

Tipping the Balance: Factors That Influence the ER Signaling Network in Breast Cancer

by

Jeff S Jasper

Department of Pharmacology and Cancer Biology
Duke University

Date: _____

Approved:

Donald P. McDonnell, Supervisor

Ann Marie Pendergast

David M. MacAlpine

Kris C. Wood

Jen-Tsan Ashley Chi

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor
of Philosophy in the Department of Pharmacology and Cancer Biology in the Graduate
School of Duke University

2014

ABSTRACT

Tipping the Balance: Factors That Influence the ER Signaling Network in Breast Cancer

by

Jeff S Jasper

Department of Pharmacology and Cancer Biology
Duke University

Date: _____

Approved:

Donald P. McDonnell, Supervisor

Ann Marie Pendergast

David M. MacAlpine

Kris C. Wood

Jen-Tsan Ashley Chi

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Pharmacology and Cancer Biology in the Graduate School of Duke University

2014

Copyright by
Jeff S Jasper
2014

Abstract

The estrogen receptor(ER) is a master transcriptional regulator in the breast where it plays key roles in the development and maintenance of normal breast epithelium but is also critical to the growth of luminal breast cancers. ER is also a well-defined molecular therapeutic target and anti-estrogens, such as tamoxifen, are used clinically to inhibit the mitogenic activity of ER and delay disease progression. However, despite the initial benefits to tamoxifen therapy, nearly one third of luminal breast cancer tumors eventually become resistant, limiting the therapeutic utility the drug. Mechanisms of resistance can be attributed to circumvention of ER and reliance on alternative growth pathways, or through upregulation of pathways that converge with ER to allow reactivation. Understanding the molecular determinants of resistance is a critical endeavor that demands attention in order to shape new drug developments and extend the therapeutic efficacy of anti-estrogens.

A major challenge in elucidating mechanisms of resistance is in understanding the complexities of the ER signaling program in respect to receptor occupancy and the coordinated relationship with chromatin architecture and collaborating transcription factors. This work therefore integrates the relationship between accessible chromatin, as measured by DNase-Seq, with ER occupancy and ER-mediated transcription in an in vivo derived tamoxifen resistant cell line (TamR) and a comparator group of two closely related tamoxifen sensitive cell lines. Cumulatively, these data demonstrate an enhanced

role for FOXA1 in tamoxifen resistance. Specifically, FOXA1 occupancy is greatly enriched at differential DNase hypersensitive loci in TamR cells, and expression of nearby genes are transcriptionally upregulated in a FOXA1-dependent manner. Furthermore, siRNA directed against FOXA1 can rescue the expression of these target genes to MCF7 levels. The TamR cells also have increased ER occupancy at FOXA1 overlapping sites, where ER is engaged to chromatin in a ligand-independent manner and results in enhanced activation of nearby target genes that can be repressed with the pure anti-estrogen, ICI. The increased role of FOXA1 is not due to an increase in total protein levels however and instead is manifested through increased activity that may depend on protein modifications.

Other clinical associations of resistance have been elucidated for which there is little to no mechanistic evidence currently available. The expression of HOXB13 is associated with tamoxifen therapy failure from differential microarray expression profiling of patients who relapsed compared to those that remained disease-free at the five year follow-up. The outcome of our studies elucidates the activity of HOXB13 to downregulate the expression of GATA3 which subsequently leads to loss of ER function and parallel activation of inflammatory pathways. Forced overexpression of GATA3 in this background restores ER expression, partially restores ER target gene expression and suppresses the expression of the HOXB13-induced inflammatory genes.

The present study also makes use of publicly available clinical datasets to generate an integrative database of 4885 patients from 25 independent studies. Furthermore, analytical methods and functions were also developed to allow efficient use and application of the data. Access to the breast cancer meta-set and functions are made available to end users via a web interface, GeneAnalytics. Together, the breast cancer meta-set and associated access through the GeneAnalytics web sites provides novel opportunities for researchers to integrate functional studies with tumor derived expression data to further our understanding of cancer related processes.

Collectively, our findings demonstrate that the ER signaling program is modified as tumors progress to resistance by an increased role of FOXA1 to facilitate ER binding and reprogramming, and by HOXB13 to suppress the actions of ER and promote inflammatory pathways. These mechanisms highlight distinct methods of resistance and provide rational for new therapeutic approaches to extend the utility of the anti-estrogens currently in use.

Dedication

To my family for their love and support throughout the journey of grad school.

Contents

Abstract	iv
List of Tables	xii
List of Figures	xiii
List of Abbreviations	xv
Acknowledgements	xvii
1 Introduction.....	1
1.1 Breast Cancer.....	1
1.1.1 General overview	1
1.1.2 Complex heterogeneity	2
1.2 Estrogen Receptor signaling program	4
1.2.1 Discovery and importance in breast cancer	4
1.2.2 Mechanism of ER function.....	5
1.2.3 Coregulators.....	12
1.2.4 Chromatin dynamics and collaborating factors.....	14
1.3 ER targeted therapies and resistance.....	20
1.3.1 Pharmacological intervention.....	20
1.3.2 Endocrine resistance in the clinic.....	23
1.3.3 Mechanisms of endocrine resistance	25
2 Identification of the DNA-dependent, ER associated factors in tamoxifen resistance. 28	
2.1 Introduction.....	28

2.1.1	Genomic evaluation of ER.....	29
2.2	Results	33
2.2.1	DNase-Seq identifies an increased role for FOXA family members in a cellular model of tamoxifen resistance.....	33
2.2.1.1	DNase-Seq on sensitive and resistant cell lines.....	33
2.2.1.2	Motif Enrichment Analysis	38
2.2.1.3	Comparison with mapped binding sites.....	40
2.2.1.4	Footprint analysis of top factors.....	44
2.2.1.5	Correlation with resistant-specific ER binding sites.....	45
2.2.2	Increased occupancy of FOXA1 and ER in TamR cells.....	46
2.2.3	Characterization of FOXA1 and ER function in TamR cells	48
2.2.4	Identification of FOXA1 interacting partners.....	50
2.2.5	Integrative analysis with mRNA expression.....	53
2.3	Discussion.....	57
3	The role of HOXB13 on breast cancer pathogenesis.....	62
3.1	Introduction.....	62
3.2	Results	66
3.2.1	Validation of the impact of HOXB13 on breast cancer progression and clinicopathological features	66
3.2.2	HOXB13 attenuates ER signaling.....	68
3.2.3	HOXB13 promotes inflammatory pathways.....	69
3.2.3.1	Function Coexpression Analysis identifies an interferon gene signature and loss of a subset of ER target genes	69
3.2.4	HOXB13 promotes a pro-inflammatory response	71

3.2.5	Overexpression of HOXB13 in cellular models deregulates ER signaling via loss of GATA3.....	72
3.2.6	Loss of GATA3 promotes the HOXB13-mediated inflammatory phenotype. 73	
3.2.7	HOXB13 induces AP-1 Responsive elements.....	75
3.2.8	Cardiac Glycosides downregulate HOXB13	76
3.3	Discussion.....	78
4	Development of a breast cancer meta-set and web application for discovering gene function and clinical relationships.....	80
4.1	Introduction.....	80
4.2	Results	81
4.2.1	Development of a breast cancer meta-set	81
4.2.2	Development of functional modules	86
4.2.2.1	Expression Analysis	87
4.2.2.2	Survival analysis.....	89
4.2.2.3	Forest plot.....	91
4.2.2.4	Functional co-expression analysis.....	92
4.2.3	Web application.....	94
4.3	Discussion.....	96
5	Conclusions and future directions	97
6	Materials and Methods.....	101
6.1	Chemicals.....	101
6.2	Antibodies.....	101
6.3	Plasmids.....	102

6.4	Cell Culture	102
6.5	RNA Isolation and qPCR.....	103
6.6	Transient Transfection Assays.....	105
6.7	Viral Stable Cell Generation.....	106
6.8	RNA-Seq	107
6.9	siRNA Experiments.....	108
6.10	Western Blotting.....	109
6.11	Chromatin Immunoprecipitation.....	109
6.12	Preparation of Nuclear Extracts for Mass-Spec	110
6.13	DNase-Seq Library Generation	112
6.14	DNase-Seq Analysis.....	113
6.15	Motif Analysis.....	114
	References	115
	Biography.....	141

List of Tables

Table 1-1: Tissue agonist profile of common SERMs.....	22
Table 2-1: Overview of sequencing libraries and reads.....	34
Table 2-2: Table of high confidence de-novo motif finding results from Homer for sites increased in DHS signal.....	39
Table 4-1: GSEA results from Functional co-expression analysis for AURKA.....	94
Table 6-1: qPCR primers used.....	103
Table 6-2: siRNA sequences.....	108
Table 6-3: CHIP primers used.....	110

List of Figures

Figure 1-1: Intrinsic subtypes stratify prognosis.	3
Figure 1-2: ER structure and homology.....	9
Figure 1-3: ER signaling mechanisms.	11
Figure 1-4: Pioneer factors facilitate chromatin opening.....	16
Figure 1-5: Overlap of pioneer factor binding.	20
Figure 2-1: Differential DHS sites between cell lines.	35
Figure 2-2: Plots of significant DHS sites.....	37
Figure 2-3: Distribution of DHS sites.	38
Figure 2-4: Identified motifs in differential DHS sites.....	40
Figure 2-5: Heatmap of binding profiles.	42
Figure 2-6: Differential binding profiles for ER and FOXA1.	43
Figure 2-7: Footprint analysis of FOXA1.	45
Figure 2-8: Comparison of DNase data with 448 sites found in tamoxifen resistance.....	46
Figure 2-9: ChIP-qPCR of ER and FOXA1.....	47
Figure 2-10: Heatmap of FOXA1 activated targets.....	49
Figure 2-11: FOXA1 repressed genes.	50
Figure 2-12: Mass-Spec identifies putative interacting partners.	52
Figure 2-13: GSEA enrichment map.	54
Figure 2-14: Network plot of factors identified by Mass-Spec.	57
Figure 3-1: Clinical association of HOXB13 in a breast cancer meta-set.	67
Figure 3-2: Waterfall Plot of HOXB13 in TCGA data.....	68

Figure 3-3: HOXB13 attenuates ER signaling.....	69
Figure 3-4: Association of HOXB13 with notable factors.	71
Figure 3-5: HOXB13 promotes an inflammatory response.	72
Figure 3-6: HOXB13 expression leads to loss of ER target genes.....	73
Figure 3-7: Adenovirus of GATA3 ameliorates the HOXB13-induced inflammatory genes.	74
Figure 3-8: HOXB13-mediated activation of AP-1 reporters is sensitive to MEK1/2 inhibition.	76
Figure 3-9: Cardiac glycosides suppress HOXB13 in BT474 cells.	78
Figure 4-1: MDS plot of data before and after merging.....	84
Figure 4-2: Histogram of ER expression.	85
Figure 4-3: Expression Analysis of AURKA.....	88
Figure 4-4: Survival analysis of G0:0046605 genes.	90
Figure 4-5: Survival analysis of AURKA.	91
Figure 4-6: Forest plot of GO:0046605 genes.	92
Figure 4-7: Screen shot of GeneAnalytics website.....	95

List of Abbreviations

4OHT	4-hydroxytamoxifen
AF (AF-1, AF-2)	Activation function
AI	Aromatase inhibitor
AIB1	Amplified in breast cancer 1
ANOVA	One way analysis of variance
AP-1	Activator protein 1
AR	Androgen receptor
BMD	Bone mineral density
CBP/p300	Cyclic adenosine monophosphate response element binding (CREB) binding protein
CCDC137	Coiled-coil domain containing protein 137
CDC2L5	cell division cycle 2-like 5
cDNA	Complementary deoxyribonucleic acid
CFS	Charcoal-stripped fetal bovine serum
ChIP	Chromatin immunoprecipitation
CXCL12	Chemokine (C-X-C motif) Ligand 12
CXCR4	Chemokine (C-X-C motif) receptor 4
CYP	Cytochrome P450
DBD	DNA binding domain
DNA	Deoxyribonucleic acid
E1	Estrone
E2	Estradiol
E3	Estriol
EGFR	Epidermal growth factor receptor
ELISA	Enzyme linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
ER	Estrogen receptor
ERBB4	v-erb-a erythroblastic leukemia viral oncogene homolog 4
ERE	Estrogen response element
ERK	Extracellular signal-regulated kinase
ERR γ	Estrogen related receptor γ
ET/HT	Estrogen /Hormone replacement therapy
FBS	Fetal bovine serum
FKBP51	FK506-Binding Immunophilin 51
FSD1	fibronectin type 3 and SPRY domain containing 1
FSH	Follicle stimulating hormone
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase

GFP/EGFP	Green fluorescent protein
GR	Glucocorticoid receptor
GRIP1	Glucocorticoid receptor interacting protein 1
H12	Helix 12
HNF	Hepatocyte nuclear factor
HSD	Hydroxysteroid dehydrogenase
Hsp90	Heat shock protein 90
ICI	ICI 182,780 / faslodex / fulvestrant
IGF-1	Insulin-like growth factor 1
IL	Interleukin
LBD	Ligand binding domain
Luc	Luciferase
M2H	Mammalian 2-hybrid
MAPK	Mitogen activated protein kinase
mRNA	Messenger ribonucleic acid
NCoR	nuclear receptor corepressor
NFκB	Nuclear Factor κ B
NR	Nuclear receptor
PBX	Pre-B-cell leukemia homeobox
PCR	Polymerase chain reaction
PI3K	Phosphatidylinositol (PI) 3-kinase
PKC	Protein kinase C
Pol II	RNA polymerase II
PPAR	Peroxisome proliferator-activated receptor
PR	Progesterone receptor
PS2	Trefoil factor 1
qPCR	Quantitative real time polymerase chain reaction
RAL	Raloxifene
SDF-1	Stromal derived factor 1
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SERD	Selective estrogen receptor down-regulator
SERM	Selective estrogen receptor modulator
siRNA	Small interfering ribonucleic acid
SMAD3	Mothers against decapentaplegic homolog 3
SMRT	silencing mediator of retinoid and thyroid hormone receptors
Sp1	Specificity protein 1
TAM	Tamoxifen
TRAP220	thyroid hormone receptor-associated protein 220
WISP2	WNT1 inducible signaling pathway protein 2
beta-gal	beta-Galactosidase

Acknowledgements

Foremost, I would like to acknowledge my dearest wife, Stephanie, for supporting my every need throughout graduate school and allowing me the time and energy to focus my attention on science related endeavors. She has supported me with constant love, devotion and incredible patience in the face of physical hardships, late nights, and tight finances-all the while effectively meeting the demands of an ever growing family. She has selflessly devoted countless hours in the development and watchful care of our beautiful children Evelyn, Olivia and the as of yet unborn baby boy with enthusiasm and great passion. I am thankful that I could have her to rely on for emotional support and that I can count on her laughter and encouragement to uplift me on the most difficult of days. We started together on this journey full-knowing what trials we might encounter, especially with family far away, and ultimately I am grateful for everything she has done to make it a success-I love you!

My kids also deserve huge credit for sacrificing precious together time for the sake of science. Playing together with them is always entertaining and their contagious laughter keeps me grounded and inspires an optimistic outlook to all aspects of life. Overall I don't know how I could have devoted the amount of time and effort towards research without the benefits that result from a balance of work and an enjoyable home life. I also wish to thank my parents, Scott and Sandy, for their long-standing support

and encouragement. Even as a kid, they taught me to aim higher, overcome fears, and to live life with a purpose; for which I will forever be grateful.

To the McDonnell lab, I am grateful for your patience and willingness to share so much of your knowledge and time whenever occasion permitted. Each person has contributed to this body of work in some way, whether it is with data, reagents, insightful discussion or through inspiring motivation. I have relied on many current and past members of the lab for advice and will remember a lot of great moments between discussions in cell culture, practical jokes and random sayings. Honestly I can't thank you all enough and no amount of words can do it justice. I am extremely grateful to Ching-Yi, for her invaluable input, thoughtful discussions and pragmatism to keep my scientific vision focused. I would be absolutely lost without her.

I also want to give a special thanks to Donald, for fostering a productive work environment, for inspiring and constantly encouraging me to do bigger and greater things, and for his boundless curiosity and love of science. His infectious enthusiasm for data, colorful graphs, and good jokes will not be forgotten. I have learned a great deal from his unique perspective on life and his mentorship has helped me discover what to aspire towards in both a scientific career and life after graduate school.

I would also like to express gratitude to my committee members: Ann Marie Pendergast, David MacAlpine, Kris Wood and Jen-Tsan Ashley Chi for their patience,

enthusiasm, and willingness to devote time to help guide my projects and direct me in successful paths.

Thank you all for making graduate school a worthwhile, enjoyable experience!

1 Introduction

1.1 Breast Cancer

1.1.1 General overview

Breast cancer is a complex, highly prevalent disease that continues to be one of the leading causes of death among women. The American Cancer Society estimates that in 2014, 232,670 new cases of breast cancer will be diagnosed and approximately 40,000 women will die as a result of breast cancer in the USA (Siegel et al., 2014). Breast cancer is second only to lung cancer for cancer-related deaths among women but has a higher incidence rate overall. The current lifetime probability of occurrence is 1 in 8 for females and current incidence rates are holding steady. Even though incidence rates are less than decades past, in part due to better detection and treatment, breast cancer remains a significant health burden worldwide and demands further scientific research.

There are number of risk factors associated with breast cancer including sex, age, hormones, BRCA1/2 mutations, and breast density. Upon diagnosis the treatment strategy is largely dictated by histopathological features, tumor grade, proliferative capacity, metastatic events and tumor subtype. Depending on tumor staging, the patient will undergo a series of treatment modalities including surgical resection, radiation therapy, cytotoxic agents, endocrine disruption and targeted therapies. Treatment timing is also critical. Patients may be given therapeutics, cytotoxics, or radiation prior to surgery (neo-adjuvant) with the intent of improving surgical success or delaying

progression. Surgical approaches vary based on risk assessment and are primarily aimed at reducing tumor burden and removing affected areas of the breast and lymphatic system to improve overall survival. Following primary surgical treatment, patients remain at risk for recurrent events; whether they occur locally, in the contralateral breast or as distant metastases and compel the need for long term therapeutic intervention. The risk of recurrence is greatest in the first five years with the peak as early as two years but varies widely depending on the tumor subtype.

1.1.2 Complex heterogeneity

Breast cancer is inherently heterogeneous with several distinct biological subtypes that vary in prognostic and therapeutic considerations. Clinicians have long appreciated that breast cancer cannot be treated as a single disease and routinely segregate tumors into manageable phenotypic classes using IHC directed against known markers in breast cancer including ER α , PR, HER2 and a proliferation marker such as Ki67.

Thanks to the explosion in genomic technologies we now have a better understanding of the molecular determinants that influence breast cancer heterogeneity beyond these few markers. In a ground breaking study, Perou and colleagues performed unsupervised clustering techniques on microarray data from patient samples and observed that the samples clustered into biologically relevant subgroups. They defined the subtypes as Luminal A, Luminal B, Basal-like, HER2-enriched, and Normal-like

reflecting their histological counterparts (Perou et al., 2000; Sørlie et al., 2001). They further postulated that the difference in expression profiles (molecular portraits) underlies the heterogeneity of breast tumors and diverse response to treatments and could be distilled down into key features, known as the PAM50 gene-set (Parker et al., 2009), to identify molecular phenotypes. Furthermore, they also showed that there were significant differences in recurrence rates between the clusters; with basal and HER2 subtypes indicating a worst prognosis and the two Luminal subtypes exhibiting more favorable outcomes as seen in Figure 1-1. Subsequent studies have confirmed distinct roles for the intrinsic subtypes on incidence, survival and response to treatments (Cheang et al., 2009; Millikan et al., 2008; Nielsen et al., 2010; Prat et al., 2010).

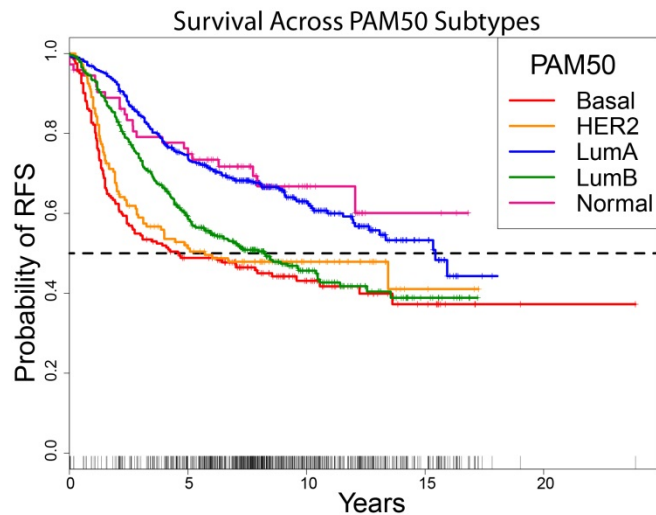


Figure 1-1: Intrinsic subtypes stratify prognosis.

Outcome probability of relapse-free survival in a cohort of 3374 patients as discussed in Chapter 4. Intrinsic subtypes were determined using a 50 gene-set classifier (PAM50) and subtypes were assigned using a nearest-centroid approach (Parker et al., 2009).

1.2 Estrogen Receptor signaling program

1.2.1 Discovery and importance in breast cancer

George Beatson, a surgeon in Scotland, was the first to publish on the association of hormones produced in the ovaries to the proliferative actions it had on the breast in 1896 (Beatson, 1897). Taking cues from farmers that removed the ovaries of lactating cows to have them produce milk indefinitely, he performed experiments on rabbits to validate the dependency (Beatson, 1899). He later detailed three case studies of advanced, inoperable breast cancer patients that entered remission following oophorectomy (surgical removal of the ovaries) (Beatson, 1896). Based on these translational results, Stanley Boyd and others initiated large clinical tests targeting the ovaries for surgical removal and observed that roughly one-third of patients benefited from ovarian ablation (Boyd, 1902). It's important to note that these studies were performed without any knowledge of the hormones involved or their mechanistic actions.

In 1923, Allen and Doisy identified the responsible estrogenic steroid hormone and Doisy later crystallized the first steroid hormone, estrone, in 1929 (Allen and Doisy, 1923). This discovery prompted research to understand how hormones signal and achieve tissue specificity. Fueled by the possibility that hormone action could be attributed to specific binding proteins--contrary to prevailing theory--Elwood Jensen administered radiolabeled estradiol to rats and found that it was predominately

sequestered to the female reproductive tract (Glascock and Hoekstra, 1959; Jensen et al., 1966). This finding prompted him to suggest that an “estrogen receptor” existed and later identified the cell fractions containing the macromolecular complex (Jensen and DeSombre, 1973). Since it had been known that only one-third of patients respond to ovarian ablation, Jensen predicted that the estrogen receptor levels was the limiting factor and hence could determine outcome. He devised a clinical test to assess ER levels and found that patients with high amounts of the receptor responded positively while those with low amounts did not. This is one the first examples of the use of a predictive biomarker in the cancer setting. Subsequent use of antibodies by Greene and Jensen allowed for greater sensitivity and enabled quick quantification for more routine use (Greene et al., 1980). The ability to stratify tumors that are more likely to respond to endocrine therapies and rapid method of testing greatly benefited the clinic and became part of the standard of care by the 1980's. This paradigm of testing laid the groundwork for tumor staging and the definition of the clinico-pathological subtypes that have evolved into the molecular subtypes in use today.

1.2.2 Mechanism of ER function

Estrogenic ligands and their cognate receptors are essential for a myriad of normal physiological processes and play pivotal roles in numerous diseases states (Burns and Korach, 2012). Estrogens are predominately produced in the ovaries and, like other steroidal hormones, are synthesized from precursors derived from a common

cholesterol backbone. The culminating step in the biosynthesis of estrogens requires the CYP19 aromatase, a P450 enzyme that converts androstenedione into estrone or testosterone into estradiol. Estrone can also be reduced to estradiol by 17 β -hydroxysteroid dehydrogenase. Estradiol is the most abundant circulating estrogen during pre-menopausal years and its levels are coupled to the menstrual cycle where it increases sharply near the end of the follicular phase just prior to ovulation. Owing to its physicochemical properties as a steroid, estrogens are lipophilic and can freely diffuse into cells where it interacts with the estrogen receptor.

The biological activities of estrogens in cells are mediated by the estrogen receptor, a ligand-dependent transcription factor and a member of the nuclear receptor superfamily. In the absence of ligand, ER resides in the cytosol in an inactive monomeric form bound to heat shock proteins. Upon binding ligand, the receptor undergoes a conformation change, dissociates from the heat shock complex, dimerizes and translocates to the nucleus where it interacts directly with DNA at a specific motif commonly known as the estrogen response element (ERE). The canonical ERE is defined as an inverted repeat of DNA half sites also known as a palindrome and is based on the *Xenopus laevis* vitellogenin A2 promoter: 5'-GGTCAnnnTGACC-3' (Gruber et al., 2004). Upon binding to DNA, ER recruits a multitude of coregulatory factors including transcriptional coactivators, histone acetyltransferases (HATs), histone

methyltransferases (HMTs), and other chromatin remodelers to promote a permissive chromatin environment for transcription.

While it was initially thought that the receptor was limited to an active or inactive conformation, it is now understood that the structure and response is not binary. Each ligand confers a unique conformation and allows for a spectrum of activities (Norris et al., 1999). The biological response therefore hinges upon the ligand-induced conformational change and dictates key facets of regulation including cofactor binding profiles, DNA binding preferences and receptor turnover (Rosenfeld and Glass, 2001).

ER is a member of the Class I nuclear receptors which function as homodimers on DNA and include the glucocorticoid receptor, androgen receptor, mineralocorticoid receptor and progesterone receptor (Bain et al., 2007; Beato and Klug, 2000; Weatherman et al., 1999). The transcriptional activity of the estrogen receptor is mediated by two key functional domains termed AF-1 and AF-2 which contain key docking surfaces for coregulators and serves also as sites for protein modifications (Figure 1-2) (Tremblay et al., 1999; Webb et al., 1998). Like other nuclear receptors, the estrogen receptor shares the typical patterning of structural domains A-F. The N-terminal A/B domain contains activating function 1 (AF-1) and its activity does not require ligand binding *per se*, but is manifest upon the delivery of the receptor to DNA. There are several phosphorylation sites (S104, S106, S118, and S167) with AF-1 that are the targets of growth factor

regulated kinases and impact the functionality of this transcriptional domain. The C domain represents the DNA-binding domain (DBD) with its two zinc fingers and plays a minor role in receptor dimerization. The D-domain or hinge region contain a nuclear localization sequence and is necessary for complete activation and synergy between the AF-1 and AF-2 domains (Zwart et al., 2010). The E domain contains the ligand-binding domain (LBD) and AF-2 the activity of which is influenced by ligand-dependent changes of the receptor. Ligand binding induces a structure which allows helix 3, 4, 5, and 12 to form a shallow hydrophobic groove that exposes a surface that recognizes a commonly conserved LXXLL motif on coactivators. The E domain also contains a strong dimerization sequence. Lastly, the carboxy-terminal F domain is reported to be necessary in mediating the effects of antagonists (Montano et al., 1995) and is important for receptor turnover (Wittmann et al., 2007).

Two molecular forms of the estrogen receptor have been identified in humans, ER α and ER β , which are encoded by separate genes ESR1 and ESR2 respectively (Dupont et al., 2000; Hewitt and Korach, 2003; Ruff et al., 2000). The receptors are transcribed from different chromosomes and while both receptors are expressed in most cells, the expression of one or both subtypes has been observed in some cells. The amino acid structure of ER α and ER β are highly homologous yet they mediate distinct biological activities (Figure 1-2). ER β is more highly expressed in the prostate, ovaries, lungs, and nervous systems while ER α is predominately expressed in

the breast, uterus, pituitary gland, liver, kidneys, and adrenal gland (Zilli et al., 2009). Both forms are expressed in the breast and their interplay and ability to oppose one another has been subject to an evolving debate on clinical implications and therapeutic strategies (Grober et al., 2011; Hall and McDonnell, 1999; Katzenellenbogen and Katzenellenbogen, 2000). ER α is a more robust transactivator and experimentally ER β has been shown to suppress the activities of ER α (Hall and McDonnell, 1999). The absolute and relative ratios are therefore important in determining the overall response. The differences in activity can be attributed to distinct sequence compositions of the receptors themselves. ER α is composed of 595 amino acids and ER β has 530. The relative differences and similarities are further exemplified in the degree of homology across the different domains (Figure 1-2). In particular, it is notable that the DBD is 97% conserved while the A/B domain containing AF-1 is only 18% conserved.

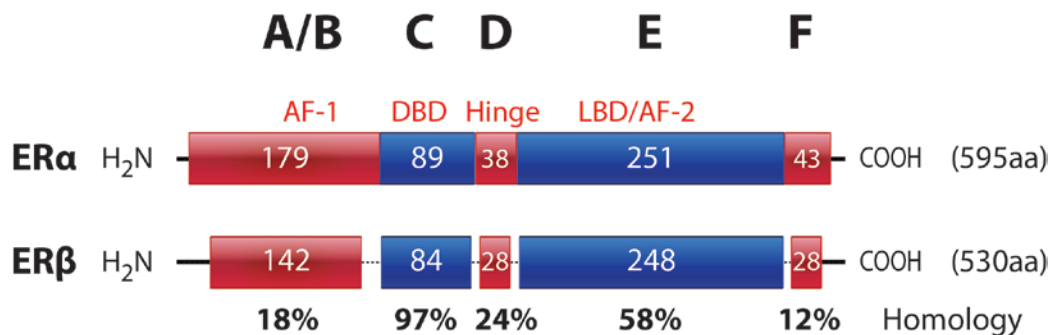


Figure 1-2: ER structure and homology.

A schematic representation of the protein structural domains and key features of the ER receptors. Listed above are the nuclear receptor derived domains A-F. The numbers inside the box refer to

the number of amino acids and total amino acids to the right. Amino acid percent homology is listed below.

Genome wide profiling of ER binding sites has confirmed that the canonical ERE is the most prominent DNA motif present within the regulatory regions of these target genes, but also that the homo-dimerized receptor binding preferences are somewhat flexible and allow for some variations beyond the canonical sequence (Hurtado et al., 2011a; Kong et al., 2011b; Ross-Innes et al., 2012a; Welboren et al., 2009). Adding another layer of complexity is the fact that many ER binding sites contain only a partial ERE and other regions do not bear any resemblance to an ERE. In addition to the classical mode of ER action, ER can also interact with DNA in an indirect manner via tethering to other transcription factors including binding to FOS and JUN at activator protein 1 (AP-1) elements, to specificity protein 1 (SP1), or to nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) (Figure 1-3). Presumably tethered complexes and other factors that stabilize ER at imperfect EREs account for the binding anomalies observed in global binding studies.

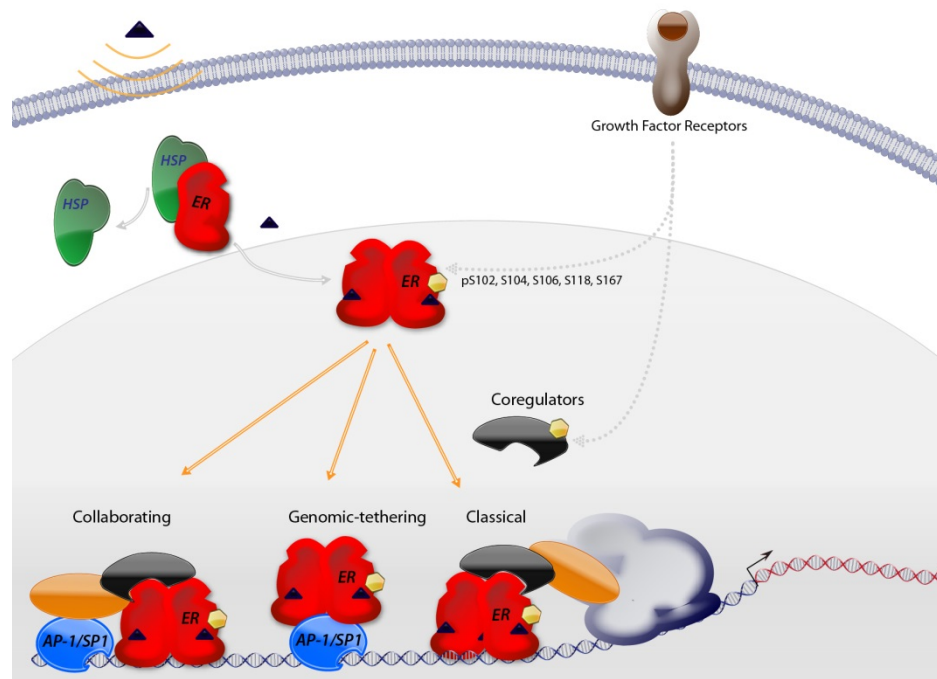


Figure 1-3: ER signaling mechanisms.

