation [92].

The increase in expression in these genes involved with mesoderm formation, despite the maintenance of key genes, may be enough to tip the scales towards differentiation within a population of cells. To confirm if Dax-1 has the strongest interaction with mesoderm genes, the candidate list can be expanded to include genes involved with this specific fate should be investigated in the future.

To further confirm if the interactions with genes of interest are direct or not, additional ChIP experiments can be carried out as described in Chapter 3. However, it should be noted that there are research groups who have already carried out large-scale ChIP assays with publically accessible results. This allows for a computer-based approach to sorting through data to confirm Dax-1 interactions based on detection in regulatory regions of genes.

Chapter 5

<u>Functional enrichment analysis of Dax-1 ChIP-seq high-throughput data</u>

Introduction

As described in Chapter 3, chromatin immunoprecipitation assays are often employed to investigate whether a protein of interest is present in complex to specific promoter regions of select target genes. The samples resulting from X-ChIP are typically used for analysis of a small number of regulatory regions, often with primers flanking a region designated by the user. On a slightly larger scale, ChIP-chip allows one to analyze the resulting DNA fragments on a chip array with a pre-selected set of primers for known regulatory regions. When scaled up even more and paired with massive parallel sequencing, conventional ChIP becomes ChIP-seq, and allows for genome wide functional analysis. Using this final method, binding sites of the protein of interest are sequenced and are associated with genes from numerous gene ontology databases using computer annotation tools. The wealth of these data is greatly untapped, and to repeat experiments when there is information available would prove to be wasteful and costly.

A well-known repository for data sets from published papers is the National Center for Biotechnology Information's Gene Expression Omnibus (GEO), online at http://www.ncbi.nlm.nih.gov/geo/, where array and sequenced based data are collected and can be searched for with keywords or an assigned accession number. New resources such as NURSA (Nuclear Receptor Signaling Atlas), online at http://www.nursa.org/, provide a way for researchers to seek specifically for

information pertaining to specific nuclear hormone receptors, with raw information organized by target rather than a massive repository to comb through.

In relation to Dax-1, a search through GEO yielded both ChIP-chip and ChIP-seq experiment results in mESC, the latter of which was obtained for further analysis [93]. The outcome of this analysis can be used not just to validate earlier studies in pluripotency, which were limited to a small set of genes in a PCR array, but also allows for new characterization of Dax-1 in mouse embryonic stem cells. The raw ChIP-seq data and other types of sequence data can be mapped onto the mouse genome through the use of computer annotation tools. These tools associate given sequences (also referred henceforth as peaks) where the protein of interest was bound to the regulatory regions of genes and calculate how accurate the enrichment of a set of genomic regions is. There are numerous gene annotation tools that use different types of statistical methods, genome databases, and gene ontology databases to map provided sequence data.

Like humans, the mouse genome is comprised of only ~5% protein-coding regions, with the remainder "non-coding" DNA consisting of untranslated regions with unknown purpose, non-protein-coding genes, chromosomal structural elements, and regulatory sections [94]. Depending on the means of statistical analysis, there is a possibility for a false association when a test allows for an extensive regulatory region to be associated with a gene. These regions are also known as "gene deserts", and depending on the database, can be as large as a million base pairs away from a gene.

Statistical Analysis

The hypergeometric method is a standard gene enrichment test used by annotation tools such as DAVID and GO Term Finder. The formula, as shown in Figure 5-1A, tests if a gene with a number of peaks in an associated region is enriched with annotations out of a full set of genes in a genome. The values from the formula are as follows: N for the number of genes in the genome, n for the number of genes in a test gene set, K_π for the number of genes in the genome with the annotation π, and k_π for the number of genes in a test gene set with annotation π. The p-value calculated by this method, using the formula in Figure 5-1A, is the probability of choosing k_π or more genes with annotation π when n genes are randomly selected from the genome. There is a heavy bias in this method to only associate peaks with genes if it falls in a proximal region, so many distal association are missed.

The binomial method is another means of statistical analysis. Genes are assigned various regulatory region size, and the p value calculated in the formula in Figure 5-1B accounts for the probability of given peaks to fall within the range of numerous regulatory regions. This can provide novel gene associations, but this bias towards distal binding peaks can lead to false links.

GREAT, which stands for genomic regions enrichment of annotations tool, is an annotation tool available online at (http://bejerano.stanford.edu/great/public-2.0.2/html/) [95]. Running initial data through GREAT analysis provides the user with various calculated values from each statistical method. Among these values, there is a false discovery rate (FDR) q-value, which is obtained by selecting for an

110

ideal p-value among results. These corrected values account for false reads from a single method and give a more accurate ranking of the strength of gene association.

In order to overcome the errors that come from each individual test even after correcting for possible false reads, provided data from GREAT are further processed to find ontology terms that are significant by both tests (B ∩ H). The extracted information is processed through R, a computer language and software environment useful for overall data analysis and graphics [96]. Users can process information through their own programming, or using available packages to implement statistical techniques.

A.

$$\sum_{i=k_\pi}^{min(n,K_\pi)} \frac{\binom{K_\pi}{i}\binom{N-K_\pi}{n-i}}{\binom{N}{n}}$$

B.

$$\sum_{i=k_\pi}^{n} \binom{n}{i} p_\pi^i (1-p_\pi)^{n-i}$$

C.

