

# Role of the GATA3 Transcription Factor on AR Signalling in Breast Cancer

A Thesis submitted to the University of Adelaide in fulfilment of the  
requirements for the degree of Doctor of Philosophy

by

Leila Hosseinzadeh

Dame Roma Mitchell Cancer Research Laboratories,  
Faculty of Health and Medical Sciences,  
School of Medicine,  
The University of Adelaide

March 2021

## Table of Contents

Abstract.....	7
List of abbreviations.....	9
Declaration.....	14
Acknowledgments .....	15
Conference abstracts and presentations.....	18
Conference attendance .....	18
1. Chapter-1: Introduction.....	20
1.1. Mammary gland development and breast cancer .....	20
1.1.1. Mammary gland development .....	20
1.1.2. Breast cancer incidence and mortality .....	25
1.1.3. Histological classification of breast cancer .....	25
1.1.4. Molecular classification of breast cancer and cellular origins.....	26
1.1.5. Endocrine therapy for ER-positive breast cancer .....	32
1.1.6. Endocrine resistance.....	33
1.2. Nuclear receptors and cancer.....	35
1.2.1. Nuclear receptor structure and classifications .....	35
1.2.2. Nuclear receptor mechanism of action .....	39
1.2.3. Technology advancements in nuclear receptor function studies.....	43
1.2.4. Estrogen receptor function in normal and malignant breast tissue.....	49
1.2.5. Androgen receptor function in normal and malignant breast .....	52
1.2.5.1. Androgen receptor gene and protein structure .....	52
1.2.5.2. Androgens and androgen receptor signalling in normal breast tissue.....	55
1.2.5.3. Androgen receptor expression in breast cancer.....	57
1.2.5.4. Androgen receptor signalling in ER+ breast cancer .....	60
1.2.5.5. Androgen receptor signalling in ER- breast cancer.....	61
1.2.6. Nuclear receptor cross-talk in cancer .....	62
1.2.7. NR co-regulators and cancer.....	66
1.3. Role of the GATA3 transcription factor in the mammary glands .....	68
1.3.1. GATA family members, their similar protein structure and DNA binding features.....	68
1.3.2. The GATA3 transcription factor .....	74
1.3.3. Role of GATA3 in normal mammary gland development and differentiation .....	78
1.3.4. GATA3 and breast cancer progression.....	80
1.3.5. Role of GATA3 in tumour invasion and metastasis.....	81
1.3.6. GATA3 cooperates with ER and FOXA1 as a definitive luminal complex in ER+ breast cancer	83

1.3.7. GATA3 mutations in breast cancer .....	86
1.4. Summary .....	93
1.5. Hypothesis and Aims.....	94
Chapter-2: AR interacts with GATA3 to induce a luminal epithelial phenotype in breast cancer through regulation of lineage-restricted genes.....	99
Abstract.....	100
Introduction .....	101
Results.....	102
GATA3 is a novel AR interacting protein in breast cancer .....	102
GATA3 nuclear translocation is induced upon hormone stimulation in breast cancer.....	105
AR agonist-induced GATA3 cis-regulatory elements are co-occupied by AR at active chromatin loci.....	106
AR agonist-induced AR-GATA3 cis-regulatory elements are associated with gene regulatory profiles involved in development and differentiation of the mammary epithelium .....	109
GATA3 acts as an AR co-regulator in breast cancer cells.....	110
AR agonist-induced modifications of the chromatin landscape at AR-GATA3 binding events associates with regulation of lineage-restricted genes .....	111
Discussion .....	113
Material and Methods .....	117
Main figures.....	126
References .....	138
Supplementary figures.....	150
Extended data .....	182
Chapter-3: GATA3 facilitates AR tumour-suppressive function in ER+ breast cancer.....	218
Abstract.....	219
Introduction .....	220
Results.....	223
Estrogen stimulation of ER+ breast cancer cells rearranges the GATA3 cistrome .....	223
Androgen stimulation of ER+ breast cancer cells also rearranges the GATA3 cistrome .....	225
AR-induced ER-reprogramming is mediated through a switch in GATA3 co-regulatory function .....	227
GATA3 knockdown reduces the growth inhibitory effects of androgens in E2-stimulated ER+ breast cancer cells .....	229
Discussion .....	231
Materials and Methods.....	234
Main figures .....	239
References .....	248

Supplementary figures.....	255
4. Chapter-4: General discussion .....	268
4.1. Introduction .....	268
4.2. Major findings of this thesis.....	269
4.2.1. GATA3 is a novel AR interacting protein independent of ER expression in normal mammary tissues and different BC subtypes. ....	269
4.2.2. Stimulation of breast cancer cells with estrogen or androgen hormones reprograms the GATA3 cistrome in breast cancer cell line models. ....	271
4.2.3. GATA3 is bound and plays a regulatory role at genomic loci associated with AR-mediated growth inhibition in ER+ BC to inhibit E2-stimulated growth. ....	272
4.2.4. A collaborative role for GATA3 and AR in promotion of luminal epithelial differentiation in breast cells regardless of ER expression.....	273
4.3. Future directions.....	275
4.3.1. Investigating the structural basis of AR and GATA3 protein-protein interaction.....	275
4.3.2. AR-GATA3 function in normal mammary glands .....	275
4.3.3. Investigating the effect of GATA3 depletion on mRNA expression level of basal lineage targets .....	277
4.3.4. Exploring the role of GATA3 in cytokinesis and cell cycle progression of breast cancer cells	278
4.3.5. Assessing the effect of GATA3 ectopic expression in ER+ and ER- breast cancer cells	279
4.4. Concluding remarks .....	280
5. Chapter-5: Appendix.....	282
5.1. Appendix – Generating stable knock-down of factors in breast cancer cell lines.....	282
5.1.1. Appendix-1: Plasmid constructs.....	282
5.1.2. Appendix-2: Generating lentiviral transfer plasmid encoding the insert of interest, viral transfection, and transduction in desired cell lines.....	294
Bibliography.....	299

## Table of figures

### Chapter-1

<b>Figure 1:</b> Ductal structure .....	22
<b>Figure 2:</b> Human breast epithelial hierarchy and cell type differentiation within the mammary epithelium.....	23
<b>Figure 3:</b> Schematic representation of a terminal end bud.....	24
<b>Figure 4:</b> Molecular classification of breast cancer subtypes .....	30
<b>Figure 5:</b> Potential relationships between stem and progenitor cells and different breast cancer subtypes.....	31
<b>Figure 6:</b> Nuclear receptor structure .....	38
<b>Figure 7:</b> Schematic illustration of androgen receptor signalling.....	41
<b>Figure 8:</b> Schematic illustration of estrogen receptor signalling .....	42
<b>Figure 9:</b> Chromatin-immunoprecipitation (ChIP)- sequencing .....	46
<b>Figure 10:</b> RNA-sequencing (RNA-seq) .....	47
<b>Figure 11:</b> Rapid immunoprecipitation mass spectrometry of endogenous protein (RIME) .....	48
<b>Figure 12:</b> Schematic illustration of estrogen receptor isoforms.....	51
<b>Figure 13:</b> Schematic illustration of the androgen receptor gene and protein structures .....	54
<b>Figure 14:</b> Chromosomal location of GATA family members .....	72
<b>Figure 15:</b> The GATA family of proteins.....	73
<b>Figure 16:</b> GATA3 Functional domains and protein structure .....	77
<b>Figure 17:</b> Distribution of GATA3 mutations in breast tumors.....	89

### Chapter-2

<b>Figure 1:</b> GATA3 is a novel AR interacting protein in breast cancer.....	126
<b>Figure 2:</b> Activated AR induces AR and GATA3 nuclear translocation .....	128
<b>Figure 3:</b> AR agonist-induced GATA3 cis-regulatory elements are co-occupied by AR at transcriptionally active chromatin loci .....	130
<b>Figure 4:</b> AR agonist-induced AR-GATA3 binding sites are implicated in development and differentiation of mammary epithelium.....	132
<b>Figure 5:</b> GATA3 acts as an AR co-regulator in breast cancer cells .....	134
<b>Figure 6:</b> AR agonist-induced AR-GATA3 loci are associated with lineage-restricted genes in breast cancer in vitro and in vivo models .....	136
<b>Supplementary 1:</b> GATA3 is a novel AR interacting protein in breast cancer cells.....	150
<b>Supplementary 2:</b> Validation of the AR and GATA3 interactions in breast cancer.....	152

<b>Supplementary 3:</b> AR and GATA3 interactions occur in both nucleus and cytoplasm of breast cancer cells, and estrogen treatment caused moderate GATA3 nuclear translocation.....	154
<b>Supplementary 4:</b> Validation of AR and GATA3 expression in GAR15-13 and HCI-005 PDX models .....	156
<b>Supplementary 5:</b> AR activation is the major source of variation in GATA3 ChIP-seq datasets... 158	
<b>Supplementary 6:</b> AR agonism caused significant enrichment of new GATA3 binding sites across all the in vitro and in vivo breast cancer models.....	160
<b>Supplementary 7:</b> Distribution of AR and GATA3 ChIP-seq data upon AR activation in breast cancer cells.....	162
<b>Supplementary 8:</b> Replicate concordance for GATA3, AR, and H3K27ac ChIP-seq experiments .	164
<b>Supplementary 9:</b> H3K27ac enrichment at AR-GATA3 binding events upon AR activation .....	166
<b>Supplementary 10:</b> Differentially up- and down-regulated genes associated with AR agonist-induced AR-GATA3 peaks in breast cancer models.....	168
<b>Supplementary 11:</b> Biological pathways associated with the differentially expressed AR-GATA3 gene targets in breast cancer in vitro and in vivo models.....	170
<b>Supplementary 12:</b> Gene set enrichment analysis (GSEA) of AR-GATA3 gene targets in breast cancer.....	172
<b>Supplementary 13:</b> AR knock-down significantly suppressed the expression of AR target genes but did not affect the GATA3 protein expression level in breast cancer cells.....	174
<b>Supplementary 14:</b> GATA3 knock-down significantly reduced the expression of AR target genes but did not affect AR protein expression level in breast cancer cells .....	176
<b>Supplementary 15:</b> AR agonist-induced AR-GATA3 loci are associated with lineage-restricted genes in breast cancer in vitro and in vivo models.....	178
<b>Supplementary 16:</b> GATA3 and AR binding at loci associated with regulation of luminal epithelial genes in breast cancer cells .....	180
<b>Extended data 1:</b> AR RIME interactome in different breast cancer cell lines .....	182
<b>Extended data 2:</b> Motifs enriched upon AR agonist stimulation in different breast cancer models .....	201
<b>Extended data 3:</b> ChIP-seq data analysis information in different breast cancer models.....	207
<b>Extended data 4:</b> List of Luminal and Basal genes .....	210
<b>Extended data 5:</b> List of siRNAs, antibodies, RT-PCT primers, and ChIP-PCR primers.....	212
<b>Chapter-3</b>	
<b>Figure 1:</b> Estrogen stimulation of ER+ breast cancer cells rearranges the GATA3 cistrome.....	239
<b>Figure 2:</b> Androgens redistribute the GATA3 binding profile in ER+ breast cancer cells ...	241

<b>Figure 3:</b> AR-induced ER-reprogramming is mediated through a switch in GATA3 co-regulatory function.....	244
<b>Figure 4:</b> GATA3 depletion reduces the growth inhibitory effects of androgens in E2-stimulated ER+ breast cancer cells .....	246
<b>Supplementary 1:</b> GATA3 cistrome in different hormone treated conditions in ER+ breast cancer cells .....	255
<b>Supplementary 2:</b> Estrogen stimulation of ZR-75-1 cells rearranges the GATA3 cistrome .....	257
<b>Supplementary 3:</b> Androgens redistribute the GATA3 binding profile in ZR-75-1 breast cancer cells .....	259
<b>Supplementary 4:</b> AR-induced ER-reprogramming is mediated through a switch in GATA3 co-regulatory function .....	262
<b>Supplementary 5:</b> Shared DEG AR signatures with AR-GATA3 targets in T-47D cells .....	264
<b>Supplementary 6:</b> List of antibodies, siRNAs and primers used in this study .....	266

## Abstract

Almost 80 % of all breast cancers (BCs) are dependent on the estrogen receptor alpha (ER) for their growth (termed ER+ BC), while the rest are negative for the expression of ER (ER- BC) (~ 20-25 %). Most ER+ BCs (more than 90 %) and around 50 % of ER- BCs express the androgen receptor (AR). AR is an important BC biomarker, with prognostic and therapeutic potential. AR is a tumor-suppressor in ER+ BC, where it inhibits estrogen-stimulated growth of tumours through down-regulation of key ER-regulated cell cycle genes and activation of good outcome genes. AR has been suggested to play a role in promoting a basal to luminal lineage transition in normal mouse mammary epithelial cells. Despite the role of AR in ER- BCs being controversial, AR positivity has been shown to be associated with improved disease-free survival and more benign clinical and pathologic factors (e.g. lower tumor grade and smaller tumor size) in ER- BCs. In the recent past, the GATA3 transcription factor (TF) has been characterized as an important regulator of ER signalling and mediates normal mammary gland development and cell lineage determination. GATA3 is expressed exclusively in the luminal epithelial cell population, plays an essential role in mammary development and specification, and actively maintains the luminal epithelial differentiation in the adult mammary gland. To date, the role of GATA3 in AR signalling in the context of BC, in the presence or absence of ER, has not been investigated. Therefore, the aim of this thesis was to further investigate and characterise the cross-talk between AR and GATA3 in the presence or absence of ER, using a variety of cell-line models, clinical samples and where possible, patient-derived xenografts (PDXs). I first investigated the functional interplay between GATA3 and AR in inducing the luminal epithelial lineage in breast cancer cells regardless of ER expression. Since the inhibitory



role of AR in ER+ BCs is well established, I then investigated whether GATA3 is involved in AR-induced growth inhibition in ER+ BC models in response to ER and/or AR activation.