Illustration of the ER signaling pathway. Key phosphorylation sites are pictured and also three distinct methods that allow ER to associate with chromatin. Classical: direct binding of EREs. Tethering: protein-protein interaction with general transcription factors. Collaborating: Stabilization of binding through adjacent factors.

A recent study leveraged the peculiarity of ER binding sites at imperfect EREs to elucidate their relationship across different cell types (Gertz et al., 2013). They compared the ER cistrome of T-47D cells (breast) and ECC-1 (endometrial) cells for the sequence determinants of binding relative to chromatin accessibility as determined by DNase-Seq. In agreement with the observation of cell-specific ER activities, ER binding sites were also diverse and subsets of cell-specific and shared (~10%) sites were identified. Roughly one-third of all sites had no discernible ERE and these were three times more likely to be found within genes that are expressed in a cell selective manner. Moreover, shared sites

were enriched in high affinity EREs while cell-specific sites contained lower affinity EREs but also were strongly associated with other transcription factor motifs that were expressed in a cell-specific manner.

1.2.3 Coregulators

Transcriptional activation by ER involves a coordinated interplay between the receptor, ligand, chromatin, and coregulators for robust activation of target genes. Upon associating with chromatin, coactivators are recruited to ER and impart transcriptional regulation by remodeling local chromatin and permitting RNA polymerase activity. The first glimpse of coactivator function was revealed with the observation that transcription factors can interfere with one another through a phenomenon known as squelching, an event attributed to a limited supply of a factor absolutely necessary for activation and results in reduced activation of both receptors. Later, a yeast two-hybrid screen recovered SRC-1 and found it to bind to both PR and ER and that over expression of SRC-1 could mitigate the squelching effects known to occur between the two receptors (Oñate et al., 1995).

Hundreds of coactivators have been identified and the majority contain a signature leucine-rich motif (LXXLL), also called the NR box, which is both necessary and sufficient for binding to ER (Heery et al., 1997). Crystallization studies of the LBD of ER and a fragment of PPAR γ have elucidated that coactivators bind to a hydrophobic cleft of the receptor comprising helices 3, 4, 5 and 12 through a specific interaction with

the LXXLL motif (Nolte et al., 1998). Flanking amino acids of the motif were also found to be important in receptor selectivity and activity (Chang et al., 1999).

The ligand-induced conformation of ER dictates coregulator binding profiles. Crystal structures of ER have elucidated that helix 12 is highly mobile in response to the nature of the bound ligand and truncation of helix 12 at 535 abolishes coregulator binding. Antagonists bound by the receptor results in a unique conformation of the receptor and repositions helix 12 effectively blocking the LXXL interaction motif thereby rendering ER activity null (Shiau et al., 1998). Antagonists, including tamoxifen, and their role on coregulator recruitment will be discussed in section 1.3.

The fundamental unit of chromatin is the nucleosome and consists of DNA wrapped around a histone octamer in a 147bp stretch. Nucleosomes play an important role in inhibiting gene expression by blocking access of regulatory proteins and RNA polymerase to DNA. Consequently, chromatin structure must be reorganized by nucleosome loss, rearrangement or displacement during transcriptional activation or preset to allow activation. The N-terminal tails of core histones are subject to modifications (acetylation, methylation, phosphorylation and ubiquitination) and play a fundamental role in structural organization of chromatin and in recruiting effector molecules for transcriptional initiation and elongation(Hebbes et al., 1988; Lee et al., 1993). Coactivators that possess enzymatic functions, such as the P160 family which contain domains for HAT activity, have an active role in this process, while others act as

scaffolds enabling the assembly of secondary protein complexes(Halachmi et al., 1994; Oñate et al., 1995). Multi-protein complexes are common and contain HATs (p300, CBP, p/CAF, and the P160 family), HMTs (PRMT1 and CARM1) and ATPases (SWI/SNF) that together destabilize nucleosomes and facilitate association with mediator complexes to promote PolIII initiation.

In addition to coactivators, corepressors are equally important in negatively regulating ER activity. ER can actively repress genes in a ligand-dependent manner by recruitment of corepressors to silence gene expression. Diametric in function, corepressors interact with ER to mitigate transcriptional activation by facilitating chromatin condensation. Interaction of the corepressors with ER is mediated through the CoRNR box of ER (Shang et al., 2000). Corepressors include NCoR, SMRT, NRIP1, SAFB1, REA, and HDACs. These rely upon secondary factors that allow them to exert repressive effects. As an example, both NCoR and SMRT associate with HDAC3 in order to achieve transcriptional silencing.

1.2.4 Chromatin dynamics and collaborating factors

The simplified model of ER signaling entails ligand bound receptor associating with chromatin where it recruits coregulators and activates target gene expression. However, genome wide chip studies have revealed that a transcription factor generally only occupies a small fraction of the potential number of high affinity binding sites available within the genome(Johnson et al., 2007). Simply stated, chromatin features

beyond DNA sequence are important determinants of whether a factor can indeed bind to DNA. The simplified model can therefore be expounded upon with the idea that ER predominately binds to regions of open chromatin and that the chromatin landscape therefore governs binding capacity. As each tissue has distinct profiles of open chromatin (Song et al., 2011a), distinct ER binding profiles are also observed and this results in unique gene expression patterns. This raises the question of how binding sites become competent and what factors are necessary for an intact ER signaling program.

Our understanding of the fundamental aspects of ER activity has evolved with the burgeoning information obtained from genome-wide studies. The use of chromatin immunoprecipitation coupled with genome wide microarrays (ChIP-chip) or next generation sequencing (ChIP-Seq and ChIP-exo) (Johnson et al., 2007; Rhee and Pugh, 2011) have allowed global identification of ER binding sites in an unbiased manner (Carroll et al., 2005; Hurtado et al., 2011a; Joseph et al., 2010; Kong et al., 2011b; Ross-Innes et al., 2012a; Welboren et al., 2009). The estrogen receptor has been subject to a large number of genome profiling studies and these have revealed that the majority of ER binding events occur in cis-regulatory distal enhancer regions as opposed to proximal promoter regions. Motif enrichment analyses of ER binding sites have also revealed that other transcription factor motifs are commonly associated with ER, namely the pioneer factor FOXA1 (Carroll et al., 2005; Ross-Innes et al., 2012b).

Pioneer factors are defined by their ability to associate to DNA elements within condensed chromatin and permit other transcription factors to bind (Figure 1-4) (Cirillo et al., 2002; Lupien et al., 2008).

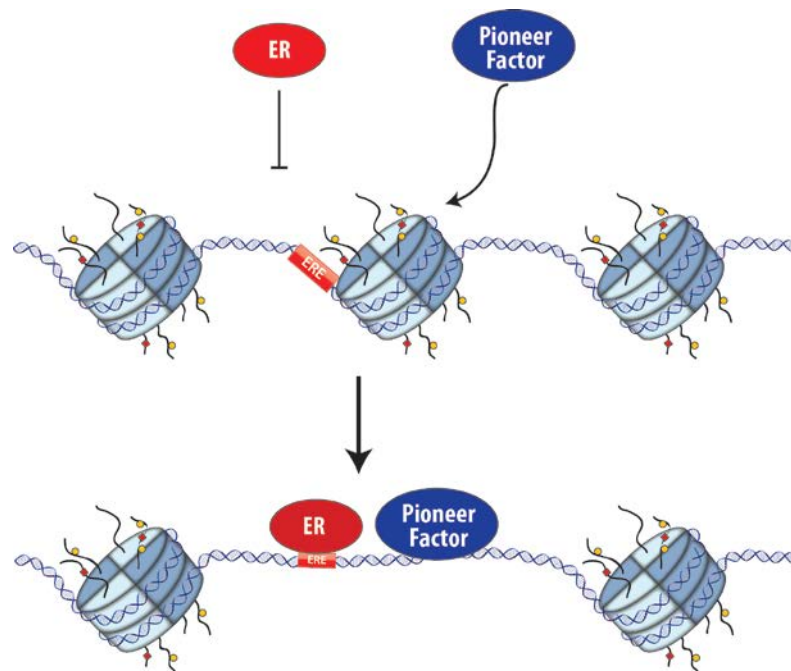


Figure 1-4: Pioneer factors facilitate chromatin opening.

Pioneer factors can directly interact with nucleosomes to allow access for other general transcription factors such as ER.

Pioneer factors are required for initiation of transcription and also play a role in maintaining competency of an enhancer by active presence. FOXA1 has been profiled by ChIP-Seq in MCF-7 cells and nearly half of all ER binding sites are coincident with FOXA1 sites (Hurtado et al., 2011b). Further substantiating ER's dependency on FOXA1, siRNA directed against FOXA1 results in loss of ER binding at these loci and concomitant decrease in target gene expression (Eeckhoutte et al., 2006; Laganière et al.,

2005). Conversely, loss of ER has little to no impact on FOXA1 binding as binding is independent of ER actions (Hurtado et al., 2011b). In another study, forced overexpression of ER, GATA3, and FOXA1 could recapitulate a complete ER signaling program in an ER negative cell line while ER expression alone could not (Kong et al., 2011b).

FOXA1 is a member of the forkhead family of transcription factors that were originally defined for their developmental role in the liver (Nagy et al., 1994). The pioneering capacity of FOXA1 is embodied in the forkhead domain, which is a variant of the helix-turn-helix structure and is composed of two large loops that resemble a winged helix. The structure of the winged helix is notable because it is highly similar to linker histone structure and also possesses DNA binding capacity allowing it to both bind nucleosome core particles and DNA simultaneously (Chaya et al., 2001). FOXA1 therefore facilitates ER binding by associating to condensed chromatin, displacing H1 linker histones and destabilizing the nucleosome allowing for other factors to engage DNA. Its ability to open condensed chromatin is ATP-independent and there is evidence that it binds nucleosomal DNA more strongly than to naked DNA (Cirillo and Zaret, 1999). FOXA1 also maintains competency of the binding site by inhibiting nucleosome condensation and is retained on DNA through mitosis acting as a bookmark for subsequent nucleosome positioning. Lastly, it is important to note that FOXA1, like

other transcription factors, does not bind to every available motif in the genome, indicating that additional levels of specificity exist.

As genomic technologies have advanced, it is now clear that chromatin dynamics are closely associated with and contingent upon DNA methylation, nucleosome positioning, and epigenetic features (Thurman et al., 2012). Despite its role as a pioneer factor, FOXA1 binding activity can be altered by certain chromatin modifications. Due to binding constraints in its forkhead domain, FOXA1 associates poorly with methylated DNA. Additional studies suggest that the binding of FOXA1 to chromatin is governed by epigenetic modifications on histone H3, specifically on the mono- and di-methylation of histone H3 at lysine 4 (H3K4me1 and H3K4me2). Experimental evidence has indicated that the demethylation of histone sites by LSD1 inhibits FOXA1 binding (Eeckhoute et al., 2009; Lupien et al., 2008; Sérandour et al., 2011), suggesting that H3K4 methylation is necessary and intricately linked to FOXA1 activity.

While FOXA1 is the prototypical pioneer factor and essential to an intact ER signaling program, there is considerable interest in identifying other factors that harbor the ability to modify chromatin and reprogram ER binding preferences. This may be especially important in disease states where ER elicits a gain of function beyond normal physiological roles, an effect that may be attributed to ER reprogramming by pioneer factors. To date, there have been a handful of factors identified through enrichment analysis procedures around ER binding sites and include GATA3, AP-2, TLE1, and PBX1

(Holmes et al., 2012; Joseph et al., 2010; Kong et al., 2011b; Magnani et al., 2011).

Intriguingly, expression of many of these factors highly correlates with ER and are part of the minimal definition of a luminal cell type (Perou et al., 2000), and further exemplifies their roles in a broad regulatory network to support ER function. There is ample evidence that FOXA1, TLE and GATA factors shapes enhancers by binding directly to condensed chromatin but as yet there is minimal evidence to support such roles for PBX1 or AP-2 factors. Nonetheless, AP-2 was found to bind almost 60% of the ER cistrome and siRNA directed against AP-2 was able to abolish ER occupancy at a subset of sites (Tan et al., 2011b). AP-2 and FOXA1 were also found to work cooperatively, both being necessary for complete activation. Altogether there is considerable overlap (Figure 1-5) in binding among the recognized pioneer factors yet specific binding profiles do exist for each. Further studies are necessary to understand the biological outcomes as they relate to the composite binding profiles and what specific subsets contribute to breast cancer progression.

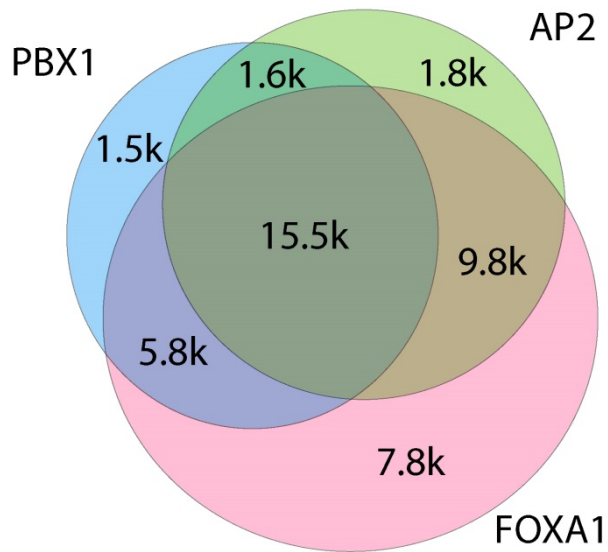


Figure 1-5: Overlap of pioneer factor binding.

Composite profiles of pioneer factor binding sites from ChIP-Seq studies. Figure adapted from Jozwik, 2012. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Cancer. ©(2012)(Jozwik and Carroll, 2012)

1.3 ER targeted therapies and resistance

1.3.1 Pharmacological intervention

The Luminal tumor subtypes account for two-thirds of all tumors and depend on estrogens signaling through the estrogen receptor for growth and survival. These tumors are therefore amenable to therapeutics that target the estrogen receptor signaling axis to reduce rates of recurrence and improve overall survival probability.

The anti-estrogen tamoxifen, first approved in 1977 for metastatic breast cancer, was the first ever targeted therapy to be used in cancer and still remains part of the standard of care of today. Its history is also a remarkable story of drug development and repurposing. ICI pharmaceuticals developed a series of non-steroidal anti-estrogens in

the 1960s to be used as anti-fertility agents. First developed as a post-coital contraceptive, ICI 46,474 (tamoxifen) worked well in rat models but in fact showed pro-fertility traits in humans and development was abandoned. One unique aspect of the drug however was the fact that it exhibited anti-estrogen activity on some tissues but agonistic activity on others all the while having a very low toxicity profile. Arthur Walpole at ICI then enlisted the help of V. Craig Jordan and tasked him with the responsibility of identifying novel applications for the failed drug that could be used clinically (Jordan and DOWSE, 1976). Jordan was able to confirm that tamoxifen did indeed bind the estrogen receptor and could occlude estradiol, effectively acting as anti-estrogen in the breast (Jordan, 1976; Jordan and Koerner, 1975; Jordan and DOWSE, 1976). He further showed that the drug delayed the progression of chemically-induced mammary tumors in rat models and that continued long-term exposure was necessary to control tumor growth. Ultimately tamoxifen became approved for metastatic breast cancer in 1977 and later as an adjuvant treatment option in conjunction with chemotherapy in 1986 (Ingle et al., 1986). Over time the use of tamoxifen became approved for more indications and is now used for all stages of ER+ breast cancer.

It is now appreciated that ligands exert a continuous spectrum of activity ranging from full agonists to antagonists. Furthermore they exhibit cell and tissue specific activities by promoting a unique conformation of ER that allows for differential cofactor interactions. Indeed the pharmacology of tamoxifen is not only tissue-specific but is also

species-specific. In rat models, tamoxifen behaves as an anti-estrogen but in contrast is estrogenic in the same mice tissues. As compounds cannot simply be called estrogens or anti-estrogens in a broad sense, a new nomenclature has emerged and the term selective estrogen receptor modulator (SERM) now describes the pharmacological ligands(McDonnell, 2000). Tamoxifen is classified as a SERM and in the context of the breast displays an antagonist profile being able to both compete with estradiol and block coactivator recruitment, but has agonist profiles in bone, uterus, and the cardiovascular system. Next generation SERMS have been developed and overall have more favorable tissue profiles (Table 1-1). In particular, tamoxifen is an agonist in uterus while bazedoxifene and raloxifene are antagonists and therefore decrease the risk of uterine hyperplasia(Delmas et al., 1997; Smith et al., 1975). While next generation SERMs have less unfavorable effects and better agonist profiles to preserve the beneficial aspects of ER in certain tissues, drug development is always in search of the “perfect SERM” to delay breast growth, prevent uterine and endometrial hyperplasia, ameliorate hot flashes, and improve bone density.

Table 1-1: Tissue agonist profile of common SERMs.

	Breast	Uterus	Bone	Vasculature	Brain
Estradiol	++	++	++	++	++
Tamoxifen	-	+	+	+	-
Raloxifene	-	-	+	+	-
Bazedoxifene	-	-	+	+	-

Another class of ligands exist for clinical use and are known as selective estrogen receptor down-regulators (SERDS). Fulvestrant (ICI 182780) is FDA approved for metastatic breast cancer and is indicated for post-menopausal women following first line anti-estrogen therapy(Howell et al., 2001; Howell et al., 2004; Howell et al., 2002; Osborne et al., 2002). Fulvestrant exhibits no partial-agonist activities, has higher affinity than tamoxifen, and promote ER degradation. However, the downside to 'pure anti-estrogens' like Fulvestrant is that they also counteract the positive effects of estrogens on bone mineral density and other estrogen-dependent tissues. Lastly utility of Fulvestrant in the clinic is hindered by its poor pharmacodynamics.

Rather than intervening at the receptor directly, the ER signaling axis can also be disrupted by reducing the concentration of the endogenous ligand estradiol. The final step in the synthesis of estrogens requires aromatase/CYP19A1 and uses testosterone or androstenedione as a precursor(Simpson et al., 1994). Thus aromatase inhibitors (AIs), such as exemestane, anastrozole, and letrozole are available for clinical use and lead to more favorable outcomes in post-menopausal patients(Chia et al., 2008; Winer et al., 2005). Similar to SERDs, AIs are contraindicated in pre-menopausal patients and lead to a decrease bone mineral density.

1.3.2 Endocrine resistance in the clinic

Although tamoxifen reduces recurrence breast cancer rates by nearly 50%, approximately one-third of patients relapse within 15 years (Early Breast Cancer

Trialists' Collaborative, 2005), a finding that highlights the need for clinical assays to stratify patient response. As tumor classifications have evolved it is now clear that there are sub-classifications that further discriminate response and reflect biological outcomes to all classes of anti-estrogens. It has only recently been appreciated that Luminal B tumors differ considerably with Luminal A in their response to tamoxifen and that certain sub-classifications of each can be teased out with high success rates to a particular therapy (Cheang et al., 2009). In a cohort of 976 patients, the 10-year relapse-free survival rates for tamoxifen treated Luminal A patients was 70% and contrasted with 53% for Luminal B patients (Cheang et al., 2009). Another study was able to identify a subset of Luminal A patients with 95% survival probability (Nielsen et al., 2010).

Building on the premise of distinct expression profiles driving a particular phenotype of interest, several multi-gene tests are now available for clinical use to stratify risk and predict therapeutic value of treatment options. Thus, genomic efforts now provide the opportunity to inform treatment decisions in a molecular subtype-dependent and gene-dependent manner. The 2013 St. Gallens international breast cancer conference touched upon the use of multi-gene signatures to further differentiate the Luminal class of patients and supported their use for both prognostic and predictive purposes (Goldhirsch et al., 2013; Goldhirsch et al., 2011). Of these, the OncotypeDX recurrence assay is FDA approved and the most widely used commercially available prognostic test available for node-negative estrogen receptor positive patients (Paik et

al., 2006). By analyzing the relative expression of 21 genes it aims to assist clinicians in predicting the likelihood of recurrence as well as the benefit of endocrine therapy given alone or in conjunction with systemic chemotherapy.

Other clinically available tests include MammaPrint dx (van't Veer et al., 2002), PAM50(Parker et al., 2009), and Theros Breast Cancer Gene Expression Ratio Assay(Ma et al., 2004). The last is derived from the ratio of HoxB13 to IL17BR (HI) otherwise known as the Two Gene Index, where high expression of HoxB13 relative to low expression of IL17BR is predictive for shortened time to recurrence and tamoxifen therapy failure. Conversely, a low ratio is indicative of positive outcome and therapy success. Compared to Mammaprint, which contains 70 genes, and OncotypeDX with 21 genes, the Two Gene Index is surprisingly accurate in predicting responders to tamoxifen and disease outcome.

1.3.3 Mechanisms of endocrine resistance

Even though tamoxifen has remained a mainstay as a breast cancer therapeutic, resistance remains a significant issue as many patients are innately resistant to the drug and others acquire resistance during the course of treatment. Innate resistance is characterized by tumors that fail to respond to hormone therapy from the onset of treatment and is typically associated with low ER levels or a loss of ER dependency as a dominant growth pathway. Acquired resistance is multifactorial and an overview of the literature reveals mechanisms centered around growth factor receptors and their

signaling cascades that converge on coregulators or directly on ER (Osborne et al., 2003). Importantly, ER remains a viable drug target in acquired resistance. Patients that fail tamoxifen as a first line therapy historically have good response rates to AIs or Fulvestrant (Dodwell et al., 2006; Robertson et al., 2003).

The ER signaling pathway overlaps with receptor tyrosine kinase (RTK) activities. This is due to overexpression or enhanced activation of the RTKs in the cancer setting. Xenograft studies have revealed a significant impact of EGFR members on tamoxifen response. (Arpino et al., 2007; Massarweh et al., 2008). Consistent with the observation of an increased role of growth factors, serine 118 on ER is phosphorylated in response to HER2 activity and this modification clinically associates with tamoxifen resistance (Chen et al., 2013; Yamashita et al., 2008).

Other mechanisms of resistance are focused on coregulators, particularly, SRC3 (AIB1 or NCOA3). Increased levels of this coactivator correspond with decreased antagonism of tamoxifen and also represent a HER2-mediated target of phosphorylation (Osborne et al., 2003). Co-expression of SRC3 and HER2 also predict poor response. In another study, PAX2 was found to compete with SRC3 for binding to the HER2 promoter and expression of PAX2 correlated with good prognosis (Hurtado et al., 2008).

Other clinical associations of resistance have been elucidated for which there is little to no mechanistic evidence currently available. Among these, is the aforementioned role of HOXB13 in predicting tamoxifen failure. First discovered using microarray

expression data from primary biopsies of ER+ patients treated with tamoxifen monotherapy, HOXB13 was found to be differentially expressed between patients who relapsed and those that remained disease-free at the five year follow-up(2004; Ma et al., 2006; Ma et al., 2004). Despite the abundant evidence as an associative biomarker, a causal role for HOXB13 in breast cancer pathogenesis has not yet been established.

Understanding the mechanisms that facilitate resistance to tamoxifen and other ER-directed therapies will aid in uncovering the biological determinants of response and will provide a much needed impetus for rational drug design and in tailoring individual therapies to those that will benefit the most.

2 Identification of the DNA-dependent, ER associated factors in tamoxifen resistance.

2.1 Introduction

The estrogen receptor is a master transcriptional regulator in the breast where it plays key roles in the development and maintenance of normal breast epithelium but is also critical to the growth of luminal breast cancers. Anti-estrogens, such as tamoxifen, are used clinically to impede the actions of ER and delay disease progression. Despite the initial benefits of tamoxifen therapy, nearly one-third of luminal breast cancer tumors eventually become resistant, limiting the therapeutic utility of the drug. While resistance was initially considered to represent a by-pass of ER signaling and reliance on alternate growth pathways, there are now several lines of evidence that ER remains a viable target and can be inhibited by other ER-targeted therapies to delay progression in a large percentage of tumors (Dodwell et al., 2006; Howell et al., 2001; Robertson et al., 2003). In line with this, ER is found to be expressed at high levels in 80% of metastases (Harrell et al., 2006). Collectively, these data indicate that the ER program is central to the growth of these tumors and that adaptive mechanisms feedback and converge with ER to allow for continued activation.

To enable the study of relevant mechanisms of resistance, our lab has generated an in vivo derived tamoxifen resistant (TamR) model by serial passage of an MCF7 xenograft tumor in a nude xenograft model under continuous tamoxifen treatment (Connor et al., 2001). The TamR model is stimulated by tamoxifen in vivo and does not

require estradiol for growth. More importantly, the growth of these tumors is still ablated by the pure anti-estrogen, ICI, and also with the second generation SERM, bazedoxifene, closely reflecting the clinical resistance setting (Wardell et al., 2013). Moreover, ER signaling is very robust and many classical ER target genes are dramatically increased, contrary to in-vitro derived models that typically lose ER expression. Thus, this model provides an experimental model that can be manipulated and evaluated in vitro to explore mechanisms of resistance in breast cancer.

2.1.1 Genomic evaluation of ER

Consistent with the idea of an ER-centered mechanism of resistance, ChIP-Seq of patient tumor samples revealed that progression and resistance is accompanied by a redistribution of ER binding sites (Ross-Innes et al., 2012a). Furthermore, motifs enriched around these ER binding sites associated with resistance suggest a role for pioneer factors in mediating the reprogramming. Specifically, GATA3 sites were indicated to be relatively lost in resistance and motifs rich in PAX2, AP-1 and FOXA1 were abundant in the tamoxifen-resistant set of ER binding events. These results are intriguing because it suggests that adaptive mechanisms may work at the level of chromatin accessibility and also of factors that permit chromatin remodeling for ER binding, an idea largely unstudied in drug resistance and particularly for tamoxifen resistance.

A major challenge in understanding the complexities of the ER signaling program is in defining receptor occupancy and the coordinated relationship with chromatin architecture and collaborating transcription factors. In order to fully elucidate the regulatory factors that permit ER to bind to novel regions on chromatin in the setting of drug resistance, candidate factors would each have to be subjected to Chromatin IP followed by deep sequencing (ChIP-Seq) which limits throughput and confines the experiment to factors known a-priori. Alternatively, regulatory elements and associated motifs can be identified in an unbiased manner using DNase-I hypersensitivity analysis (Galas and Schmitz, 1978). The enzyme, DNase-I takes advantage of chromatin dynamics and preferentially cleaves open chromatin at labile sites to map active regulatory elements. Open chromatin regions are referred to as DNase-I hypersensitive sites (DHSs) and are traditionally viewed as sites with nucleosome depletion due to transcription factor binding (Gross and Garrard, 1988).

By coupling DNase-I hypersensitivity analysis with modern technologies such as microarrays (DNase-chip) or high-throughput sequencing (DNase-Seq), genome-wide views of chromatin accessibility can be obtained (Boyle et al., 2008a; Crawford et al., 2006). Consequently, active regulatory elements including promoters, enhancers, silencers, insulators and locus control regions can be identified globally (Boyle et al., 2011; Cockerill, 2011; Heintzman et al., 2007). As part of the ENCODE project, DNase-Seq profiles of 125 human cell lines were characterized and integrated with mRNA

expression, DNA methylation, chromatin conformation and transcription factor binding sites (Consortium, 2012; Thurman et al., 2012). Collectively, these data reveal that DHSs are often distal to associated promoters, correlate strongly with transcription factor binding, are embedded in histone modified regions and cluster with similar cell-types. Additional studies have elucidated cell-specific DHS profiles and that distal interactions predominately control cell identity, while DHSs ubiquitous among cell types are located proximally to transcription start sites (Gertz et al., 2013; Sheffield et al., 2013; Song et al., 2011b). The binding of a transcription factor to DNA protects local chromatin from DNase-I cleavage and leaves a border of heightened cleavage on the periphery to produce a 'footprint'. Digital genomic footprinting via DNase-Seq leverages this concept to predict occupancy and factor identification through associated motifs in a single experiment at single nucleotide resolution (Hesselberth et al., 2009; Neph et al., 2012).

The methodology of DNase-Seq has been used previously to delineate the actions of nuclear receptors ER and AR with respect to their cognate ligands, and showed strong correlation to occupancy as determined by ChIP-Seq (He et al., 2012; Tewari et al., 2012). These studies also revealed critical differences in the binding properties of these two receptors. ER predominately binds to regions of partially open chromatin, induces further opening with ligand activation and does not promote nucleosome remodeling to a large extent. Furthermore, DHS sites alone are good predictors of ER binding and the precision of identifying a true ER binding site is

dramatically increased with the intersection of DHSs and motifs. In comparison, AR binds to both “poised” and closed chromatin regions with a concordant change in nucleosome positioning. In the course of the present study, a report was published that compared the ER chromatin landscape and ER binding sites across two distinct, estrogen-sensitive cell lines of breast and endometrial origin (Gertz et al., 2013). Overall, the study concluded that strong affinity motifs for ER are generally conserved binding regions and that ER is capable of remodeling the chromatin at these regions to a larger extent than locations with weak affinity motifs. Moreover, the cell-specific ER binding sites exhibited weak motifs but were likely available for binding due to the presence of juxtaposed cell-specific factors that bookmarked chromatin to maintain an active open state.

Therefore, the global approach of DNase-Seq is well-suited to identify ER collaborating and associated DNA-bound factors that inform novel ER binding profiles. While precedence of ER actions across cell lines and response to ligand has been set, the molecular mechanisms that underlie tamoxifen resistance are scarcely studied on the level of chromatin and demand attention. Elucidation of these features will improve patient outcomes by further stratifying risk, will enable clinicians to make better-informed decisions in regards to therapy, and hold the potential to inform novel drug development programs. To this end, we undertook the present study to investigate the consequences of chromatin architecture and transcription factors in tamoxifen resistance

and with the intent of identifying and characterizing factors as they relate to ER function.

2.2 Results

2.2.1 DNase-Seq identifies an increased role for FOXA family members in a cellular model of tamoxifen resistance.

2.2.1.1 DNase-Seq on sensitive and resistant cell lines

To assess the relationship between accessible chromatin and the ER signaling program we performed DNase-Seq on a tamoxifen resistant cell line (TamR) and a comparator group of two closely related tamoxifen sensitive cell lines: MCF7 and a subline of MCF7 that has been adapted for xenograft growth conditions named MARCO. Two independent biological replicates of MCF7 cells (sensitive), MARCO cells (sensitive) and TamR cells (resistant) were plated in charcoal stripped serum for 48 hours and then treated with vehicle or 4-hydroxytamoxifen (4-OHT) at 100nM for 24 hours. Using previously published methodologies from the Crawford Lab (Boyle et al., 2008b), libraries were prepared and submitted for 50bp SR Illumina sequencing on the HiSeq platform at Duke resulting in an average of 70 million unique mapped reads per sample (Table 3-1). Additional details of data processing and filtering are available in the materials and methods.

Table 2-1: Overview of sequencing libraries and reads.

Cell Line	Treatment	Replicate	Unique Mapped Reads
MCF7	Vehicle	1	67,199,476
MCF7	Vehicle	2	68,986,891
MARCO	Vehicle	1	85,401,169
MARCO	Vehicle	2	76,555,399
TAMR	Vehicle	1	75,331,103
TAMR	Vehicle	2	63,442,543
MCF7	4-OHT	1	60,291,199
MCF7	4-OHT	2	70,235,124
MARCO	4-OHT	1	74,614,947
MARCO	4-OHT	2	76,358,470
TAMR	4-OHT	1	62,928,784
TAMR	4-OHT	2	69,496,954

DNase-I hypersensitive sites (DHSs) were defined using the peak caller F-SEQ which uses a kernel density estimation procedure appropriately suited for the background distribution of DNase-Seq data(Boyle et al., 2008b). Taking the union set (full outer join) of DHSs from every condition, 508,533 sites in total were defined representing 4.3% base coverage of the genome. Peaks falling within +/- 500bp of the TSS were also separated to distinguish between general transcriptional effects and bona fide TF binding. The number of sequence reads that mapped to each feature of the union set of DHSs was quantified and differential testing was performed using DESeq(Anders and Huber, 2010) in R/bioconductor with a significance threshold of $p < 0.01$.

The number of changes due to (4-OHT)-(Veh) in any of the cell lines was quite modest--typically less than 50 DHSs being statistically different. It has been previously

reported that tamoxifen induces DNA binding in a similar manner to E2 but at much lesser a degree in a short time frame and only in a subset of regions, and so it was not surprising that the number of changes in MCF7 cells were insignificant at 24 hours. On the other hand, there were dramatic differences between any two cell lines in the untreated state. This was intriguing because these cell lines are of the same genetic origin and the differences likely represent the divergent phenotypes they display in response to 4-OHT. Comparing the sensitive group to resistant TamR cells under vehicle treatment, 7,399 DHSs were identified as significantly different. Among these, 82% or 6,016 DHSs have an increased signal in TamR cells and 18% or 1,383 DHSs are decreased (Figure 2-1).