Binomial test over genomic regions
n = 11 total genomic regions
p_π = 0.75 fraction of genome annotated with π
k_π = 11 genomic regions annotated with π
P-value = 0.04

Hypergeometric test over genes
N = 6 total genes
K_π = 3 genes annotated with π
n = 3 genes with an associated genomic region
k_π = 3 genes annotated and with a genomic region
P-value = 0.05
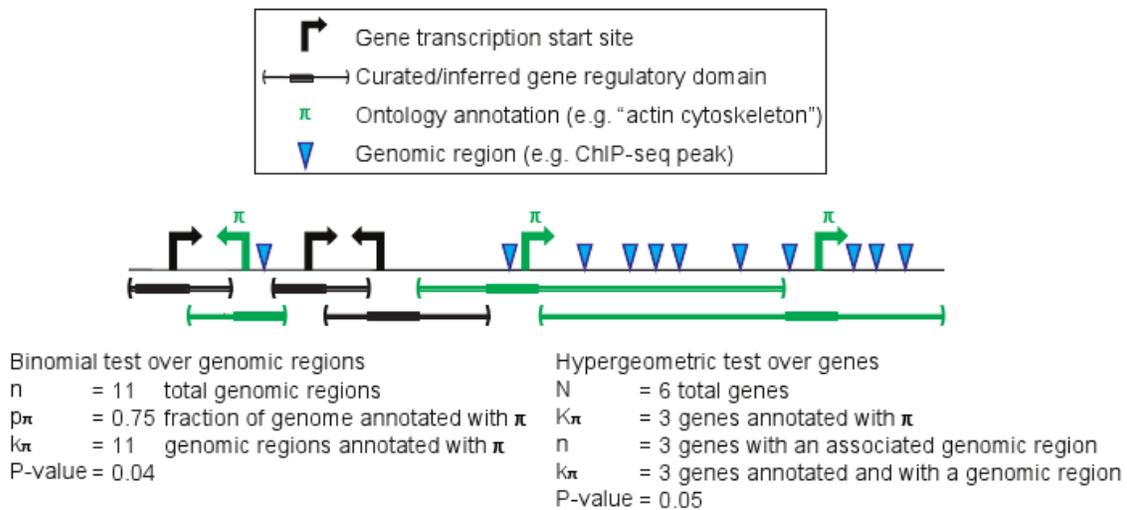
Figure 5-1. Statistical analysis of ChIP-seq data using the GREAT online annotation software. This software utilizes the formula of two tests, A. hypergeometric and B. binomial. C. GREAT carries out each method separately in order to associate a genomic peak or number of peaks with a single gene, obtained from an ontology annotation [97].

Materials and Methods

In collaboration with Vaishali Chaudhuri, a student enrolled in HS617 (Projects in Statistical Computer for Biomedical Data Analytics) at the University of San Francisco, publically available data of Dax-1 ChIP-seq data on NCBI's Gene Expression Omnibus (GEO) database were analyzed. Data sets associated ascension numbers GSM1183116 and GSM1183117 pertaining to Dax-1 binding events in J1 mouse embryonic stem cells were used [93]. The available data online are in .wig format, which is an older format containing data such as GC percent, probability scores, and transcriptome data. BED files of the replicate experiments carried out encoding detected Dax-1 peaks for the entire mouse genome was provided by the Kim lab at UT Austin.

Analysis was carried out using the GREAT annotation tool, available online at (http://bejerano.stanford.edu/great/public-2.0.2/html/), where the data are provided as a BED file for input. The program associates the provided sequences with regulatory regions of genes, both proximal and distal, using numerous gene ontology databases and the two distinct statistical methods described in the background section. GREAT version 3.0.0 was used to analyze the Dax-1 ChIP-seq data from both BED files with parameters listed in Table 5-1, in order to find the association of provided peaks with gene ontologies.

Table 5-1. Parameters for GREAT version 3.0.0 online analysis.

Species Assembly	Mouse: NCBI build 38 (UCSC mm10, Dec/2011)
Test regions	DAX_BED1, DAX_BED2 [93]
Background Regions	Whole Genome
Association rule settings	Basal plus extension Proximal: 5.0 kb upsteam, 1.0 kb downstream Distal: 1000.0 kb

The GREAT website then gives an output of associated genes grouped by the different gene ontologies, ranked by the calculated enrichment through both the binomial test and hypergeometric test. The results from the selected ontologies, as shown in Figure 5-2 were compiled from individual .tsv files into a single .csv file. This output was then run through three distinct packages in the R-language program: hierarchial cluster analysis [98], K means cluster analysis through ggplot [99], and principal component analysis [100].

Figure 5-2. Selected Ontology results from processing DAX1_BED1 file through the online GREAT program.

Results

Following processing through the R programming language, the top five ontologies were selected for further analysis (Figure 5-3). These ontologies had high p-values in both statistical methods, and therefore show the greatest confidence of Dax-1 influence within the category. The overall descriptions of associated ontologies are described in Table 5-2. The highest ranking ontology encompasses genes that are expressed in differentiated cells that are part of the trophectoderm, followed by genes that lose their expression before birth or at birth, once the embryo no longer needs to develop. The next three categories are more broad, with two (Gen_V6.5 and Gen J1) containing genes that were down-regulated in late stages of differentiation in two distinct mouse embryonic stem cell lines, and one involving genes that participate in the platelet-derived growth factor receptor signaling pathway.

The genes within the selected ontologies were further narrowed down in Table 5-3 based on the distance of Dax-1 binding from the transcription start site (TSS). Genes were arbitrarily selected if the binding occurred within 6,000 base pairs before the start sites, and up to 2,000 base pairs after the start site. Genes that were previously studied in Chapter 4 and that appeared within the top ontologies, namely Nanog, Pou5f1, SOX2, and T, were also selected in Table 5-4.