Findings of this PhD thesis showed:

- GATA3 is a novel AR interacting protein independent of ER expression in normal mammary tissues and different BC subtypes.
- Stimulation of BC cells with estrogen or androgen hormones reprograms the GATA3 cisome in ER+ and ER- BC cell lines and ER+ PDX models.
- GATA3 and AR co-regulate the expression of essential luminal epithelial markers in both ER+ and ER- cell lines and ER+ PDXs, indicating a key collaborative role for GATA3 and AR in luminal epithelial differentiation of the breast cells regardless of the ER expression.
- AR requires the presence of GATA3 at genomic loci associated with AR-mediated growth inhibition in ER+ BC to inhibit the E2-stimulated growth in these models.

Collectively, the generated data from this thesis suggest a cooperative role for AR and GATA3 TFs in suppressing the tumour growth of ER+ BCs and in driving the luminal-lineage identity in mammary glands and BC contexts. Also, the findings of this thesis provide novel insight into the cross-talk between ER, GATA3, and AR in BC, highlight the signalling complexity of TFs in this disease and provide the basis for further investigations into AR and GATA3 co-operative genomic activity and the direct consequences of this for BC progression and differentiation.

## List of abbreviations

**3C** Chromosome conformation capture

**AF-1** Activation function domain-1

**AF-2** Activation function domain-2

**AI** Aromatase inhibitor

**AR** Androgen receptor

**ARE** Androgen response element

**ARKO** AR-knockout

**ATAC-seq** Assay for transposase-accessible chromatin using sequencing

**BC** Breast cancer

**BL** Basal like

**CCND1** Cyclin-D1

**cDNA** Complementary DNA

**ChIP** Chromatin immunoprecipitation

**ChIP-exo** ChIP-exonuclease

**ChIP-seq** ChIP-sequencing

**CK** Cytokeratin

**Co-IP** Co-immunoprecipitation

**CREB** Cyclic-amp responsive element binding protein

**CRISPR** Clustered regularly interspaced short palindromic repeats

**CTD** C-terminal domain

**CTE** Carboxyl terminal extension

**DBD** DNA-binding domain

**DCIS** Ductal carcinoma in situ

**DEGs** Differentially expressed genes

**DEX** Dexamethasone

**DFS** Disease-free survival

**DHT** 5- $\alpha$ -dihydrotestosterone

**DNMT** DNA methyl transferase

**DNMT3b** DNA methyl transferase 3b

**E2** Estrogen, 17 $\beta$ -oestradiol

**EcR** Ecdysone receptor

**EMT** Epithelial to mesenchymal transition

**ER-** ER-negative

**ER** Estrogen receptor

**ER+** ER-positive

**EREs** estrogen response elements

**ERKO** ER-knockout

**ER $\alpha$**  Estrogen receptor alpha, ESR1

**ER $\beta$**  Estrogen receptor beta, ESR2

**ESCC** Esophageal squamous cell carcinoma

**FOG1** Friend-of-GATA proteins 1

**FOXA1** Forkhead box protein A1

**FUSCC** Fudan university shanghai cancer centre

**GO** Gene ontology

**GOF** Gain of function

**GR** Glucocorticoid receptor

**GREs** Glucocorticoid receptor elements

**GSEA** Gene set enrichment analysis

**H** Hinge region

**HAT** Histone acetyltransferase

**HDAC** Histone deacetylase

**HDR syndrome** Hypoparathyroidism, sensorineural deafness and renal disease

**HER2** Human epidermal growth factor receptor-2

**HREs** Hormone response elements

**HSP** Heat shock protein

**IDC** Invasive ductal carcinoma

**IF** Immunofluorescence

**IGV** Integrative genomics viewer

**ILC** Invasive lobular carcinoma

**IP** Immunoprecipitation

**Kb** kilobases

**KDa** Kilo dalton

**KO** Knock out

**LAR** Luminal androgen receptor

**LBD** Ligand binding domain

**LCIS** Lobular carcinoma in situ

**LCR** Locus control region

**LOF** Loss of function

**M** Mesenchymal

**MA** Molecular apocrine

**MaSC** Mammary stem cell

**MECs** Mammary epithelial cells

**MET** Mesenchymal to epithelial transition

**METABRIC** Molecular taxonomy of breast cancer international consortium

**miR** MicroRNA

**MMTV** Mouse mammary tumor virus

**MMTV-Luc** Mouse mammary tumor virus promoter with a luciferase reporter

**MR** Mineralocorticoid receptor

**MSL** Mesenchymal stem Like

**NGS** Next generation of sequencing

**NLS** Nuclear localisation sequence

**NR** Nuclear receptor

**NTD** Amino terminal domain

**OS** Overall survival

**P4** Progesterone

**PCA** Principle component analysis

**PCa** Prostate cancer

**PCR** Polymerase chain reaction

**PD** Progression of disease

**PDX** Patient-derived xenograft

**PLA** Proximity ligation assay

**PolyG** Polyglycine

**PolyQ** Polyglutamine

**PPI** Protein-protein interaction

**PR** Progesterone receptor

**PRE** Progesterone response element

**PTMs** Post translational modifications

**RAR** Retinoic acid receptor

**RIME** Rapid immunoprecipitation mass spectrometry of endogenous protein

**RIPA** Radioimmunoprecipitation assay

**RNA-seq** RNA-sequencing

**ROR** Receptor tyrosine kinase-like orphan receptor

**SARM** Selective androgen receptor modulator

**SERD** Selective estrogen receptor degrader

**SERM** Selective estrogen receptor modulator

**TAD** Transactivation domains

**TCGA** The cancer genome atlas

**TEB** Terminal end bud

**TF** Transcription factor

**T<sub>H</sub>2** T helper cell type 2

**TNBC** Triple negative breast cancer

**TR** Thyroid hormone receptor

**TSS** Transcriptional start site

**VDR** Vitamin D receptor

**Wt** Wild-type

**ZnFn** Zinc finger motif

## Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Leila Hosseinzadeh;

March, 2021

## Acknowledgments

This thesis would not have been possible without the support and guidance of numerous people.

First and foremost, I would like to express my deep and sincere gratitude to my Principal Supervisor Assoc. Prof. Theresa Hickey for her continuous support of my PhD study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me throughout my research studies and writing of this thesis. I could not have imagined having a better supervisor and mentor for my PhD journey.

I wish to express my gratitude to the rest of my thesis committee: Firstly, Prof. Wayne Tilley for giving me the opportunity to conduct research in his fabulous “Dame Roma Mitchell Cancer Research Laboratories” (DRMCRL) research group and for providing invaluable guidance throughout this period, for his insightful comments, and the challenging questions; and secondly, to Dr Amy Dwyer for her valuable edits on my drafts and publications. Her dynamism, vision, continuous positivity and endless motivation have deeply inspired me.

I would also wish to express my deepest appreciation to Dr Iza Denis for helping with the initial experimental designs of this project.

I am exceptionally grateful to Prof. Richard Iggo (University of Bordeaux, Bordeaux, France) for his expertise and advice in generating several transfected breast cancer lines and his ideas and enthusiasm for my PhD project.

I wish to thank Professor Elgene Lim (Garvan Institute, Sydney, Australia) for providing the *in vivo* models (patient-derived xenografts) and Professor Jason Carroll (University of Cambridge, Cambridge, United Kingdom) for collaborating with DRMCRL in generating the RIME proteomic data, respectively.



I would like to extend my gratitude to all my research colleagues at the DRMCRL including Elizabeth Kuczek, Stephen Martin Pederson, Geraldine Laven-Law, Zoya Kikhtyak and Marie Pickering who aided my PhD project.

On a personal note, I wish to thank my best friends in our lovely office, Ebtihal Hashem Mustafa, Raj Kumar Shrestha and Mohammadreza Alizadeh Ghodsi for being so kind, supportive and helpful and for all our great chats and enjoyable moments.

Reza, thank you for being such a fantastic brother and for all your unforgettable kindness.

My appreciation also goes out to my friends, Vida, Paniz and Anita for all our chats and laughter that made Adelaide even more beautiful to me.

I wish to acknowledge my spouse, Dr Mohammad R. Ardani deeply from my heart, for always being so encouraging and supportive, for his endless love, shiny smile and warm heart that inspire me to be strong and successful all the time.

Very special thanks must also go to my family: My parents for their endless love and unsurpassed support throughout my life, I would not be me without you, and to my little sister, for being such a sweet angel in my sky.

Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

## Conference abstracts and presentations

- 1) **Leila Hosseinzadeh**, Amy R. Dwyer, Stephen M. Pederson, Geraldine Laven-Law, Zoya Kikhtyak, Elgene Lim, Wayne D. Tilley, Theresa E. Hickey, “AR interacts with GATA3 to induce a luminal epithelial phenotype in breast cancer through regulation of lineage-restricted genes”, E-poster + talk, *Lorne Cancer Conference*, Australia, February, 2021
- 2) **Leila Hosseinzadeh**, Iza Denis, Wayne Tilley, Theresa Hickey, “The GATA3 transcription factor mediates AR signalling in ER-positive Breast Cancer”, Poster, *Annual Florey Postgraduate Research Conference*, Australia, September, 2019
- 3) **Leila Hosseinzadeh**, Iza Denis, Wayne Tilley, Theresa Hickey, “Role of the GATA3 transcription factor on AR signalling in ER-positive breast cancer”, Poster, *Australian Society for Medical Research (ASMR)*, Australia, June, 2019
- 4) **Leila Hosseinzadeh**, Iza Denis, Wayne Tilley, Theresa Hickey, “GATA3 transcription factor in AR signalling in ER-positive breast cancer”, Poster, *Adelaide Protein Group (APG)*, Australia, June, 2019

## Conference attendance

- 1) *Lorne Cancer conference*, virtual, Feb 2021
- 2) *15<sup>th</sup> Indo-Australian Biotechnology Conference*, Nov 2019, held by Dame Roma Mitchell Cancer Research Laboratories (DRMCRL), Adelaide, South Australia, Australia
- 3) *Adelaide Cancer Research Symposium*, March 2019, held by Dame Roma Mitchell Cancer Research Laboratories (DRMCRL), Adelaide, South Australia, Australia
- 4) *7<sup>th</sup> International PacRim Breast and Prostate Cancer Meeting*, Mar 2019, Barossa Valley, Adelaide, South Australia, Australia