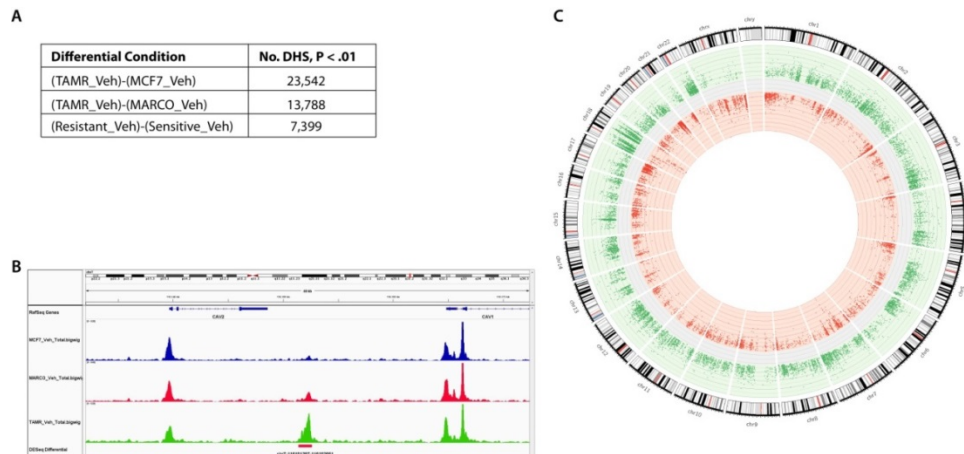


Figure 2-1: Differential DHS sites between cell lines.

Differential DHS sites between cell lines. (A) Table of the number of sites identified as significantly differential by DESeq with $p < 0.01$ for each testing condition. (B) Track view of CAV1 locus and continuous DNase-Seq signal in MCF7 (Blue), MARCO (Red) and TamR (Green). (C) Circos map of the whole-genome distribution of significant DHS sites from TamR relative to MCF7. A normal karyotype ideogram is shown on the outside. Inner layer is a scatter plot of the \log_2 fold change from DESeq. Green=up in TamR, Red=down in TamR for each DHS.

While each tested differential condition likely epitomizes the biological features unique to each cell type along with the resistant mechanisms, the multi-factorial condition of (Resistant-Sensitive) was chosen for downstream analyses. This condition represents the compilation of sites that are both necessary and sufficient for the resistant phenotype and is a more conservative approach than comparison against either of the two sensitive cell lines independently. The 7,399 sites of (Resistant-Sensitive) and relative differences in DNase signal across cell lines is visualized in Figure 2-2.

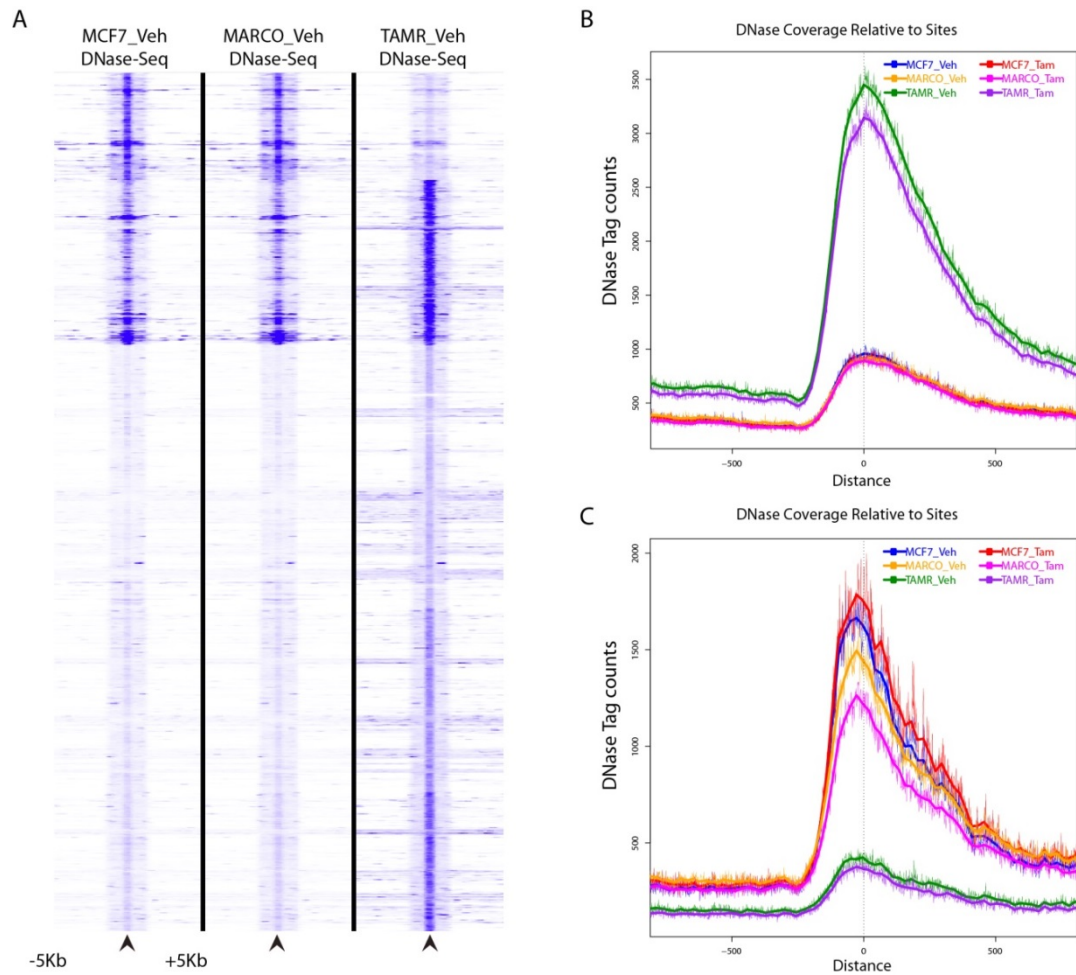


Figure 2-2: Plots of significant DHS sites.

Plots of significant DHS sites. (A) Heatmap of DNase signal in a 10kb window of 7,399 significant different DHSs between sensitive and resistant cells. Each row is centered on an individual DHS site and data is clustered using k-means. (B) Histogram of the 6,016 DHS sites that are more open in TamR cells. (C) Histogram of the 1,383 DHS sites that are more closed in TamR cells.

We also compared the distribution of DHSs relative to genomic features (3' UTR, 5' UTR, CpG-Islands, Distal, Exon, Intron, Promoter, Satellite, and TTS) as both an absolute number and additionally by the relative enrichment when normalized by DNA content. Even though intron and distal sites are more represented overall, they are not

nearly enriched compared to CpG-Islands or promoters, which relatively contain more DHSs when size of the feature is taken into account. Compared to all DHSs, the 7,399 significant set of significantly differential DHSs was similar in trend with minor differences. The differential DHS sites were even more enriched in intron and distal features, indicative of transcriptional enhancers (Figure 2-3).

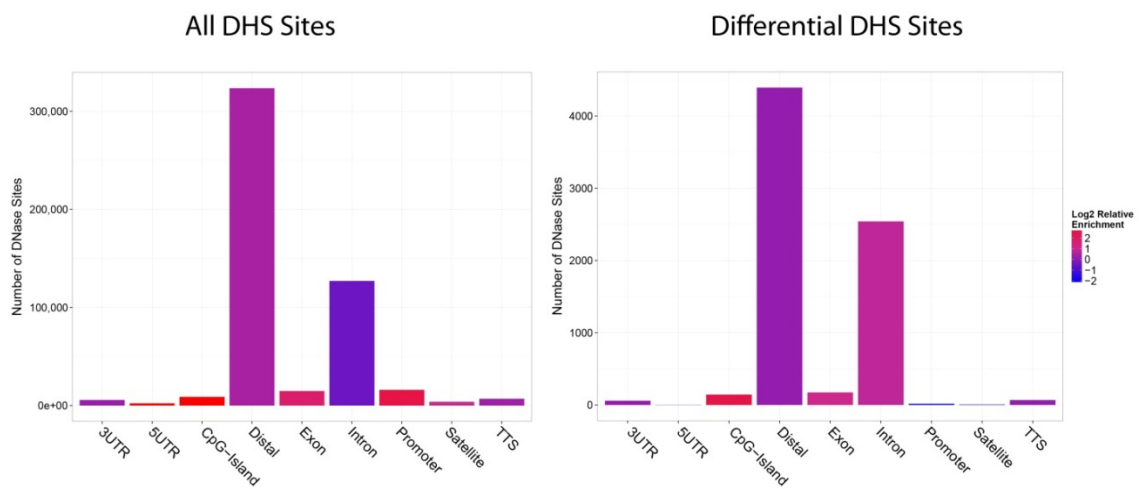


Figure 2-3: Distribution of DHS sites.

Plots representing the distribution of DHS sites in each genomic feature for (left) all DHS sites from the union set and (right) differential DHS sites from (Resistant)-(Sensitive). Bar colors represent the relative enrichment of features normalized to DNA content.

2.2.1.2 Motif Enrichment Analysis

A major objective of these studies was to identify the DNA-bound factors with differential binding profiles in resistance. Therefore, differential DHSs for (Resistant)-(Sensitive) were interrogated for motif enrichment analysis with a two pronged approach to minimize bias. It's important to note that of the ~2000 identified transcription factors, only a few hundred have high quality motifs mapped and de novo

motif finding is essential to identifying the factors involved (Vaquerizas et al., 2009). Motifs were identified using a local installation of MEME-ChIP (Machanick and Bailey, 2011) with a “zoops” model for motif finding with a focus on the number of hits relative to background, high percentage of hits, and central distribution of the motif within the DHS sites. Parallel to this approach, motif finding for factors with known motifs and de-novo motif finding was also performed using Homer (Heinz et al., 2010) (Table 3-2). Several factors were coincident among all motif finding methods and best matched known motifs for AP-1, AP-2, FOXA, ELF5, ER, and TEAD families. Discordant motifs were also found and most closely resembled IRX5, GRHL2, KLF4, MED-1, SPDEF and ETS family members.

Table 2-2: Table of high confidence de-novo motif finding results from Homer for sites increased in DHS signal.

Rank	P-value	% of Targets	% of Background	Best Match/Details
1	1e-464	18.53%	3.37%	Jun-AP1(bZIP)/K562-cJun-ChIP-Seq/Homer
2	1.00E-246	24.19%	9.39%	AP2gamma(AP2)/MCF7-TFAP2c-ChIP-Seq/Homer
3	1.00E-186	19.63%	7.81%	FOXA1(Forkhead)/MCF7-FOXA1-ChIP-Seq/Homer
4	1.00E-158	21.78%	9.99%	PH0086.1_Irx5/Jaspar
5	1.00E-143	23.15%	11.42%	ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer
6	1.00E-95	21.73%	12.14%	TEAD4(TEA)/Tropoblast-Tead4-ChIP-Seq(GSE37350)/Homer

Motifs were also elucidated for sites that decrease in DNase signal in TamR cells. The number of sites used for querying was lower in this group but motif enrichment analysis still yielded a strong hit for CTCF and weakly associated motif for AP-1 and AP-2.

Motifs matching all criteria for inclusion as outlined above, in both increased and decreased sites, are represented in Figure 2-4.

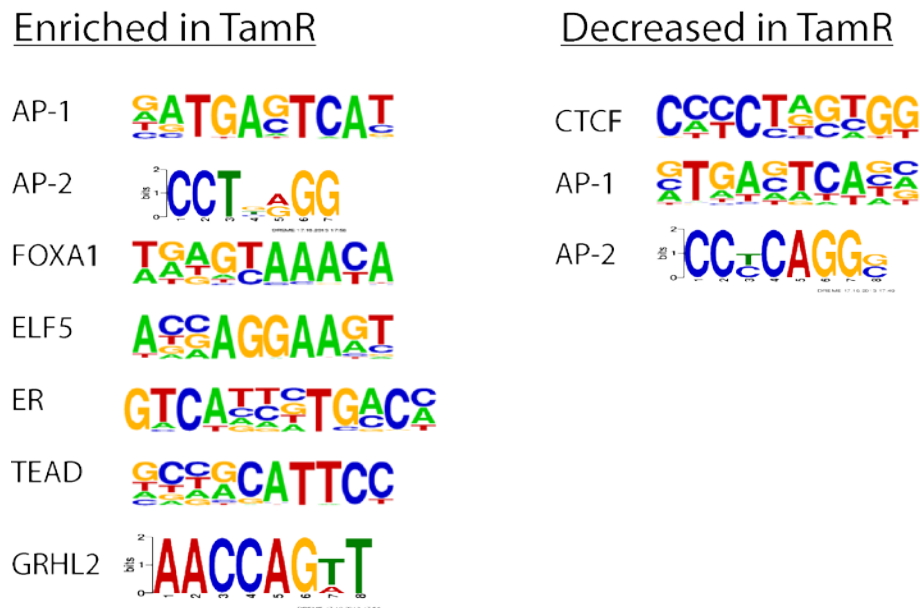


Figure 2-4: Identified motifs in differential DHS sites.

2.2.1.3 Comparison with mapped binding sites.

One important limitation of relying on motifs for identification is that DNA sequence is not the sole determinant of transcription factor occupancy. Even though motif enrichment procedures have merit, direct evidence of factor binding is often more reliable. Therefore, to supplement our motif analysis, we additionally downloaded and

processed publicly available ChIP-Seq and ChIP-exo data for breast cancer cell lines (GSE25710, GSE48930, GSE40129) and ENCODE factors (GSE32465, GSE31477), to identify co-enriched binding profiles. Binding coordinates for each factor were analyzed to identify overlapping profiles and correlation of co-localization for every factor pair, generating a large correlation matrix of Spearman's coefficients. This data was then clustered and visualized as a heatmap (Figure 2-5). The regions of increased DNase signal (DNase_Up) clustered most strongly with FOXA1 regions and secondarily with ER, closely mirroring the outcomes from the motif analysis. Regions of decreased DNase signal clustered with CTCF and RAD21, again coinciding with the motif finding results. One shortcoming to these analyses is that not all factors identified by motif analysis have global binding data and thus cannot be evaluated by this approach. Another caution is that multiple cell lines are represented and may not be reflective of the de facto binding sites of cells used in our studies.

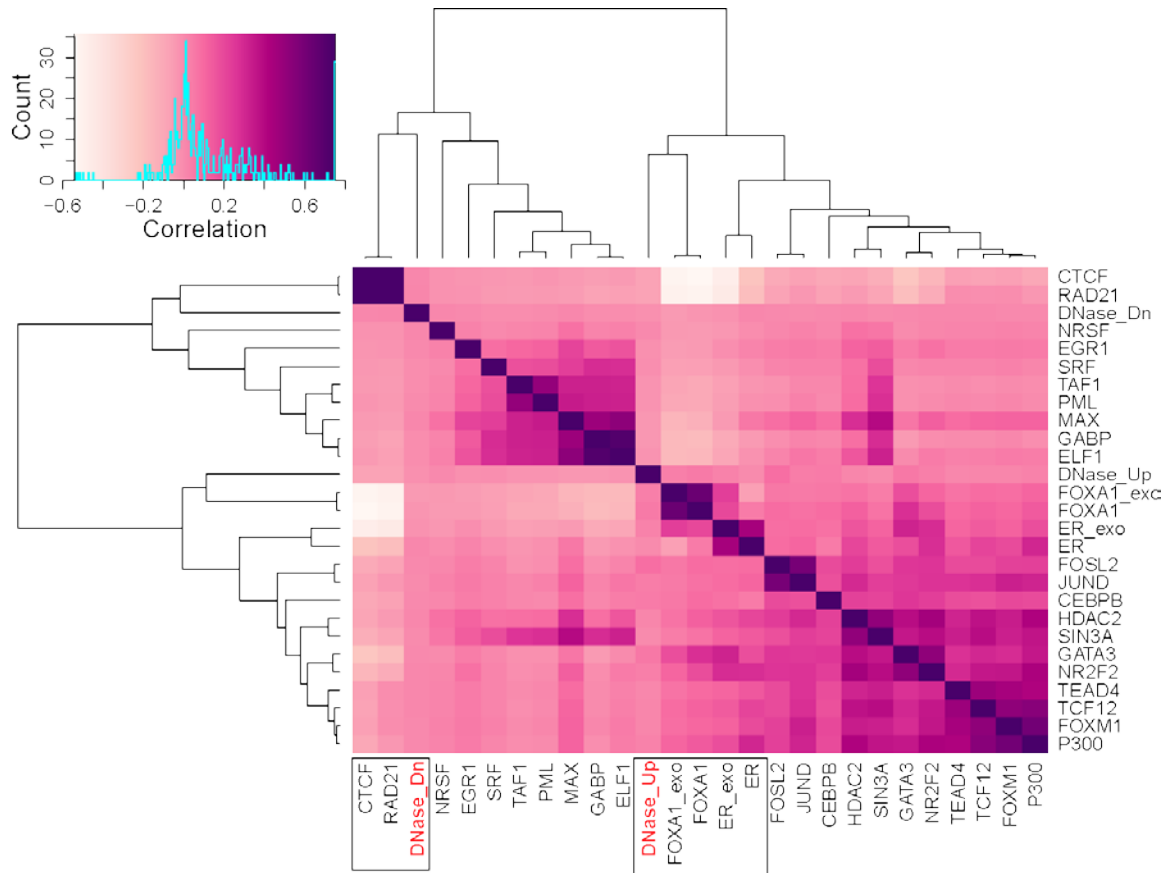


Figure 2-5: Heatmap of binding profiles.

Heatmap representation of the clustering of different transcription factors binding coordinates and their correlation to sites with increased signal (DNase_Up) and decreased signal (DNase_Dn).

Given the preponderance of evidence that FOXA1 and other pioneer factors cooperate with ER and co-localize on DNA, we also performed differential binding analyses to identify regions where factors uniquely exist (factor-unique sites) and where they overlap with other factors (factor-shared) to better elucidate each factor's contribution. As a binary example, FOXA1 has 70,915 known binding sites while ER has 21,700. These two factors coincide at 9,580 sites, leaving 61,335 FOXA1-unique sites and

12,120 ER-unique sites. Each factor-unique subset of binding sites can then be compared with the DNase data for evidence of factor enrichment in TamR cells (Figure 2-6). These data substantiate a role for FOXA1 on both ER-independent sites and ER-dependent sites and bear further evidence that known ER-unique sites (devoid of FOXA1) are unaffected in TamR cells.

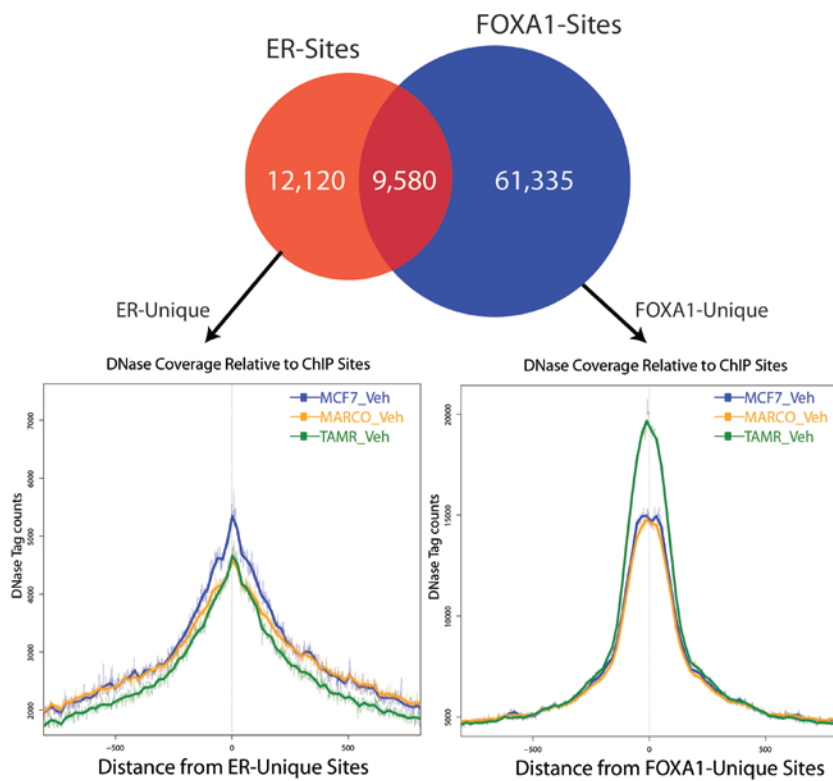


Figure 2-6: Differential binding profiles for ER and FOXA1.

Schematic representing factor-unique profiles for both ER and FOXA1. Venn diagram of ChIP-Seq binding sites with overlap (above). Histograms for each factor-unique subset (below).

As there is considerable overlap with ER, FOXA1 and GATA3, unique binding profiles of each can also be derived in this context. Without such subsets, GATA3 and other factors show apparent enrichment with DNase signal, but when FOXA1

overlapping sites are removed, the apparent increase is mitigated. In this manner, several interesting factors were profiled: GATA3, AP-2gamma, CTCF, TCFL2, and PBX1 (data not shown). These results largely corroborated the motif enrichment results and highlighted both AP-2gamma and CTCF as factors that are lost in the intermediate, yet sensitive MARCO lineage which may represent a necessary but insufficient step in tamoxifen resistance.

2.2.1.4 Footprint analysis of top factors

Evidence of a well-defined footprint for a specific factor in differential DHS sites is another method to verify motif results and prioritize the number of possible candidates. In comparison to correlation of ChIP-Seq data, footprint analysis can additionally account for de novo binding sites which are not present in MCF7 cells. For this purpose, genome-wide motifs were identified with a strict log odds ratio and only sites that overlapped with differential DHS regions were used for analysis.

Among the 6,016 differential DHS sites, 4,295 high confidence motifs for FOXA1 were identified. Sequencing tags were quantified relative to motif start site in a strand specific manner and plotted as a histogram at single base pair resolution (Figure 2-7). TamR cells are highly enriched in DNase signal in these regions, similar to Figure 2-2B, and more importantly exhibit a strong footprint central to the distribution of the data. Moreover, the characteristic footprint observed in the differential DHS sites is consistent with the footprint obtained from the union set of DHS sites.

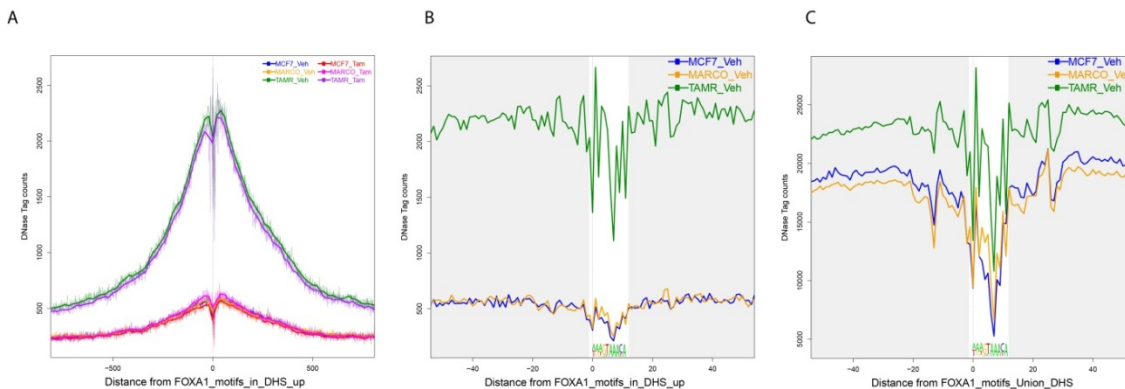


Figure 2-7: Footprint analysis of FOXA1.

Aggregate profiles of DNase signal relative to FOXA1 motifs. (A) Histogram of increased differential DHS sites across a 1.5kb window. (B) Same data as in A, but view is focused on 50bp each direction of FOXA1 motif. (C) FOXA1 footprint in the union set of DHS sites.

The number of confident ER motifs in the increased DHS sites was much smaller due to motif degeneracy with 635 sites, yet an enrichment of signal in TamR cells was observed in addition to a palindromic footprint. Intriguingly, there was also a noticeable difference of base pair dependency in the central degenerate linker region of the two half sites, suggesting additional contacts of stabilization specific to ER in TamR cells.

In this manner, top motifs were profiled by footprint analysis. The association of AP-2 to differential sites was also noteworthy with strong relative enrichment and associated DNase protection at the central motif region but was not as robust as FOXA1.

2.2.1.5 Correlation with resistant-specific ER binding sites.

In order to evaluate the relevance of the TamR model in resistance we relied on a previously published study that found differential binding of ER in patients exhibiting resistance to tamoxifen (Ross-Innes et al., 2012b). The study specifically identified 448

sites unique in resistance and the DNase data from our study showed similar trends of increased DNase signal at these resistant-specific regions (Figure 2-8).

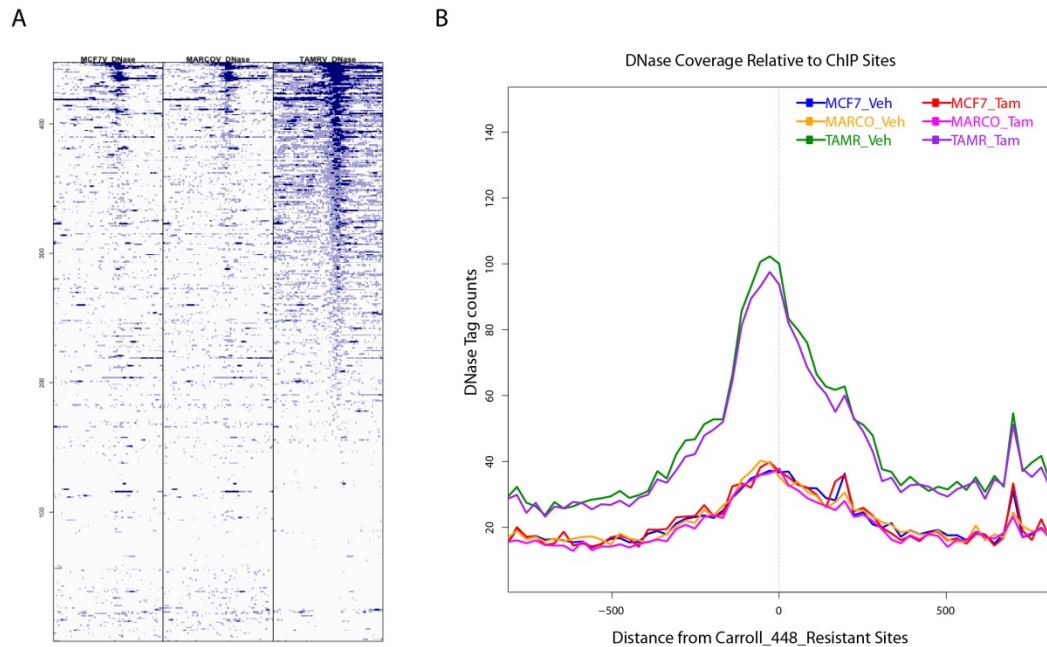


Figure 2-8: Comparison of DNase data with 448 sites found in tamoxifen resistance.

The 448 of ER binding sites unique in tamoxifen resistance were profiled for their relationship with the DNase data in MCF7, MARCO and TamR cellular models. (A) Heatmap of DNase-Seq data at 448 binding sites where TamR DNase intensities are ranked from high to low. (B) Aggregated histogram of the same data.

2.2.2 Increased occupancy of FOXA1 and ER in TamR cells.

FOXA1 emerged from the motif enrichment, footprint analysis and ChIP-Seq comparative analyses as a likely candidate to occupy regions within differential DHS sites. To validate a biological role for FOXA1 and associated binding of ER to these sites, ChIP-qPCR was performed at 15 candidate loci. Primer sequences are available in

materials and methods. Four of these sites are presented in Figure 2-9. Two of these sites are ER-independent, AGR2 enhancer 2 (AGR2-2) and WISP1 enhancer 2 (WISP1-2), while S100A9 enhancer 2 (S100A9-2) and XBP1 enhancer 3 (XBP1-3) are both ER-dependent and FOXA1-dependent. These data clearly highlight an enhanced recruitment of FOXA1 to these sites and also enhanced binding of ER at ER-dependent sites. The level of ER occupancy at these two loci also reveals that ER binding is maximal in TamR cells under basal conditions and there is little enrichment with E2, contrasting the role of ER in MCF7 cells.

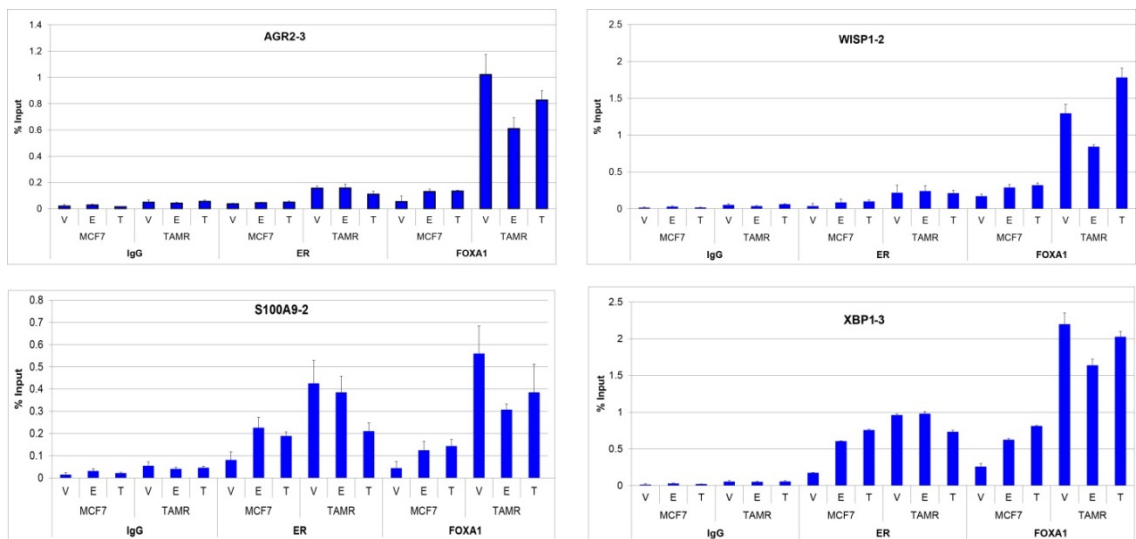


Figure 2-9: ChIP-qPCR of ER and FOXA1.

Occupancy of ER and FOXA1 was analyzed in MCF7 and TamR cells treated with either vehicle(V), 10nM estradiol(E) or 100nM 4-OHT(T) for 45 minutes. Cells were cross-linked and immunoprecipitated with either control IgG, ER or FOXA1 antibodies. All data is represented as the percent input from total chromatin.

2.2.3 Characterization of FOXA1 and ER function in TamR cells

To substantiate the role of increased FOXA1 signaling and its effects on ER, robust activated targets of FOXA1 were assessed by qPCR. In these experiments we used siRNA directed against FOXA1 in combination with ER ligands to understand the role of each factor on these genes. These approaches are far from comprehensive and are not meant to elucidate every gene and its regulation by these factors, but rather to verify a role for FOXA1 on ER targets. The qPCR data is presented as a heatmap in Figure 2-10. These data indicate that of the FOXA1 targets profiled, expression is relatively increased in TamR cells and remains sensitive to siFOXA1 ablation. Furthermore, these genes are comparatively insensitive to E2 in TamR cells but using the pure anti-estrogen ICI, gene expression is inhibited on some genes suggesting that certain subsets are more dominated by FOXA1 while others require both FOXA1 and ER for complete activation.

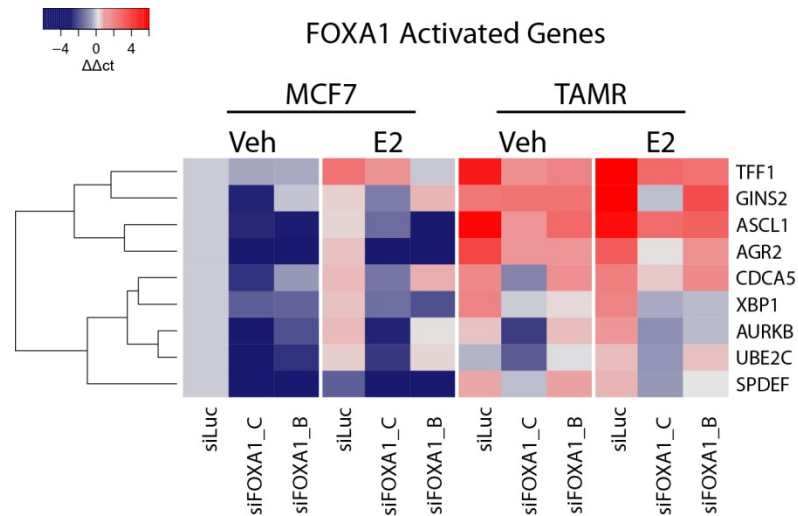


Figure 2-10: Heatmap of FOXA1 activated targets.

Heatmap of mRNA expression of FOXA1 target genes as determined by qPCR. Cells were reverse-transfected with control siRNA (siLuc), and two siFOX1 (siFOX1_C, siFOX1_B) for 72 hours and then induced with E2 for 24 hours prior to harvesting. Data points are represented as delta delta Ct values from the mean of triplicate experiments.