Top Five Interactions in DAX1 CHIP-Seq Interactions
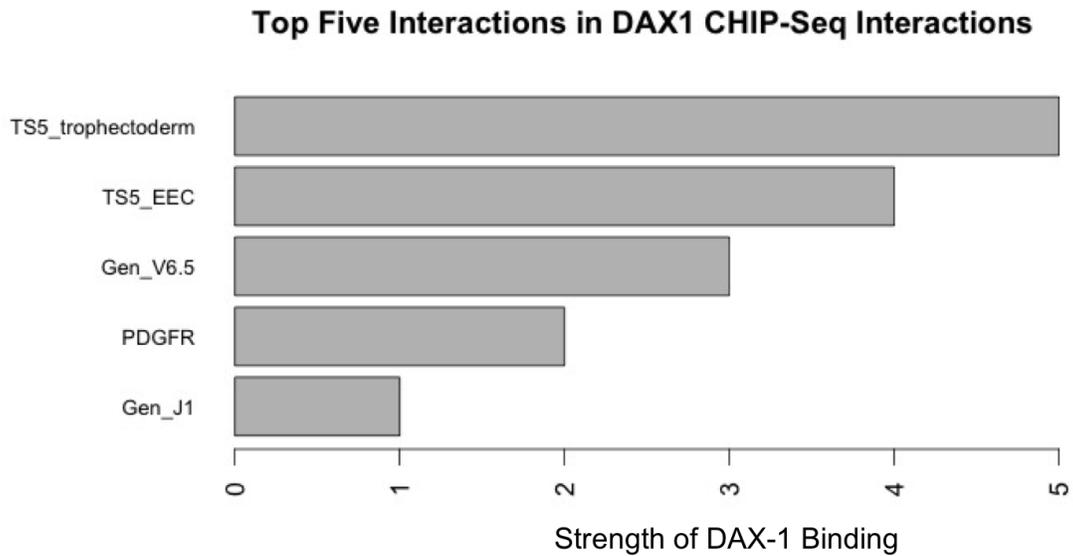
Figure 5-3. Top five interactions in Dax-1 ChIP-seq interactions obtained through analysis of ontology data from GREAT by Dynamic Tree Cut, ggplot2, and PCA packages in R language. Image generated by V. Chaudhuri.

Table 5-2. Top five Interactions in Dax-1 ChIP-seq interactions.

Rank	Biological Process	Description	Regions	Genes	Ontology Source
1	TS5_Trophectoderm	These genes are expressed and play a role in differentiated cells that are part of the trophectoderm.	205	63	MGI Expression: Detected
2	TS5_EEC	Extraembryonic components consist of genes that may be lost before birth or will be discarded at birth, or when the embryo becomes an independent organism.	214	67	MGI Expression: Detected
3	Gen_V6.5	Genes down-regulated during late stages of differentiation of embryoid bodies from V6.5 embryonic stem cells.	203	83	MSigDB Oncogenic Signatures
4	PDGFR	Platelet-derived growth factor receptor signaling pathway; The series of molecular signals generated as a consequence of a platelet-derived growth factor receptor binding to one of its physiological ligands.	73	16	GO Biological Process
5	Gen_J1	Genes down-regulated during late stages of differentiation of embryoid bodies from J1 embryonic stem cells.	196	82	MSigDB Oncogenic Signatures

Table 5-3. Selected genes from the top five ontologies. Regions with negative values are peaks before the gene transcription start site (TSS), and positive base pair values are distances after the gene TSS. Genes that showed Dax-1 peaks within 6,000 base pairs before TSS and 2,000 base pairs after TSS are shown.

Gene	Region (bp)	Ontology
Acacb	+1,355, +5,318	(5) Gen_J1
Ankrd10	-5,182	(5) Gen_J1
Atp1b1	-94,421, -92,777, -64,303, -60,187, -45,391, -36,143, -25,414, -2,739, +12,409, +79,738, +81,235	(2) EEC
Avpi1	-7,164	(3) Gen_V6.5
Bcl3	-798, +7,504	(3) Gen_V6.5
Cbx7	-1,159, +27,487, +30,030	(3) Gen_V6.5
Cdh1	-3,617, +3,593, +11,996	(2) EEC
Csrnp1	-276, +25,671	(3) Gen_V6.5
Dppa2	-1,381	(3) Gen_V6.5
Dppa5a	-6,686, -3,389, -1,095, +3,296, +7,485, +8,252	(2) EEC
Dpys	+893, +76,433	(5) Gen_J1
Eras	+1,451	(5) Gen_J1
Etv4	-557, +7,171, +8,185, +12,655	(3) Gen_V6.5
Fbxo15	-851	(3) Gen_V6.5
Fgfr2	+1,368. +12,718, +14,001	(2) EEC
Frs2	-3,980	(1) Trophectorderm, (2) EEC
Gbx2	-2,861	(3) Gen_V6.5, (5) Gen_J1
Gli2	-353,788, 351,438, -345,108, -294,626, -251,793, -248,600, -112,886, -98,929, -94,242, -28,753, -22,877, -1,856, +4,415, +69,593, +405,274, +410,581	(3) Gen_V6.5, (5) Gen_J1
Gpa33	-4,837, +5,088, +12,317	(3) Gen_V6.5
Gtf2ird1	-71,300, +55, +11,081, +17,489, +116,355, +119,412, +123,640, +127,110	(2) EEC
Hspg2	-6,950, +1,653, +11,424, +20,048, +28,713	(2) EEC
Icam1	-1,113	(3) Gen_V6.5, (5) Gen_J1
Iqgap1	-40,825, -19,882, +163, +19,419, +56,666	(2) EEC
Jam2	-4,230, +5,299	(3) Gen_V6.5, (5) Gen_J1
Kank2	-409, +30,112	(2) EEC
Kat2b	-5,197	(3) Gen_V6.5
Kat7	-2,063	(2) EEC
Kdm5b	-4,567, +3,786, +10,410	(2) EEC