# Chapter-1: Introduction

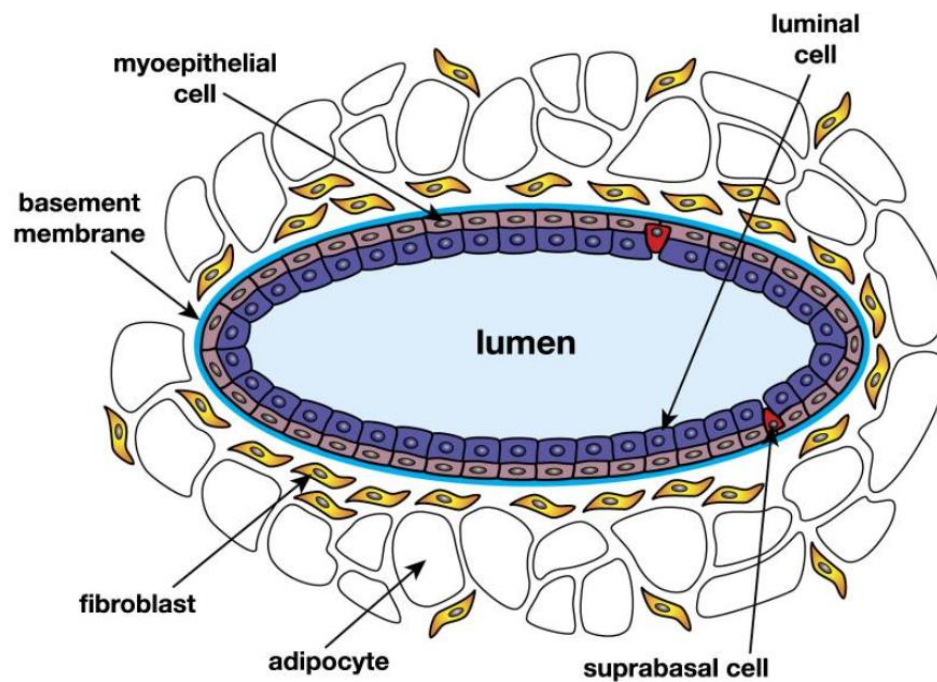
## 1. Chapter-1: Introduction

### 1.1. Mammary gland development and breast cancer

#### 1.1.1. Mammary gland development

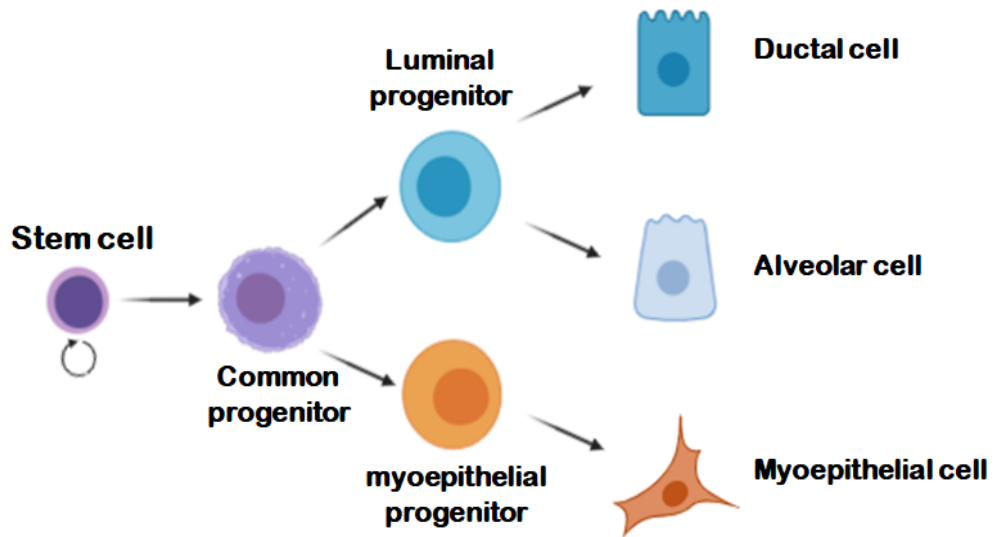
The human mammary gland is a dynamic glandular organ composed of two major mammary epithelial cell lineages. The inner layer of polarized luminal epithelial cells lines the lumen of the mammary ducts and alveoli, while basal cells form the outer myoepithelial layer ([Figure 1](#)) (Cardiff & Wellings 1999; Smalley & Ashworth 2003). During lactation, production of milk depends on the action of the ductal and alveolar luminal cells that make up the ductal network and secrete the milk components into the ductal lumen, while myoepithelial cells allow the expulsion of milk (Inman et al. 2015). Transplantation and genetic lineage tracing studies prove there is a differentiation hierarchy of stem and progenitor cells that create and maintain the mammary epithelium in both mice and humans (reviewed in (Feng et al. 2018)). Mammary stem cells (MaSCs) exist as a very small proportion of bi- or multi-potent cells in the mammary gland. MaSCs are undifferentiated cells with self-renewal ability that can undergo differentiation into various committed progenitors and differentiated cells (Lloyd-Lewis et al. 2017). MaSCs are located in a suprabasal niche (MaSC-specific cellular microenvironment) within the mammary gland and are poised to respond to extracellular cues that regulate MaSC fate and differentiation to maintain the self-renewal and multilineage differentiation capacity (Lloyd-Lewis et al. 2017; Oakes, Gallego-Ortega & Ormandy 2014). According to the model of mammary epithelial hierarchy, a common progenitor cell becomes committed to the luminal or myoepithelial lineage, then further differentiates into differentiated ductal, alveolar or myoepithelial cells ([Figure 2](#)) (Lloyd-Lewis et al. 2017; Shackleton et al. 2006). Different mammary cell lineages are identified by expression of specific markers such as cytokeratin

(CK)18, CK8, and CK19 for luminal cells, and CK5, CK14 and p63 for myoepithelial cells (O'Hare et al. 1991). During early development, the polarized luminal epithelial cells end in terminal ducts with their associated acinar structures, termed the terminal end buds (TEBs), comprising an outer layer of cap cells (myoepithelial progenitors) and a multilayer inner core of body cells, which contains the luminal cell progenitors. During puberty, these ducts invade and elongate through the fatty stroma in a process known as branching morphogenesis ([Figure 3](#)) (Chou, Provot & Werb 2010). The mammary stroma consists of a heterogeneous cell population of fibroblasts, adipocytes, macrophages, eosinophils, and mast cells, which collectively play an important role in facilitating branching and ductal elongation (Chou, Provot & Werb 2010). During the lifetime of the female, the mammary gland epithelium and surrounding stroma continually undergo drastic remodelling in structure and function during oestrous cycling (mice) or menstrual cycling (human), pregnancy and lactation, allowing the mammary epithelium to continually adapt to meet its physiological requirements. During these reproductive stages, mammary cells retain many properties important during development including proliferation, invasion, differentiation, resistance to apoptosis, and angiogenesis that are also associated with breast carcinogenesis and tumour progression (Inman et al. 2015).



**Figure 1: Ductal structure (Visvader 2009)**

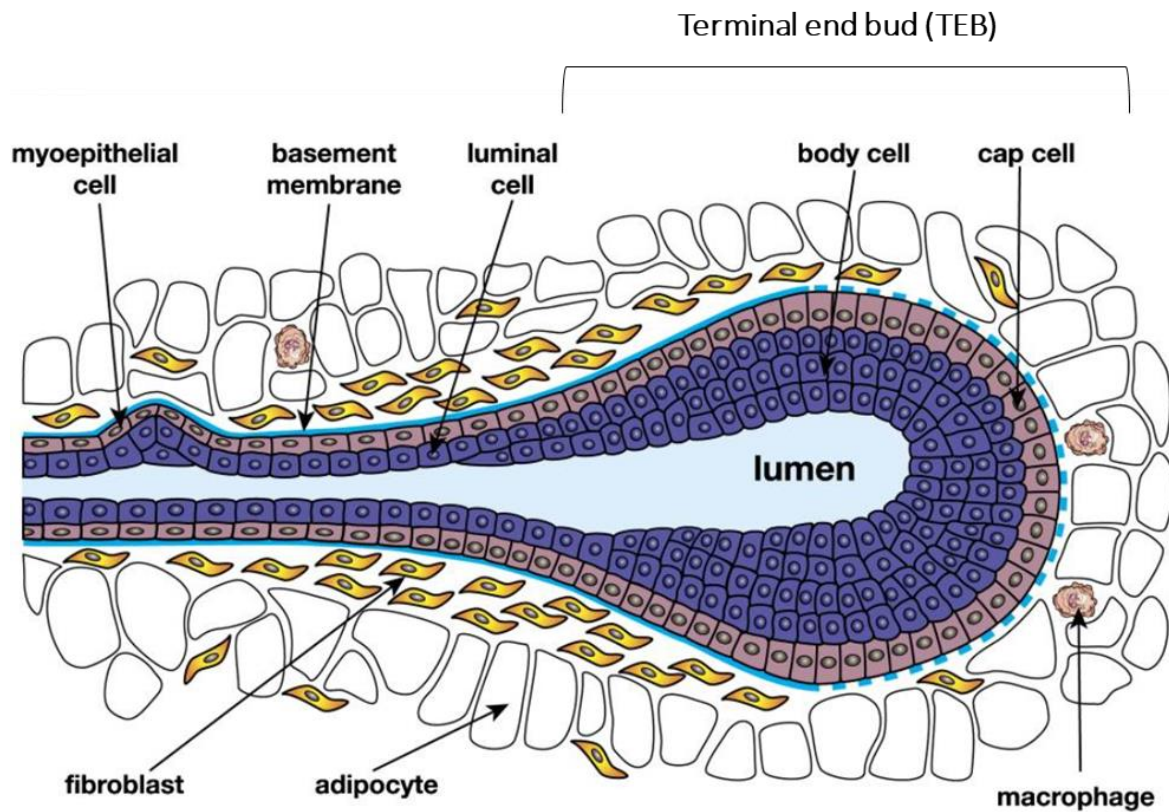
Image shows the two different epithelial cell subtypes resident in the mammary ducts, including luminal and myoepithelial cells. These epithelial cells are surrounded by the basement membrane, which separates them from the stroma containing a heterogeneous cell population of fibroblasts, adipocytes, macrophages, eosinophils, and mast cells that play an important role in facilitating branching and ductal elongation. Suprabasal stem cells are located in the myoepithelial layer, but do not reach the lumen. The ductal lumen is lined with luminal epithelial cells.



**Figure 2: Human breast epithelial hierarchy and cell type differentiation within the mammary epithelium (Created with BioRender.com)**

The hierarchical organisation from a breast stem cell and progenitor cells is shown schematically. MaSCs are bi- or multi-potent cells with the capability of self-renewal by dividing. Luminal and myoepithelial progenitor cells are derived from a common progenitor cell. Transformation into luminal and myoepithelial progenitor lineage occur(s) subsequently, whereby the luminal progenitor cells differentiate into either ductal or alveolar cells and the myoepithelial progenitors transform into mature myoepithelial cells.





**Figure 3: Schematic representation of a terminal end bud (Visvader 2009)**

Schematic of a developing mammary duct and its corresponding stroma. The mammary ducts are lined by epithelium that is composed of an inner luminal epithelial cell layer and an outer myoepithelial cell layer. The immature ducts terminate in structures called the terminal end buds (TEBs), which invade through the fat pad during pubertal development. The cap cells at the tip of the TEB differentiate into myoepithelial lineage on the outer side of the TEB. The body cells generate transit cells of a luminal epithelial lineage to form the central TEB mass. As the TEB invades into the fat pad, the ductal lumen is progressively formed as central body cells undergo apoptosis and the remaining outer cells differentiate into luminal epithelial cells.

### **1.1.2. Breast cancer incidence and mortality**

Breast cancer is the most common malignancy worldwide, affecting millions of women annually in both developed and developing countries (Bray et al. 2018). It is estimated that 19,974 new cases of breast cancer (167 males and 19,807 females) will be diagnosed in 2020, in Australia (Australia 2020). Furthermore, 3,031 deaths from breast cancer will likely be recorded by the end of 2020 (33 males and 2,997 females) (Australia 2020). In 2016, breast cancer was the second most commonly diagnosed cancer in Australia and the most commonly diagnosed cancer among females. However, according to WHO, breast cancer has now overtaken lung cancer as the world's mostly commonly-diagnosed cancer up to 2020 (WHO 2020). This indicates that despite advances in treatment, breast cancer is still one of the leading causes of cancer-related mortality among women, and that its incidence is continuously rising (Australia 2020).

### **1.1.3. Histological classification of breast cancer**

Breast cancer occurs in any of the cells comprising the mammary gland and exhibits a wide scope of morphological features, different immunohistochemical profiles, and unique histopathological subtypes that have a specific clinical course and outcome. Based on which cell-of-origin is involved, breast cancers can be divided into two broad classifications: carcinomas and sarcomas. Carcinomas are breast cancers arising from the epithelial component of the breast and sarcomas are a much rarer form of breast cancer (<1 % of primary breast cancer) arising from the stromal components of the breast. Most breast cancers are carcinomas, which are further subclassified into non-invasive (or *in situ*), locally invasive (90-95 % of all breast cancer types), or metastatic, based on their

invasiveness relative to the primary tumour sites. Two main types of breast tissues are ducts and lobules. *In situ* breast cancers classified into two groups of “Ductal carcinoma in situ, (DCIS)” and “Lobular carcinoma in situ, (LCIS)”. DCIS, or intra-ductal carcinoma, is breast cancer in the lining of the milk ducts that has not yet invaded nearby tissues. LCIS or lobular neoplasia arise from the lobules of the milk-producing glands of the breast, but are limited within the walls of the lobules. Invasive carcinomas are composed of two groups: Invasive Ductal Carcinoma (IDC), which accounts for ~80 % of all breast cancers (most common type of breast cancer), and Invasive Lobular Carcinoma (ILC), that accounts for approximately 10–15 % of all breast cancers (reviewed in (Feng et al. 2018)).