TamR cells also reflect increased FOXA1 activity on FOXA1 repressed genes:

ANXA1, KLK6, WISP2, and S100A9. A large percentage of the FOXA1 repressed genes are also induced by ER in MCF7 cells, and the regulation of these genes are divergent in TamR cells. In MCF7 cells, both mechanisms are intact and activation by ER requires FOXA1 while FOXA1-mediated repression is independent of ER. This tug of war phenomenon results in a class of genes that are driven by ER and upregulated, and a second class of genes that are further repressed by FOXA1 and consequently downregulated in TamR cells (Figure 2-11). As an example, KRT13 can be induced by ER while siFOX1 de-represses gene expression, meaning both mechanisms are intact in MCF7 cells. However, TamR cells have lost FOXA1-mediated repression of this gene

and its regulation is instead dominated by ER leading to a large increase in gene expression (Figure 2-11). In the other class, WISP2 is both ER-activated and FOXA1-repressed in MCF7 cells but is dominated by the FOXA1 repression mechanism in TamR cells.

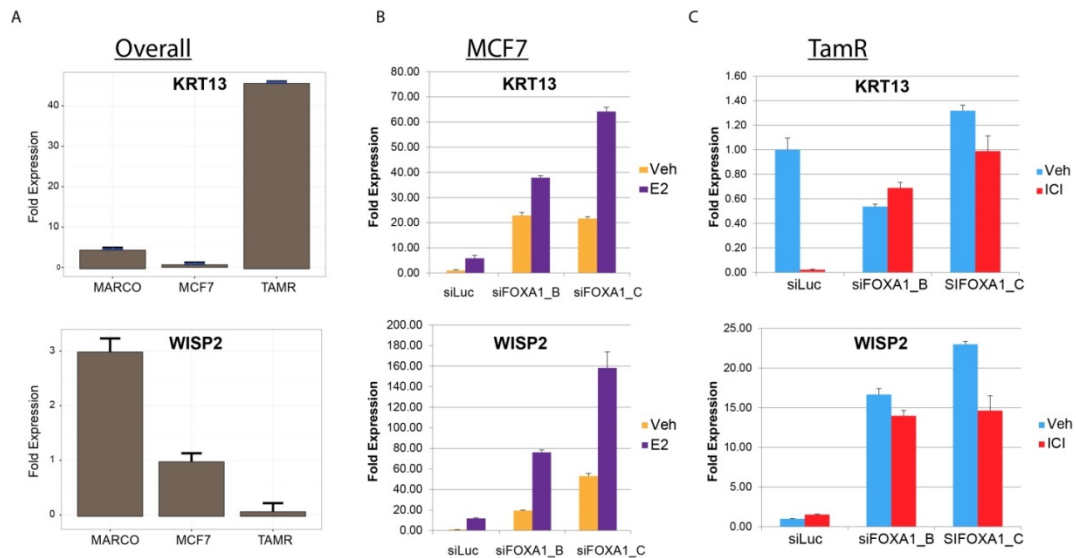


Figure 2-11: FOXA1 repressed genes.

Two distinct classes of FOXA1 repressed genes are present in TamR cells. (A) Overall gene expression across cell types for both KRT13 and WISP2. (B) Expression in MCF7 cells with control siLUC or two siRNAs of FOXA1. (C) Expression in TamR cells with control siLUC or two siRNAs of FOXA1.

2.2.4 Identification of FOXA1 interacting partners

It was observed that the overall protein levels of FOXA1 was not significantly different between MCF7 and TAMR cells, and therefore we hypothesized that the activity of FOXA1 was being modified by either an epigenetic mechanism or by other factors modifying and/or binding to FOXA1. To address the latter hypothesis, we performed a set of experiment to identify PTMs and interacting partners of FOXA1 by

using Mass Spectrometry. To this end it was necessary to chemically crosslink the antibody to agarose beads with DMP permitting co-immunoprecipitation (CoIP) experiments without eluting large contaminating IgG fragments. These experiments were performed in MCF7 and TamR cells in normal FBS conditions. Nuclear extracts were isolated and normalized to 10mg of protein for each and immunoprecipitated with control IgG and FOXA1 antibodies cross-linked to A/G beads. Washing conditions were optimized to efficiently remove non-specific proteins while maintaining adequate FOXA1 levels. Eluted fractions were submitted to the Duke Proteomics Core and LC-MS/MS was used to identify putative FOXA1 interacting partners and separately to identify PTMs of FOXA1. These runs resulted in identification of a total of 229 high-confidence proteins with 1% FDR, of which 23 of these are putative binding partners of FOXA1 in other cell types. Interestingly, two of the factors identified by CoIP were also identified by motif analysis above, namely AP-2 and GRHL2 (Figure 2-12). HDAC1 and SIN3A were also specifically found in the TamR subset and may indicate that corepressor complexes have an increased role in the FOXA1 signaling paradigm in TamR cells. A network plot and table of interactors is presented later in this chapter. Approximately 40% sequence coverage of FOXA1 was achieved, which is as much as could be expected with trypsin digestion. Of the modifications identified, only a phosphorylation site at S331 was found to be differential, with it preferentially being lost

in TamR cells. Studies are currently underway with GRHL2 to understand if and how it participates in FOXA1 signaling.

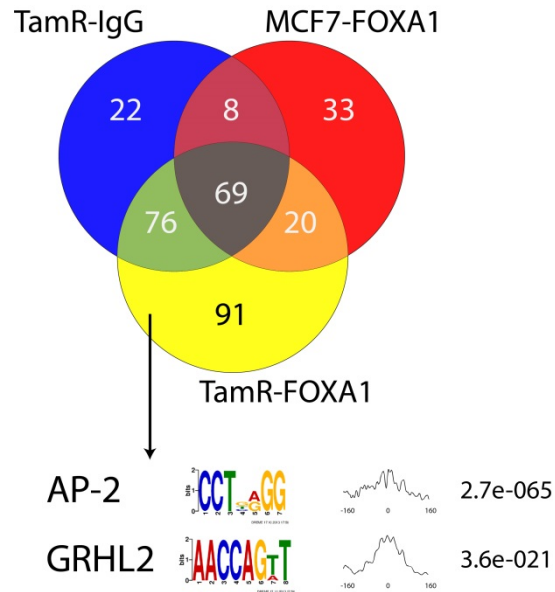


Figure 2-12: Mass-Spec identifies putative interacting partners.

Of the 229 identified proteins, 91 proteins were found specific to FOXA1 in TamR cells. Among these two were also identified from motif enrichment analysis in 1.2.1.2. Motif results (below) are the outputs from the MEME Suite and include identified motif, distribution in differential DHS sites and significance score.

In the course of doing the CoIP experiments it was observed that FOXA1 partitions between the nuclear and cytosolic compartments and shuttles depending on the growth conditions of the cells. This partitioning is enhanced in FBS and FOXA1 levels are more enriched in the nuclear compartment in TamR cells. In addition to partitioning, the size of FOXA1 is slightly shifted upward on the gel in the nuclear extract samples, possibly reflecting a PTM induced mobility shift or the size difference could also be due to cleavage of the protein product in the cytosolic extracts. Mass-spec

analysis of the protein only had 40% coverage and so we could not confidently map all regions to identify specific modifications. Therefore, to determine if phosphorylation of FOXA1 was responsible for the upshift, we incubated the prepared samples in lambda phosphatase and while the control samples had mobility reversed, the FOXA1 samples were unchanged indicating the shift was not due to phosphorylation. FOXA1 pulldowns were also performed and immunoblotted with phosphorylation and acetylation antibodies for presence of these modification with no success. At present there is no evidence to explain these results.

2.2.5 Integrative analysis with mRNA expression

Genome-wide expression was also performed on the MCF7, MARCO, and TamR cells with vehicle or 4-OHT for 24 hours with RNA-Seq. These treatment conditions were identical to the DNase-Seq data but were performed at different times and in independent cell passages. Reads from 50bp paired-end Illumina HiSeq were aligned to the human reference genome hg19. Concordant read pairs were then quantified over Ensembl transcripts and differential expressed transcripts were identified using edgeR(Robinson et al., 2010). Differential transcripts were identified for (4-OHT)-(veh) in every cell type and a general linear model was defined to identify the multifactorial condition of interaction of response (sensitive and resistant) and treatment (vehicle or 4-OHT). The differential condition, (Resistant_veh)-(Sensitive_veh), is used for

downstream integrative analysis and relates to the DNase-Seq data (Resistant-Sensitive) outlined above.

Pathway analysis on the differential transcripts was performed with GSEA and resulting enriched gene-sets were organized into a topological similarity network (Figure 2-13).

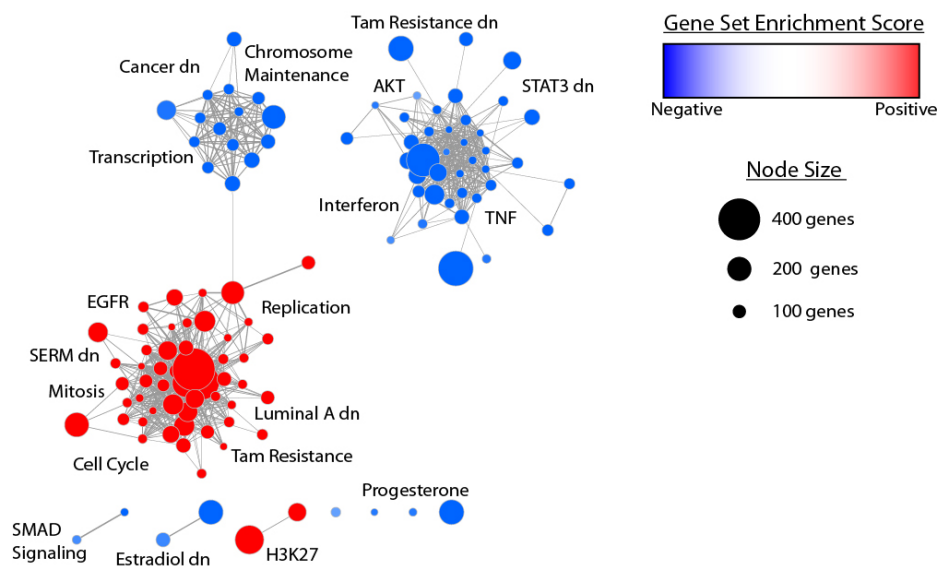


Figure 2-13: GSEA enrichment map.

A similarity network of GSEA results, where nodes represent gene-sets, and edge length represents similarity of the genes between the gene-sets. Size of the nodes indicates the number of genes within a gene-set and color of the node reflects the normalized enrichment score. Red is positively correlated with resistance and blue is negative correlated. Associated terms were derived from the Cytoscape plugin, WordCloud.

One way of integrating expression data into the DNase-Seq data is to identify all differentially expressed transcription factors and presence, if any, in the identified motif enrichment analysis. Expression levels of a factor may be the simplest explanation as to how a factor has an increased role in TamR cells. For this purpose, the gene ontology

term, 0003700: sequence-specific DNA binding transcription factor activity, was used to subset the available significantly differentially expressed genes to a total of 217 TFs.

Notable, robustly upregulated TFs included : GATA4, FOXA2, NKX2.2, ELF5, TFAP2B (AP-2beta), NOTCH1, and STAT6. Downregulated TFs included: SMAD4, PGR, SMAD2, ETV5, STAT1, and HOXA10.

Among these, FOXA2 is of primary interest because it is highly homologous to FOXA1, binds to a similar DNA consensus sequence and interacts with FOXA1 as a heterodimer (Bochkis et al., 2012). Genetic studies also indicate partial redundant roles in liver, foregut and pancreas during development (Gao et al., 2008). In mice knockout studies, FOXA1 or FOXA2 deletions had no obvious effect on liver morphology, while double knockouts led to complete loss of liver formation (Lee et al., 2005). In our system, siRNA against FOXA2 failed to elucidate a role for it in the before mentioned FOXA1 target genes or on biological assays measuring viability or growth in soft agar (data not shown). Nonetheless, we have not yet ruled out a role for it in establishing enhancer competency of the de-novo FOXA1 sites.

Other interesting candidates include ELF5 and AP-2. Curiously, two AP-2 motifs were identified with slightly differing binding preferences and one was associated with a loss of DNase and the other with a gain of DNase in DHSs in TamR cells. The CHIP-Seq comparison of AP-2gamma also revealed loss of DNase signal at these sites. In light of the fact that AP-2beta is highly upregulated the possibility exists that isoform bias

may account for some of the upregulated DHS sites and requires further attention. ELf5 was also recently reported to suppress the actions of ER in T-47D cells (Kalyuga et al., 2012).

Histone modifiers are also of utmost importance given their correlation with FOXA1 activity (Lupien et al., 2008). In this case H3K4me1/2 positively correlates with FOXA1 binding and activity. Histone modifying genes were subset using the GO term, 0016570: histone modification, yielding 48 differentially expressed genes. These subsets of genes could not explain increase in HAT activity or a loss of a histone demethylase. The increase in FOXA1 activity could also be explained through a mechanism of coregulator activity by potentially shifting the balance towards activation. Notable repressors downregulated include TLE1, RIP140, ID1, ID2, and ID3. The number of coactivators is quite large and without knowing exactly which bind to FOXA1 through further proteomic techniques, it is difficult to ascertain a model of resistance at this point.

Lastly, the Mass-Spec revealed binding partners of FOXA1 that are MCF7-specific, TamR-specific or shared. A network plot of the putative factors was generated to visualize the data. Cytoscape was used to color the nodes according to differential expression of (Resistance)-(Sensitive) and the outer stroke color of the node to represent average expression. When plotted with expression it is apparent that increased differential expression is not a major determinant of association (Figure 2-14).

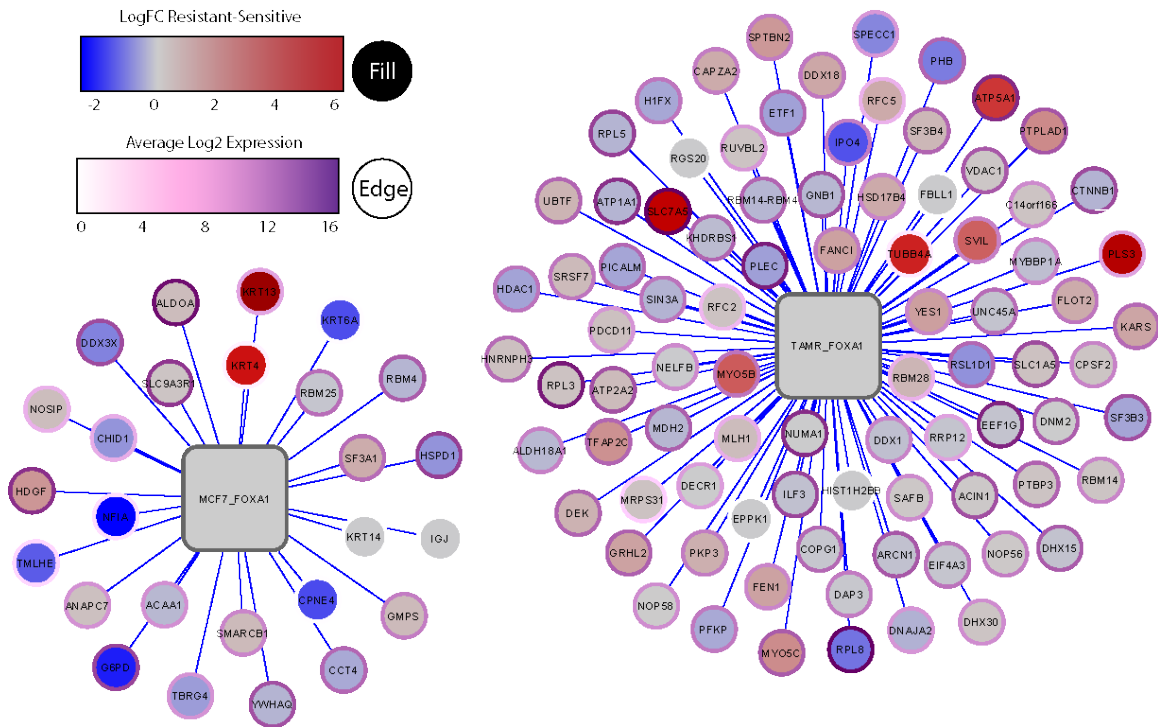


Figure 2-14: Network plot of factors identified by Mass-Spec.

Cytoscape plot of cell-specific FOXA1 associated proteins. Nodes are colored according to the differential Log2 expression of (Resistance)-(Sensitive) and average expression is indicated with the outer stroke color of the node.

2.3 Discussion

In this study we have identified several transcription programs that are differentially active in a cellular model of tamoxifen resistance. In particular, FOXA1 is enriched in resistant-specific chromatin accessible loci and regulates target gene transcription in both an ER-dependent and an ER-independent manner. The increased role of FOXA1 is not due to an increase in total protein levels however and instead is manifested through increased activity and potentially by localization.

There are multiple independent lines of evidence that ER transcriptional activity drives cell growth in endocrine resistance. The ER signaling program is plastic and can be reprogrammed in a cellular context by chromatin accessibility, epigenetics, coregulators and other transcriptional initiators including pioneer factors. Indeed, ER negative cells can be forced to house an ER program if ER and other pioneer factors are co-expressed, but not in the context of ER expression alone (Kong et al., 2011a). Since the discovery of FOXA1 as both a pioneer factor and implicit member of the ER signaling program in breast tissues, additional studies have uncovered several factors with reported licensing activity for ER and correlative association in clinical data (Magnani et al., 2011; Tan et al., 2011a; Theodorou et al., 2013). While many studies have probed ER and co-enriched motifs, evidence of functionality has been lacking. Furthermore, there is currently an absence of evidence as to the role of these ER-associated factors in drug resistance or in the role of reprogramming ER in tumor progression. Identification of the factors that endow ER with transcriptional competency in a model of resistance compared to sensitive cells was the primary focus of this study. Integrative analysis of DNase-Seq and enriched motifs with differential binding profiles of dozens of factors have elucidated a potential role for FOXA1 in mediating resistance. These results are substantiated by a dramatic increase of FOXA1 occupancy at these regions and experimental evidence validating a heightened functional role for FOXA1 in adjacent target gene transcription.

The role of FOXA1 in resistance has been controversial with opposing views in the literature. As FOXA1 is essential to the ER signaling program, its expression is highly correlated with ER and makes up the minimal definition of a luminal subtype. Analysis of tissue microarray first elucidated the correlation of FOXA1 with ER and association in low grade tumors (Wolf et al., 2007). Subsequent studies confirmed these associations in more than 5000 patients. Furthermore, we and others have shown FOXA1 expression levels of primary tumors to correlate with good prognosis (Badve et al., 2007; Thorat et al., 2008). FOXA1 was also found to be a good prognostic marker and that it had independent prognostic and predictive ability in line with Oncotype DX (Ademuyiwa et al., 2010). FOXA1 could also be considered favorable because loss of which could lead to basal characteristics and worse outcomes (Bernardo et al., 2013); underlying the need to selectively inhibit FOXA1.

On the other hand, FOXA1 binding sites are enriched in regions of novel ER binding sites in tumors resistant to tamoxifen (Ross-Innes et al., 2012b) and overlap with ~50% of ER binding sites total; imparting some control over ER binding preferences. Our results also reveal a functional difference of FOXA1 activity in TamR cells compared to MCF7 cells and suggests that a switch occurs to allow FOXA1 to take on oncogenic functions. One potential means of increased activation may rest on the ability of FOXA1 to respond to upstream signaling cascades. Phosphorylation mapping of FOXA1 revealed a loss of phosphorylation at S331 in TamR cells and functionality of this site is

currently being tested. In addition, our sequencing results found MCF7 cells to harbor the FOXA1 variant S448N, rs33984772, at a low frequency (~10%), but TamR cells carry the variant allele at a higher frequency (~58%); suggestive of biological functionality of S448. Another report identified FOXA1 to be acetylated by P300 and consequently diminishes DNA binding (Kohler and Cirillo, 2010), an event which may be due to growth factor signaling (Katika et al., 2013).

One interesting aspect to these results is that FOXA1 levels in TamR cells do not change and therefore undermines primary testing of FOXA1 levels as a prognostic indicator of its activity. In this regard, an activity signature of FOXA1 in this context is absolutely required. Using the robust targets of 30 experimentally validated target genes of FOXA1, a negative prognostic association of FOXA1 activity on relapse-free survival and distant-free metastasis is observed, highlighting the need for a thorough evaluation of target gene regulation to associate the effect on survival more broadly. While the importance of FOXA1 in ER action has been implicated in the clinic, the TamR model is the first observation of heightened FOXA1 activity in a model of resistance. Despite the scarcity of a FOXA1-mediated role in tumor progression or resistance, FOXA1-dependent genes are commonly identified as negative prognostic markers and notably include PS2, SPDEF, KRT13, XBP1, and AGR2.

Targeting FOXA1 in cancer should be approached with caution. As stated before, FOXA1 does have a role in maintaining an active ER program and elucidation of the

features that allow growth in a resistant state should be meticulously defined to inform drug development on the positive and negative facets of FOXA1. As stated before, tamoxifen resistant tumors and TamR cells are still sensitive to ER inhibition and alternative anti-estrogens inhibit growth. With this in mind, identification of the SERM profiles and their intrinsic differences of inhibiting ER on unique subsets within these FOXA1 loci is an intriguing approach that could have a real impact on the clinic in the near term. Also, an understanding of the SERM profiles in relation to FOXA1 activity will elucidate novel biomarkers for tracking progression and inform key molecular targets in drug discovery.

In summary, several independent experimental data and orthologous techniques elucidate a novel role for FOXA1 in driving tamoxifen resistance by means of reactivation of ER pathways. Experimental evidence indicates an altered usage of FOXA1 localization, modification and DNA occupancy in a tamoxifen resistant model. Thus, FOXA1 may be susceptible to modification as tumors progress to resistance. A further understanding of the role of FOXA1 in progression to and maintenance of tamoxifen resistance will determine whether this essential factor can be effectively targeted to delay tumor progression and improve the efficacy of ER-targeted therapies.

3 The role of HOXB13 on breast cancer pathogenesis

3.1 Introduction

With the abundance of gene expression data and relatively high rates of tamoxifen therapy failure, it's not surprising that several groups have used gene signatures to stratify risk by assessing the predictive value of gene expression on therapeutic outcome. One of these signatures is the two gene ratio of HOXB13:IL17RB, where the high expression of HOXB13 relative to a low IL17RB was found to differentiate patients that relapsed to those that remained disease-free at the five-year follow-up (Ma et al., 2004). More importantly this ratio correlates with poor prognosis, shortened time to recurrence, and tamoxifen therapy failure in ER+ breast cancers. The significance of the prognostic and predictive value of this ratio has been studied extensively and confirmed by multiple investigators in several cohorts of patients (Goetz et al., 2006; Habel et al., 2013; Jansen et al., 2007; Jerevall et al., 2008; Jerevall et al., 2011; Kok et al., 2009; Ma et al., 2006; Ma et al., 2008; Ma et al., 2004; Sgroi et al., 2013; Wang et al., 2007). This ratio is also remarkably confident when compared with the larger 21 gene set and 81 gene signature profiles when predicting prognosis and tamoxifen responsiveness (Kok et al., 2009). Even more impressive is the fact that HOXB13 alone has been shown to be an independent predictor of tamoxifen efficacy (Jerevall et al., 2008). Subsequent studies have confirmed these findings in breast cancer at both the mRNA level and additionally in a cohort of 912 tissue microarray slides by IHC but

have yet to elucidate a mechanism or establish a causal role for HOXB13 in mediating resistance (Jerevall et al., 2010). Despite all that is known about its utility as a predictive marker, a causal role for HOXB13 in breast cancer pathogenesis has yet to be established. A potential role for this protein in cell migration was hinted at in preliminary studies performed in the MCF-10 breast cancer cell line but has not since been expanded upon in the published literature (Ma et al., 2004).

HOXB13 is a member of the Homeobox superfamily of transcription factors that contain the conserved 61 amino acid DNA binding Homeodomain (Zeltser et al., 1996). These proteins are temporally and spatially regulated in development and play a central role in embryonic morphogenesis and determining position along the anterior-posterior axis. The 39 mammalian Hox proteins are organized into 4 genomic clusters (ABCD), each residing on a separate chromosome with each cluster containing 13 paralog groups. The DNA binding sites of these proteins have been defined by oligo arrays revealing a surprising lack of complexity for the specificity in binding (Berger et al., 2008). This implies that there may be additional factors that regulate DNA binding specificity and affinity or secondary DNA binding domains on the protein. In the adult, Hox proteins serve a role in differentiation and cell growth and their expression is tightly regulated, with their expression only being manifest in a tissue and cell specific manner. However, their expression is frequently dysregulated in cancer (Shah and Sukumar, 2010). Normal expression of HOXB13 is limited to the tail bud region including the Prostate, Colon,

and Urogenital Sinus but is misexpressed in a number of malignancies including that of breast cancer.

The expression of HOXB13 has been shown to be regulated by several steroid hormones in cellular models of breast cancer. Of note is the observation that it is suppressed by E2 in an ER dependent manner and that tamoxifen abrogates this regulation(Wang et al., 2007). The opposing role of estrogens suppressing HOXB13 and inducing IL17RB expression illustrates how this particular gene signature could merely reflect ER action and tamoxifen responsiveness. Interestingly, HOXB13 expression also correlates with tumor aggressiveness and closely tracks with other indicators of negative outcome, namely HER2 (Wang et al., 2007). This correlation to HER2 is also noteworthy given the crosstalk between ER and HER2 that has previously been implicated in tamoxifen resistance (Shou et al., 2004). Dysregulation of estrogen signaling pathways in breast cancer is a well-recognized event in tumor initiation and the loss of repression on HOXB13 by estrogens could underlie aspects of tumorigenesis.

This project arose as an opportunistic spinoff of an ongoing project in the McDonnell laboratory aimed at defining the factors expressed in prostate cancer that impact androgen signaling (Norris et al., 2009a). It was the dramatic effect of HOXB13 expression on the pharmacology of AR ligands and on AR target gene specificity that led us to consider that HOXB13 may play an important role in breast cancer outside of it being a useful marker of tamoxifen sensitivity. Specifically, in these now published

studies we used a microarray approach to evaluate AR target gene transcription in cells in which HOXB13 had been overexpressed or knocked down. In brief, we observed that HOXB13 can exhibit dramatic positive and negative effects on AR dependent transcription. In general, those genes upregulated by HOXB13 have roles in proliferation and those downregulated are involved in differentiation. The determinants of such diverse regulation by HOXB13 are based on both promoter context and its capacity to work in concert with AR through a novel tethering mechanism. The ability of nuclear receptors to tether to promoter bound transcription factors have been previously reported with AP-1, NfκB, and SP1 transcription factors (Zhou et al., 2007). These tethering factors are thought to play a key role in priming the chromatin landscape to allow other signal dependent transcription factors such as nuclear receptors to activate particular gene expression profiles.

In regards to therapeutic intervention, recent studies have identified cardiac glycosides as being able to downregulate HOXB13 (Johnson et al., 2002) and this class of drugs may provide an alternative therapeutic opportunity for this patient population.

While the molecular mechanisms by which HOXB13 associates with clinical outcome have not been defined, the aforementioned studies highlight an intriguing and potentially important role for it in breast cancer progression. It is therefore crucial to gain a better understanding of the mechanistic basis of HOXB13 to elucidate novel facets of resistance to ER targeted therapies, efficacy of targeting HOXB13, and to enable

clinicians to better-understand the value of HOXB13 expression in primary breast tumors. To that end, we undertook a project to evaluate the potential causal roles for HOXB13 on tamoxifen resistance and progression through convergence with ER signaling and also through alternate ER-independent mechanisms.

3.2 Results

3.2.1 Validation of the impact of HOXB13 on breast cancer progression and clinicopathological features

As a beginning to the project, the role of HOXB13 on survival and other relevant clinical features was confirmed with publicly available data. For this analysis, we utilized individual breast cancer datasets obtained from GEO(Edgar et al., 2002) and a breast cancer meta-set comprising more than 5000 patients from 27 individual datasets as described in Chapter 4. The expression of HOXB13 was interrogated for its association to available clinical parameters within each dataset and across the meta-set. Foremost, expression of HOXB13 was not significantly different across the molecular subtypes, as defined by subtype-specific gene expression program. Furthermore the impact of HOXB13 on relapse-free survival (RFS) and distant-metastasis free survival (DMFS) indicated that the role of this protein in disease pathogenesis was likely to be complex. In confirmation of previous published data, we noted that the high expression of HOXB13 had a significant interaction with tamoxifen therapy failure. However, the effect of HOXB13 on RFS and DMFS is also markedly pronounced in the untreated state suggesting the interaction of HOXB13 on survival may not be tamoxifen-specific but

may in fact indicate a more aggressive type of tumor (Figure 3-1). The effect on survival was tumor subtype-specific with worse prognosis in Luminal A with a hazard ratio of 1.37. In contrast to the luminal tumors, HOXB13 actually associates with prolonged survival in basal tumors and further exemplifies the tumor-specific roles a factor can display within the heterogeneity of breast cancer.

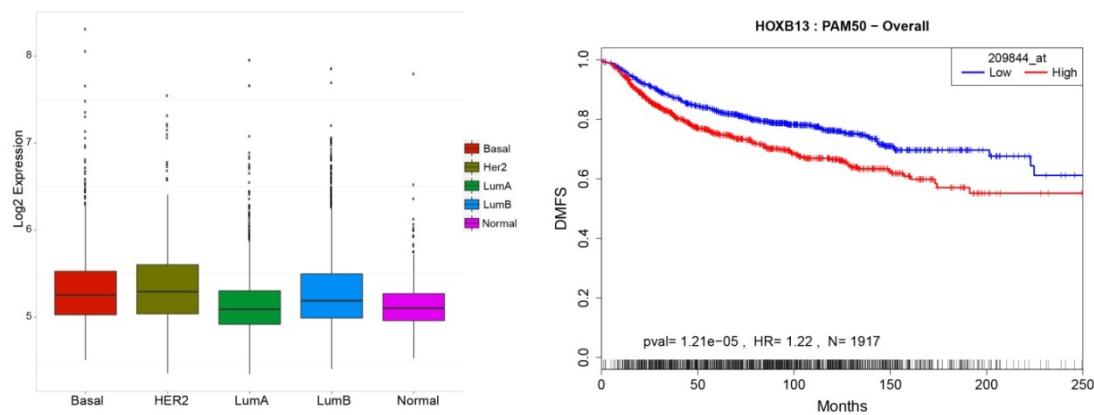


Figure 3-1: Clinical association of HOXB13 in a breast cancer meta-set.

HOXB13 expression is not significantly different across PAM50 tumor subtypes (left). Kaplan-Meier analysis of HOXB13 indicates worse prognosis in HOXB13 high tumors. Curves were generated within R using the survival package. Gene expression was split into tertiles and only the Low (1st tertile) and High (3rd tertile) data were plotted. Reported p-values were calculated using the log-rank method.

The newly released breast cancer TCGA dataset contains paired samples from matched normal-adjacent and primary tumor biopsies. We therefore utilized this paired-sample data to demonstrate in a case-by-case manner that the expression of HOXB13 is specific to primary tumors when compared to expression in normal adjacent sites (Figure 3-2). Consequently, we proceeded to evaluate the possibility that HOXB13 was a

biologically relevant regulator of ER action in the breast. Similar to the meta-set, expression is not limited to luminal tumors as it extends across all subtypes.

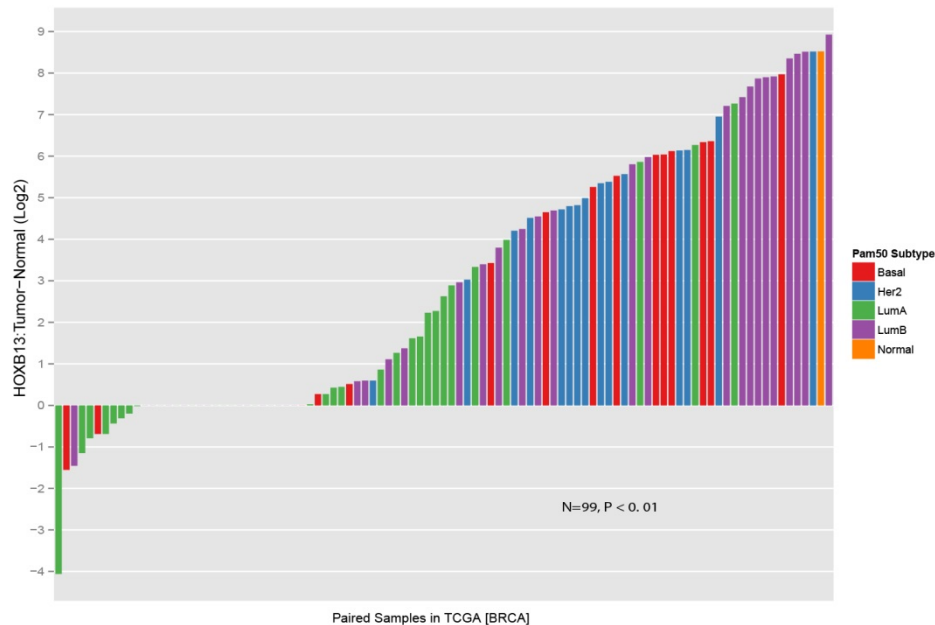


Figure 3-2: Waterfall Plot of HOXB13 in TCGA data.