Mras	+1,698, +12,426, +16,551, +19,849, +45,946	(3) Gen_V6.5
Msrb2	-70,042, +118, +9,270	(3) Gen_V6.5
Mybl2	-22,846, -2,379, +12,610, +20,795	(2) EEC
Myc	-6,505, +24,556, +25,544, +107,426, +176,743, +973,579	(3) Gen_V6.5
Nanog	-44,193, -17,124, -4,791, +34	(1) Trophectoderm, (2) EEC
Notum	-2,804	(3) Gen_V6.5, (5) Gen_J1
Nr0b1	-4,311, +1,854	(3) Gen_V6.5, (5) Gen_J1
Ntn1	-355, +6,644, +52,524	(5) Gen_J1
Pcolce	-2,199, +555	(3) Gen_V6.5, (5) Gen_J1
Pdgfc	-13,859, -3,326	(4) PDGFR
Pdgfrb	-2,085	(4) PDGFR
Pou5f1	-20,232, -15,487, -3,252, -1,985, -994, -100	(1) Trophectorderm, (2) EEC
Ppcs	-295	(5) Gen_J1
Prmt6	-495	(5) Gen_J1
Rif1	-733	(1) Trophectorderm, (2) EEC
Sall4	-86,804, -84,580, -74,561, -24,142, -19,141, -12,691, -2,340, +2,037	(2) EEC
Sema4b	-56,334, -611, +34,730	(3) Gen_V6.5
Slc39a4	-798	(1) Trophectorderm, (2) EEC
Slc4a3	+129, +55,687, +258,304, +390,568, +482,891	(5) Gen_J1
Socs3	-6,503, -3,615, +6,947, +36,854, +80,486	(3) Gen_V6.5, (5) Gen_J1
Sox13	-134,922, -99,881, -4,606, -3,529, +25,571	(1) Trophectorderm, (2) EEC
Sox2	-3,574, +18,648, +85,086, +106,847, +108,032, +110,438, +111,406, +787,897	(1) Trophectorderm, (2) EEC
Spry4	-1,087, +25,127, +27,298, +79,109	(3) Gen_V6.5
Syncrip	-2,124	(5) Gen_J1
T	-19,833, -4,154	(1) Trophectorderm, (2) EEC
Tcea3	-1,457	(1) Trophectorderm, (2) EEC, (3) Gen_V6.5, (5) Gen_J1
Tcf15	-1,242, +34,026, +35,029, +35,952, +36,925, +46,163	(3) Gen_V6.5, (5) Gen_J1

Tdgf1	-1,994	(5) Gen_J1
Tet1	-1,158, +681, +1,996, +3,457 +6,027	(5) Gen_J1
Tfap2c	-53,052, -32,751, -28,800, -4,830, +27,812, +44,801, +64,252, +278,876, +322,038	(1) Trophectorderm, (2) EEC
Tmem79	-2,377	(3) Gen_V6.5
Trh	-2,924, -1,883	(5) Gen_J1
Ttc39b	-5,907, -4,517, +7,385	(3) Gen_V6.5
Tubb2b	-1,447, +20,578	(5) Gen_J1
Tubb3	+923	(3) Gen_V6.5, (5) Gen_J1
Vegfc	-976	(5) Gen_J1
Zbtb45	-4,675	(3) Gen_V6.5
Zfp219	-1,921	(1) Trophectorderm, (2) EEC
Zfp296	-1,463	(1) Trophectorderm, (2) EEC
Zfp42	-26,631, -13,963, -140	(3) Gen_V6.5
Zfp57	-6,871, +27	(3) Gen_V6.5

Table 5-4. Genes from the top five ontologies that overlap with mESC pluripotency project. Regions with negative values are peaks before the gene transcription start site (TSS), and positive base pair values are distances after the gene TSS. Possible significant sites that are within a reasonable distance from the TSS are underlined.

Gene	Region (bp)
Nanog	-44,193, -17,124, <u>-4,791</u>, +34
Pou5f1	-20,232, -15,487, <u>-3,252, -1,985, -994, -100</u>
SOX2	<u>-3,574</u>, +18,648, +85,086, +106,847, +108,032, +110,438, +111,406, +787,897
T	-19,833, <u>-4,154</u>

Discussion

Analyzing the results of ChIP-seq allowed for a greater understanding of the broad influence Dax-1 has within the mouse embryonic stem cell model. The presence of the Dax-1 protein in a regulatory region of a target gene reveals possible direct interactions and influence on transcription. The resulting gene list (Table 5-3) does not encompass all of the genes associated with Dax-1 binding within the selected ontologies, because the GREAT program allows for extremely distal binding. While distal binding accounts for more complex interactions, more proximal sites are ultimately easier to investigate and confidently draw any conclusions from.