#### **1.1.4. Molecular classification of breast cancer and cellular origins**

Breast cancer is a heterogeneous disease with respect to its molecular and cellular attributes. Therefore, establishing tumour classifications that could be clinically useful in terms of prognosis or prediction of the disease is very challenging. Breast cancers also vary in terms of prognosis and clinical outcomes (Ellsworth et al. 2017; Perou et al. 2000). Based on the fact that the biological diversity of tumours may be derived from differing transcriptional programs, gene expression profiling by microarray was introduced for the first time by Perou *et al.*, in 2000, to provide prognostic information beyond standard histological assessment (Perou et al. 2000). Very soon after that, Sorlie *et al.*, (2001) refined the molecular classification of breast carcinomas into 4 main categories (Normal-like, Luminal epithelial, Basal-like and human epidermal growth factor receptor 2 (HER2)-overexpressing) based on the variation in gene expression patterns derived from complementary DNA (cDNA) microarrays (Sørliie et al. 2001; Sorlie et al. 2003). With the

advance of gene expression profiling techniques, intrinsic molecular subtypes have been further refined for better understanding of the biological complexity of breast cancer and to increase its clinical utility. In 2009, Parker *et al.*, developed the PAM50 gene set, which added significant prognostic and predictive value to standard clinical variables (including tumour size, histologic grade and node status) (Parker et al. 2009) and classified breast tumours into Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal-like subgroups ([Figure 4](#)) (Parker et al. 2009).

Luminal A (accounting for 50-60 %) and Luminal B (accounting for 15-20 %) of all breast cancers express the estrogen receptor (ER) and are the most differentiated breast cancer subtypes with the best prognoses. The luminal A subtype also expresses the progesterone receptor (PR), but does not over-express HER2, and has low levels of Ki-67 (a proliferative marker), so they are more indolent with a slower growth over time in comparison with all the other molecular subtypes. Luminal A tumours also show less frequent and less extensive lymph nodal involvement. Conversely, the luminal B subtype has variable expression of PR (low level of expression or no expression (+/-)) and is sometimes HER2+ (overexpressed with or without gene amplification), and has higher expression of Ki-67, consistent with their faster tumour growth compared to luminal A. These molecular variations partially explain why the luminal B subtype is associated with a worse prognosis compared with luminal A. Also, luminal B tumours are associated with a greater potential risk of local-regional recurrence compared to luminal A (Fragomeni, Sciallis & Jeruss 2018; Reis-Filho & Pusztai 2011). ER is considered the main oncogenic driver of Luminal A and Luminal B breast cancers. Therefore, targeting ER is the main therapeutic strategy. The anti-estrogen Tamoxifen has been clinically used over the last 30 years as a central treatment component of ER+ breast cancer subtypes (Castrellon 2017).

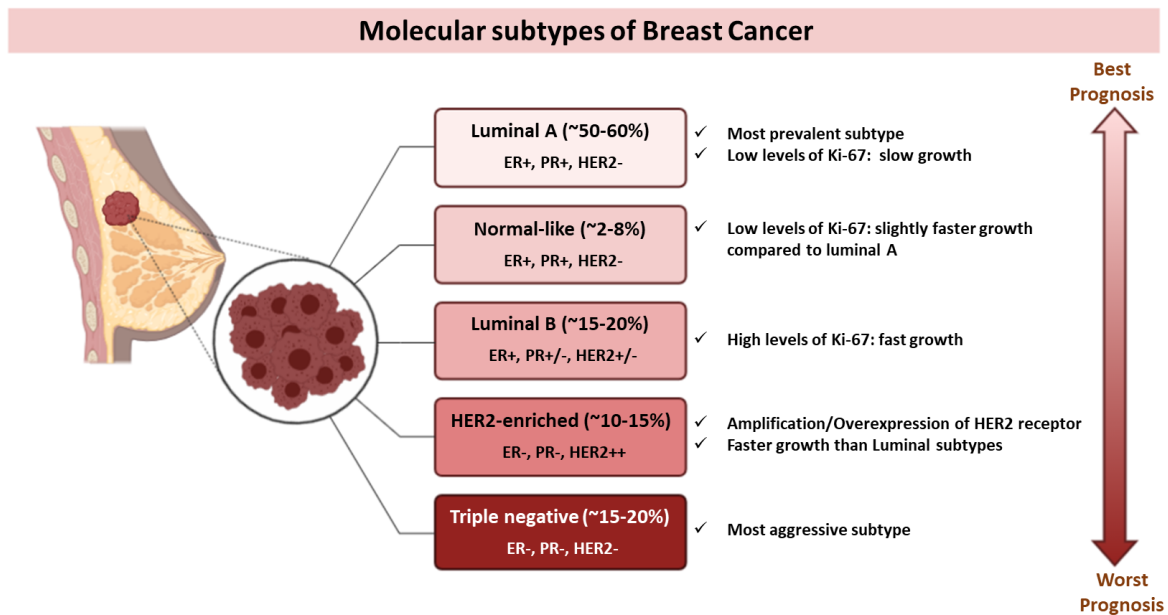
More recently, aromatase inhibitors (AI) have also become standard-of-care, especially in post-menopausal women (EBCTCG 2015).

The HER2-enriched subtype makes up 10-15 % of all breast cancers. These are characterized by the absence of ER and PR expression, high expression of HER2 and proliferation-associated genes, and low expression of the luminal and basal genes. HER2-enriched cancers grow faster than luminal cancers and have a worse prognosis. They can be treated with targeted therapies aimed at the HER2 protein, such as Herceptin (or Trastuzumab) (reviewed in (Feng et al. 2018)).

Basal-like breast cancers (15-20 % of all cases) are characterised as ER-negative, PR-negative and HER2-negative and are consequently termed Triple Negative Breast Cancer (TNBC). TNBC behaves more aggressively than other breast cancer subtypes, making it a high-grade breast cancer with very poor outcomes. Unlike other breast cancer subtypes that have the possibility of targeted therapies, TNBC's non-surgical treatment has been limited to conventional chemotherapy, until recent approval of the PARP inhibitor Olaparib for BRCA1 and BRCA2 mutation carriers, who are more likely to develop TNBC (Feng et al. 2018).

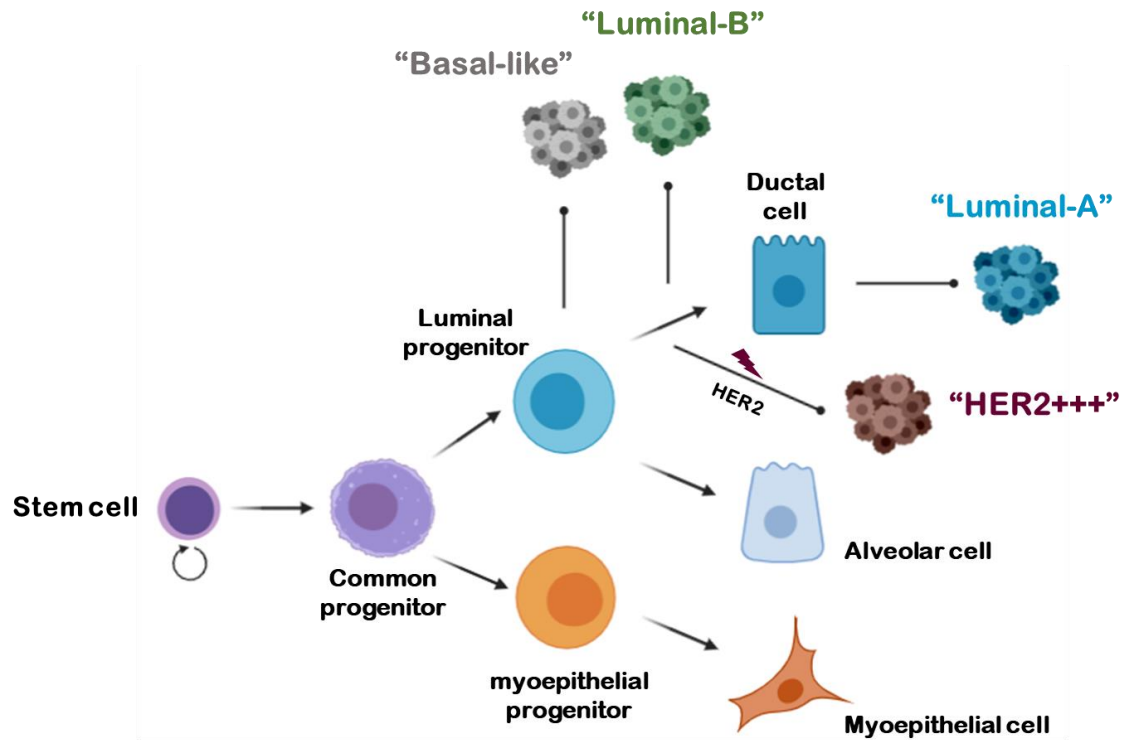
The normal-like subtype has high expression of genes known to be expressed by adipose tissue and other non-epithelial cell types. This subtype is similar to luminal A, with the same pattern of receptor expression (ER+, PR+, HER2-) and low levels of Ki-67. Although luminal A and normal-like breast cancer subtypes are associated with the best disease outcomes, the prognosis of normal-like breast cancer is still slightly worse than luminal A, maybe due to the different gene expression patterns or lack of PR (ER+, PR±, HER2-, Low Ki67) (Feng et al. 2018).

Even though molecular subtyping of breast tumours predicts the outcome of the disease, patient response to a specific targeted therapy or to chemotherapy remains variable, possibly due to intra-tumoural heterogeneity and the evolution of treatment resistance. Also, each breast cancer subtype originates from distinct breast epithelial lineages that serve as the 'cell-of-origin' ([Figure 5](#)). Therefore, a better understanding of both breast tumour heterogeneity and cellular origin of the tumour cells may improve breast cancer therapies (Visvader 2009).



**Figure 4: Molecular classification of breast cancer subtypes (Created with BioRender.com)**

Breast cancer is a heterogeneous disease and the intrinsic biological subgroups have been identified and are classified into five molecular subtypes with different prognoses as illustrated: 1) Luminal A; 2) Normal-like 3) Luminal B; 4) HER2-enriched; and 5) Basal-like.



**Figure 5: Potential relationships between stem and progenitor cells and different breast cancer subtypes (Created with BioRender.com)**

The diagram depicts the proposed association between the cell-of-origin for different tumour types with their closest normal epithelial counterpart based on gene expression analyses (based on information from (Visvader 2009)).



### 1.1.5. Endocrine therapy for ER-positive breast cancer

Luminal breast cancers (~80 % of all cases) are considered 'hormone responsive' as defined by expression of ER, PR, or both. Drugs that prevent ER activity, generally known as endocrine therapy, are standard-of-care treatment for ER-positive (ER+) breast cancer patients (Castrellon 2017). One of the first anti-estrogen strategies to directly target ER are a class of therapies called selective estrogen receptor modulators (SERMs), such as Tamoxifen. These drugs compete with estrogens by binding to ER with high affinity. This leads to a conformational change in the structure of the ER protein, preventing ER transcriptional activity in breast epithelial cells (Fisher et al. 2005). Evidence has shown that ER+/PR+ patients benefit from targeted endocrine therapies more than ER+/PR- patients (Higgins & Stearns 2009); the Tamoxifen response rate in ER+/PR+ disease is about 80 %, however for ER+/PR- this is decreased to approximately 40 % (Higgins & Stearns 2009). It has been shown that 5 years of adjuvant Tamoxifen therapy alone in ER+ patients reduces the annual breast cancer death rate by 31 %, independent of the use of chemotherapy, patient age, PR status, or other tumour characteristics such as nodal status (EBCTCG 2005). Despite its clinical success, resistance to Tamoxifen therapy is common. About two thirds of Tamoxifen-resistant tumours maintain ER expression and their growth still depends on ER function (Ali et al. 2016; Kilker et al. 2004). Therefore, other hormone therapies are used to control tumour growth. AIs prevent estrogen production by blocking the aromatase enzyme from converting androgen pre-cursors into estrogen. AIs, including Anastrozole and Letrozole, are considered more effective than Tamoxifen in postmenopausal women because they reduce the risk of recurrent disease, especially at distant sites (Thürlimann et al. 2005). For instance, in metastatic ER+ breast tumours, Anastrozole is currently considered the first-line treatment for postmenopausal women

(Miller 2003). However, Due to the large amount of aromatase substrate present in the ovary of premenopausal women, treatment with AIs is less effective in inhibiting ovarian estrogen production in these women (Winer 2005). Fulvestrant is one of a newer generation of ER antagonists known as selective estrogen receptor degraders (SERDs). Fulvestrant is a complete degrader and works by binding to the ER, inducing destabilization and degrading the ER protein (Howell 2005; Howell et al. 2004). Data from two phase III trials demonstrated that Fulvestrant 250 mg (once-monthly intramuscular injection) is at least as effective as Anastrozole (1 mg/day orally) in the treatment of postmenopausal women with Tamoxifen-resistant advanced breast cancer and is well tolerated (Osborne et al. 2002; Robertson et al. 2003). Despite the benefits of these different endocrine therapies, all are associated with development of resistance where the treatment becomes non-effective and patients undergo relapse months or years after the initial treatment. Endocrine resistance is a major clinical issue and the cause of most breast cancer-related death.