RNA-Seq data and clinical information from the TCGA data portal and paired-sample data was used for this analysis. Data is presented as the difference in Log2 expression data from the primary tumor minus normal adjacent tissue in a case by case manner. Adjusted p-value is shown.

3.2.2 HOXB13 attenuates ER signaling

Our initial studies focused on HOXB13 as a coregulator of ER and were a logical extension of previous work centered on AR. To this end we used luciferase reporter assays to elucidate the impact of HOXB13 on ER-mediated transactivation of a synthetic reporter containing three tandem repeats of estrogen response elements(3x-ERE-TATA-Luc) (Chang et al., 1999) and additionally on an endogenous promoter of PS2, 1088 base

pairs in length (Lu et al., 2001). For these analyses, we titrated the levels of HOXB13 from no vector to 25ng of expression construct and found a dose-response relationship with either activation of 3X-ERE-TATA-Luc or repression of PS2-Luc (Figure 3-3).

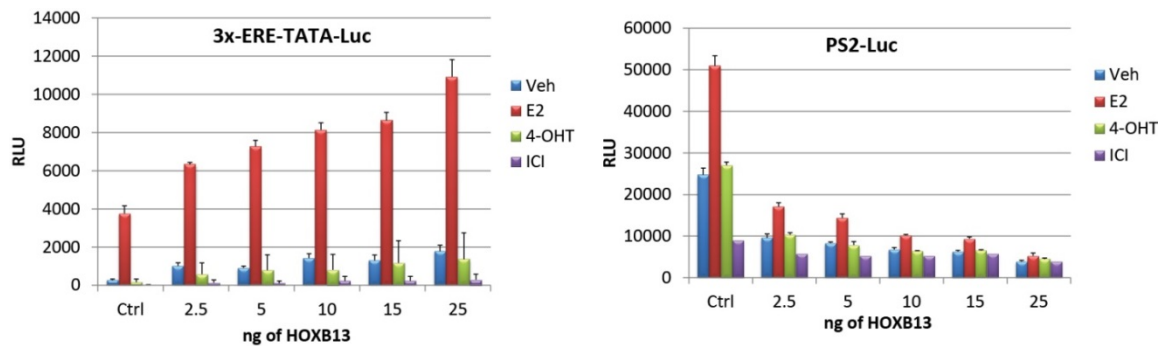


Figure 3-3: HOXB13 attenuates ER signaling.

Reporter assays of HepG2 transfected cells. Cells were plated in a 96-well plate, transfected with luciferase reporters and relevant expression constructs, and treated for 18 hours with indicated treatments. Luciferase signal was normalized to beta-gal activity to control for transfection efficiency.

3.2.3 HOXB13 promotes inflammatory pathways

3.2.3.1 Function Coexpression Analysis identifies an interferon gene signature and loss of a subset of ER target genes

Using the assembled breast cancer meta-set and TCGA data, we performed correlation analyses against HOXB13 in all the tumor samples and within each of the different PAM50 subtypes. The resulting gene list of ranked Spearman Coefficients was then submitted to GSEA from the Broad Institute and interrogated with gene-sets from the C2 and C5 databases to identify potentially meaningful biological insights of HOXB13 action. The enrichment analysis revealed a correlative loss of certain subsets of

ER target genes, but also showed a positive correlation of HOXB13 to an interferon signature of inflammatory genes. In particular, the CXCR3 family of chemokine ligands: CXCL9, CXCL10 and CXCL11 and other inflammatory genes MMP1, CCL5, and CCL18 were tightly associated. The CXCR3 ligands are pro-inflammatory chemokines and their action are linked to the development of autoimmune diseases such as Graves' disease, systemic lupus erythematosus, rheumatoid arthritis, and ulcerative colitis (Antonelli et al., 2010; Egesten et al., 2007; Lacotte et al., 2009). In breast cancer, the actions of these inflammatory gene products are also implicated in promoting metastases to the lung (Ma et al., 2009a; Ma et al., 2009b).

A subset of ER target genes, PR, CTSD, and SCUBE2, also negatively correlated with HOXB13. The inverse relationship with PR is noteworthy since loss of PR, as proven by IHC, identifies a subclass of Luminal B patients with worse outcome and a significantly decreased response to anti-estrogens (Arpino et al., 2007; Canello et al., 2013; Cui et al., 2005). These data may indicate that the expression programs that underlie resistance associated with HOXB13 are congruent with the associated mechanisms of resistance associated with PR loss. HOXB13 expression also strongly correlated with AURKA, a reliable indicator of proliferating cancer cells (Figure 3-4).

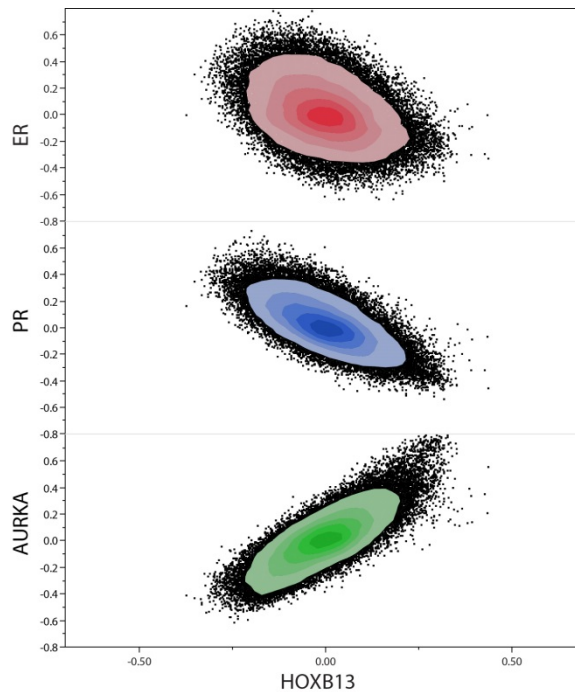


Figure 3-4: Association of HOXB13 with notable factors.

A scatter density plot of HOXB13 against ER, PR and AURKA in TCGA data.

3.2.4 HOXB13 promotes a pro-inflammatory response

To confirm the association of HOXB13 with the inflammatory gene subsets identified by GSEA, MCF7 stable cell lines were engineered to constitutively express either control Gal4 or HOXB13 by lentivirus. Two different levels of HOXB13 expressing cells were made, one for very low expression equivalent to the lower third tier of HOXB13 expression in patients and a higher expressing cell line. Intriguingly, addition of HOXB13 to MCF7 cells caused growth inhibition and for a period of several weeks no growth was observed while control cells grew as normal. As soon as cells began to grow 4 weeks later, the identified inflammatory targets including the CXCR3 ligands were

assessed by qPCR in these cells. In line with the clinical association, the genes, CXCL9, CXCL10 and CXCL11 were found to be dramatically upregulated relative to control cells (Figure 3-5).

The CXCR3 ligands are highly expressed in type 1 macrophages. These polarized macrophages have an activated IRF-3 transcription program which results in interferon expression and subsequent STAT1/3 activation leading to CXCL9, CXCL10, and CXCL11 upregulation. In relation to macrophages, IFN-alpha and IFN-beta were found to be upregulated in HOXB13 expressing MCF7 cells, and while STAT1 phosphorylation was not differential, STAT3 phosphorylation was increased to a minor extent. Other STAT genes were not assessed.

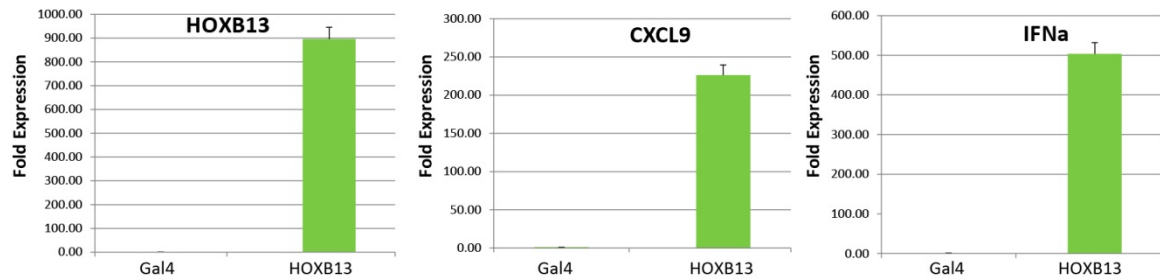


Figure 3-5: HOXB13 promotes an inflammatory response.

MCF7 stable cells were generated from Lentivirus infection of pLenti-CMV vectors containing a control Gal4 or HOXB13. Cells were seeded in a 6-well dish at 300k cells per well in CS-FBS media. Two days later cells were harvested and qPCR was used to quantify mRNA.

3.2.5 Overexpression of HOXB13 in cellular models deregulates ER signaling via loss of GATA3

While HOXB13 may have some direct actions on ER, as observed with transactivation assays, the most pronounced results on ER signaling in MCF7 cells

expressing HOXB13 are likely attributed to loss of the pioneer factor GATA3. GATA3 has previously been shown to positively regulate ER levels in MCF7 cells (Eeckhoutte et al., 2007), and indeed ER levels were decreased along with a subset of ER target genes (Figure 3-6). However not all genes were decreased and so as a control, SDF1 (CXCL12) is shown.

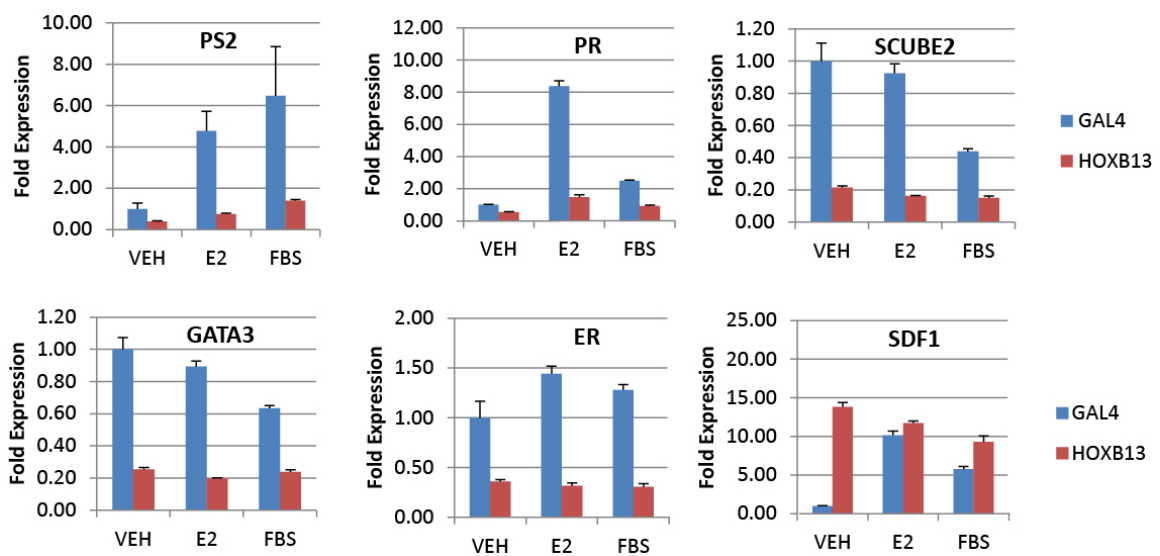


Figure 3-6: HOXB13 expression leads to loss of ER target genes.

qPCR results of prominent ER target genes. Cells were seeded in a 6-well dish at 300k cells per well in CS-FBS media. Two days later cells were treated for 18 hours with vehicle, estradiol (E2) or FBS media and harvested for qPCR.

3.2.6 Loss of GATA3 promotes the HOXB13-mediated inflammatory phenotype.

Beyond its role as a pioneer factor, GATA3 plays a central role in TH1 polarization. T cell fate is determined by the interplay of two opposing transcription factors, GATA3 and T-bet which are activated by IFN and Il-4 respectively. Activation of

one pathway also feeds back and negatively inhibits the other in order to exert dominance and cell commitment. The GATA3 program promotes TH1 cell differentiation and subsequently induces chemokines CXCL9, CXCL10 and CXCL11. In our system, expression of HOXB13 leads to loss of GATA3 and increases T-Bet expression. We hypothesized that the loss of GATA3 and associated increased Interferon response was mimicking the signaling paradigm manifested in T cells and that re-expression of GATA3 could suppress the inflammatory response. Indeed, adenovirus-induced expression of GATA3 was able to markedly decrease the majority of the inflammatory genes relative to control virus and rescue ER expression (Figure 3-7).

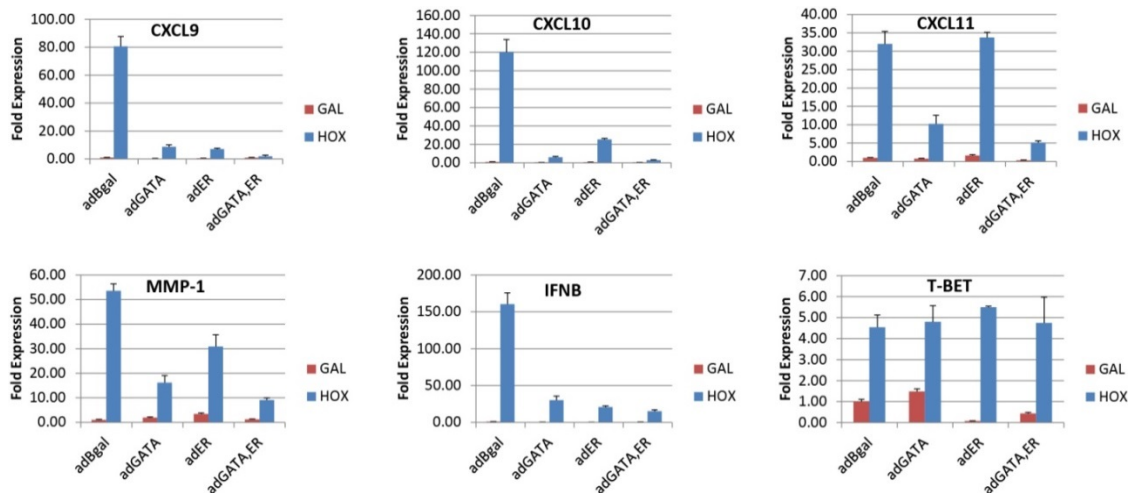


Figure 3-7: Adenovirus of GATA3 ameliorates the HOXB13-induced inflammatory genes.

Rescue of inflammatory genes. Cells were seeded in a 6-well dish at 300k cells per well in CS-FBS media and infected with adenovirus at an MOI of 10. 48 hours after infection cells were harvested for qPCR as before.

3.2.7 HOXB13 induces AP-1 Responsive elements

In the course of assessing the relationship of HOXB13 with ER it was observed that HOXB13 activated genes responsive to AP-1 and reporter constructs with an active AP-1 site, particularly the inflammatory gene MMP1 and associated luciferase reporter pCOL (MMP1). To verify this association with AP-1, we cloned an oligonucleotide containing three tandem repeats of the canonical AP-1 element, TGAGTCA upstream of a minimal promoter in PGL4.26. The phorbol myristate acetate (PMA) is a known activator of AP-1 by mediating phosphorylation of c-Jun and c-Fos through a PKC/Raf/MEK dependent mechanism (Brinckerhoff et al., 1979; Sharma and Richards, 2000). Therefore PMA was used to induce AP-1 activity in the presence of a control, wild type HOXB13 and a DNA binding mutant of HOXB13 (HOXB13-3A) on these reporters (Figure 3-8A). These experiments revealed that HOXB13 alone induces a minimal AP-1 reporter to similar levels of PMA-induced activity and that this activity required the DNA binding domain of HOXB13. Moreover, the transactivation by HOXB13 was selectively ablated with MEK1/2 inhibitors (U0126, AZD6244, PD184161) among a panel of inhibitors (Figure 3-8B). The association of HOXB13 and MEK are intriguing in regard as to how the two factors synergize. Protein levels and phosphorylation of MEK and ERK are unchanged after 24 and 48 hours post-transfection of HOXB13. Furthermore, PMA-induced and HOXB13-induced activity

absolutely requires MEK activity, implying a dependence of HOXB13 on downstream effectors of MEK.

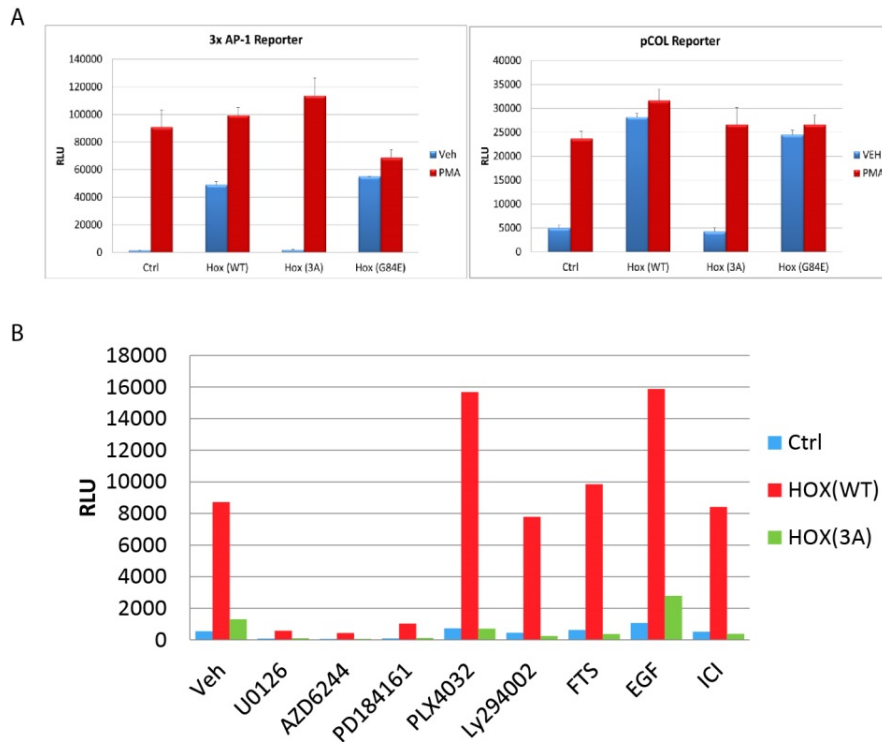


Figure 3-8: HOXB13-mediated activation of AP-1 reporters is sensitive to MEK1/2 inhibition.

Reporter assays of HepG2 transfected cells. Cells were plated in a 96-well plate, transfected with luciferase reporters and relevant expression constructs, and treated for 18 hours with indicated treatments. Luciferase signal was normalized to beta-gal activity to control for transfection efficiency.

3.2.8 Cardiac Glycosides downregulate HOXB13

The expression of HOXB13 is robustly suppressed by Ouabain and other cardiac glycosides, which provides a potential for therapeutic intervention of tumors with HOXB13. Specifically, Johnson et al., used a multiplex assay to target candidate genes in PC-3 prostate cancer cells with 9000 compounds and found cardiac glycosides to inhibit

HOXB13(Johnson et al., 2002). Cardiac glycosides are used clinically for congestive heart failure by inhibiting the sodium-potassium ATPase pump and forcing intracellular calcium pools to increase, resulting in a stronger force of contraction by cardiac myocytes. Epidemiologic evidence for cardiac glycosides suggests their use is associated with lower mortality rates and cancer incidence (Mijatovic et al., 2007). In particular, Stenkvist performed a 22 year follow-up on breast cancer patients and found that patients taking digitalis had a 28% lower death rate (Stenkvist et al., 1982). Indeed other studies have also found cardiac glycosides to exhibit anti-cancer properties(Haux, 1999).

The sodium-potassium pump is a major signaling hub and is critical for cell growth, differentiation and cell survival. Furthermore, it acts as a scaffold for several major protein interactions such as SRC, and binding of cardiac glycosides induces several major signal transduction pathways including the MAP kinase pathway(Haas et al., 2002; Li and Xie, 2009). Other signaling cascades are also activated in response to ionic imbalance across the membrane.

In order to assess if cardiac glycosides can also suppress HOXB13 in breast cancer cells, we used BT474 cells which have high levels of endogenous HOXB13. Several cardiac glycosides (Ouabain, Digoxin, Proscillaridin A, and Lanatoside C) were able to inhibit HOXB13 at both the mRNA and protein level following an 18 hour treatment (Figure 3-9). Moreover, treatment of the cardiac glycosides over a ten day period resulted in inhibition of growth. However, forced expression of a CMV driven

HOXB13 could not rescue the growth phenotype of cardiac glycosides and is somewhat expected given the fact that these molecules simultaneously activate so many pathways.

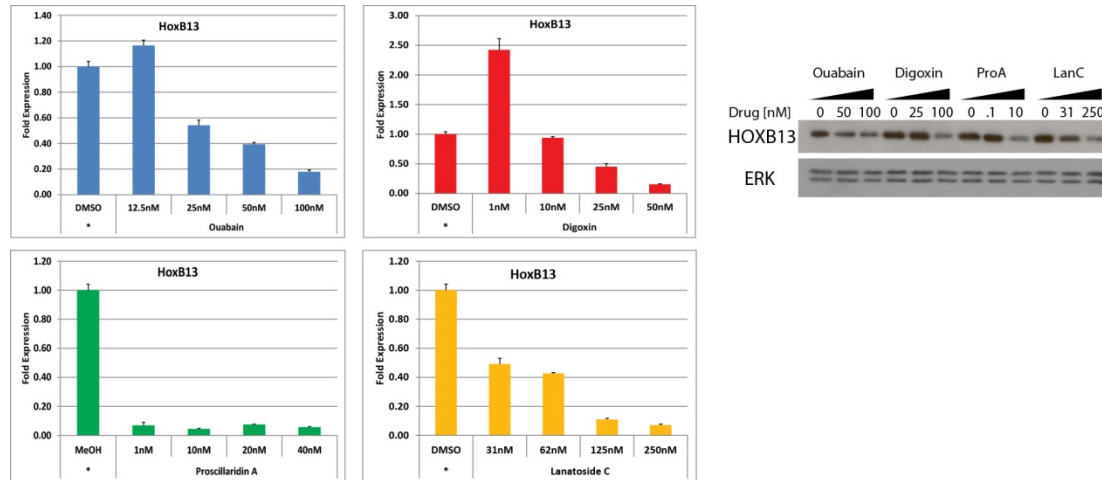


Figure 3-9: Cardiac glycosides suppress HOXB13 in BT474 cells.

BT474 cells were seeded in 12-well dishes at 250,000 cells per well. After 48 hours cells were treated with Vehicle, Ouabain, Digoxin, Proscillaridin A, or Lanatoside C at indicated concentrations for 18 hours (left). Western blot of protein run in parallel with qPCR assays.

3.3 Discussion

Expression of HOXB13 has been indicated to predict tamoxifen therapy outcome and shortened time to recurrence in patients, establishing it as a potentially useful biomarker. Previous studies have elucidated that HOXB13 is repressed by E2 and inversely correlates with ER levels in the clinic. However, the molecular mechanisms by which HOXB13 associates with clinical outcomes are undefined.

In this study we identify a critical node of HOXB13 action in that it suppresses the expression of the pioneer factor GATA3 and leads to loss of a subset of ER target genes and partial loss of ER itself. Beyond its role as a pioneer factor, GATA3 is also

well known to modulate inflammatory signaling and in this case, loss of GATA3 promotes a pro-inflammatory program that relies on STAT3 phosphorylation to activate CXCL9, CXCL10, CXCL11, CXCL12, CCL18, CCL15, and MMP1. Furthermore, the dependency on MEK1/2 and relative impact on AP-1 signaling can be explained by STAT3 activity. STAT3 has been previously shown to interact with AP-1 on the MMP1 promoter by (Zugowski et al., 2011) and that binding of STAT3 was necessary for transcriptional regulation by AP-1.

In the course of confirming a role for STAT3 activation as a means to subvert the repressive effects of tamoxifen, Shah et al. published a report that HOXB13 induces IL-6 and promotes STAT3 activation in a Jak-independent manner by way of mTOR (Shah et al., 2013). These results have not been confirmed in our studies.

In-vitro studies have failed to show any significant resistance/desensitization to tamoxifen by HOXB13 when assayed by cell proliferation or soft agar assays. It is highly plausible that tamoxifen resistance would only be manifest in an intact environment allowing tumor-stromal interactions, especially since the defining characteristic of HOXB13 is that of promoting inflammatory genes which are involved in cell to cell communication signaling pathways. Future studies should therefore define the impact of HOXB13 on tumor growth in a syngeneic model using C57BL/6 mice with E0771 cells stably expressing HOXB13 or control.

4 Development of a breast cancer meta-set and web application for discovering gene function and clinical relationships

4.1 Introduction

Breast cancer is a diverse, highly heterogeneous disease with vast differences in outcome and response to therapeutic strategies. Gene expression profiling of tumors has been instrumental in defining the molecular determinants of tumor subtypes and cancer related genes. Wide-spread use of expression profiling of tumors have precipitated an abundance of gene-signatures and mixture models to assess patient survival, therapeutic response and identification of underlying biological processes. These top-down approaches are now well-established across disease states and are built around the paradigm of comprehensive feature analysis to aid in the diagnosis and treatment of cancer through data-driven discovery. With this vision, large-scale and focused studies have generated copious amounts of publicly available data comprising thousands of tumor samples. These information-rich data sets provide unprecedented opportunities for scientific investigation and hypothesis-driven research through bottom-up approaches. Leveraging this data is not without difficulties however, and requires programming skills and bioinformatics expertise; often discouraging effective use by investigators.

The barriers to effective application of clinical data are numerous and common within the research setting. The process requires collection of raw data, quality checks,

normalization, analytical methods and visualization techniques. Related to this process is the idea of generating statistically robust, reproducible results where the workflow is transparent and well-documented. Another point to consider is that much of the expression data is generated with small patient sizes and cross-validation approaches or data merging is necessary to elucidate the significance more broadly. Nonetheless, bottom-up approaches are essential in providing biologically meaningful functional studies to associated human tumor data. Integration of these two distinct types of data is necessary to fully understand the connection of biological mechanisms and impact on breast cancer outcome. Therefore, an application built around an established framework that can perform high level analytical methods on one or a handful of genes is essential in empowering biologists with the tools to quickly and efficiently query data to add to the scientific knowledge domain.

The focus of this study can be divided into three major points: generation of breast cancer meta-set to leverage the number of patient data, development of functional modules for high level analyses, and design of an investigator friendly user interface.

4.2 Results

4.2.1 Development of a breast cancer meta-set

As breast cancer is a heterogeneous disease, large sample sizes are critical in generating meaningful, subtype-specific data with high confidence. While the recent large, multi-center effort of The Cancer Genome Atlas (TCGA) consist of ~1000 patients

presently, long-term clinical outcome is yet to be realized and thus survival analysis requires alternative data. Early gene expression profiles relied on Affymetrix microarrays and these clinical data only consist of a few hundred patients per study at most and a few dozen on the low end. Thus it is necessary to cross-validate results across many studies or to perform a meta-analysis to infer significance. For this purpose, data sets are often pooled to increase sample size to extend results obtained from single studies. However, combining data in this manner, especially in breast cancer, poses a number of statistical difficulties and often results in gross errors. These challenges arise from the fact that studies are often performed on different platforms by different facilities with different methods of tumor extraction protocols. Moreover, clinical cohorts in these studies are often divergent on ethnicity, tumor subtype, and other patient characteristics central to the associated trial. Thus, non-biological experimental variations or “batch effects” are often hard to account for while conserving true biological relationships. Consequently, a simple merging of data sets is unadvised. For this purpose, normalization procedures which allow for gene-wise linear pairing and cross-platform methodologies have been developed (Benito et al., 2004; McCall et al., 2010; Shabalina et al., 2008). In particular, frozen Robust Multi-array (fRMA)(McCall et al., 2010) is a normalization procedure similar to RMA that parameterizes the signal for a given platform using “frozen vectors” and normalizes the data such that each array can

be processed independently while still yielding consistent results when data is from different sources.

In order to collect as many data sets as possible, the gene expression omnibus (GEO) was queried for breast cancer datasets that were performed on HGU133A or HGU133Plus2 Affymetrix platforms. In total, 25 non-redundant data sets were identified comprising 4885 patients. Datasets used: GSE10780, GSE11121, GSE12093, GSE12276, GSE1456, GSE16391, GSE16446, GSE17705, GSE17907, GSE19615, GSE20194, GSE2034, GSE20685, GSE20711, GSE2109, GSE21653, GSE22093, GSE24185, GSE25066, GSE3494, GSE5460, GSE6532, GSE6532, GSE7390, and GSE9195. The raw data was downloaded from GEO, and each dataset was normalized with fRMA to remove platform-specific batch effects. Note that the two Affymetrix platforms are not directly comparable using fRMA and further data merging techniques are required. Therefore, the data was then combined using the COMBAT algorithm implemented in the sva package within R (Leek et al., 2012) with a design matrix to account for known covariates including data source and platform. Each tumor was then classified into PAM50 (Parker et al., 2009) molecular subtypes using *genefu* (Haibe-Kains et al., 2012). To confirm normalization, a Multi-dimensional Scaling (MDS) plot was used to visually inspect the data in relation to platform and tumor subtype (Figure 4-1). It is clear that the data prior to merging is grouped relative to platform, while after merging is grouped strongly with tumor subtype.

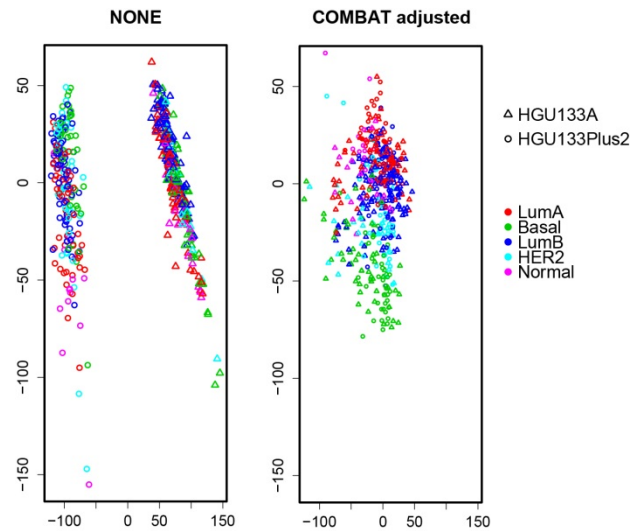


Figure 4-1: MDS plot of data before and after merging.

MDS plot of 500 random patient data before adjustment (left) and after COMBAT adjustment (right). Tumor subtypes and Affymetrix platform are represented by color and shape respectively. Both plots are on the same scale.

Additionally, ER levels were assessed to ensure that a bimodal distribution was present after merging the data with COMBAT (Figure 4-2A). Notably, when the data is plotted according to its originating data source (Figure 4-2B), the histogram in red is situated between the two bimodal distributions. However, this data set (GSE10780) is comprised of histologically normal tissue and profiles accordingly.

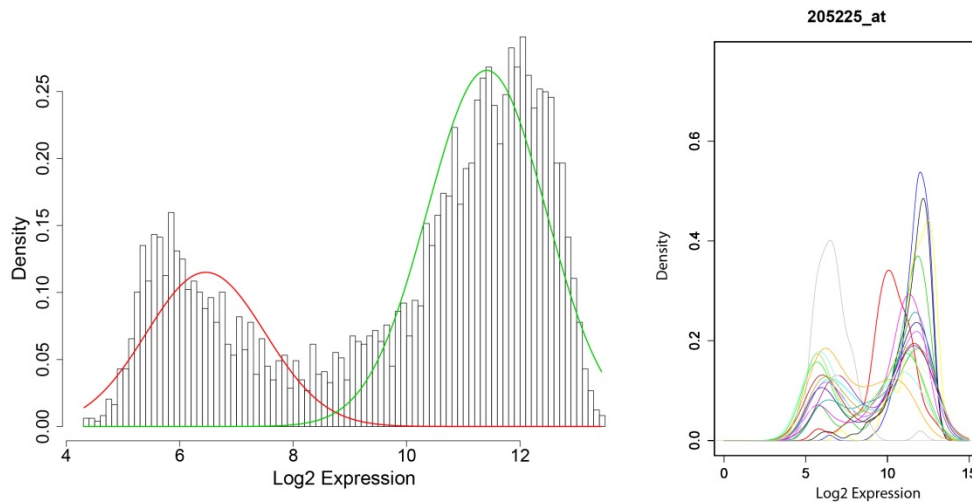


Figure 4-2: Histogram of ER expression.