Overall, the results shed some light on why Dax-1 appears to have a strong but not absolute influence on pluripotency. While it has some interactions with master regulators, it also appears to affect a subset of genes involved in a specific differentiation route towards maturation into the trophectoderm. Genes involved with development of the trophectoderm are among the first to be activated, since these cells are required for implantation of the embryo and eventually give rise to some of the placenta [101]. Therefore, modification of Dax-1 seems to heavily influence only one route, and other cells with transcription factors favoring any other route of differentiation may be less influenced by the Dax-1 absence.

While the genes investigated in Chapter 4 were selected from previous research, they were but a small number out of the total number of identified interactions. Cross-referencing of candidate genes with the targets found through ChIP-seq analysis made up for the lack of labor-intensive ChIP experiments for each

individual gene. Among the candidates from Chapter 4, Dax-1 was found to be present within the regulatory regions of Nanog, Pou5f1 (aka OCT3/4), Sox2, and T (Table 5-4).

Lastly, there are genes that were not previously identified as candidate targets through any previous array or project. There are those that fall within the realm of pluripotency and expand on the interactions within that phenomenon in stem cells, but additional ontologies available through the GREAT program revealed Dax-1 interactions with genes involved in a range of diseases. Manipulation of the statistical analyses can account for similar ontologies with great overlap of genes, and allow for different ontologies to gain a higher relative rank. Given that the mouse model is often used to further understand human models, interactions with possible novel genes can answer a myriad of questions.

End Summary

DAX-1 serves as a great protein of interest, with unique properties as an orphan nuclear hormone receptor. Its specific expression in cell types, especially stem cells and certain cancer cells, gives rise to questions about what molecular pathways in which it could possibly be involved.

These interactions were investigated in a human model with two types of breast cells by changing the endogenous expression of DAX-1, based on studies that revealed DAX-1 involvement in cell survival pathways. Apoptosis appeared to be the process of interest, due to changes in key genes and a similar result to that of a known small molecular inhibitor treatment. Following qPCR assays to observe mirrored expression in cell culture models that were treated to have opposing expression of DAX-1, candidate genes were selected for further analysis. BCL-2, an anti-apoptotic member of the BCL-2 protein family, was found to have DAX-1 present along its promoter, and it is likely that this regulation is through interaction with ERα. Therefore, one way that DAX-1 influences apoptosis is through BCL-2 transcriptional regulation.

In mouse embryonic stem cells, Dax-1 had been found to play a role in pluripotency. Using two different knock-down methods, siRNA and CRISPR-Cas9, candidate genes selected from earlier studies within the Tzagarakis-Foster Lab were tested for changes in gene expression. With the exception of the T gene, changes in expression for both types of knock-down treatments were similar and strongly suggested that Dax-1 interacts with subsets of genes involved in specific differentiation pathways.

And lastly, the scope of Dax-1 regulated genes was expanded through statistical analysis of available Dax-1 ChIP-seq data, using the R-language and various packages to manipulate data. After first being run through an online program GREAT to find ontologies that are statistically significant through two analysis methods, binomial and hypergeometric, the results from each method were further processed to find overlapping results to confirm Dax-1 association. Resulting genes reveal interactions with not only candidate genes from small scale experiments, but also open up studies for uninvestigated and novel Dax-1 transcriptional interactions.

Based on these results, DAX-1 is not limited to involvement with a single biological process, but is restricted enough to be a therapeutic target in an individual cell type. Apoptotic observations in the human breast cells were not seen in the mouse embryonic stem cells, and vice versa. The interactions within the breast cancer cells, not only with ERα but other key steroid receptors such as AR [102], strongly connect DAX-1 as a gene target within cancer. DAX-1 and other orphan nuclear hormone receptors are increasingly becoming not just therapeutic targets [103], but also markers of prognosis and how successful therapy is [104]. Hopefully, these findings add to the rich story that continues to develop in the DAX-1 related research field.

References

1. Sonoda, J., Liming, P., and Evans, R.M. (2008). *Nuclear Receptors: Decoding Metabolic Disease.* FEBS Lett. *582*, 2-9.
2. Germain, P., Staels, B., Dacquet, C., Spedding, M., and Laudet, V. (2006). Overview of nomenclature of nuclear receptors. *Pharmacological reviews 58*, 685-704.
3. Sladek, F.M. (2011). What are nuclear receptor ligands? *Molecular and cellular endocrinology 334*, 3-13.
4. Kumar, R., and Thompson, E. (1999). The structure of the nuclear hormone receptors. *Steroids 64*, 310-319.
5. Heery, D.M., Kalkhoven, E., Hoare, S., and Parker, M. (1997). A signature motif in transcription co-activators mediates binding to nuclear receptors. *Nature 387*.
6. Plevin, M.J., Mills, M.M., and Ikura, M. (2005). The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends in Biochemical Sciences 30*, 66-69.
7. Hu, X., and Lazar, M.A. (1999). The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. *Nature 402*, 93-96.
8. Sáez, P.J., Lange, S., Pérez-Acle, T., and Owen, G. (2010). Nuclear Receptor Genes: Evolution. I*n Encyclopedia of Life Sciences.* (John Wiley & Sons, Ltd.).
9. King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature 451*, 783-788.
10. Aranda, A., and Pascual, A. (2001). Nuclear Hormone Receptors and Gene Expression. *Physiological Reviews 81*, 1269-1304.
11. Niakan, K.K., Davis, E.C., Clipsham, R.C., Jiang, M., Dehart, D.B., Sulik, K.K., and McCabe, E.R. (2006). Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. *Molecular Genetics and Metabolism 88*, 261-271.
12. Ehrlund, A., and Treuter, E. (2012). Ligand-independent actions of the orphan receptors/corepressors DAX-1 and SHP in metabolism, reproduction and disease. *The Journal of Steroid Biochemistry and Molecular Biology 130*, 169-179.
13. Lalli, E. (2014). Role of Orphan Nuclear Receptor DAX-1/NR0B1 in Development, Physiology, and Disease. Advances in Biology *2014*, 1-19.
14. Iyer, A., and McCabe, E. (2004). Molecular mechanisms of DAX1 action. *Molecular Genetics and Metabolism 83*, 60-73.
15. Zhang, H., Thomsen, J.S., Johansson, L., Gustafsson, J.A., and Treuter, E. (2000). DAX-1 functions as an LXXLL-containing corepressor for activated estrogen receptors. *The Journal of Biological Chemistry 275*, 39855-39859.
16. Sablin, E.P., Woods, A., Krylova, I.N., Hwang, P., Ingraham, H.A., and Fletterick, R.J. (2008). The structure of corepressor Dax-1 bound to its target nuclear receptor LRH-1. P*roceedings of the National Academy of Sciences of the United States of America 105*, 18390-18395.