#### **1.1.6. Endocrine resistance**

Although many patients benefit from endocrine therapy, *de novo* (primary endocrine resistance) and/or acquired resistance occurs in approximately 40 % of all patients with ER+ breast cancer (Davies et al. 2011). *De novo* resistance is described as relapse in the first 2 years of adjuvant endocrine therapy or progression of disease (PD) in the first 6 months of first-line endocrine therapy for metastatic ER+ breast cancer. Acquired (secondary) resistance is described as relapse while on adjuvant endocrine therapy after the first 2 years, or PD after 6 months from starting endocrine therapy for

metastatic ER+ breast cancer (Cardoso et al. 2014). Several mechanisms of both types of endocrine resistance have been hypothesized:

- 1) Loss of ER expression (15-20 % of resistant patients). Such tumours are no longer driven by estrogen, so it is obvious that the lack of ER expression will result in *de novo* resistance to antiestrogen therapy (Fan, Chang & Fu 2015; Osborne & Schiff 2011).
- 2) ER mutations resulting in constitutive activity of ER or reduced Tamoxifen binding affinity. These mutations allow tumour growth and confer resistance to Tamoxifen. Although ER mutation rates appear to be rare in primary breast cancers (TCGA 2012), they are more frequent among metastatic and pre-treated ER+ breast cancers (Murphy & Dickler 2016; Toy et al. 2013). Also, ER mutations emerge after prolonged periods of AI therapy in over 20 % of patients with early-stage disease years or decades after diagnosis (Davies et al. 2011).
- 3) Upregulation of growth factor pathways such as EGFR, HER2, PI3K-AKT-mTOR and RAF/MEK/ERK (Masoud & Pagès 2017; Murphy & Dickler 2016; Osborne & Schiff 2011).
- 4) Cell cycle checkpoint alterations. For instance, amplification of Cyclin D1, which is involved in progression through the G1-S phase, is a common event in ER+ tumours, identified in 29 % of luminal A cancers and 58 % of luminal B cancers (Murphy & Dickler 2016).

There is widespread on-going investigation aimed at identifying new therapeutic strategies to circumvent the emergence of endocrine resistance and to minimize side effects. For instance, recent studies indicate that reprogramming, rather than blocking ER

activity, could be an alternative strategy for primary ER+ disease (Lim et al. 2016; Mohammed et al. 2015; Peters et al. 2009). This new strategy will be discussed further in later sections.

## **1.2. Nuclear receptors and cancer**

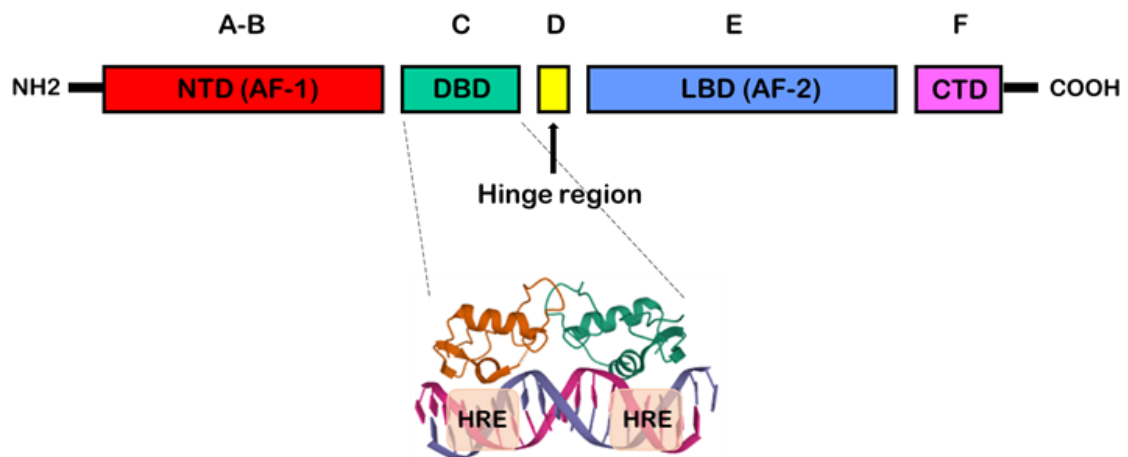
### **1.2.1. Nuclear receptor structure and classifications**

Estrogens, progesterone, and androgens are sex steroid hormones that act biologically through binding to specific intracellular nuclear receptors (NR). NRs are a large superfamily of ligand-activated TFs that directly bind to a specific DNA sequence near target genes to modulate their transcriptional activity. Some NRs (e.g. ER, PR, and androgen receptor (AR)) have key roles as pro-proliferative TFs that can drive tumorigenesis (Stallcup et al. 2003). Almost all NRs are structurally similar, having distinct well-defined functional domains (Figure 6): 1) the amino-terminal domain (NTD) (A-B region), which is highly variable in both sequence and size, has an unknown 3D structure and is involved in transcriptional activation by binding to transcriptional co-regulators; 2) the DNA-binding domain (DBD) (C region), which is the most highly conserved domain among NRs; 3) C-terminal ligand binding domain (LBD) (E region), which is recognised by specific ligands that bind to its hydrophobic ligand binding pocket; 4) the hinge region (H) (D region), which links the DBD and LBD domains to provide stability and flexibility to the nuclear receptor; and 5) a highly variable C-terminal domain (CTD) (F region) that is poorly understood (Henderson, Ponder & Ross 2003). The CTD may be involved in transcriptional activity, ligand-receptor interactions, or interaction with co-regulators among various NRs (Pawlak, Lefebvre & Staels 2012). The DBD recognises hormone response elements (HREs)

on DNA and is also required for the formation of receptor dimers (Edwards 2000). The hinge region is not just a flexible bridge between the DBD and LBD, but it often contains DNA minor groove binding residues at the C-terminal end of the DBD that flank HREs and directly participate in DNA binding. Therefore, the hinge region is also termed as a carboxyl terminal extension (CTE) of the DBD. The hinge region also contains a nuclear localisation sequence (NLS) comprised predominantly of lysine residues, which plays key role in translocation of the receptors from the cytoplasm to the nucleus (Edwards 2000). The LBD contains a domain termed the activation function domain-2 (AF-2). In most NRs, the AF-2 domain is essential for recruiting transcriptional co-activators (e.g. p160/SRC). AF-2 is also required for releasing co-repressor (e.g. NCOR, SMRT) interactions with co-regulators (e.g. p300/CBP and CARM1) and chromatin remodelling proteins (e.g. Histone Deacetylase (HDACs)) in a ligand-dependent manner, favouring transcriptional activation (Bain et al. 2007; Melmed et al. 2015; Rastinejad 2001). The NTD region contains a potent transcriptional activation domain called activation function domain-1 (AF-1). In contrast to AF-2, the ligand-independent AF-1 sequence demonstrates weak sequence conservation within the NR superfamily (Takimoto et al. 2003).

NRs can be divided into 3 distinct categories based on sensitivity to their ligands: 1) Ligand-regulated receptors have high affinity for their ligands and include receptors for steroid hormones, thyroid hormones, and the fat-soluble vitamins A and D. The ligands of steroid hormone receptors (e.g. glucocorticoid (GR), mineralocorticoid (MR), estrogen (ER), androgen (AR), and progesterone (PR) receptors) are exclusively synthesised from endogenous endocrine sources. These classic endocrine receptors bind to DNA as homodimers and their ligands are regulated by negative-feedback control of the hypothalamic-pituitary axis (Chawla et al. 2001; Melmed et al. 2015). These receptors are

subclassified into two classes based on dimerization capabilities, and the ability to recognise specific hormone response elements: a) Class I includes AR, PR, GR, and MR, which undergo homodimerization, and recognise structurally similar HREs due to the highly conserved sequence that comprises the DBD of each receptor (De Vos et al. 1993; Denayer et al. 2010; Henderson, Ponder & Ross 2003). Importantly, it has also been documented that AR has the ability to bind as a monomer to single half-site motifs in chromatin (Massie et al. 2007); b) Class II includes estrogen receptor-alpha (ER $\alpha$ ) and estrogen receptor-beta (ER $\beta$ ) isoforms; which bind to a distinct set of HREs and can homo or heterodimerise (Glass 1994; Pace et al. 1997); 2) Orphan receptors are so called since no endogenous ligand has been yet discovered for them, such as the Rev-Erb receptor, RORs (receptor tyrosine kinase-like orphan receptors), NGFI-B, NURR1, and NOR1 (Mazaira et al. 2018); and 3) Adopted receptors were once orphan receptors, but are now adopted by their newly established ligands. Adopted receptors that have low affinity to their ligands include PPAR, LXR, FXR, PXR, CAR, thyroid hormone (TR), retinoic acid (RAR), vitamin D (VDR) and ecdysone (EcR) receptors (Chawla et al. 2001; Rastinejad 2001).



**Figure 6: Nuclear receptor structure**

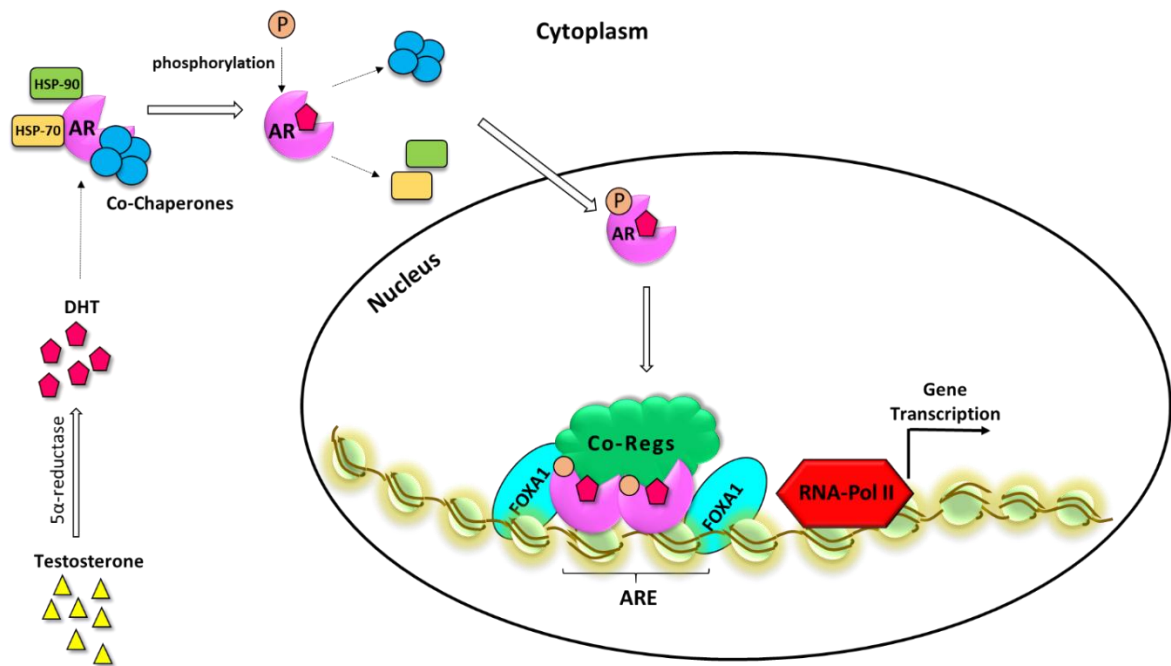
The schematic structure of a typical nuclear receptor is shown here: The N-terminal region (A-B domain) of nuclear receptors can harbor a transcriptional activation function domain known as Activation Function domain-1 (AF-1). The DNA binding domain (DBD) (C domain) functions in recognition of HREs on DNA and formation of receptor homodimers. The ligand binding domain (LBD) (E domain) interacts with the specific ligands of a receptor. Also, the second activation function domain (AF-2) is located at the C-terminal region of the LBD, which acts in recruiting the transcriptional activators in a ligand-dependent manner, receptor dimerization, and interaction with co-regulators. The DBD and LBD are linked through a hinge region (D domain). The C-terminal domain known as the F domain possesses a highly variable sequence and has unknown functions.