Density distribution plot of ER expression levels of probe 205225_at. Density fitting curve for bimodal distribution was also calculated with the R package mixtools.

Another formidable task in data set merging is in aggregating the associated clinical data, especially in light of the fact that there is not a standard format of annotation used in the studies. Data from GEO, published reports, author's websites, and additionally by means of communication, was collected and manually curated. In this manner, patient characteristics including age, node involvement, neo-adjuvant treatment, adjuvant treatment, survival end points and other relevant data was obtained.

After merging gene expression data with clinical information, it was apparent that some patients were represented in multiple studies and therefore these duplicates were removed to form a database of non-redundant patient samples. Furthermore, some arrays failed to pass quality checks and were also removed.

4.2.2 Development of functional modules

One of the key aims of this project was to develop a set of functions that allow rapid generation of data with minimal input to address key questions. Establishment of routine procedures leads to reproducible results and instills confidence to others about the integrity of the data being generated. Other problems related to use of clinical expression data is the fact that one-off data sets misrepresent the underlying biology and meta-analytical procedures are necessary to confirm data concordance. In this regard, metrics were instituted into the analytical functions to report dataset-specific results.

Two independent data sources are used by the functions. The aforementioned breast cancer meta-set is used in all functions while TCGA data is used for everything except survival analysis. These data are stored in a SQLite database format for rapid retrieval of the data of interest. Also, the R environment is used for data retrieval, processing and graphing. Three different algorithms to determine tumor subtype are incorporated: PAM50 (Parker et al., 2009), MOD1, and MOD2 (Desmedt et al., 2008; Wirapati et al., 2008). Functions developed include survival analysis of a single factor, survival analysis of a gene-set, expression of a single factor, expression of multiple, and correlation analysis with subsequent pathway enrichment. Each set of functions are described below with representative output. All code is available at (<https://github.com/jasper1918/GeneAnalytics>).

These functions require input of a gene identifier and can be queried either with an Affymetrix probe set, (e.g. 2055225_at), or with a gene symbol, (e.g. ESR1). In the latter method, gene symbols are mapped to associated Affymetrix probe sets and this often results in multiple matching probe sets. In the case of multiple matching probe sets, typically probe set intensities are averaged or the most invariant probe set used, but this can lead to unreliable results. To illustrate the point, nine probe sets for ER are represented on the HGU133A platform and only one, 205225_at, has been shown to correlate with ER status by IHC. To this end, a series of metrics related to gene-specificity of the probe, transcript coverage, and robustness can be used to identify the highest scoring probe set as previously published (Li et al., 2011).

4.2.2.1 Expression Analysis

The identification of tumor subtypes is based on the relative differences of a subset of genes across tumors. Thus, profiling a single factor or handful of genes across tumor subtypes can be used to inform biological functionality or to identify which tumors are more likely to respond to therapeutics targeting the factor of interest. Two functions are available for expression analysis: single gene expression and a heatmap for visualizing multiple genes. Parameters available to the user include gene identifier and choice of subtype algorithm.

Genes from GO term 0046605 representing 'regulation of centrosome cycle' were used as input for expression analysis in a demonstrative example below (Figure 4-3A).

Both Aurora Kinase A (AURKA) and NIMA-related kinase 2 (NEK2) show strong differential expression across the subtypes in these results. The kinase, AURKA is involved several steps of mitosis and has been implicated in disease progression and tumor aggressiveness in breast cancer (Fu et al., 2007). Further analysis of AURKA expression reveals that it is highly expressed in Basal-like HER2-enriched, and Luminal B tumors, but is much less represented in Luminal A and Normal-like tumors (Figure 4-3B).

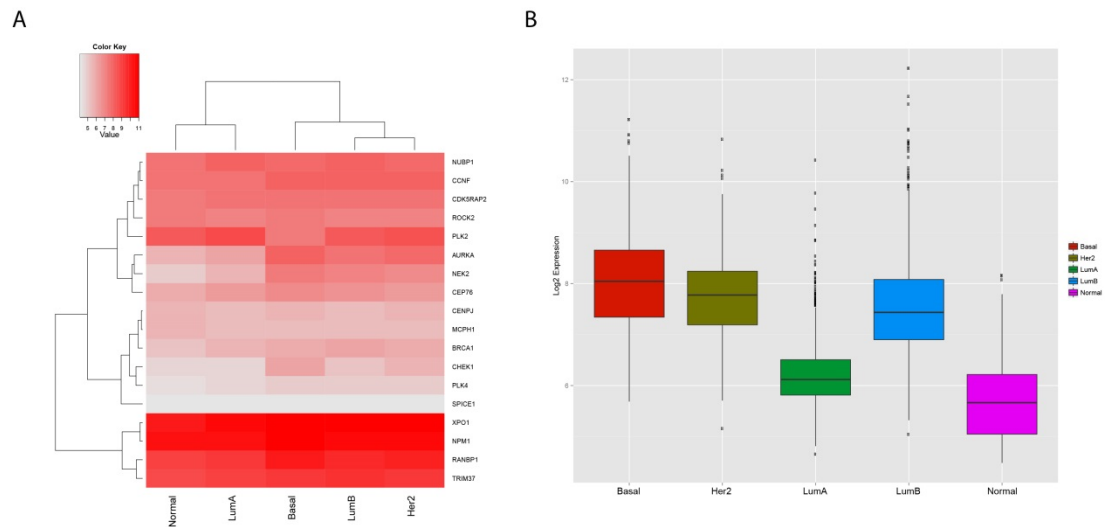


Figure 4-3: Expression Analysis of AURKA.

Plots produced from the breast cancer meta-set using the expression heatmap function with GO term 0046605 (A), and single gene expression of AURKA across PAM50 subtypes (B).

4.2.2.2 Survival analysis

Kaplan-Meier curves are used to assess the impact of a gene or gene-set on survival. In this approach relapse-free survival and distant metastasis-free survival data from the breast cancer meta-set are used in conjunction with censor information to build a cox proportional hazard model using the `surv` package in R. Parameters available to the user include gene identifier, months to plot, choice of tumor subtype algorithm, method of splitting low and high groups, whether to include tamoxifen or chemotherapy treated patients, and relevant survival end point. Log-rank p-values are reported on all plots. Concordance-Index (c-index) scores (Schröder et al., 2011) are also generated for each data set in order to report on the relative impact of each study on survival outcomes for each subtype.

Survival analysis can be run with either a single gene or with multiple genes to form a gene-set. With a gene-set, signature scores are calculated for each patient and subsequently split into low and high groups for plotting. Continuing with the example set of genes from go term 0046605, survival curves of the signature scores are profiled across all tumor subtypes using the combination of RFS and DMFS endpoints (Figure 4-4). Furthermore, to address data consistency a plot of the c-index scores of each dataset is produced for each tumor subtype. The Luminal B plot of c-index scores is shown in Figure 4-4D.

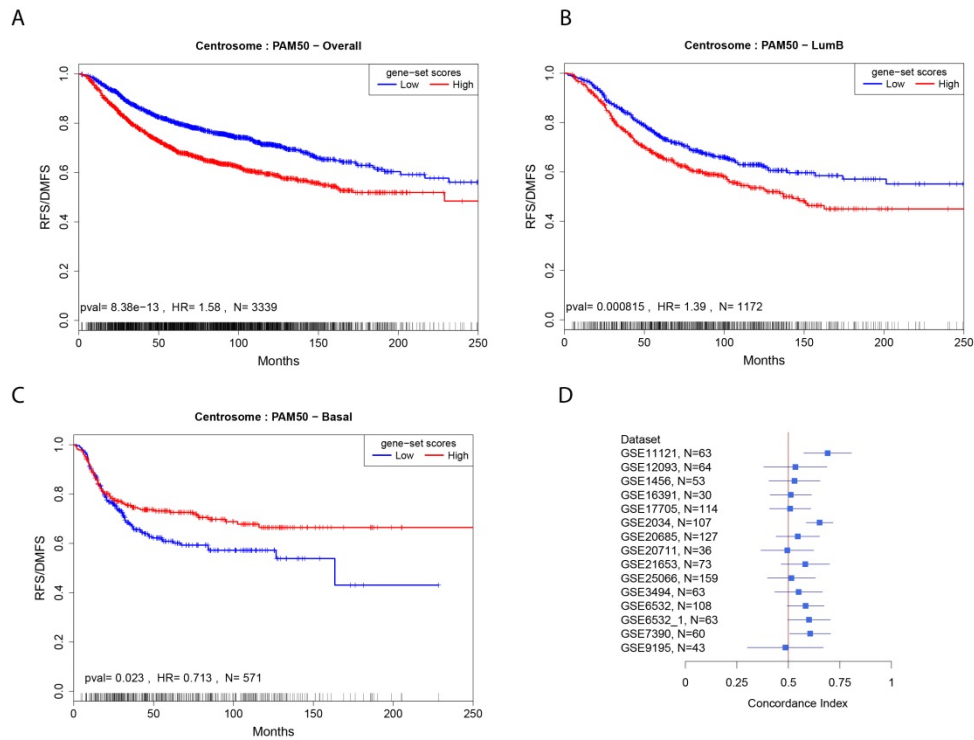


Figure 4-4: Survival analysis of G0:0046605 genes.

Plots produced from the breast cancer meta-set using the gene-set survival function. Kaplan-Meier curves are generated for all tumors (A), Luminal B tumors (B) and Basal-like tumors (C). Plot of the Concordance-Index scores for each data set used. Gene expression was split on the median into low and high classifications and Kaplan-Meier curves were plotted in R.

In a similar manner, AURKA is profiled across all tumors and subtypes using the combination of RFS and DMFS as an end point (Figure 4-5). Data from the breast cancer meta-set reveals that while it is expressed in high levels across both ER negative and ER positive subtypes, a negative association of the gene-set or AURKA on prognosis is only relevant to ER positive groups (not all data shown). The plot of c-index scores for Luminal B reveals that 13 of the 15 datasets used to generate the merged survival curve for AURKA are trending together while two show opposing effects.

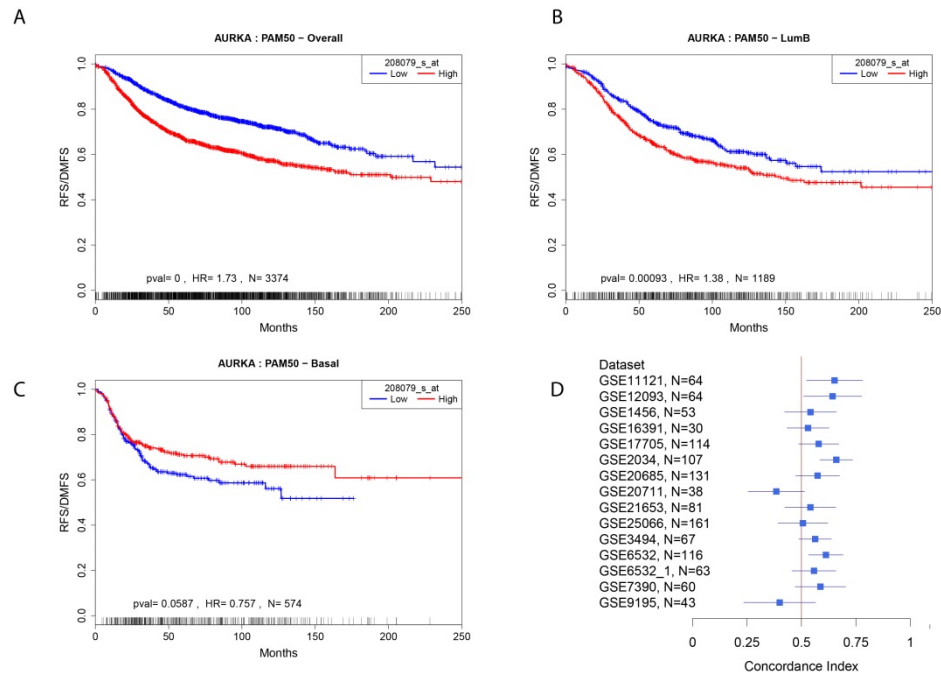


Figure 4-5: Survival analysis of AURKA.

Plots produced from the breast cancer meta-set using the single gene survival function. Kaplan-Meier curves are generated for all tumors (A), Luminal B tumors (B) and Basal-like tumors (C). Plot of Concordance-Index scores for each data set used. Gene expression was split on the median into low and high classifications and Kaplan-Meier curves were plotted in R.

4.2.2.3 Forest plot

The results above suggest that genes represented in the example using GO term 0046605 representing ‘regulation of centrosome cycle’ are highly correlated with tumor progression in Luminal B tumors. While AURKA profiles very similar to the gene set overall, a separate function can be used to assess the effect of each gene within a gene set on survival. The forest plot function requires the same set of parameters from the gene-set survival. Results with GO term 0046605 reveal that AURKA is indeed highly

associated with a negative outcome among the gene-set along with NEK and CENPJ genes.

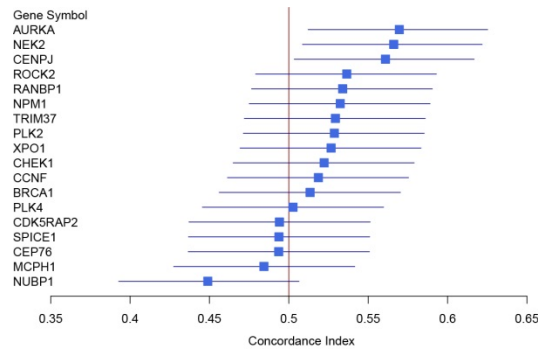


Figure 4-6: Forest plot of GO:0046605 genes.

Plots produced from the breast cancer meta-set using the forest plot function for genes in GO term 0046605.

4.2.2.4 Functional co-expression analysis

With several thousand samples of patient derived expression data it is possible to utilize a systems biology approach to infer biological insight and functionality of a gene from a co-expression network and topology approaches. These discovery approaches also hold the potential in identifying co-regulatory gene-networks, transcription factor programs, and nodes of convergence for pathways that cross-talk. Gene co-expression networks have been used previously to elucidate intersecting genes to build prognostic signatures, identify transcriptional networks, and discover gene complementarity (Carter et al., 2004; Ruan et al., 2010; Xiang et al., 2012; Zhang and Horvath, 2005).

This function takes a gene-centric approach to build a co-expression network utilizing Spearman's rank correlation coefficients of a single gene of interest with every other gene represented in the data set. This is done across all tumors and additionally in each tumor subtype. Thus, the only two parameters necessary are a gene identifier and choice of a tumor subtype algorithm.

A table of correlation coefficients is returned for each tumor subtype and across all tumor datasets. The function then takes the top 50 genes that highly correlate and the bottom 50 that negatively correlate with the factor of interest to build a correlation network of all gene pairs and produces plots of correlation matrices as corelograms and node-edge graphs. Furthermore, the resultant rank list of correlation coefficients is used as input for pathway enrichment procedures using GSEA, GAGE, and against individual databases of pathways including KEGG and Reactome. Each of these approaches produces a table and figure of the most significantly associated relationships.

As a proof of concept, AURKA was used as input for Functional co-expression analysis. GSEA results of the C2 database (Table 4-1) identified the AURKA ranked correlation list to be relatively enriched in gene-sets central to cancer progression, mitosis, and proliferation; strongly coinciding with its described role from the published literature. Other pathway approaches also strongly indicated a role for AURKA in mitosis, proliferation, cell cycle, and cancer-specific pathologies.

Table 4-1: GSEA results from Functional co-expression analysis for AURKA.

<i>Gene Set Name</i>	<i>p.geomean</i>	<i>stat.mean</i>	<i>p.val</i>	<i>q.val</i>	<i>set.size</i>	<i>expl</i>
SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_A6	3.20E-89	22.95752	3.20E-89	1.33E-85	446	3.20E-89
SOTIRIOU_BREAST_CANCER_GRADE_1_VS_3_UP	2.51E-44	19.19713	2.51E-44	2.08E-41	147	2.51E-44
ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER	1.06E-41	18.36404	1.06E-41	7.33E-39	140	1.06E-41
TARTE_PLASMA_CELL_VS_PLASMABLAST_DN	6.38E-40	14.2299	6.38E-40	3.79E-37	300	6.38E-40
GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_UP	6.40E-38	13.44016	6.40E-38	3.33E-35	445	6.40E-38
REACTOME_CELL_CYCLE	8.43E-38	13.63092	8.43E-38	3.90E-35	342	8.43E-38
KOBAYASHI_EGFR_SIGNALING_24HR_DN	1.11E-37	13.93454	1.11E-37	4.60E-35	247	1.11E-37
BERENJENO_TRANSFORMED_BY_RHOA_UP	1.77E-36	13.09974	1.77E-36	6.68E-34	479	1.77E-36
GARY_CD5_TARGETS_DN	2.31E-36	13.28923	2.31E-36	7.99E-34	362	2.31E-36
REACTOME_CELL_CYCLE_MITOTIC	2.27E-35	13.28379	2.27E-35	7.27E-33	281	2.27E-35
BENPORATH_PROLIFERATION	8.67E-35	15.11737	8.67E-35	2.58E-32	131	8.67E-35
BOYALT_LIVER_CANCER_SUBCLASS_G3_UP	3.93E-34	13.67147	3.93E-34	1.09E-31	185	3.93E-34
LINDGREN_BLADDER_CANCER_CLUSTER_3_UP	5.11E-34	12.98491	5.11E-34	1.33E-31	275	5.11E-34
ZHANG_TLX_TARGETS_60HR_DN	1.07E-33	13.09477	1.07E-33	2.62E-31	230	1.07E-33
REACTOME_DNA_REPLICATION	2.37E-33	13.58798	2.37E-33	5.47E-31	173	2.37E-33

All correlation matrices and pathway enrichment results are conveniently compressed into a single file for download and provide further opportunities to leverage the data for subtype-specific analysis, differential correlation networks, and cluster-specific approaches.

4.2.3 Web application

In order to make the aforementioned breast cancer meta-set and associated functions accessible and easy to use, a web-based interface was developed. The web application, named GeneAnalytics, is built on HTML with ajax and PHP helper functions to auto-fill gene identifiers and map gene symbols to probe set identifiers for quick input. Each function is associated with an individual web form that when

submitted, calls a Perl script via a Common Gateway Interface (CGI) to extract user parameters and prepare necessary file requirements. Perl then calls an R script to assemble necessary data, run the described analysis and generate the appropriate files and figures. All results are then presented back to the user with a link to download the files for permanent record keeping. It is also important to note that all figures are generated in PDF format to maintain high quality, high resolution images which can be adapted for publication. On average, each process takes about 10 seconds to run with exception to the functional co-expression analysis, which takes closer to a minute to finish. A sample screen shot can be seen in Figure 4-7.

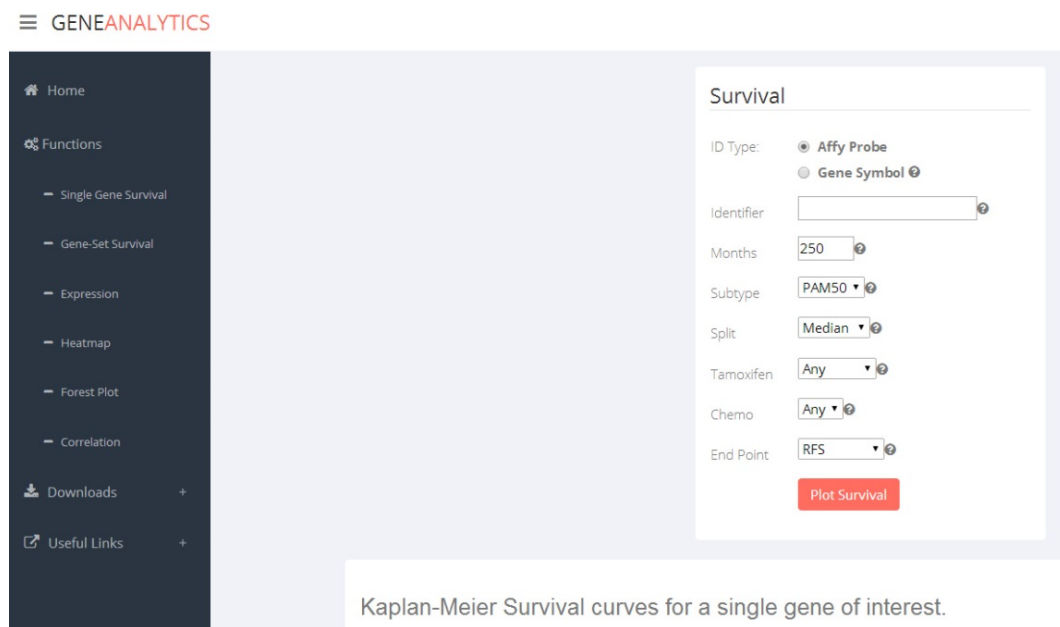


Figure 4-7: Screen shot of GeneAnalytics website.

A screen shot of the single gene survival function made available at GeneAnalytics.

4.3 Discussion

The present study makes use of the publicly available datasets to generate an integrative database of 4885 patients from 25 independent studies. Furthermore, analytical methods and functions were developed to efficiently use and apply the data. Access to the breast cancer meta-set and functions are made available to end users via a web interface.

Clinical data sets are abundant and their use in in discovery and validation contexts is well documented (Chang et al., 2011; Haibe-Kains et al., 2012; Magnani et al., 2011; Parker et al., 2009; Perou et al., 2000; Shi et al., 2010; Xiang et al., 2012). Nonetheless, use of gene expression data for these purposes remains a challenging task within the broad scientific community due to lack of bioinformatics training and the fact that no established frameworks exist. While researches can access gene expression relative to clinical features in websites such as OncoPrint (Rhodes et al., 2007), several limitations exist. The GeneAnalytics web application is the first to offer access to a breast cancer meta-set and offer high level analytical functions to apply the data in a meaningful way to the research setting.

In summary, the breast cancer meta-set and associated access through the GeneAnalytics web sites provides novel opportunities for researchers to integrate functional studies with tumor derived expression data to further our understanding of cancer related processes.

5 Conclusions and future directions

Despite initial success, resistance to both anti-estrogens and aromatase inhibitors is common, and these tumors continue to express ER. Furthermore, there is abundant evidence that ER transcriptional activity drives cell growth in endocrine resistance in the clinic. Likewise, the growth of TamR xenografts is blunted by ICI and the anti-estrogen bazedoxifene. The ER signaling program is plastic and can be reprogrammed in a cellular context by chromatin accessibility, epigenetics, coregulators and other transcriptional initiators including pioneer factors. The results of these studies have identified several transcription programs that are important to the development of tamoxifen resistance in breast cancer. In particular, FOXA1 binding is enriched in differentially sensitive DNase regions in the TamR model where it drives expression of key target genes whose actions are well-known in the pathogenesis of breast cancer.

The FOXA1 transcriptional program is related to ER signaling in that approximately 50% of its binding sites overlap with ER binding events. The increased role of FOXA1 in mediating the actions of ER is prominent in the TamR cell model and underlies resistance by establishing competent binding sites that allow ER to bind to novel regions and also to facilitate ER to engage chromatin in the absence of ligand at levels similar to ligand-induced activation. This phenomenon known as ligand-independent activation may require other upstream signaling mechanisms to modify ER, but the relevance of these studies is that FOXA1 obviates the need for ligands to

promote interactions of ER with coregulators to modify the chromatin environment. Efforts to understand the mechanism by which tumors adapt to SERDs, or AIs is an area that requires further discovery to understand the context to which they overlap or are divergent. From which, the FOXA1 adaptive program that promotes resistance may be better suited to a particular, SERM, SERD or AI. However we do have limited evidence that AIs would lead to a similar program from analysis of long-term estrogen deprived MCF7 cells that adopt a similar expression program as TamR cells. Future studies are necessary to identify whether FOXA1 alone is sufficient to support basal activities of ER, whether ER coregulators are necessary and to what extent other ER-directed therapies can inhibit ER at these FOXA1-dependent loci.

Our results also reveal a functional difference of FOXA1 activity in TamR cells compared to MCF7 cells and suggests that FOXA1 may be modified to co-opt a growth program. One potential means of increased activation may rest on the ability of FOXA1 to respond to upstream signaling cascades. Phosphorylation mapping of FOXA1 revealed a loss of phosphorylation at S331 in TamR cells and functionality of this site is currently being tested. In addition, our sequencing results identified the variant S448N, rs33984772, at a low frequency (~10%) in MCF7 cells, but TamR cells carry the variant allele at a higher frequency (~58%); suggestive of biological functionality of S448.

Another report identified FOXA1 to be acetylated by P300 and consequently diminishes DNA binding, an event which may be due to growth factor signaling (Katika et al.,

2013). Mutational studies of FOXA1 are ongoing to determine the functionality of these identified sites and the relevance to which they promote FOXA1 binding, stabilization and enhanced transcriptional activity.

There are many aspects of ER signaling and cross-regulatory actions of FOXA1 signaling program that remain to be fully elucidated. Although there has been success with anti-estrogens, an emerging paradigm in oncology field is focused on targeting co-dependent pathways to sustain inhibition of adaptive mechanisms that take over to reactivate a transcriptional program. In the context of breast cancer, a better understanding of the convergent pathways and adaptive mechanisms that allow ER to escape tamoxifen-mediated inhibition to maintain a central role in driving tumor growth is necessary. Targeting FOXA1 directly may prove beneficial in this regard but there is also a potential downside in that long term therapy may promote tumors to adopt an ER negative basal-like expression program, a tumor subtype that has unfavorable outcomes. Thus, inhibitors of kinases or key nodes of pathways upstream of FOXA1 may be useful. However, it is currently unclear what pathways and signaling cascades are necessary to mediate this switch. Therefore, screening platforms using small ligands, kinase inhibitors, or shRNA libraries may be able to identify these convergent pathways. The TamR model is perfectly suited for this case as it is the first observation of heightened FOXA1 activity in a model of resistance and allows for a differential screen of the TamR model compared to MCF7 cells to separate the negative and positive aspects of FOXA1.

For this purpose gene expression of TamR-specific FOXA1 regulated genes can be used as a measured outcome and end point in determining therapeutic utility.

The mass-spec results have revealed ~100 factors that preferentially bind in TamR cells. These factors further represent novel drug candidates to delay the FOXA1-dependent resistant mechanism. However, confirmation of these interactions in the context of cell signaling is necessary, and it will be important to ascertain their role in pathologies associated with ER and FOXA1.

Taken together, these studies advance our understanding of both ER and the adaptive mechanisms that promote resistance to anti-estrogens. This work has identified a novel role for FOXA1 in driving tamoxifen resistance by means of reactivation of ER pathways and ER-independent gene expression. Experimental evidence indicates an altered usage of FOXA1 localization, modification and DNA occupancy in a tamoxifen resistant model. Additional studies will be necessary to elucidate the positive and negative impacts of FOXA1 on biological processes and whether a FOXA1 therapeutic may be used to delay recurrence. A further understanding of the role of FOXA1 in progression to and maintenance of tamoxifen resistance will determine whether this essential factor can be effectively targeted to delay tumor progression and improve the efficacy of ER-targeted therapies.

6 Materials and Methods

6.1 Chemicals

ER ligands obtained from Sigma (St. Louis, MO) include: 17 β -Estradiol [50-28-2] (E8875), Fulvestrant (ICI) [129453-61-8] (I4409), and 4-hydroxytamoxifen [68047-06-3] (H7904). Cardiac glycosides used were also purchased from Sigma: Digoxin [20830-75-5] (D6003), Ouabain, [11018-89-6] (O3125), Lanatoside C [17575-22-3] (L2261), and Proscillaradin A [466-06-8] (P2428). Other inhibitors used include: AZD6244 [606143-52-6] from Selleckchem, U0126 [109511-58-2] from Sigma (U12), PD184161 [212631-67-9] from Cayman, LY294002 [934389-88-5] from Sigma (L908), and Farnesylthiosalicylic acid (FTS) [1092521-74-8] from SCBT (sc-223986). PCR and qPCR reagents were purchased from Bio-Rad (Hercules, CA), Qiagen (Valencia, CA), Integrated DNA Technologies (Coralville, IA), and Sigma.

6.2 Antibodies

The following antibodies were purchased from Santa Cruz Biotechnology: HOXB13 antibody (H-80): sc-66923, HOXB13 antibody (F-9): sc-28333, and FOXA1 antibody (C-20): sc-6553. Remaining antibodies were purchased from Abcam: FOXA1 antibody (ab23738), FOXA2 antibody (ab60721), STAT1 antibody (ab2415), STAT1 (phospho Y701) [M135] antibody (ab29045), STAT3 (phospho Y705) antibody [EP2147Y] (ab76315).

6.3 Plasmids

Expression plasmids for HOXB13 were previously generated in the lab (Norris et al., 2009b). Reporter construct for PS2-Luc (Lu et al., 2001), and the 3XERE-TATA-Luc (Hall and McDonnell, 1999) have been described previously. Luciferase reporters for pCOL (MMP1) were generated from constructs described elsewhere (Fan et al., 1996) and made as follows: pCOL-FL (F-atcgctcgagGAAGAGCCACCGTAAA, R-cgataagcttGCCTTTGTCTTCTTTC), pCOL-1:292 (F-atcgctcgagGAAGAGCCACCGTAAA, R-cgataagcttcaacaagatgtgtg), pCOL-292:380 (F-atcgctcgagaagttaatcatgacat, R-cgataagcttgctatgaatagactag), pCOL-380:452 (F-atcgctcgagtaatcaagaggatgtt R-cgataagcttatatagatgccttg), and pCOL-452:538 (F-atcgctcgagacagaggagcttct, R-cgataagcttGCCTTTGTCTTCTTTC). The 3x AP-1 reporter was cloned with the insert tcgagGTTGAGTCACGGTTGAGTCACGGTTGAGTCACGa. FOXA1 was cloned into pLenti_Puro using an ORF obtained from Open Biosystems, clone id:52698380. FOXA2 was also cloned into Plenti_Puro using a clone from PlasmID Database at Harvard, clone id: HsCD00044936.

6.4 Cell Culture

The MCF7, MARCO and TamR cell lines were maintained in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12). TamR cells were also kept under constant selection with 100nM 4-OHT. BT474 Cells were grown in RPMI 1640. HEPG2 cells and BT483 cells were grown in DMEM. All cell lines were supplemented

with 8% fetal bovine serum (FBS) or charcoal stripped FBS (CS-FBS) (Hyclone Laboratories, Logan, UT), 0.1 mM non-essential amino acids (NEAA) and 1 mM sodium pyruvate (NaPyr).

6.5 RNA Isolation and qPCR

Cells were seeded in 6-well plates in phenol red-free media containing 8% CS-FBS for 48 hours and treated with ligands as indicated. After the indicated time period, cells were harvested and total RNA was isolated using the Aurum™ Total RNA Mini Kit (Bio-Rad, Hercules, CA). One microgram of purified RNA was reverse transcribed using the iScript cDNA synthesis kit (Bio-Rad). The Bio-Rad CFX384 Touch Realtime PCR Detection System was used to amplify and quantitate levels of target gene cDNA. Reactions for qPCR were performed with 1 μ L cDNA, 10 μ M specific primers, and iQ SYBRGreen supermix (Bio-Rad). Data are normalized to RPLP0 (36B4) housekeeping gene and presented as fold expression relative to controls. Data is presented as the mean \pm SEM for triplicate amplification reactions from one representative experiment. Unless otherwise indicated, experiments were performed at least three independent times with coincident results. Human qPCR primer sequences are listed in Table 6-1.

Table 6-1: qPCR primers used.