17. Zanaria, E., Muscatelli, F., Bardoni, B., Stom, T., Guioli, S., Guo, W., Lalli, E., Moser, D., Walker, A., McCade, E., et al. (1994). An unusual member of the nuclear hormone receptor superfamily responsible for X-linked adrenal hypoplasia congenita. *Nature 372*, 635-641.

18. Zanaria, E., Bardoni, B., Dabovic, B., Calvari, V., Fraccaro, M., Zuffardia, O., Giovanna, C., J., G., and McLaren, A. (1995). Xp Duplications and Sex Reversal. *Philosophical Transactions of the Royal Society of London, B Biological Sciences. 350*, 291-296.

19. Guo, W., Burris, T.P., and McCabe, E.R. (1995). Expression of DAX-1, the gene responsible for X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism, in the hypothalamic-pituitary-adrenal/gonadal axis. *Biochem. Mol. Med. 56*, 8-13.

20. Simandi, Z., Cuaranta-Monroy, I., and Nagy, L. (2013). Nuclear receptors as regulators of stem cell and cancer stem cell metabolism. *Seminars in Cell and Developmental Biology 24*, 10-12.

21. Lalli, E., and Alonso, J. (2010). Targeting DAX-1 in embryonic stem cells and cancer. *Expert Opinion on Therapeutic Targets 14*, 169-177.

22. Subik, K., Lee, J., Baxter, L., Strzepek, T., Costello, D., Crowley, P., ZXing, L., Hung, M., Bonfiglio, T., Hicks, D., et al. (2010). The Expression Patterns of ER, PR, HER2, CK5/6, EGFR, Ki-67 and AR by Immunohistochemical Analysis in Breast Cancer Cell Lines. *Breast Cancer: Basic and Clinical Research 4*, 35-41.

23. Ramachandran, P.V., and Ignacimuthu, S. (2013). RNA interference--a silent but an efficient therapeutic tool. *Applied Biochemistry and Biotechnology 169*, 1774-1789.

24. Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell 118*, 57-68.

25. Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell 123*, 631-640.

26. Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. *Nature Methods 10*, 957-963.

27. Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science 327*, 167-170.

28. Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology 9*, 467-477.

29. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science 337*, 816-821.

30. Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS 109*, E2579-E2586.

31. Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating, and targeting genomes. *Nature Biotechnology 32*, 347-355.

32. Moore, J.K., and Haber, J.E. (1996). Cell Cycle and Genetic Requirements of Two Pathways of Nonhomologous End-Joining Repair of Double-Strand Breaks in Saccharomyces cerevisiae. *Molecular and Cellular Biology 16*, 2164-2173.

33. Jasin, M., and Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harbor perspectives in biology 5*, a012740.

34. Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell 157*, 1262-1278.

35. Horii, T., Morita, S., Kimura, M., Kobayashi, R., Tamura, D., Takahashi, R.U., Kimura, H., Suetake, I., Ohata, H., Okamoto, K., et al. (2013). Genome engineering of mammalian haploid embryonic stem cells using the Cas9/RNA system. *PeerJ 1*, e230.

36. Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.F., Sontheimer, E.J., and Thomson, J.A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proceedings of the National Academy of Sciences of the United States of America 110*, 15644-15649.

37. Zhou, J., Shen, B., Zhang, W., Wang, J., Yang, J., Chen, L., Zhang, N., Zhu, K., Xu, J., Hu, B., et al. (2014). One-step generation of different immunodeficient mice with multiple gene modifications by CRISPR/Cas9 mediated genome engineering. *The International Journal of Biochemistry & Cell Biology 46*, 49-55.

38. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science 339*, 819-823.

39. Cho, M., Cho, T.J., Lim, J.M., Lee, G., and Cho, J. (2013). The establishment of mouse embryonic stem cell cultures on 96-well plates for high-throughput screening. *Molecules and cells 35*, 456-461.

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

41. Simstein, R., Burow, M., Parker, A., Weldon, C., and Beckman, B. (2003). Apoptosis, Chemoresistance, and Breast Cancer: Insights From the MCF-7 Cell Model System. *Experimental Biology and Medicine 228*, 995-1003.