### 1.2.2. Nuclear receptor mechanism of action

Nuclear receptors are transcriptional regulators that function dynamically in the cell to either stimulate or repress gene transcription. The maturation of some nuclear receptors (AR, GR, and MR) occurs in the cytoplasm, where they are in a hetero complex with Heat Shock Proteins (HSP) (e.g. hsp70, hsp90) and other co-chaperones (e.g. p23) (Echeverria & Picard 2010). The first step of signal transduction is the direct binding of ligands to the LBD domain of their receptors. Ligand binding causes an allosteric change in the receptor structure, which causes dissociation of the NR from co-chaperones. This leads to stabilization and activation of the receptor, followed by dimerization and phosphorylation events leading to translocation from the cytoplasm to the nucleus (Figure 7) (Echeverria & Picard 2010). In contrast, ER and PR can undergo translocation from the cytoplasm into the nucleus in an un-liganded state (Echeverria & Picard 2010). The canonical action of ER and PR involve binding to their ligands with a concomitant dissociation from heat shock and chaperone proteins, followed by receptor homo- or hetero-dimerization (e.g. ER $\alpha$  with ER $\beta$ ; PRA with PRB) inside the nucleus (Figure 8) (Echeverria & Picard 2010). Once activated in the nucleus, the DBD domains of activated NRs recognise and bind to specific DNA response elements (such as androgen response elements (AREs), estrogen response elements (EREs), and progesterone receptor elements (PREs) for AR, ER, and PR, respectively) in the regulatory regions of target genes inside the nucleus (Echeverria & Picard 2010). The transcriptional output is mediated through the interplay between co-regulators and the transcriptional machinery, which form nuclear receptor co-regulatory binding complexes. Normal NR signalling controls the development and homeostasis of reproductive tissues, which can be co-opted in pathological conditions such as cancer (Lim et al. 2016; Perou et al. 2000). Therefore, mutations to, or aberrant

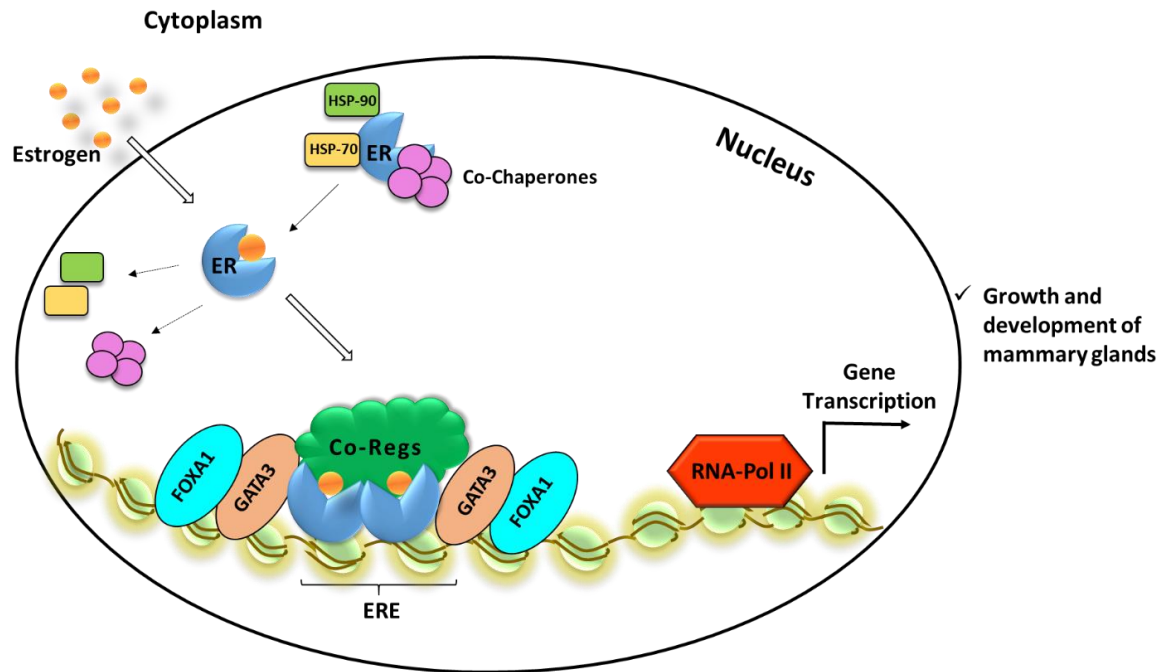


expression of, nuclear receptors and their co-regulatory factors can also participate in the development and progression of cancers.



**Figure 7: Schematic illustration of androgen receptor signalling**

Testosterone is converted to its more potent metabolite 5- $\alpha$ -dihydrotestosterone (DHT) via the 5- $\alpha$ -reductase enzyme. Both testosterone and DHT can bind to and activate AR, but DHT has greater affinity and potency. In the absence of ligand, AR is predominantly localised in the cytoplasm in an inactive multi-protein complex with heat shock proteins such as Hsp-90 and Hsp-70 and co-chaperone proteins such as p23. The canonical action of the AR involves binding to its ligand with a concomitant dissociation from heat shock and chaperone proteins, and phosphorylation of the receptor. Activated receptors translocate to the nucleus and homo-dimerize, then bind to AREs located near the target genes. AR then forms a multi-protein complex along with co-regulatory interactors and pioneer factors (e.g. FOXA1, that is known to facilitate AR binding at AREs) and recruits the transcription machinery to regulate gene transcription.



**Figure 8: Schematic illustration of estrogen receptor signalling**

In the absence of estrogen (E2), ER is predominantly localised in the nucleus in an inactive multi-protein complex with heat shock proteins such as Hsp-90 and Hsp-70 and co-chaperone proteins such as p23. The canonical action of the ER involves binding to its ligand with a concomitant dissociation from heat shock and chaperone proteins, receptor homo-dimerization and binding to EREs located within the promoter/enhancer regions of target genes. ER then forms a multi-protein complex along with co-regulatory interactors and pioneer factors (e.g. FOXA1 and GATA3 that are known ER pioneer factors) to regulate transcription.

### 1.2.3. Technology advancements in nuclear receptor function studies

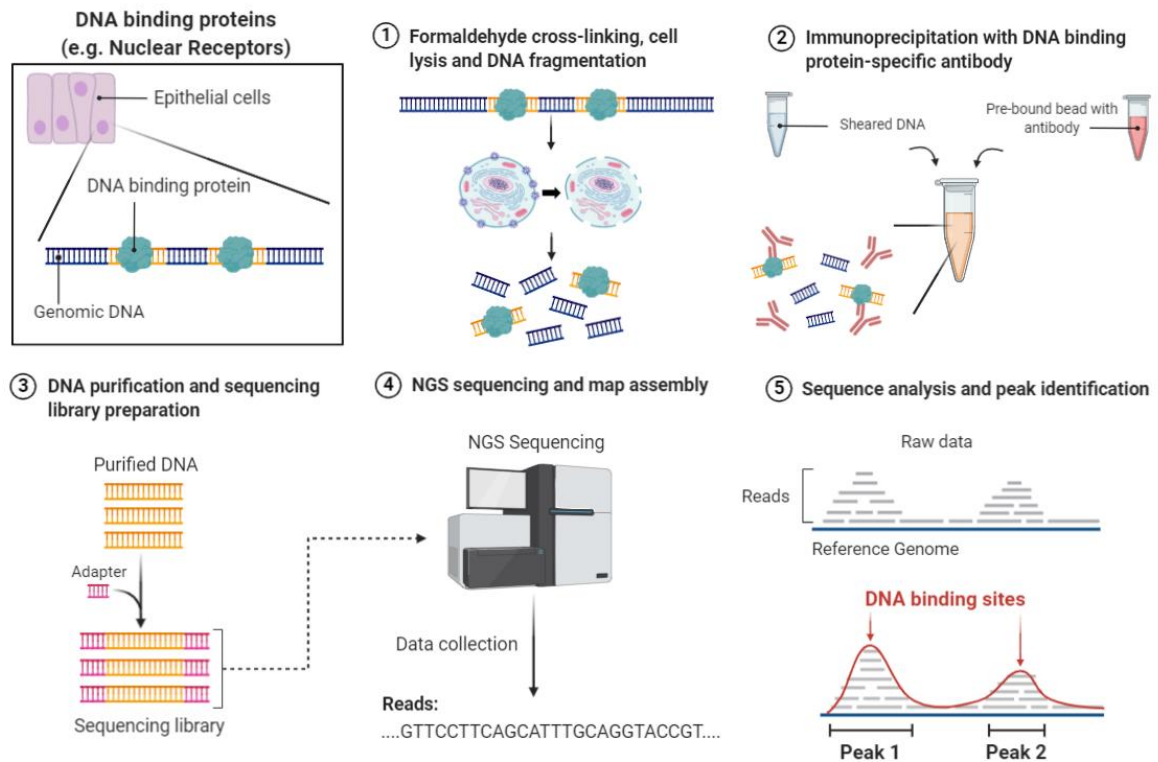
Understanding of nuclear receptor control of gene regulation has remarkably advanced during the two last decades. Contemporary technologies such as chromatin-immunoprecipitation (ChIP) combined with tiled DNA oligonucleotide microarray analysis -on-ChIP and ChIP-sequencing (ChIP-seq) ([Figure 9](#)) has allowed genome-wide mapping of NR DNA binding (cistrome) and function (Carroll et al. 2005; Carroll et al. 2006; Lupien & Brown 2009; Massie et al. 2007; Wang, Q et al. 2007). For instance, it was historically thought that ER binds to EREs located exclusively in the promoter region of target genes; however, with the advent of ChIP-seq technology, which reveals genome-wide cis-regulatory binding events of any DNA-binding factor, we now know that ER is preferentially recruited to enhancer regions of target genes upon estrogen stimulation (Carroll et al. 2005). These enhancer elements modulate target gene expression by forming chromatin loops with promoter containing chromatin regions (Sanyal et al. 2012). Similarly, it has been shown that while AR binds to AREs located in close proximity (10-15 bp) of the transcriptional start site (TSS) of its target genes, AREs are more often located in distal upstream genomic regions (100-200 kb) (Lupien & Brown 2009; Massie et al. 2007; Wang, Q et al. 2007). Furthermore, sequencing technologies have led to discovery of multiple binding motifs for other TF co-regulators in the vicinity of NR binding sites, such as Forkhead box protein A1 (FOXA1), GATA2, and Oct-1 that bind in a non-promoter specific manner in order to make regulatory loops with NRs (Carroll et al. 2005; Wang, Q et al. 2007). FOXA1 is a member of the Forkhead class of DNA-binding proteins that have been well studied in breast and prostate cancers, playing a central role in ER and AR signalling, respectively as a common ER and AR co-regulator (Lupien et al. 2008; Robinson, JL et al. 2014). It has been shown that nearly 60 % of the ER cistrome in MCF-7 (breast) and 70 %

of the AR cistrome in LN-CaP (prostate) cancer cell lines overlap with FOXA1 binding events. FOXA1 binds to compacted chromatin and opens it to facilitate chromatin accessibility to ER and AR (Lupien et al. 2008; Robinson, JL & Carroll 2012; Robinson, JL et al. 2014). Studies to date indicate that ER binding to DNA is more reliant on FOXA1 than AR binding to DNA (Bernardo et al. 2010; Carroll et al. 2005; Laganière et al. 2005; Robinson, JL & Carroll 2012; Wang, Q et al. 2007; Wang, Q et al. 2009).

RNA-sequencing (RNA-seq) is another recently-developed technique which uses next generation of sequencing (NGS) to profile the whole set of cellular RNA transcripts, termed as the transcriptome (Wang, Z, Gerstein & Snyder 2009). A huge number of studies currently use the RNA-seq approach in determining the differentially expressed genes upon various experimental conditions, and measuring the levels of transcripts and their isoforms (Wang, Z, Gerstein & Snyder 2009). Collectively, ChIP-seq and RNA-seq approaches provide valuable, comprehensive information with regards to NR cistromes and transcriptomes ([Figure 10](#)).