<i>Identifier</i>	<i>Symbol</i>	<i>Primer</i>	<i>Sequence</i>
NM_005409	CXCL11	CXCL11-F	ACTCCTTCC AAGAAGAGCAGCA
NM_005409	CXCL11	CXCL11-R	CCATGCCCTTCACACTCATGTT
NM_002416	CXCL9	CXCL9-F	GTAGTGAGAAAGGGTCGCTGT

NM_002416	CXCL9	CXCL9-R	AGGGCTTGGGGCAAATTGTT
NM_001565	CXCL10	CXCL10-F	GTGGCATTCAAGGAGTACCTC
NM_001565	CXCL10	CXCL10-R	TGATGGCCTTCGATTCTGGATT
NM_001504	CXCR3A	CXCR3A-F	ACCCAGCAGCCAGAGCACC
NM_001504	CXCR3A	CXCR3A-R	TCATAGGAAGAGCTGAAGTTCTCCA
NM_001142797	CXCR3B	CXCR3B-F	TGCCAGGCCTTTACACAGC
NM_001142797	CXCR3B	CXCR3B-R	TCGGCGTCATTTAGCACTTG
NM_002988	CCL18	CCL18-F	CTCCTTGTCTCGTCTGCAC
NM_002988	CCL18	CCL18-R	TCAGGCATTAGCTTCAGGT
NM_024013	IFNA2	IFNA2-F	CAGAGTCACCCATCTCAGCA
NM_024013	IFNA2	IFNA2-R	CACCACCAGGACCATCAGTA
NM_000619	IFNG	IFNG-F	TCAGCTCTGCATCGTTTTGG
NM_000619	IFNG	IFNG-R	GTTCCATTATCCGCTACATCTGAA
NM_002051	GATA3	GATA3-F	GCGGGCTCTATCACAAAATGA
NM_002051	GATA3	GATA3-R	GCTCTCCTGGCTGCAGACAGC
NM_006361	HOXB13	HOXB13-F	GTGCTGCCCGCTGGAGTC
NM_006361	HOXB13	HOXB13-R	AGTTACCTGGACGTGTCTGTGG
NM_001145938	MMP1	MMP1-F	GGTCTCTGAGGGTCAAGCAG
NM_001145938	MMP1	MMP1-R	CCAGGTCCATCAAAGGAGA
NM_004217	AURKB	AURKB_480	ATCAGCTGCGCAGAGAGATC
NM_004217	AURKB	AURKB_629	AGCTCTTCTGCAGCTCCTTG
NM_003225	TFF1	TFF1_41	TGGCCACCATGGAGAACAAG
NM_003225	TFF1	TFF1_180	CGTGACACCAGGAAAACCAC
NM_016095	GINS2	GINS2_460	CCAAACTCCGAGTGTCTGCT
NM_016095	GINS2	GINS2_588	TGTACATGTGGTTGAGCGCT
NM_012259	HEY2	HEY2_499	TCAGGCAACAGGGGGTAAAG
NM_012259	HEY2	HEY2_593	GCGCAACTTCTGTTAGGCAC
NM_012391	SPDEF	SPDEF_1079	AAAGAGCGGACTTCACCTGG
NM_012391	SPDEF	SPDEF_1201	CTTGAGGAACTGCCACAGGT
NM_153321	PMP22	PMP22_279	CGCAACTGATCTCTGGCAGA
NM_153321	PMP22	PMP22_395	ATCGACAGGATCATGGTGCC
NM_004316	ASCL1	ASCL1_1151	CCCCAACTACTCCAACGAC
NM_004316	ASCL1	ASCL1_1267	TGAAGTCGAGAAGCTCCTGC
NM_002275	KRT15	KRT15_830	CGAGTCCTGGATGAGCTGAC
NM_002275	KRT15	KRT15_952	CTGGCTGCTGAACTCCTTCA
NM_021810	CDH26	CDH26_434	CCAGATGCCACAATGCACAG
NM_021810	CDH26	CDH26_570	CCTTCTCGCTGTAGACGTG
NM_181803	UBE2C	UBE2C_246	CGAGCTCTGGAAAAACCCCA
NM_181803	UBE2C	UBE2C_360	AAGACGACACAAGGACAGGC
NM_004316	ASCL1	ASCL1_1563	GAGCAACTGGGACCTGAGTC

NM_004316	ASCL1	ASCL1_1685	TCAGCTGTGCGTGTTAGAGG
NM_024915	GRHL2	GRHL2_496	GTGATGAGGACAGTGCTGCT
NM_024915	GRHL2	GRHL2_596	TTGGCTGTCACTTGCTTTGC
NM_001032280	TFAP2A	TFAP2A_1724	ACCGACAACATTCCGATCCC
NM_001032280	TFAP2A	TFAP2A_1815	CAGCAGGTCGGTGAACCTCTT
NM_003221	TFAP2B	TFAP2B_174	ACCTCCTAGAGACCAGGCTG
NM_003221	TFAP2B	TFAP2B_323	CTCGAGTAGGGTCCTTGGGA
NM_003222	TFAP2C	TFAP2C_993	CTGTCCCCACCTGAATGCTT
NM_003222	TFAP2C	TFAP2C_1095	TCTTGTCCAATTCTCCCCGC
uc010tpz.2	FOXA1_1	FoxA1_Iso1_125	AAAACGCGTATTGGAACCTGC
uc010tpz.2	FOXA1_1	FoxA1_Iso1_226	GCCTGAGTTCATGTTGCTGA
uc001wuf.3	FoxA1_2	FoxA1_Iso2_327	GAAGATGGAAGGGCATGAAA
uc001wuf.3	FoxA1_2	FoxA1_Iso2_423	GCCTGAGTTCATGTTGCTGA

6.6 Transient Transfection Assays

For reporter gene assays, cells were seeded in 96-well plates in phenol red-free MEM containing 8% charcoal-stripped fetal bovine serum (CS-FBS) (Hyclone, Logan, UT), 0.1 mM NEAA and 1 mM NaPyr 24 h before transfection. DNA was introduced into the cells using Lipofectin (Invitrogen) or Fugene 6 (Promega) -mediated transfection as described by the manufacturer. Briefly, triplicate transfections were performed using 500ng of total DNA; within each experiment, the total amount of DNA used to transfect each plate was kept constant by addition of the corresponding empty expression vector DNA lacking a cDNA insert. Cells were incubated with the DNA-transfection mixture for 24 h. Next, the transfection mix was replaced with fresh media containing the appropriate ligands. Following overnight treatment, luciferase and β -galactosidase (β -gal) activities were assayed on a Fusion Alpha-FP HT Universal Microplate Reader

(Perkin Elmer, Danvers Grove, IL). Results are expressed as relative luciferase activity (normalized to β -gal for transfection efficiency) for one representative experiment performed in triplicate, and error bars indicate the standard error of the mean (SEM) for the triplicate wells.

6.7 Viral Stable Cell Generation

Adenoviruses expressing β -gal, ER and GATA3 were generated using the ViraPower Adenoviral Expression System (Invitrogen) and were amplified and purified by CsCl₂ centrifugation.

For retrovirus generation, 293TS cells were plated at 2E6 in a 10cm dish in DMEM 8% FBS. The next day 8ug of target vector, 8ug of PCL10A1 was mixed with 600 ul OPTI-MEM for 5 minutes and 36 ul of Fugene 6 was added to the tube. After 30 minutes, transfection mix was added dropwise to the cells using a 1ml pipette. The following day, media was replaced with DMEM containing 2mM caffeine. 72 hours after transfection, virus was harvested by filtration through a .45uM filter and polybrene was added at 8ug/ml prior to infecting target cells. MCF7 cells were selected at 1ug/ml of Puromycin and BT474 cells were selected at .5ug/ml.

Lentiviral constructs were made from PLenti_Puro (Campeau et al., 2009) and transcript cDNA was cloned using LR Clonase II. For viral generation, 293FT cells were plated at 3E6 in a 10cm dish in DMEM 8% FBS. The next day 6ug of target vector, 4.5ug of dr8.2 dVPR, and 1.5ug VSVG was mixed with 600ul OPTI-MEM for 5 minutes and

36ul of Fugene 6 was added to the tube. After 30 minutes, transfection mix was added dropwise to the cells using a 1ml pipette. The following day, media was replaced with DMEM containing 2mM caffeine. 72 hours after transfection, virus was harvested by filtration through a .45uM filter and polybrene was added at 8ug/ml prior to infecting target cells. MCF7 cells were selected at 1ug/ml of Puromycin and BT474 cells were selected at .5ug/ml.

6.8 RNA-Seq

Two independent biological replicates for each cell line treated with Vehicle or 4-OHT were harvested as described above for RNA. Libraries were generated using Illumina TruSeq Stranded mRNA kit according to manufacturer protocols. Libraries were submitted to 100bp PE sequencing on Illumina HiSeq. FASTQ files were pre-processed to remove adapters, and low-quality 3' reads and aligned to the hg19 genome. Resulting BAM files were filtered to remove multi-mapped reads and rRNA reads.

Reads were quantified using easyRNASeq (Delhomme et al., 2012) over ensemble transcripts in R and genes with more than 2cpm in at least two conditions were brought forward to edgeR(Robinson et al., 2010) for differential expression. A multi-factorial design was incorporated to account for resistance and sensitive cells in each treatment state. TMM normalization was used and genes with $p < .01$ were considered differentially expressed for subsequent analyses.

Pathway analyses was accomplished with GAGE(Luo et al., 2009) and GSEA(Subramanian et al., 2005). GSEA results were further processed using the “Enrichment Map” plugin.

6.9 siRNA Experiments

For experiments involving transient transfection of small interfering RNA (siRNA), validated Stealth siRNA or siRNA control were obtained from Invitrogen (see Table 5.3 for siRNA sequences). Cells were plated in phenol red-free DMEM containing 8% CS-FBS, 0.1 mM NEAA and 1 mM NaPyr in the presence of 40 nM siRNA or siRNA control using DharmaFECT-1 (Dharmacon, Lafayette, CO) as the transfection agent according to the manufacturer's recommendations. After 48 h of knockdown, cells were serum-starved in phenol red-free DMEM containing 0.1% CS-FBS, 0.1 mM NEAA and 1 mM NaPyr for 24 h, then treated with the appropriate ligand and harvested for qPCR analysis as detailed above.

Table 6-2: siRNA sequences.

<i>Company</i>	<i>Reference</i>	<i>Name</i>	<i>Sequence</i>
Sigma	SASI_Hs01_00168403	siFOXA1_1	CACACAAACCAAACCGUCA
Sigma	SASI_Hs01_00168404	siFOXA1_2	CGUACUACCAAGGUGUGUA
Sigma	SASI_Hs01_00168405	siFOXA1_3	CGCCUUACGGCUCUACGUU
Invitrogen	FOXA1HSS104878	siFOXA1_A	GAACAGGCACUGCAAUACUCGCCUU
Invitrogen	FOXA1HSS104880	siFOXA1_B	CAUGAAACCAGCGACUGGAACAGCU
Invitrogen	FOXA1HSS179280	siFOXA1_C	CAGCAUAAGCUGGACUUCAAGGCAU
Invitrogen	FOXA2HSS142471	siFOXA2_A	GCCGUCCGACUGGAGCAGCUACUAU
Invitrogen	FOXA2HSS142473	siFOXA2_B	GCGGGCUCCAUGAACAUGUCGUCGU

Invitrogen	FOXA2HSS179282	siFOXA2_C	CACCCUGACUCGGGCAACAUGUUCG
-------------------	----------------	-----------	---------------------------

6.10 Western Blotting

Cells were seeded in 6-well plates in phenol red-free DMEM containing 8% CS-FBS, 0.1 mM NEAA and 1 mM NaPyr for 48 h and treated as indicated. Following treatment for the indicated time periods, cells were harvested in ice-cold PBS and lysed in RIPA Buffer [50 mM Tris-HCl pH 8.0, 200 mM NaCl, 1.5 mM MgCl₂, 1% Triton X-100, 1 mM EDTA, 10% glycerol, 50 mM NaF, 2 mM Na₃VO₄, and 1X protease inhibitor mixture (EMD Chemicals, Inc, San Diego, CA)] while rotating at 4°C for 30 min. 20 µg of whole-cell extract was resolved by SDS-PAGE, transferred to a PVDF membrane (Bio-Rad) and probed with the appropriate antibodies.

6.11 Chromatin Immunoprecipitation

Cells were seeded in a 15cm dishes with appropriate media described above. Cells were grown to 90% confluence in phenol red-free media supplemented with 8% CS-FBS, 0.1 mM NEAA and 1 mM NaPyr for 48 h, after which the cells were treated as indicated. Following treatment with the appropriate ligand for the indicated time periods, cells were subjected to ChIP analysis. ChIP was performed as described elsewhere (DuSell et al., 2008) with modifications. Each plate of cells were cross-linked with 540ul of 37% formaldehyde (1% final concentration) for 10 minutes exactly and quenched with 2ml of 2.5M glycine (250mM final) for 5 minutes. Cells were then

harvested with cold PBS and snap frozen for storage at -80 degrees. qPCR analysis was performed in 384-well format with 4ul reactions and data is normalized to the input for the immunoprecipitation. ChIP primer sequences are listed in Table 6-2.

Table 6-3: ChIP primers used.

<i>Name</i>	<i>Chr</i>	<i>Start</i>	<i>Stop</i>
TFF1-1	chr21	43786717	43786865
S100A9-1	chr1	1.53E+08	1.53E+08
S100A9-2	chr1	1.53E+08	1.53E+08
S100A9-3	chr1	1.53E+08	1.53E+08
CAV1-1	chr7	1.16E+08	1.16E+08
CAV1-2	chr7	1.16E+08	1.16E+08
CAV1-3	chr7	1.16E+08	1.16E+08
WISP1-1	chr8	1.34E+08	1.34E+08
WISP1-2	chr8	1.34E+08	1.34E+08
WISP1-3	chr8	1.34E+08	1.34E+08
AGR2-1	chr7	16825998	16826146
AGR2-2	chr7	16844873	16845021
AGR2-3	chr7	16851302	16851450
XBP-1	chr22	29209832	29209980
XBP-2	chr22	29215585	29215733
XBP-3	chr22	29218960	29219108
KRT13-1	chr17	39571364	39572125
KRT13-2	chr17	39619973	39621791
KRT13-3	chr17	39811719	39812650
ASCL1-1	chr12	1.03E+08	1.03E+08
SPDEF-1	chr6	34494339	34494461
SPDEF-2	chr6	34524091	34524239
SPDEF-3	chr6	34566308	34566935

6.12 Preparation of Nuclear Extracts for Mass-Spec

Antibody (60ug) was incubated with 90ul Protein A/G beads (Pierce #20421) in PBS overnight. The next morning beads were washed three times in .2M sodium borate,

pH 9.0. Complexed beads were conjugated with dimethylpimelimidate (DMP) with .0259g DMP and 5ml of .2M sodium borate to make a 20mM solution. Beads were incubated for 40min with end over end rocking at room temperature. After, beads were washed in .2M ethanolamine (pH 8.0) to quench residual DMP and then suspended in .2M ethanolamine for additional 1 hour incubation. Uncoupled antibody was then washed three times using .58% acetic acid with 150mM NaCl. Beads were stored in PBS with sodium azide at 4 degrees.

Nuclear extracts of MCF7 and TamR cells was performed for Mass-Spec analysis. Briefly, fifteen 10cm plates were washed with PBS and harvested with .25% trypsin. Cells were scraped into a conical tube and spun for 5 min at 1500g in a pre-cooled centrifuge. Cell pellets were suspended in 5 times the cell pellet volume in hypotonic buffer for 5 min and thereafter checked every minute with trypan blue until greater than 90% of cells stained positive. NP40 was added to .1% (10ul/ml of 10% stock) and vortexed on mid setting for 10 seconds and immediately centrifuged at 3000g for 1 minute. Resulting supernatant, predominately consisting of cytosolic extract was set aside and nuclear pellet was suspended in a half a cell pellet volume of low salt buffer [20mM HEPES pH 7.9, .02M KCl, .2mM EDTA, 25% Glycerol 1mM DTT, beta-glycerophosphate, protease inhibitors, NaF, NaV, and NaB] being careful not to break nuclei. A half bed volume of high salt buffer [20mM HEPES pH 7.9, 1M KCl, .2mM EDTA, 25% Glycerol 1mM DTT, beta-glycerophosphate, protease inhibitors, NaF, NaV,

and NaB] was added gently and tube was rocked for 1 hour in cold room. Nuclear debris was pelleted at 14000g for 15 min and supernatant was dialyzed against dialysis buffer [20mM HEPES pH 7.9, 100mM KCl, .2mM EDTA, 20% Glycerol] to normalize salts for two changes at 1 hour each.

Nuclear extracts were normalized at 10mg each and pre-cleared with A/G beads for 1 hour and then incubated overnight with the prepared conjugated beads. The next day beads were washed 5 times for minutes each with wash buffer [50mM HEPES pH 7.9, .2% NP40, 150mM KCl, 1mM EGTA, 1mM DTT, beta-glycerophosphate, protease inhibitors, NaF, NaV, and NaB]. Final washes were performed using PBS three times and 50mM Ammonium Bicarbonate twice before submitting the Duke Proteomics core.

6.13 DNase-Seq Library Generation

DNase-Seq was performed as previously described(Song and Crawford, 2010; Song et al., 2011b). Briefly, cells were plated in CS-FBS in 15cm plates and treated 48 hours later with either vehicle or 4-OHT for 24 hours before harvesting. Two independent biological replicates of each condition were prepared. Nuclei were extracted and digested with optimal concentrations of DNaseI enzyme. After confirmation of adequate digestion, DNaseI-digested ends were blunt ended, and a biotinylated linker was ligated to these ends. Fragments with linker attached were isolated, digested with MmeI, and captured using streptavidin-conjugated magnetic beads. A second linker was ligated to the MmeI-digested end, and then the fragments

were amplified and subsequently purified via gel electrophoresis. The libraries were sequenced using 50bp SR on Illumina HiSeq.

6.14 DNase-Seq Analysis

FASTQ files were aligned to the human female genome hg19 from UCSC using BWA (Li and Durbin 2009). Reads were filtered from the SAM file that do not align based on the 0X004 flag, align to multiple locations, align to more than two ambiguous location, and fall of chromosome boundaries using the chrom.sizes file from UCSC. Additional alignments were also filtered to remove problematic repetitive regions such as alpha satellites and sequence artifacts as defined by ENCODE. Biological replicates were compared for reproducibility and correlation. Final base-pair resolution signal as a Wig file was generated using F-Seq at 300bp signal bandwidth (Boyle et al., 2008b) and converted to bigwig using the UCSC utility, WigtoBigWig . Peaks were called by F-Seq and significance of the peaks were determined by fitting DNase-Seq signal data to a gamma distribution and then determining the signal value that corresponded to a p-value < 0.05.

Sequencing tags were quantified in each DNase peak for each condition using multicov from the bedtools suite(Quinlan and Hall, 2010). Reads overlapping within 500bp each direction from the TSS were also subset to discriminate against sites likely unrelated to TF binding using the Refseq annotated genes. To identify regions of significant change across cell lines and treatments in DNase-Seq data, we used the

bioconductor package, edgeR (Robinson et al., 2010) and DESeq(Anders and Huber, 2010). These packages use a negative binomial distribution to determine differential expression across treatment groups by fitting a generalized linear model. Both approaches yielded very similar results with the DNase data.

6.15 Motif Analysis

To determine motif enrichment adjacent to differential DHS sites we utilized used three distinct algorithms: MEME-ChIP(Machanick and Bailey, 2011), and CentDist (Zhang et al., 2011), and HOMER(Heinz et al., 2010). De novo motif enrichment analysis was performed using MEME-ChIP which combines the MEME suite: MEME and DREME for motif discovery, Centrimo for central enrichment and TomTom for motif matching. Notable parameters (-meme-p 6 -meme-nmotifs 20 -meme-mod zoops -meme-minw 8 centrimo-local). A markov model was determined from the supplied differential sequences and additionally with a background defined from the union DHS sites. HOMER was run as both a de novo motif finding algorithm and separately against all known PWMs defined in the program. Background models were defined from differential DHS sites and again using union DHS sites to further discriminate the list. CentDist was run from the website with default parameters. Reported motifs were considered concordant if found by at least two approaches and discordant if found only by one of the methods.

References

1. (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931-945.
2. Ademuyiwa, F.O., Thorat, M.A., Jain, R.K., Nakshatri, H., and Badve, S. (2010). Expression of Forkhead-box protein A1, a marker of luminal A type breast cancer, parallels low Oncotype DX 21-gene recurrence scores. *Modern Pathology* 23, 270-275.
3. Allen, E., and Doisy, E.A. (1923). An ovarian hormone: Preliminary report on its localization, extraction and partial purification, and action in test animals. *Journal of the American Medical Association* 81, 819-821.
4. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* 11, R106.
5. Antonelli, A., Ferrari, S.M., Fallahi, P., Ghiri, E., Crescioli, C., Romagnani, P., Vitti, P., Serio, M., and Ferrannini, E. (2010). Interferon-alpha, beta and gamma induce CXCL9 and CXCL10 secretion by human thyrocytes: modulation by peroxisome proliferator-activated receptor-gamma agonists. *Cytokine* 50, 260-267.
6. Arpino, G., Gutierrez, C., Weiss, H., Rimawi, M., Massarweh, S., Bharwani, L., De Placido, S., Osborne, C.K., and Schiff, R. (2007). Treatment of Human Epidermal Growth Factor Receptor 2-Overexpressing Breast Cancer Xenografts With Multiagent HER-Targeted Therapy. *Journal of the National Cancer Institute* 99, 694-705.
7. Badve, S., Turbin, D., Thorat, M.A., Morimiya, A., Nielsen, T.O., Perou, C.M., Dunn, S., Huntsman, D.G., and Nakshatri, H. (2007). FOXA1 expression in breast cancer—correlation with luminal subtype A and survival. *Clinical cancer research* 13, 4415-4421.

8. Bain, D.L., Heneghan, A.F., Connaghan-Jones, K.D., and Miura, M.T. (2007). Nuclear receptor structure: implications for function. *Annu. Rev. Physiol.* 69, 201-220.
9. Beato, M., and Klug, J. (2000). Steroid hormone receptors: an update. *Human reproduction update* 6, 225-236.
10. Beatson, G. (1896). ON THE TREATMENT OF INOPERABLE CASES OF CARCINOMA OF THE MAMMA: SUGGESTIONS FOR A NEW METHOD OF TREATMENT, WITH ILLUSTRATIVE CASES.< sup> 1</sup>. *The Lancet* 148, 104-107.
11. Beatson, G.T. (1897). Extirpation of the Ovaries as a Cure for Cancer. *British medical journal* 2, 54.
12. Beatson, G.T. (1899). On the Etiology of Cancer, with a Note of some Experiments. *British medical journal* 1, 399.
13. Benito, M., Parker, J., Du, Q., Wu, J., Xiang, D., Perou, C.M., and Marron, J.S. (2004). Adjustment of systematic microarray data biases. *Bioinformatics* 20, 105-114.
14. Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., and Chan, E.T. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 133, 1266-1276.
15. Bernardo, G.M., Bebek, G., Ginther, C.L., Sizemore, S.T., Lozada, K.L., Miedler, J.D., Anderson, L.A., Godwin, A.K., Abdul-Karim, F.W., Slamon, D.J., *et al.* (2013). FOXA1 represses the molecular phenotype of basal breast cancer cells. *Oncogene* 32, 554-563.
16. Bochkis, I.M., Schug, J., Ye, D.Z., Kurinna, S., Stratton, S.A., Barton, M.C., and Kaestner, K.H. (2012). Genome-Wide Location Analysis Reveals Distinct

Transcriptional Circuitry by Paralogous Regulators Foxa1 and Foxa2. *PLoS Genet* 8, e1002770.

17. Boyd, S. (1902). Oöphorectomy In Cancer Of The Breast. *British medical journal* 1, 110.
18. Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008a). High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132, 311-322.
19. Boyle, A.P., Guinney, J., Crawford, G.E., and Furey, T.S. (2008b). F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics* 24, 2537-2538.
20. Boyle, A.P., Song, L., Lee, B.K., London, D., Keefe, D., Birney, E., Iyer, V.R., Crawford, G.E., and Furey, T.S. (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res* 21, 456-464.
21. Brinckerhoff, C.E., Mcmillan, R.M., Fahey, J.V., and Harris, E.D. (1979). Collagenase production by synovial fibroblasts treated with phorbol myristate acetate. *Arthritis & Rheumatism* 22, 1109-1116.
22. Burns, K., and Korach, K. (2012). Estrogen receptors and human disease: an update. *Arch Toxicol* 86, 1491-1504.
23. Campeau, E., Ruhl, V.E., Rodier, F., Smith, C.L., Rahmberg, B.L., Fuss, J.O., Campisi, J., Yaswen, P., Cooper, P.K., and Kaufman, P.D. (2009). A versatile viral system for expression and depletion of proteins in mammalian cells. *PloS one* 4, e6529.
24. Canello, G., Maisonneuve, P., Rotmensch, N., Viale, G., Mastropasqua, M.G., Pruneri, G., Montagna, E., Iorfida, M., Mazza, M., Balduzzi, A., *et al.* (2013). Progesterone receptor loss identifies Luminal B breast cancer subgroups at higher risk of relapse. *Annals of Oncology* 24, 661-668.

25. Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., and Geistlinger, T.R. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122, 33-43.
26. Carter, S.L., Brechbühler, C.M., Griffin, M., and Bond, A.T. (2004). Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* 20, 2242-2250.
27. Chang, C.-y., Kazmin, D., Jasper, J.S., Kunder, R., Zuercher, W.J., and McDonnell, D.P. (2011). The metabolic regulator $ERR\alpha$, a downstream target of HER2/IGF-1R, as a therapeutic target in breast cancer. *Cancer cell* 20, 500-510.
28. Chang, C.-y., Norris, J.D., Grøn, H., Paige, L.A., Hamilton, P.T., Kenan, D.J., Fowlkes, D., and McDonnell, D.P. (1999). Dissection of the LXXLL Nuclear Receptor-Coactivator Interaction Motif Using Combinatorial Peptide Libraries: Discovery of Peptide Antagonists of Estrogen Receptors α and β . *Molecular and Cellular Biology* 19, 8226-8239.
29. Chaya, D., Hayamizu, T., Bustin, M., and Zaret, K.S. (2001). Transcription factor FoxA (HNF3) on a nucleosome at an enhancer complex in liver chromatin. *Journal of Biological Chemistry* 276, 44385-44389.
30. Cheang, M.C., Chia, S.K., Voduc, D., Gao, D., Leung, S., Snider, J., Watson, M., Davies, S., Bernard, P.S., and Parker, J.S. (2009). Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *Journal of the National Cancer Institute* 101, 736-750.
31. Chen, M., Cui, Y.-K., Huang, W.-H., Man, K., and Zhang, G.-J. (2013). Phosphorylation of estrogen receptor α at serine 118 is correlated with breast cancer resistance to tamoxifen. *Oncology letters* 6, 118-124.
32. Chia, S., Gradishar, W., Mauriac, L., Bines, J., Amant, F., Federico, M., Fein, L., Romieu, G., Buzdar, A., and Robertson, J.F. (2008). Double-blind, randomized placebo controlled trial of fulvestrant compared with exemestane after prior nonsteroidal aromatase inhibitor therapy in postmenopausal women with

hormone receptor–positive, advanced breast cancer: results from EFACT. *Journal of clinical oncology* 26, 1664-1670.

33. Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular cell* 9, 279-289.
34. Cirillo, L.A., and Zaret, K.S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Molecular cell* 4, 961-969.
35. Cockerill, P.N. (2011). Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J* 278, 2182-2210.
36. Connor, C.E., Norris, J.D., Broadwater, G., Willson, T.M., Gottardis, M.M., Dewhirst, M.W., and McDonnell, D.P. (2001). Circumventing tamoxifen resistance in breast cancers using antiestrogens that induce unique conformational changes in the estrogen receptor. *Cancer research* 61, 2917-2922.
37. Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.
38. Crawford, G.E., Davis, S., Scacheri, P.C., Renaud, G., Halawi, M.J., Erdos, M.R., Green, R., Meltzer, P.S., Wolfsberg, T.G., and Collins, F.S. (2006). DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat Methods* 3, 503-509.
39. Cui, X., Schiff, R., Arpino, G., Osborne, C.K., and Lee, A.V. (2005). Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy. *Journal of Clinical Oncology* 23, 7721-7735.
40. Delhomme, N., Padiou, I., Furlong, E.E., and Steinmetz, L.M. (2012). easyRNASeq: a bioconductor package for processing RNA-Seq data. *Bioinformatics* 28, 2532-2533.

41. Delmas, P.D., Bjarnason, N.H., Mitlak, B.H., Ravoux, A.-C., Shah, A.S., Huster, W.J., Draper, M., and Christiansen, C. (1997). Effects of raloxifene on bone mineral density, serum cholesterol concentrations, and uterine endometrium in postmenopausal women. *New England Journal of Medicine* 337, 1641-1647.
42. Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., Delorenzi, M., Piccart, M., and Sotiriou, C. (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clinical Cancer Research* 14, 5158-5165.
43. Dodwell, D., Wardley, A., and Johnston, S. (2006). Postmenopausal advanced breast cancer: Options for therapy after tamoxifen and aromatase inhibitors. *The Breast* 15, 584-594.
44. Dupont, S., Krust, A., Gansmuller, A., Dierich, A., Chambon, P., and Mark, M. (2000). Effect of single and compound knockouts of estrogen receptors alpha (ERalpha) and beta (ERbeta) on mouse reproductive phenotypes. *Development* 127, 4277-4291.
45. DuSell, C.D., Umetani, M., Shaul, P.W., Mangelsdorf, D.J., and McDonnell, D.P. (2008). 27-hydroxycholesterol is an endogenous selective estrogen receptor modulator. *Mol Endocrinol* 22, 65-77.
46. Early Breast Cancer Trialists' Collaborative, G. (2005). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 365, 1687-1717.
47. Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research* 30, 207-210.
48. Eeckhoute, J., Carroll, J.S., Geistlinger, T.R., Torres-Arzayus, M.I., and Brown, M. (2006). A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. *Genes & development* 20, 2513-2526.

49. Eeckhoute, J., Keeton, E.K., Lupien, M., Krum, S.A., Carroll, J.S., and Brown, M. (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor α expression in breast cancer. *Cancer Research* 67, 6477-6483.
50. Eeckhoute, J., Lupien, M., Meyer, C.A., Verzi, M.P., Shivdasani, R.A., Liu, X.S., and Brown, M. (2009). Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome research* 19, 372-380.
51. Egesten, A., Eliasson, M., Olin, A.I., Erjefält, J.S., Bjartell, A., Sangfelt, P., and Carlson, M. (2007). The proinflammatory CXC-chemokines GRO- α /CXCL1 and MIG/CXCL9 are concomitantly expressed in ulcerative colitis and decrease during treatment with topical corticosteroids. *International journal of colorectal disease* 22, 1421-1427.
52. Fan, J.D., Wagner, B.L., and McDonnell, D.P. (1996). Identification of the sequences within the human complement 3 promoter required for estrogen responsiveness provides insight into the mechanism of tamoxifen mixed agonist activity. *Molecular Endocrinology* 10, 1605-1616.
53. Fu, J., Bian, M., Jiang, Q., and Zhang, C. (2007). Roles of Aurora kinases in mitosis and tumorigenesis. *Molecular Cancer Research* 5, 1-10.
54. Galas, D.J., and Schmitz, A. (1978). DNAase footprinting a simple method for the detection of protein-DNA binding specificity. *Nucleic acids research* 5, 3157-3170.
55. Gao, N., LeLay, J., Vatamaniuk, M.Z., Rieck, S., Friedman, J.R., and Kaestner, K.H. (2008). Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes & development* 22, 3435-3448.
56. Gertz, J., Savic, D., Varley, Katherine E., Partridge, E.C., Safi, A., Jain, P., Cooper, Gregory M., Reddy, Timothy E., Crawford, Gregory E., and Myers, Richard M. (2013). Distinct Properties of Cell-Type-Specific and Shared Transcription Factor Binding Sites. *Molecular Cell* 52, 25-36.

57. Glascock, R., and Hoekstra, W. (1959). Selective accumulation of tritium-labelled hexoestrol by the reproductive organs of immature female goats and sheep. *Biochemical Journal* 72, 673.
58. Goetz, M.P., Suman, V.J., Ingle, J.N., Nibbe, A.M., Visscher, D.W., Reynolds, C.A., Lingle, W.L., Erlander, M., Ma, X.-J., and Sgroi, D.C. (2006). A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. *Clinical cancer research* 12, 2080-2087.
59. Goldhirsch, A., Winer, E.P., Coates, A.S., Gelber, R.D., Piccart-Gebhart, M., Thürlimann, B., Senn, H.-J., and members, P. (2013). Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of Oncology* 24, 2206-2223.
60. Goldhirsch, A., Wood, W., Coates, A., Gelber, R., Thürlimann, B., and Senn, H.-J. (2011). Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of Oncology* 22, 1736-1747.
61. Greene, G.L., Nolan, C., Engler, J.P., and Jensen, E.V. (1980). Monoclonal antibodies to human estrogen receptor. *Proceedings of the National Academy of Sciences* 77, 5115-5119.
62. Grober, O.M.V., Mutarelli, M., Giurato, G., Ravo, M., Cicatiello, L., De Filippo, M.R., Ferraro, L., Nassa, G., Papa, M.F., and Paris, O. (2011). Global analysis of estrogen receptor beta binding to breast cancer cell genome reveals an extensive interplay with estrogen receptor alpha for target gene regulation. *BMC genomics* 12, 36.
63. Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* 57, 159-197.