42. Lanzino, M., Maris, P., Sirianni, R., Barone, I., Casaburi, I., Chimento, A., Giordano, C., Morelli, C., Sisci, D., Rizza, P., et al. (2013). DAX-1, as an androgen-target gene, inhibits aromatase expression: a novel mechanism blocking estrogen-dependent breast cancer cell proliferation. C*ell Death & Disease 4*, e724.

43. Scandurra, A. (2014). Examining the Role of DAX-1 in Regulation of Cell Proliferation in Human Breast Cells. In Biology, Volume Master of Science. (University of San Francisco), p. 106.

44. Elmore, S. (2007). Apoptosis: A Review of Programmed Cell Death. *Toxicology Pathology 35*, 495-516.

45. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell 144*, 646-674.

46. Martinvalet, D., Zhu, P., and Lieberman, J. (2005). Granzyme A induces caspase-independent mitochondrial damage, a required first step for apoptosis. *Immunity 22*, 355-370.
47. Hengartner, M.O. (2000). The biochemistry of apoptosis. *Nature 407*, 770-776.
48. McIlwain, D.R., Berger, T., and Mak, T.W. (2013). Caspase Functions in Cell Death and Disease. *Cold Spring Harbor perspectives in Biology 5*, a008656.
49. Cohen, G.M. (1997). Caspases: the executioners of apoptosis. *Biochem J 326*, 1-16.
50. Rai, N.K., Tripathi, K., Sharma, D., and Shukla, V.K. (2005). Apoptosis: a basic physiologic process in wound healing. *Int J Low Extrem Wounds 4*, 138-144.
51. Enari, M., Sakahira, H., Yokoyama, H., Okawa, K., Iwamatsu, A., and Nagata, S. (1998). A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature 391*, 43-50.
52. Wallach, D., Varfolomeev, E., Malinin, N., Goltsev, Y., Kovalenko, A., and Boldin, M. (1999). Tumor necrosis factor receptor and Fas signaling mechanisms. *Annual Review of Immunology 17*, 331-367.
53. Kischkel, F., Lawrence, D., Chuntharapai, A., Schow, P., Kim, K., and Ashkenazi, A. (2000). Apo2L/TRAIL-Dependent Recruitment of Endogenous FADD and Caspase-8 to Death Receptors 4 and 5. *Immunity 12*, 611-620.
54. Adams, J., and Cory, S. (2007). The Bcl-2-regulated apoptosis switch: mechanism and therapeutic potential. *Current Opinion Immunology 19*, 488-496.
55. Sattler, M., Liang, H., Nettesheim, D., Meadows, R.P., Harlan, J.E., Eberstadt, M., Yoon, H.S., Shuker, S.B., Chang, B.S., Minn, A.J., et al. (1997). Structure of Bcl-xL-Bak Peptide Complex: Recognition Between Regulators of Apoptosis. *Science 275*, 983-986.
56. Shamas-Din, A., Brahmbhatt, H., Leber, B., and Andrews, D.W. (2010). BH3-only proteins: Orchestrators of apoptosis. *Biocheimica et Biophysica Acta 1813*, 508-520.
57. Youle, R., and Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. *Nature Reviews Molecular Cell Biology 9*, 47-59.
58. Baker, S.J., and Reddy, E.P. (1998). Modulation of life and death by the TNF receptor superfamily. *Oncogene 17*, 3261-3270.
59. Jackson-Bernitsas, D.G., Ichikawa, H., Takada, Y., Myers, J.N., Lin, X.L., Darnay, B.G., Chaturvedi, M.M., and Aggarwal, B.B. (2007). Evidence that TNF-TNFR1-TRADD-TRAF2-RIP-TAK1-IKK pathway mediates constitutive NF-kappaB activation and proliferation in human head and neck squamous cell carcinoma. *Oncogene 26*, 1385-1397.
60. Gompei, D., Somai, S., Chaouat, M., Kazem, A., Kloosterboar, H.J., Beusman, I., Forgez, P., Miimoun, M., and Rostene, W. (2000). Hormonal regulation of apoptosis in breast cells and tissues. *Steroids 65*, 593-598.
61. Gutti, R.K., Tsai-Morris, C.H., and Dufau, M.L. (2008). Gonadotropin-regulated testicular helicase (DDX25), an essential regulator of spermatogenesis,

prevents testicular germ cell apoptosis. *The Journal of Biological Chemistry 283*, 17055-17064.

62.   Seto, M., Jaeger, U., Hockett, R., Graninger, W., Bennett, S., Goldman, P., and Korsmeyer, S. (1988). Alternative promoters and exons, somatic mutation and deregulation of the Bcl-2-Ig fusion gene in lymphoma. *The EMBO Journal 7*, 123-131.

63.   Willimott, S., and Wagner, S. (2010). Post-transcriptional and post-translational regulation of Bcl2. *Biochemical Society Transactions 38*, 1571-1575.

64.   Conde, I., Alfaro, J., Fraile, B., Ruiz, A., Paniagua, R., and Arenas, M. (2004). DAX-1 expression in human breast cancer: comparison with estrogen receptors ER-alpha, ER-beta and androgen receptor status. *Breast Cancer Research 6*, R140-148.

65.   Silva, D., and Smith, A. (2008). Capturing Pluripotency. Cell *132*, 532-536.

66.   Rossant, J. (2008). Stem cells and early lineage development. *Cell 132*, 527-531.

67.   Boiani, M., and Scholer, H.R. (2005). Regulatory networks in embryo-derived pluripotent stem cells. Nature reviews. Molecular cell biology *6*, 872-884.