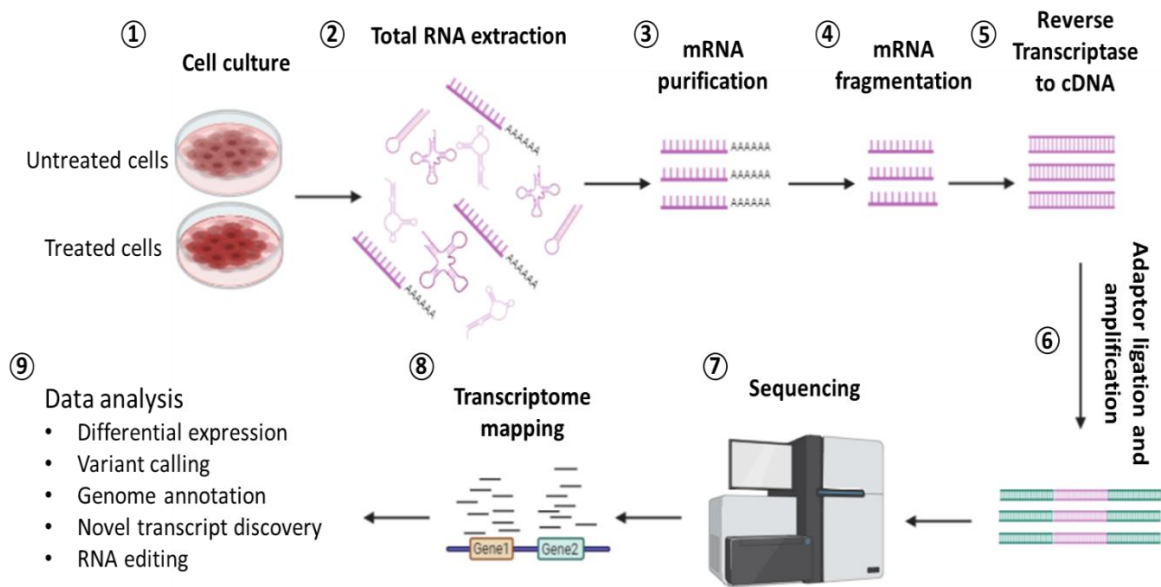
Nuclear receptors can also interact with DNA in an indirect manner through tethering to other partner proteins (Sanchez et al. 2002). Rapid immunoprecipitation mass spectrometry of endogenous protein (RIME) is a novel method that has been introduced recently to enable the study of protein-protein interactions (PPIs), especially for TF complexes on chromatin (Mohammed et al. 2016). RIME can be used in parallel with ChIP-seq experiments to provide information on both the cistrome and interactome (a complete set of molecular interactions in a particular cell, specifically protein-protein interactions (De Las Rivas & Fontanillo 2010) for a given protein of interest (Mohammed et al. 2016) ([Figure 11](#)).

My host lab, the Dame Roma Mitchel Cancer Research Laboratory (DRMCRL), has expertise in ChIP-seq, RNA-seq and RIME technologies, providing the opportunity to test my hypothesis with regards to investigation of the complexity of hormonal cross-talk between ER, AR, and GATA3 TFs in breast cancer, during my PhD journey.



**Figure 9: Chromatin-immunoprecipitation (ChIP)- sequencing (Created with BioRender.com)**

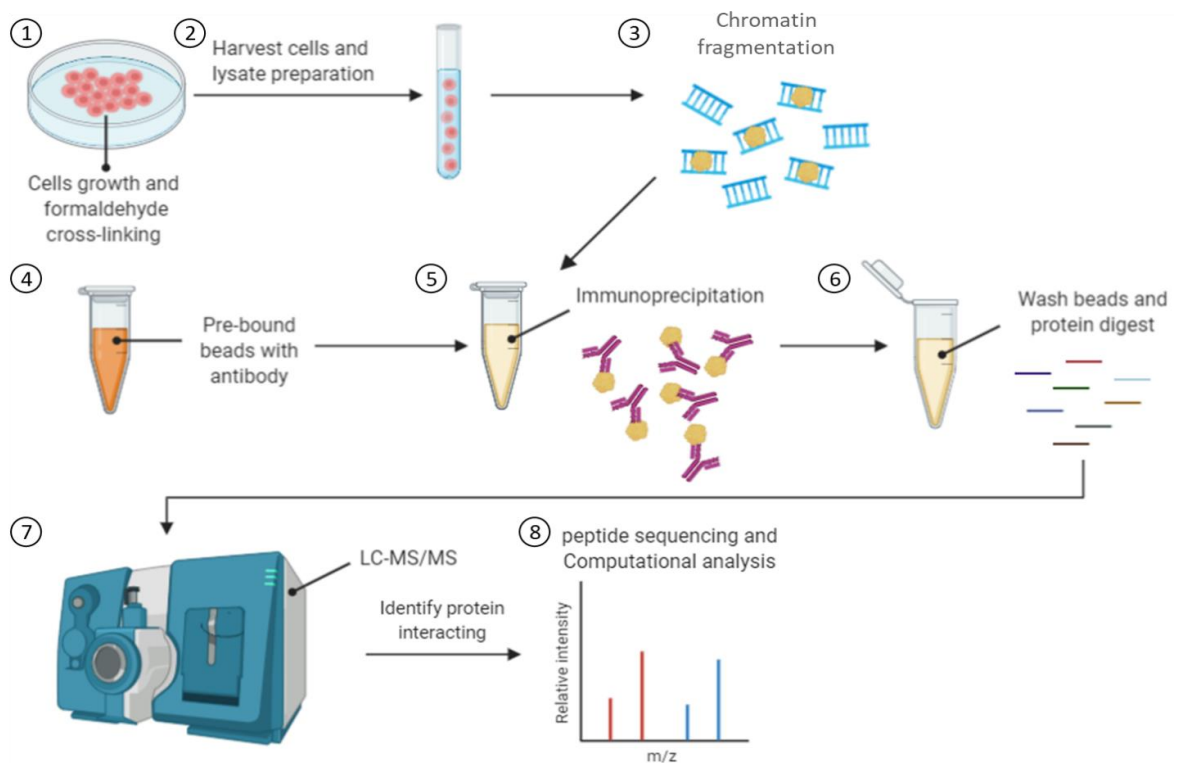
Different steps of ChIP-seq are described briefly in this figure as follows: 1) Cells are expanded and treated as per experimental protocol, cross-linked with formaldehyde and quenched with glycine. Cells are harvested, lysed, nuclear fractionated in specific buffers and pelleted; 2) chromatin is fragmented using a sonication machine and magnetic beads are prebound with the antibody against the protein of interest, then the sonicated DNA is incubated with magnetic beads prebound with antibody and immunoprecipitation (IP) is conducted overnight; 3) cross-linked chromatin is reverse cross-linked to release the DNA, which is purified and; 4) sequenced using next generation sequencing (NGS) technologies, 5) Raw fastq files are generated and bioinformatically processed to create BAM, BED, and Bigwig files to allow data visualization and analysis of peaks representing NR binding events.



**Figure 10: RNA-sequencing (RNA-seq) (Created with BioRender.com)**

RNA-seq steps are described briefly in this figure as follows: 1) Cells are grown and treated according to experimental design, and harvested; 2) total RNA is extracted; and 3) mRNAs (or sometimes total RNA) are purified; and 4) fragmented; 5) a cDNA library is generated from isolated RNA through reverse transcription; 6) Adapter sequences are ligated to the 3' and 5' end of the double-stranded cDNA and the final cDNA library is amplified by polymerase chain reaction (PCR) using parts of the adapter sequences as primers; 7) then sequenced; 8) reads are mapped against a reference genome; 9) Raw data is processed and analysed using various computational tools depending on the goal of the experiment for differential expression assessment or transcript discovery for instance.





**Figure 11: Rapid immunoprecipitation mass spectrometry of endogenous protein (RIME) (Created with BioRender.com)**

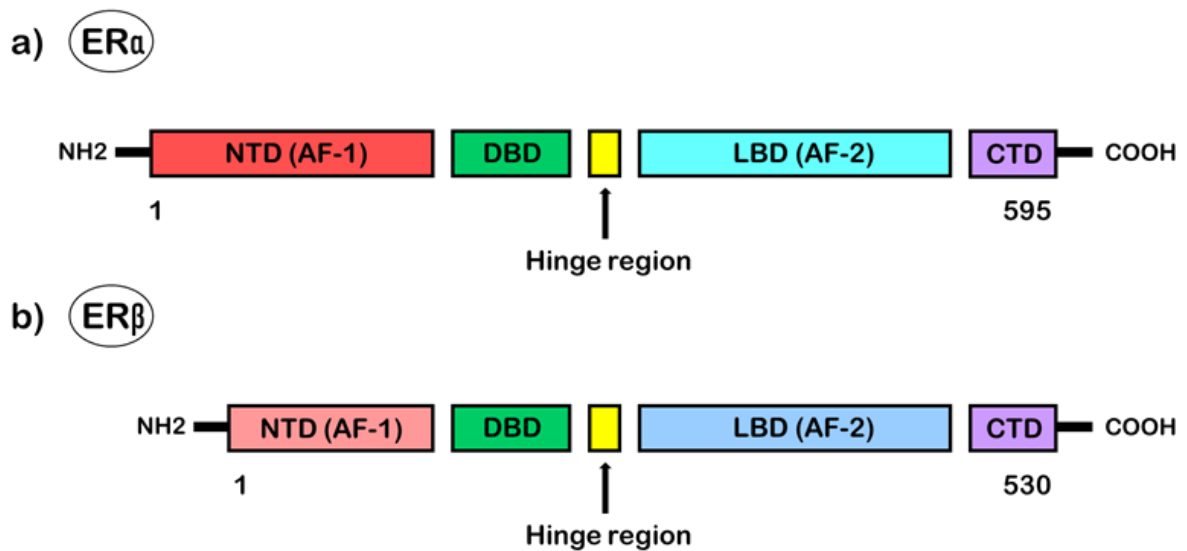
Different steps of the RIME technique are described briefly in this figure as follows: 1) Cells are grown in their normal (or experimental) condition, then cross-linked with formaldehyde and quenched by glycine; 2) Cells are washed and harvested in PBS. The nuclear fraction and pellet preparation are done in specific buffers; 3) chromatin is fragmented using a sonication machine; 4) Magnetic beads are prebound with the antibody against the protein of interest; 5) Sonicated chromatin is then incubated with magnetic beads prebound with antibody; 6) immunoprecipitation (IP) is conducted; 7) The beads are washed and bead-bound proteins are digested; 8) Digested peptide mixtures are diluted and analysed by Nano-LC-MS/MS; 9) Raw MS data files are processed to achieve the detailed results including peptide sequences, peptide scores, ion scores, expect values, and Mascot scores.

#### 1.2.4. Estrogen receptor function in normal and malignant breast tissue

Unlike other human organs, mammary gland development mainly occurs with the approach of puberty (Russo & Russo 2004). Hormones impinge on a subset of luminal mammary epithelial cells (MECs) that express hormone receptors and act as sensor cells translating and amplifying systemic signals into local stimuli. The ovarian hormones,  $17\beta$ -oestradiol (E2) (the most potent estrogen) and progesterone (P4) (a potent ligand of PR) play a pivotal role in mammary gland development, differentiation and proliferation of MECs (Brisken & O'Malley 2010). E2 exerts its biological functions through specific ligand-inducible ER $\alpha$  (Figure 12.a) and ER $\beta$  (Figure 12.b) receptors. These receptors are encoded by separate genes located on different chromosomes, yet share considerable sequence homology (Nilsson et al. 2001). The gene for ER $\alpha$  (*ESR1*) was cloned in 1986 by two groups (Green et al. 1986; Greene et al. 1986) and the gene for ER $\beta$  (*ESR2*) was cloned in 1996 (Kuiper et al. 1996). The ER $\alpha$  gene has 8 exons and is localised to chromosome 6q24-27. The ER $\alpha$  protein is composed of 595 amino acids with a molecular weight of 65 kilo dalton (KDa) (Edwards 2000). The ER $\beta$  gene is located on chromosome 14q23.2, spanning ~61.2 kb. The ER $\beta$  protein is produced from eight exons. The full-length human ER $\beta$  protein includes 530 amino acids with an estimated molecular weight of 59.2 kDa (Ogawa et al. 1998).