64. Gruber, C.J., Gruber, D.M., Gruber, I.M.L., Wieser, F., and Huber, J.C. (2004). Anatomy of the estrogen response element. *Trends in endocrinology & metabolism* 15, 73-78.
65. Haas, M., Wang, H., Tian, J., and Xie, Z. (2002). Src-mediated inter-receptor cross-talk between the Na⁺/K⁺-ATPase and the epidermal growth factor receptor relays the signal from ouabain to mitogen-activated protein kinases. *Journal of Biological Chemistry* 277, 18694-18702.
66. Habel, L.A., Sakoda, L.C., Achacoso, N., Ma, X.-J., Erlander, M.G., Sgroi, D.C., Fehrenbacher, L., Greenberg, D., and Quesenberry Jr, C.P. (2013). HOXB13: IL17BR and molecular grade index and risk of breast cancer death among patients with lymph node-negative invasive disease. *Breast Cancer Research* 15, R24.
67. Haibe-Kains, B., Desmedt, C., Loi, S., Culhane, A.C., Bontempi, G., Quackenbush, J., and Sotiriou, C. (2012). A three-gene model to robustly identify breast cancer molecular subtypes. *Journal of the National Cancer Institute* 104, 311-325.
68. Halachmi, S., Marden, E., Martin, G., MacKay, H., Abbondanza, C., and Brown, M. (1994). Estrogen receptor-associated proteins: possible mediators of hormone-induced transcription. *Science* 264, 1455-1458.
69. Hall, J.M., and McDonnell, D.P. (1999). The Estrogen Receptor β -Isoform (ER β) of the Human Estrogen Receptor Modulates ER α Transcriptional Activity and Is a Key Regulator of the Cellular Response to Estrogens and Antiestrogens 1. *Endocrinology* 140, 5566-5578.
70. Harrell, J.C., Dye, W.W., Allred, D.C., Jedlicka, P., Spoelstra, N.S., Sartorius, C.A., and Horwitz, K.B. (2006). Estrogen Receptor Positive Breast Cancer Metastasis: Altered Hormonal Sensitivity and Tumor Aggressiveness in Lymphatic Vessels and Lymph Nodes. *Cancer Research* 66, 9308-9315.
71. Haux, J. (1999). Digitoxin is a potential anticancer agent for several types of cancer. *Medical hypotheses* 53, 543-548.

72. He, H.H., Meyer, C.A., Chen, M.W., Jordan, V.C., Brown, M., and Liu, X.S. (2012). Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. *Genome Res.*
73. Hebbes, T.R., Thorne, A.W., and Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. *The EMBO journal* 7, 1395.
74. Heery, D.M., Kalkhoven, E., Hoare, S., and Parker, M.G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387, 733-736.
75. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311-318.
76. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* 38, 576-589.
77. Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., *et al.* (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* 6, 283-289.
78. Hewitt, S.C., and Korach, K.S. (2003). Oestrogen receptor knockout mice: roles for oestrogen receptors alpha and beta in reproductive tissues. *Reproduction* 125, 143-149.
79. Holmes, K.A., Hurtado, A., Brown, G.D., Launchbury, R., Ross-Innes, C.S., Hadfield, J., Odom, D.T., and Carroll, J.S. (2012). Transducin-like enhancer protein 1 mediates estrogen receptor binding and transcriptional activity in breast cancer cells. *Proceedings of the National Academy of Sciences* 109, 2748-2753.

80. Howell, A., Howell, S.J., Clarke, R., and Anderson, E. (2001). Where do selective estrogen receptor modulators (SERMs) and aromatase inhibitors (AIs) now fit into breast cancer treatment algorithms? *The Journal of steroid biochemistry and molecular biology* 79, 227-237.
81. Howell, A., Robertson, J.F., Abram, P., Lichinitser, M.R., Elledge, R., Bajetta, E., Watanabe, T., Morris, C., Webster, A., and Dimery, I. (2004). Comparison of fulvestrant versus tamoxifen for the treatment of advanced breast cancer in postmenopausal women previously untreated with endocrine therapy: a multinational, double-blind, randomized trial. *Journal of Clinical Oncology* 22, 1605-1613.
82. Howell, A., Robertson, J.F., Albano, J.Q., Aschermannova, A., Mauriac, L., Kleeberg, U.R., Vergote, I., Erikstein, B., Webster, A., and Morris, C. (2002). Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment. *Journal of Clinical Oncology* 20, 3396-3403.
83. Hurtado, A., Holmes, K.A., Geistlinger, T.R., Hutcheson, I.R., Nicholson, R.I., Brown, M., Jiang, J., Howat, W.J., Ali, S., and Carroll, J.S. (2008). Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. *Nature* 456, 663-666.
84. Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011a). FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature genetics* 43, 27-33.
85. Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011b). FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* 43, 27-33.
86. Ingle, J.N., Krook, J.E., Green, S.J., Kubista, T.P., Everson, L.K., Ahmann, D.L., Chang, M.N., Bisel, H.F., Windschitl, H.E., and Twito, D.I. (1986). Randomized trial of bilateral oophorectomy versus tamoxifen in premenopausal women with metastatic breast cancer. *Journal of Clinical Oncology* 4, 178-185.

87. Jansen, M.P., Sieuwerts, A.M., Look, M.P., Ritstier, K., Meijer-van Gelder, M.E., van Staveren, I.L., Klijn, J.G., Foekens, J.A., and Berns, E.M. (2007). HOXB13-to-IL17BR expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study. *Journal of clinical oncology* 25, 662-668.
88. Jensen, E., Jacobson, H., Flesher, J., Saha, N., Gupta, G., Smith, S., Colucci, V., Shiplacoff, D., Neumann, H., and DeSombre, E. (1966). Estrogen receptors in target tissues. *Steroid dynamics*, 133-157.
89. Jensen, E.V., and DeSombre, E.R. (1973). Estrogen-Receptor Interaction Estrogenic hormones effect transformation of specific receptor proteins to a biochemically functional form. *Science* 182, 126-134.
90. Jerevall, P.-L., Brommesson, S., Strand, C., Gruvberger-Saal, S., Malmström, P., Nordenskjöld, B., Wingren, S., Söderkvist, P., Fernö, M., and Stål, O. (2008). Exploring the two-gene ratio in breast cancer – independent roles for HOXB13 and IL17BR in prediction of clinical outcome. *Breast cancer research and treatment* 107, 225-234.
91. Jerevall, P.-L., Jansson, A., Fornander, T., Skoog, L., Nordenskjöld, B., and Stål, O. (2010). Predictive relevance of HOXB13 protein expression for tamoxifen benefit in breast cancer. *Breast Cancer Res* 12, R53.
92. Jerevall, P.-L., Ma, X.-J., Li, H., Salunga, R., Kesty, N.C., Erlander, M.G., Sgroi, D., Holmlund, B., Skoog, L., and Fornander, T. (2011). Prognostic utility of HOXB13: IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *British journal of cancer* 104, 1762-1769.
93. Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* 316, 1497-1502.
94. Johnson, P.H., Walker, R.P., Jones, S.W., Stephens, K., Meurer, J., Zajchowski, D.A., Luke, M.M., Eeckman, F., Tan, Y., and Wong, L. (2002). Multiplex Gene Expression Analysis for High-Throughput Drug Discovery: Screening and

Analysis of Compounds Affecting Genes Overexpressed in Cancer Cells.
Molecular cancer therapeutics 1, 1293-1304.

95. Jordan, V. (1976). Effect of tamoxifen (ICI 46,474) on initiation and growth of DMBA-induced rat mammary carcinomata. *European Journal of Cancer* (1965) 12, 419-424.
96. Jordan, V., and Koerner, S. (1975). Tamoxifen (ICI 46,474) and the human carcinoma 8S oestrogen receptor. *European Journal of Cancer* (1965) 11, 205-206.
97. Jordan, V.C., and DOWSE, L.J. (1976). Tamoxifen as an anti-tumour agent: effect on oestrogen binding. *Journal of Endocrinology* 68, 297-303.
98. Joseph, R., Orlov, Y.L., Huss, M., Sun, W., Li Kong, S., Ukil, L., Fu Pan, Y., Li, G., Lim, M., and Thomsen, J.S. (2010). Integrative model of genomic factors for determining binding site selection by estrogen receptor- α . *Molecular systems biology* 6.
99. Jozwik, K.M., and Carroll, J.S. (2012). Pioneer factors in hormone-dependent cancers. *Nature Reviews Cancer* 12, 381-385.
100. Kalyuga, M., Gallego-Ortega, D., Lee, H.J., Roden, D.L., Cowley, M.J., Caldon, C.E., Stone, A., Allerdice, S.L., Valdes-Mora, F., Launchbury, R., *et al.* (2012). ELF5 Suppresses Estrogen Sensitivity and Underpins the Acquisition of Antiestrogen Resistance in Luminal Breast Cancer. *PLoS Biol* 10, e1001461.
101. Katika, M., Martin, R., Siv, G., Børresen-Dale, A.L., Lopez, S., and Hurtado, T. (2013). Abstract B124: Selective inhibition of HER2 signaling pathway reveals novel functions of FOXA1 in breast cancer. *Molecular Cancer Therapeutics* 12, B124-B124.
102. Katzenellenbogen, B.S., and Katzenellenbogen, J.A. (2000). Estrogen receptor transcription and transactivation: Estrogen receptor alpha and estrogen receptor beta-regulation by selective estrogen receptor modulators and importance in breast cancer. *Breast Cancer Research* 2, 335.

103. Kohler, S., and Cirillo, L.A. (2010). Stable chromatin binding prevents FoxA acetylation, preserving FoxA chromatin remodeling. *Journal of biological chemistry* 285, 464-472.
104. Kok, M., Linn, S.C., Van Laar, R.K., Jansen, M.P., Van den Berg, T.M., Delahaye, L.J., Glas, A.M., Peterse, J.L., Hauptmann, M., and Foekens, J.A. (2009). Comparison of gene expression profiles predicting progression in breast cancer patients treated with tamoxifen. *Breast cancer research and treatment* 113, 275-283.
105. Kong, S.L., Li, G., Loh, S.L., Sung, W.-K., and Liu, E.T. (2011a). Cellular reprogramming by the conjoint action of ER[alpha], FOXA1, and GATA3 to a ligand-inducible growth state. *Mol Syst Biol* 7.
106. Kong, S.L., Li, G., Loh, S.L., Sung, W.K., and Liu, E.T. (2011b). Cellular reprogramming by the conjoint action of ER α , FOXA1, and GATA3 to a ligand-inducible growth state. *Molecular systems biology* 7.
107. Lacotte, S., Brun, S., Muller, S., and Dumortier, H. (2009). CXCR3, inflammation, and autoimmune diseases. *Annals of the New York Academy of Sciences* 1173, 310-317.
108. Laganière, J., Deblois, G., Lefebvre, C., Bataille, A.R., Robert, F., and Giguere, V. (2005). Location analysis of estrogen receptor α target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proceedings of the National Academy of Sciences of the United States of America* 102, 11651-11656.
109. Lee, C.S., Friedman, J.R., Fulmer, J.T., and Kaestner, K.H. (2005). The initiation of liver development is dependent on Foxa transcription factors. *Nature* 435, 944-947.
110. Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* 72, 73-84.

111. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882-883.
112. Li, Q., Birkbak, N.J., Györfy, B., Szallasi, Z., and Eklund, A.C. (2011). Jetset: selecting the optimal microarray probe set to represent a gene. *BMC bioinformatics* 12, 474.
113. Li, Z., and Xie, Z. (2009). The Na/K-ATPase/Src complex and cardiotoxic steroid-activated protein kinase cascades. *Pflügers Archiv-European Journal of Physiology* 457, 635-644.
114. Lu, D., Kiriya, Y., Lee, K.Y., and Giguère, V. (2001). Transcriptional Regulation of the Estrogen-inducible pS2 Breast Cancer Marker Gene by the ERR Family of Orphan Nuclear Receptors. *Cancer Research* 61, 6755-6761.
115. Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D., and Woolf, P.J. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. *BMC bioinformatics* 10, 161.
116. Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* 132, 958-970.
117. Ma, X.-J., Dahiya, S., Richardson, E., Erlander, M., and Sgroi, D.C. (2009a). Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res* 11, R7.
118. Ma, X.-J., Hilsenbeck, S.G., Wang, W., Ding, L., Sgroi, D.C., Bender, R.A., Osborne, C.K., Allred, D.C., and Erlander, M.G. (2006). The HOXB13: IL17BR expression index is a prognostic factor in early-stage breast cancer. *Journal of clinical oncology* 24, 4611-4619.
119. Ma, X.-J., Salunga, R., Dahiya, S., Wang, W., Carney, E., Durbecq, V., Harris, A., Goss, P., Sotiriou, C., and Erlander, M. (2008). A five-gene molecular grade index

and HOXB13: IL17BR are complementary prognostic factors in early stage breast cancer. *Clinical Cancer Research* 14, 2601-2608.

120. Ma, X.-J., Wang, Z., Ryan, P.D., Isakoff, S.J., Barmettler, A., Fuller, A., Muir, B., Mohapatra, G., Salunga, R., and Tuggle, J.T. (2004). A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer cell* 5, 607-616.
121. Ma, X., Norsworthy, K., Kundu, N., Rodgers, W.H., Gimotty, P.A., Goloubeva, O., Lipsky, M., Li, Y., Holt, D., and Fulton, A. (2009b). CXCR3 expression is associated with poor survival in breast cancer and promotes metastasis in a murine model. *Molecular cancer therapeutics* 8, 490-498.
122. Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27, 1696-1697.
123. Magnani, L., Ballantyne, E.B., Zhang, X., and Lupien, M. (2011). PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLoS genetics* 7, e1002368.
124. Massarweh, S., Osborne, C.K., Creighton, C.J., Qin, L., Tsimelzon, A., Huang, S., Weiss, H., Rimawi, M., and Schiff, R. (2008). Tamoxifen Resistance in Breast Tumors Is Driven by Growth Factor Receptor Signaling with Repression of Classic Estrogen Receptor Genomic Function. *Cancer Research* 68, 826-833.
125. McCall, M.N., Bolstad, B.M., and Irizarry, R.A. (2010). Frozen robust multiarray analysis (fRMA). *Biostatistics* 11, 242-253.
126. McDonnell, D.P. (2000). Selective estrogen receptor modulators (SERMs): a first step in the development of perfect hormone replacement therapy regimen. *Journal of the Society for Gynecologic Investigation* 7, S10-S15.
127. Mijatovic, T., Van Quaquebeke, E., Delest, B., Debeir, O., Darro, F., and Kiss, R. (2007). Cardiotonic steroids on the road to anti-cancer therapy. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1776, 32-57.

128. Millikan, R.C., Newman, B., Tse, C.-K., Moorman, P.G., Conway, K., Smith, L.V., Lobbok, M.H., Geradts, J., Bensen, J.T., and Jackson, S. (2008). Epidemiology of basal-like breast cancer. *Breast cancer research and treatment* 109, 123-139.
129. Montano, M.M., Muller, V., Trobaugh, A., and Katzenellenbogen, B.S. (1995). The Carboxy-Terminal F-Domain of the Human Estrogen-Receptor - Role in the Transcriptional Activity of the Receptor and the Effectiveness of Antiestrogens as Estrogen Antagonists. *Molecular Endocrinology* 9, 814-825.
130. Nagy, P., Bisgaard, H.C., and Thorgeirsson, S.S. (1994). Expression of hepatic transcription factors during liver development and oval cell differentiation. *The Journal of cell biology* 126, 223-233.
131. Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., and Johnson, A.K. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489, 83-90.
132. Nielsen, T.O., Parker, J.S., Leung, S., Voduc, D., Ebbert, M., Vickery, T., Davies, S.R., Snider, J., Stijleman, I.J., and Reed, J. (2010). A Comparison of PAM50 Intrinsic Subtyping with Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-Treated Estrogen Receptor-Positive Breast Cancer. *Clinical Cancer Research* 16, 5222-5232.
133. Nolte, R.T., Wisely, G.B., Westin, S., Cobb, J.E., Lambert, M.H., Kurokawa, R., Rosenfeld, M.G., Willson, T.M., Glass, C.K., and Milburn, M.V. (1998). Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor- γ . *Nature* 395, 137-143.
134. Norris, J.D., Chang, C.-Y., Wittmann, B.M., Kunder, R.S., Cui, H., Fan, D., Joseph, J.D., and McDonnell, D.P. (2009a). The homeodomain protein HOXB13 regulates the cellular response to androgens. *Molecular cell* 36, 405-416.
135. Norris, J.D., Chang, C.Y., Wittmann, B.M., Kunder, R.S., Cui, H., Fan, D., Joseph, J.D., and McDonnell, D.P. (2009b). The homeodomain protein HOXB13 regulates the cellular response to androgens. *Mol Cell* 36, 405-416.

136. Norris, J.D., Paige, L.A., Christensen, D.J., Chang, C.-Y., Huacani, M.R., Fan, D., Hamilton, P.T., Fowlkes, D.M., and McDonnell, D.P. (1999). Peptide antagonists of the human estrogen receptor. *Science* 285, 744-746.
137. Oñate, S.A., Tsai, S.Y., Tsai, M.-J., and O'Malley, B.W. (1995). Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* 270, 1354-1357.
138. Osborne, C., Pippen, J., Jones, S., Parker, L., Ellis, M., Come, S., Gertler, S., May, J., Burton, G., and Dimery, I. (2002). Double-blind, randomized trial comparing the efficacy and tolerability of fulvestrant versus anastrozole in postmenopausal women with advanced breast cancer progressing on prior endocrine therapy: results of a North American trial. *Journal of Clinical Oncology* 20, 3386-3395.
139. Osborne, C.K., Bardou, V., Hopp, T.A., Chamness, G.C., Hilsenbeck, S.G., Fuqua, S.A.W., Wong, J., Allred, D.C., Clark, G.M., and Schiff, R. (2003). Role of the Estrogen Receptor Coactivator AIB1 (SRC-3) and HER-2/neu in Tamoxifen Resistance in Breast Cancer. *Journal of the National Cancer Institute* 95, 353-361.
140. Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F.L., Watson, D., and Bryant, J. (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of Clinical Oncology* 24, 3726-3734.
141. Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., *et al.* (2009). Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *Journal of Clinical Oncology* 27, 1160-1167.
142. Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., and Akslén, L.A. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747-752.
143. Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 12, R68.

144. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
145. Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147, 1408-1419.
146. Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B.B., Barrette, T.R., Anstet, M.J., Kincead-Beal, C., and Kulkarni, P. (2007). OncoPrint 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* (New York, NY) 9, 166.
147. Robertson, J.F.R., Osborne, C.K., Howell, A., Jones, S.E., Mauriac, L., Ellis, M., Kleeberg, U.R., Come, S.E., Vergote, I., Gertler, S., *et al.* (2003). Fulvestrant versus anastrozole for the treatment of advanced breast carcinoma in postmenopausal women. *Cancer* 98, 229-238.
148. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
149. Rosenfeld, M.G., and Glass, C.K. (2001). Coregulator codes of transcriptional regulation by nuclear receptors. *Journal of Biological Chemistry* 276, 36865-36868.
150. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., and Green, A.R. (2012a). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389-393.
151. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., *et al.* (2012b). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389-393.

152. Ruan, J., Dean, A.K., and Zhang, W. (2010). A general co-expression network-based approach to gene expression analysis: comparison and applications. *BMC systems biology* 4, 8.
153. Ruff, M., Gangloff, M., Wurtz, J.M., and Moras, D. (2000). Estrogen receptor transcription and transactivation: structure-function relationship in DNA-and ligand-binding domains of estrogen receptors. *Breast Cancer Research* 2, 353.
154. Schröder, M.S., Culhane, A.C., Quackenbush, J., and Haibe-Kains, B. (2011). survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* 27, 3206-3208.
155. Sérandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., and Métivier, R. (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome research* 21, 555-565.
156. Sgroi, D.C., Carney, E., Zarrella, E., Steffel, L., Binns, S.N., Finkelstein, D.M., Szymonifka, J., Bhan, A.K., Shepherd, L.E., and Zhang, Y. (2013). Prediction of late disease recurrence and extended adjuvant letrozole benefit by the HOXB13/IL17BR biomarker. *Journal of the National Cancer Institute* 105, 1036-1042.
157. Shabalin, A.A., Tjelmeland, H., Fan, C., Perou, C.M., and Nobel, A.B. (2008). Merging two gene-expression studies via cross-platform normalization. *Bioinformatics* 24, 1154-1160.
158. Shah, N., Jin, K., Cruz, L.-A., Park, S., Sadik, H., Cho, S., Goswami, C.P., Nakshatri, H., Gupta, R., and Chang, H.Y. (2013). HOXB13 mediates tamoxifen resistance and invasiveness in human breast cancer by suppressing ER α and inducing IL-6 expression. *Cancer research* 73, 5449-5458.
159. Shah, N., and Sukumar, S. (2010). The Hox genes and their roles in oncogenesis. *Nature Reviews Cancer* 10, 361-371.

160. Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A., and Brown, M. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* 103, 843-852.
161. Sharma, S.C., and Richards, J.S. (2000). Regulation of AP1 (Jun/Fos) Factor Expression and Activation in Ovarian Granulosa Cells: RELATION OF JunD AND Fra2 TO TERMINAL DIFFERENTIATION. *Journal of Biological Chemistry* 275, 33718-33728.
162. Sheffield, N.C., Thurman, R.E., Song, L., Safi, A., Stamatoyannopoulos, J.A., Lenhard, B., Crawford, G.E., and Furey, T.S. (2013). Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome research* 23, 777-788.
163. Shi, Z., Derow, C.K., and Zhang, B. (2010). Co-expression module analysis reveals biological processes, genomic gain, and regulatory mechanisms associated with breast cancer progression. *BMC systems biology* 4, 74.
164. Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., and Greene, G.L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95, 927-937.
165. Shou, J., Massarweh, S., Osborne, C.K., Wakeling, A.E., Ali, S., Weiss, H., and Schiff, R. (2004). Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. *Journal of the National Cancer Institute* 96, 926-935.
166. Siegel, R., Ma, J., Zou, Z., and Jemal, A. (2014). Cancer statistics, 2014. *CA: a cancer journal for clinicians* 64, 9-29.
167. Simpson, E.R., Mahendroo, M.S., Means, G.D., Kilgore, M.W., Hinshelwood, M.M., Graham-Lorence, S., Amarneh, B., Ito, Y., Fisher, C.R., and Michael, M.D. (1994). Aromatase Cytochrome P450, The Enzyme Responsible for Estrogen Biosynthesis*. *Endocrine reviews* 15, 342-355.

168. Smith, D.C., Prentice, R., Thompson, D.J., and Herrmann, W.L. (1975). Association of exogenous estrogen and endometrial carcinoma. *New England journal of medicine* 293, 1164-1167.
169. Song, L., and Crawford, G.E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* 2010, pdb prot5384.
170. Song, L., Zhang, Z., Gräf, S., Huss, M., and Keefe, D. (2011a). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome research* 21, 1757-1767.
171. Song, L., Zhang, Z., Gräf, S., Huss, M., Keefe, D., *et al.* (2011b). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* 21, 1757-1767.
172. Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., and Jeffrey, S.S. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences* 98, 10869-10874.
173. Stenkvist, B., Bengtsson, E., Dahlqvist, B., Eriksson, O., Jarkrans, T., and Nordin, B. (1982). Cardiac glycosides and breast cancer, revisited. *The New England journal of medicine* 306, 484.
174. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15545-15550.
175. Tan, S.K., Lin, Z.H., Chang, C.W., Varang, V., Chng, K.R., Pan, Y.F., Yong, E.L., Sung, W.K., and Cheung, E. (2011a). AP-2gamma regulates oestrogen receptor-

- mediated long-range chromatin interaction and gene transcription. *EMBO J* 30, 2569-2581.
176. Tan, S.K., Lin, Z.H., Chang, C.W., Varang, V., Chng, K.R., Pan, Y.F., Yong, E.L., Sung, W.K., and Cheung, E. (2011b). AP-2 γ regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. *The EMBO journal* 30, 2569-2581.
177. Tewari, A.K., Yardimci, G.G., Shibata, Y., Sheffield, N.C., Song, L., Taylor, B.S., Georgiev, S.G., Coetzee, G.A., Ohler, U., and Furey, T.S. (2012). Chromatin accessibility reveals insights into androgen receptor activation and transcriptional specificity. *Genome biology* 13, R88.
178. Theodorou, V., Stark, R., Menon, S., and Carroll, J.S. (2013). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome research* 23, 12-22.
179. Thorat, M.A., Marchio, C., Morimiya, A., Savage, K., Nakshatri, H., Reis-Filho, J.S., and Badve, S. (2008). Forkhead box A1 expression in breast cancer is associated with luminal subtype and good prognosis. *Journal of clinical pathology* 61, 327-332.
180. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., and Vernot, B. (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75-82.
181. Tremblay, A., Tremblay, G.B., Labrie, F., and Giguère, V. (1999). Ligand-independent recruitment of SRC-1 to estrogen receptor β through phosphorylation of activation function AF-1. *Molecular cell* 3, 513-519.
182. van't Veer, L.J., Dai, H., Van De Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., and Witteveen, A.T. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *nature* 415, 530-536.

183. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 10, 252-263.
184. Wang, Z., Dahiya, S., Provencher, H., Muir, B., Carney, E., Coser, K., Shioda, T., Ma, X.-J., and Sgroi, D.C. (2007). The prognostic biomarkers HOXB13, IL17BR, and CHDH are regulated by estrogen in breast cancer. *Clinical Cancer Research* 13, 6327-6334.
185. Wardell, S.E., Nelson, E.R., Chao, C.A., and McDonnell, D.P. (2013). Bazedoxifene exhibits antiestrogenic activity in animal models of tamoxifen-resistant breast cancer: Implications for treatment of advanced disease. *Clinical Cancer Research* 19, 2420-2431.
186. Weatherman, R.V., Fletterick, R.J., and Scanlan, T.S. (1999). Nuclear-receptor ligands and ligand-binding domains. *Annual review of biochemistry* 68, 559-581.
187. Webb, P., Nguyen, P., Shinsako, J., Anderson, C., Feng, W., Nguyen, M.P., Chen, D., Huang, S.-M., Subramanian, S., and McKinerney, E. (1998). Estrogen receptor activation function 1 works by binding p160 coactivator proteins. *Molecular Endocrinology* 12, 1605-1618.
188. Welboren, W.J., van Driel, M.A., Janssen-Megens, E.M., van Heeringen, S.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. (2009). ChIP-Seq of ER α and RNA polymerase II defines genes differentially responding to ligands. *The EMBO journal* 28, 1418-1428.
189. Winer, E.P., Hudis, C., Burstein, H.J., Wolff, A.C., Pritchard, K.I., Ingle, J.N., Chlebowski, R.T., Gelber, R., Edge, S.B., and Gralow, J. (2005). American Society of Clinical Oncology technology assessment on the use of aromatase inhibitors as adjuvant therapy for postmenopausal women with hormone receptor-positive breast cancer: status report 2004. *Journal of clinical oncology* 23, 619-629.
190. Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., and Schutz, F. (2008). Meta-analysis of

gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* 10, R65.

191. Wittmann, B.M., Sherk, A., and McDonnell, D.P. (2007). Definition of Functionally Important Mechanistic Differences among Selective Estrogen Receptor Down-regulators. *Cancer Research* 67, 9549-9560.
192. Wolf, I., Bose, S., Williamson, E.A., Miller, C.W., Karlan, B.Y., and Koeffler, H.P. (2007). FOXA1: Growth inhibitor and a favorable prognostic factor in human breast cancer. *International journal of cancer* 120, 1013-1022.
193. Xiang, Y., Zhang, C.-Q., and Huang, K. (2012). Predicting glioblastoma prognosis networks using weighted gene co-expression network analysis on TCGA data. *BMC bioinformatics* 13, S12.
194. Yamashita, H., Nishio, M., Toyama, T., Sugiura, H., Kondo, N., Kobayashi, S., Fujii, Y., and Iwase, H. (2008). Low phosphorylation of estrogen receptor α (ER α) serine 118 and high phosphorylation of ER α serine 167 improve survival in ER-positive breast cancer. *Endocrine-related cancer* 15, 755-763.
195. Zeltser, L., Desplan, C., and Heintz, N. (1996). Hoxb-13: a new Hox gene in a distant region of the HOXB cluster maintains colinearity. *Development* 122, 2475-2484.
196. Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 4, 1128.
197. Zhang, Z., Chang, C.W., Goh, W.L., Sung, W.K., and Cheung, E. (2011). CENTDIST: discovery of co-associated factors by motif distribution. *Nucleic Acids Res* 39, W391-399.
198. Zhou, Y., Yau, C., Gray, J.W., Chew, K., Dairkee, S.H., Moore, D.H., Eppenberger, U., Eppenberger-Castori, S., and Benz, C.C. (2007). Enhanced NF κ B

and AP-1 transcriptional activity associated with antiestrogen resistant breast cancer. *BMC cancer* 7, 59.

199. Zilli, M., Grassadonia, A., Tinari, N., Di Giacobbe, A., Gildetti, S., Giampietro, J., Natoli, C., and Iacobelli, S. (2009). Molecular mechanisms of endocrine resistance and their implication in the therapy of breast cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1795, 62-81.
200. Zugowski, C., Lieder, F., Müller, A., Gasch, J., Corvinus, F.M., Moriggl, R., and Friedrich, K. (2011). STAT3 controls matrix metalloproteinase-1 expression in colon carcinoma cells by both direct and AP-1-mediated interaction with the MMP-1 promoter. *Biological chemistry* 392, 449-459.
201. Zwart, W., de Leeuw, R., Rondaij, M., Neefjes, J., Mancini, M.A., and Michalides, R. (2010). The hinge region of the human estrogen receptor determines functional synergy between AF-1 and AF-2 in the quantitative response to estradiol and tamoxifen. *Journal of Cell Science* 123, 1253-1261.

Biography

Jeffery Scott Jasper

Date of birth: July 20, 1981

Place of birth: Provo, UT

Education:

2004 – 2008: Bachelors of Science (Biochemistry) University of Utah

2008 – 2014: Doctoral Program in Pharmacology and Cancer Biology,
Laboratory of Dr. Donald P. McDonnell, Duke University

Publications:

1. Nelson, E. R.; Wardell, S. E.; **Jasper, J. S.**; Park, S.; Suchindran, S.; Howe, M. K.; Carver, N. J.; Pillai, R. V.; Sullivan, P. M.; Sondhi, V.; Umetani, M.; Geradts, J.; McDonnell, D. P., 27-Hydroxycholesterol Links Hypercholesterolemia and Breast Cancer Pathophysiology. *Science* 2013, 342 (6162), 1094-1098.
2. Shatnawi, A., Norris, J., Chaveroux, C., **Jasper, J.S.**, Sherk, A., McDonnell, D., & Giguère, V. (2013). ELF3 is a repressor of androgen receptor action in prostate cancer cells. *Oncogene*.
3. Pollock, J. A., Larrea, M. D., **Jasper, J. S.**, McDonnell, D. P., & McCafferty, D. G. (2012). Lysine-Specific Histone Demethylase 1 Inhibitors Control Breast Cancer Proliferation in ER α -Dependent and-Independent Manners. *ACS Chemical Biology*, 7(7), 1221-1231.
4. Chang, C. Y., Kazmin, D., **Jasper, J. S.**, Kunder, R., Zuercher, W. J., & McDonnell, D. P. (2011). The metabolic regulator ERR α , a downstream target of HER2/IGF-1R, as a therapeutic target in breast cancer. *Cancer cell*, 20(4), 500-510.

Works in progress:

1. **Jasper, J.S.**, McDonnell, D. P. (2014). PrimerBot!-A web application for high-throughput qPCR primer design. *Manuscript in preparation*.
2. **Jasper, J.S.**, Wright, T.M., Stice, J.P., Wardell, S.E., McDonnell, D. P. (2014). Tipping the Balance: Factors that influence the ER signaling program in . *Manuscript in preparation*
3. Glass, O.K., Inman, B.A., Broadwater, G., Courneya, K.S., Mackey, J.R., Goruk, S., Field, C.J., Joy, A.A., Nelson, E.R., **Jasper, J.S.**, et al. (2014). Aerobic Training Modulates the Host Systemic Milieu in Patients with Solid Tumors and Alters Cancer Cell Phenotypes In Vitro. *Under review at Cancer Research*.

4. Martz, C. A., Ottina, K. A., **Jasper, J. S.**, Peraza-Penton, A., Wang, T., Rathmell, J. C., Wargo, J.r A., McDonnell, D. P., Sabatini, D. M. and Wood, K. C. (2014). Systematic identification of signaling pathways with potential to confer anticancer drug resistance. *Under review at Science*.
5. Wright, T.M., Wardell, S. E., **Jasper, J. S.**, Stice, J.P., Safi, R.S., Nelson, E. R., McDonnell, D. P. (2014). Delineation of a FOXA1/ER α /AGR2 regulatory loop that is dysregulated in endocrine resistant breast cancer. *Under review at Molecular Cancer Research*.
6. Mahmoud Al-Faqih, **Jasper, J.S.**, McDonnell, D. P. (2014) The 27 hydroxycholesterol axis is a novel therapeutic target in castrate resistant prostate cancer. *Manuscript in preparation*.
7. Alexander, P.B., Gong, C., Yuan, L., **Jasper, J.S.**, Yang, P., Xu, X., Sun, T., McDonnell, D.P., Song, E., and Wang, X.F. (2014). Trastuzumab therapy induces breast cancer cellular senescence: molecular determinants underlying sensitivity and resistance. *Manuscript in preparation*.
8. **Jasper, J.S.**, McDonnell, D. P. (2014). Gene-Analytics, A high level web interface for discovering gene function and clinical relationships. *Manuscript in preparation*.