68.   Ogawa, K., Matsui, H., Ohtsuka, S., and Niwa, H. (2004). A novel mechanism for regulating clonal propagation of mouse ES cells. *Genes to cells : devoted to molecular & cellular mechanisms 9*, 471-477.

69.   Tamm, C., Pijuan Galito, S., and Anneren, C. (2013). A comparative study of protocols for mouse embryonic stem cell culturing. *PloS one 8*, e81156.

70.   Williams, L.R., Hilton, D.J., Pease, S., Willson, T., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N., and Gough, N. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature 336*, 684-687.

71.   Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature 444*, 364-368.

72.   J, K., Chu, J., Shen, X., Wang, J., and Orkin, S. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell 132*.

73.   Niwa, H., Miyazaki, J., and Smith, A. (2000). Quantitative expression of Oct-3:4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics 24*, 372-376.

74.   Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes and Development 17*, 126-140.

75.   Chambers, I., Coldby, D., Robertson, M., Nichols, J., Lee, S., Tweedle, S., and Smith, A. (2003). Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell 113*, 643-655.

76.   Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., and Yokota, T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *The EMBO Journal 18*, 4261-4269.

77.   Cartwright, P., McLean, C.Y., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development 132*, 885-896.

78.   Nishiyama, A., Sharov, A.A., Piao, Y., Amano, M., Amano, T., Hoang, H.G., Binder, B.Y., Tapnio, R., Bassey, U., Malinou, J.N., et al. (2013). Systematic repression of transcription factors reveals limited patterns of gene expression changes in ES cells. *Scientific Reports 3*, 1390.

79.   Orkin, S.H. (2005). Chipping away at the embryonic stem cell network. *Cell 122*, 828-830.

80.   Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell 126*, 663-676.

81.   Khalfallah, O., Rouleau, M., Barbry, P., Bardoni, B., and Lalli, E. (2009). Dax-1 knock-down in mouse embryonic stem cells induces loss of pluripotency and multilineage differentiation. *Stem Cells 27*, 1529-1537.

82.   Torres, A. (2013). The Role of Dax-1 in Regulating Pluripotency in mouse Embryonic Stem Cells. In Biology, Volume Master of Science. (University of San Francisco).

83.   Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell 113*, 631-642.

84.   Radzisheuskaya, A., Chia Gle, B., dos Santos, R.L., Theunissen, T.W., Castro, L.F., Nichols, J., and Silva, J.C. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nature Cell Biology 15*, 579-590.

85.   Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R., and Grosveld, F. (1999). The transcription factor GATA6 is essential for early extraembryonic development. *Development 126*, 723-732.

86.   Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y., and Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *The International Journal of Biochemistry & Cell Biology 39*, 2195-2214.

87.   Sekido, R., and Lovell-Badge, R. (2013). Genetic control of testis development. *Sex Dev. 7*, 21-32.

88.   Showell, C., Binder, O., and Conlon, F. (2004). T-box Genes in Early Embryogenesis. *Dev Dyn. 229*, 201-218.

89.   Meno, C., Yukio, S., Fuhii, H., Ikeda, M., Yokoyama, T., Yokoyama, M., Toyoda, Y., and Hamada, H. (1996). Left–right asymmetric expression of the TGFbeta-family member lefty in mouse embryos. *Nature 381*, 151-155.

90.   Durlinger, A., Visser, J., and Themmen, A. (2002). Regulation of ovarian function: the role of anti-Müllerian hormone. *Reproduction 124*, 601-609.

91.   Kelly, V.R., and Hammer, G.D. (2011). LRH-1 and Nanog regulate Dax1 transcription in mouse embryonic stem cells. *Molecular and Cellular Endocrinology 332*, 116-124.

92. Wilkinson, D.G., Bhatt, S., and Herrmann, B.G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature 343*, 657-659.
93. Beck, S., Lee, B.K., Rhee, C., Song, J., Woo, A.J., and Kim, J. (2014). CpG island-mediated global gene regulatory modes in mouse embryonic stem cells. *Nature Communications 5*, 5490.
94. Waterson, R.H., Lindblad-Toh, K., and Birney, E. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature 420*, 520-562.
95. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol. 28*, 495-501.
96. Ihaka, R., and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics 5*, 299-314.
97. Berjerano, G. Statistics, GREAT Documentation.
98. Langfelder, P., Zhang, B., and Horvath, S. (2007). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics *24*, 719-720.
99. Wickham, H. (2009). ggplot2: elegant graphics for data analysis. (Springer New York).
100. Zurich, E.T.H. Principal Components Analysis R-Language Package.
101. Boroviak, T., and Nichols, J. (2014). The birth of embryonic pluripotency. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 369*.
102. Chae, B.J., Lee, A., Bae, J.S., Song, B.J., and Jung, S.S. (2011). Expression of nuclear receptor DAX-1 and androgen receptor in human breast cancer. *Journal of Surgical Oncology 103*, 768-772.
103. Roshan-Moniri, M., Hsing, M., Butler, M.S., Cherkasov, A., and Rennie, P.S. (2014). Orphan nuclear receptors as drug targets for the treatment of prostate and breast cancers. *Cancer Treatment Reviews 40*, 1137-1152.
104. Riggins, R.B., Mazzotta, M.M., Maniya, O.Z., and Clarke, R. (2010). Orphan nuclear receptors in breast cancer pathogenesis and therapeutic response. *Endocrine-related Cancer 17*, R213-231.