The growth stimulatory effect of ER $\alpha$  on normal breast epithelial cells and breast cancer cells has been well studied. For instance, an *in vitro* study showed that a selective ER $\alpha$  agonist increased the proliferation of the HC11 normal mouse breast cell line (Helguero et al. 2005). Moreover, knockout (KO) mouse studies have clearly demonstrated that ER $\alpha$  is indispensable for the postnatal structural and functional development of the mammary gland, while ER $\beta$  is not (Feng et al. 2007). Mammary glands in female ER $\alpha$ KO

mice did not develop with the onset of puberty, and TEB formation, ductal elongation and side-branching were prevented (Curtis Hewitt, Couse & Korach 2000; Feng et al. 2007). In addition to ER $\alpha$ , as a critical TF in regulating epithelial cell proliferation and ductal morphogenesis during postnatal mammary gland development, knockout mouse studies have also revealed several ER $\alpha$  co-regulators that are known to play crucial role in mammary gland development (Lonard, David M & O'Malley 2007) such as SRC-3/AIB1 (Onate et al. 1995), cAMP responsive element binding protein (CREB) (CBP/p300) and ATP-dependent chromatin remodelling complexes like SWI/SNF (Lonard, D. M. & O'Malley 2006). More than 200 co-regulators, each with a specific expression level, have been identified for ER $\alpha$  and are considered critical mediators of cell-specific regulation of ER $\alpha$  target gene expression (Smith & O'Malley 2004). Notably, the expression level of each co-regulator may influence its ability to modulate the transcriptional potential of ER $\alpha$  in response to E2 (Brisken & O'Malley 2010). The critical contribution of ER $\alpha$  to breast cancer progression has also been studied well in the literature. E2 treatment of ER $\alpha$ + breast cancer cell lines (MCF-7 and T-47D) increased proliferation, which was then reversed following treatment with the anti-estrogen Tamoxifen (Aspinall et al. 2004). Clinical studies have identified that ER $\alpha$  expression in women diagnosed with breast cancers is higher than its expression in normal epithelium of women with no disease (Khan et al. 1998). Another study has also reported an increase in the level of ER $\alpha$  expression at the early stages of breast cancer progression, with a further increase in pre-malignant lesions (Shoker et al. 1999). This thesis will focus on ER $\alpha$ , referred to herein as ER for simplicity.



**Figure 12: Schematic illustration of estrogen receptor isoforms**

The estrogen receptor has two isoforms, ER $\alpha$  (a) and ER $\beta$  (b). The DBDs of the ER isoforms have very high homology (96 %), while their LBDs are only about 59 % homologous. The main differences between ER isoforms are in the NTDs, which share only 30 % of sequence identity. The isoform ER $\alpha$  has an extended NTD, which allows for more transcriptional activation, whereas, ER $\beta$  has a shorter NTD and so a shorter AF-1, which makes the ER $\beta$  less transcriptionally active compared with ER $\alpha$ . However, some studies have shown that the AF-1 differences alone are not sufficient to explain these behaviors (reviewed at (Souza et al. 2017)).

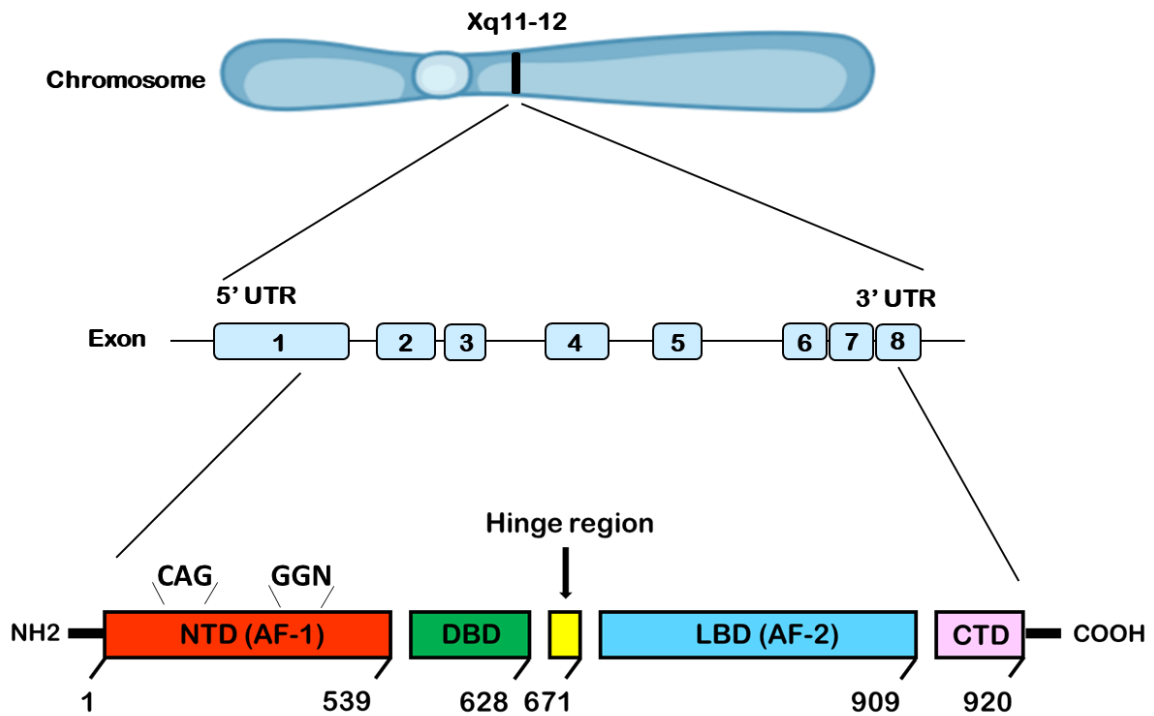
### **1.2.5. Androgen receptor function in normal and malignant breast**

In general, current understanding of androgenic activity and AR signalling has been developed through investigating the role of androgens and AR in male reproductive physiology and prostate cancer. Androgens have a crucial role in male reproductive biology, being essential for the development and maintenance of the male urogenital organs such as prostate, seminal vesicles and epididymis (Murashima et al. 2015). Despite the fact that androgens are usually considered male hormones, they are also present in the circulation of women at physiologically relevant levels (Burger 2002; McNamara, K et al. 2010), playing critical biological roles in normal breast physiology and breast carcinogenesis (Dimitrakakis & Bondy 2009; Yeh et al. 2003).

#### **1.2.5.1. Androgen receptor gene and protein structure**

The intracellular mediator of the genomic effects of androgens is the AR, located on chromosome Xq11-12 (Brown et al. 1989). The human AR gene was first cloned in the late 1980s (Tilley et al. 1989). The AR coding sequence is comprised of 8 exons, which encode approximately 917 amino acids, depending on the length of polymorphic CAG and GGN (N for any nucleotide) microsatellite regions within exon 1. The molecular weight of AR is approximately 98.8 KDa (varies depending on size of the microsatellites). Post translational modifications (e.g. phosphorylation) may increase the molecular weight of AR up to 110 KDa when analysed by SDS polyacrylamide gel electrophoresis. The full-length AR is comprised of the same functional domains as ER (Tilley et al. 1989) ([Figure 13](#)). Two key functional sub-domains of the AR are located in the NTD domain including the polyglutamine (poly-Q) and polyglycine (poly-G) tracts (Henderson, Ponder & Ross 2003).

Interestingly, it has been shown that the longer poly-Q repeat region is associated with reduced AR transactivation capacity and increased risk of developing breast cancer in women over 40 years old of age, in comparison with a shorter poly-Q repeat (Elhaji et al. 2001; Giguère et al. 2001).



**Figure 13: Schematic illustration of the androgen receptor gene and protein structures**

The full-length androgen receptor gene and protein structures are shown in this figure. The AR coding sequence is comprised of 8 exons, which encodes up to 920 amino acids depending on the length of polymorphic CAG and GGN (N for any nucleotide) regions, with an approximate molecular weight of 98.8 KDa. The full-length AR is comprised of the same functional domains as ER.

#### 1.2.5.2. Androgens and androgen receptor signalling in normal breast tissue

In contrast with the established growth stimulatory effects of estrogens in normal and malignant breast epithelial cells, direct androgen action in the normal growth and development of the mammary gland is still controversial and poorly defined. Androgens have been shown to play a protective role in female breast tissue through repression of the basal and estrogen-stimulatory effects on breast epithelial growth and proliferation (Castellano et al. 2010; Dimitrakakis & Bondy 2009; Gonzalez-Angulo et al. 2009; Peters et al. 2009). For instance, a 40 % reduction in E2-stimulated growth of mammary epithelial cells of ovariectomised rhesus monkeys has been observed following treatment with Testosterone (the major circulating androgen hormone) (Zhou et al. 2000). In a study by Peters *et al* 2009, DHT (the most potent natural AR agonist) treatment of female C57/BL6 mice from mid-puberty (5-12 weeks of age) reduced ductal extension of the mammary glands by 40 % compared with the placebo-treated mice (Peters et al. 2011). However, DHT treatment post-puberty (12-21 weeks of age) had no effect on ductal growth of the mammary gland (Peters et al. 2011). Anti-androgens reversed the anti-growth effects of androgens on normal breast epithelial cells (Peters et al. 2011). An increase in cellular proliferation and branching in response to Flutamide (an AR antagonist) treatment throughout the mammary gland of post-pubertal mice has been observed in the same study by Peters *et al* (Peters et al. 2011). Likewise, treatment of intact cycling monkeys with Flutamide also resulted in a significant increase in mammary epithelial cell proliferation (Dimitrakakis et al. 2003). The importance of direct androgen signalling in the growth and development of the normal mammary gland has been further illustrated by several AR-knockout (ARKO) female mouse studies (Shiina et al. 2006; Yeh et al. 2003). In particular, it has been shown that AR is essential for normal fertility in the female



population since ARKO mice showed dysfunctional ovulation and impaired follicular growth (Walters, Simanainen & Handelsman 2010). In a study by Yeh *et al.*, a Cre-lox systemic knockout strategy was used to generate female ARKO mice, genotyped as  $AR^{-/-}$  ACTB Cre<sup>+</sup>, in which estrogen target organs, including the mammary gland, ovary, oviduct, and uterus of adult mice with homologous deletion of AR weighed 15-23 % less as compared with their age-matched control mice (4-, 6-, and 12-week old mice) (Yeh et al. 2003). The phenotype of these pre-pubertal female ARKO mice at the age of 4 and 6 weeks demonstrated a 30-50 % reduction in ductal extension, reduced numbers and size of TEBs, and a 50 % reduction in epithelial cell proliferation compared to the AR<sup>+/+</sup> wild-type mice (Yeh et al. 2003). Also, reduced secondary and tertiary branching was observed in the mature mammary glands of 8-20-week old ARKO mice, indicating the important role of AR in mammary gland development (Yeh et al. 2003). The phenotype observed in the model used in the above study is almost exactly the same as the PR null mouse (Lydon et al. 1995), suggesting that the ovarian defects of ARKO mice could be the explanation for the observed mammary gland phenotype. Another potential explanation for the phenotype observed in the above study is a reduction in MAPK and ER signalling in the mammary glands. In an independent study, Shiina *et al.*, (2006) observed that 8-week-old female homozygous ARKO (*CMV-Cre*) mice in comparison with wild-type (Wt) female mice exhibited a substantial reduction in mammary ductal branching and elongation (Shiina et al. 2006). During pregnancy, the retarded ductal branch numbers were partially restored but still contained less milk-producing alveoli than Wt glands. These observations indicated that mammary ductal morphogenesis was impaired during pubertal development, pregnancy, and lactation in ARKO (*CMV-Cre*) mice, whereas AR functions as a stimulator of mammary ductal differentiation (Shiina et al. 2006). Therefore, in female mice, AR is

thought to be required for normal development of the mammary glands by modulating ductal branching and epithelial cell proliferation.

Although breast cancer research has mainly focused on ER, the molecular profiles of subsets of ER and PR negative primary breast carcinomas has been shown to be still hormonally regulated (Doane et al. 2006). This suggests the involvement of other steroids and/or their receptors such as the most abundantly expressed NR in breast cancer, AR.

#### **1.2.5.3. Androgen receptor expression in breast cancer**

ER+ breast cancers, which are currently defined by the detection of ER in at least 1 % of cells within the tumour biopsy (Hammond et al. 2010), are more likely to have AR expression compared with other sub-types of breast cancer (Hu, R et al. 2011; Loibl et al. 2011; Park et al. 2011), with luminal A cancers expressing AR more frequently than luminal B (Yu et al. 2011). The co-expression of ER and AR in the majority (80-90 %) of breast tumour cells suggests that the effects of estrogen and androgen on growth and survival of breast cancer cells are critically integrated (Peters et al. 2009; Santagata et al. 2014). AR has been detected in up to 85 % of primary breast cancers and up to 75 % of metastatic lesions (McNamara, KM et al. 2014; Moinfar et al. 2003; Park et al. 2010) and is typically present in a greater proportion of breast tumours (>90 %) (Agoff et al. 2003; Park et al. 2010) than ER (>75 %) (Allred, Brown & Medina 2004). Importantly, AR has been introduced as an independent predictor of breast cancer survival (Ricciardelli et al. 2018) with a favourable prognostic indication (reviewed at (Hickey, T et al. 2012; Ricciardelli et al. 2018)) that is significantly associated with reduced risk of relapse, longer survival (Yu et al. 2011) and better outcomes among patients with ER+ tumours (Park et al. 2011).

In general, ER- breast cancers are more aggressive with poorer survival prognoses compared with ER+ breast cancers (sub-chapter 1.1.4) (Barcellos-Hoff 2013). ER- breast cancers can be generally divided into HER2+ and TNBC subgroups. Molecular apocrine (MA) is a subgroup of HER2+/ER- breast cancers that is derived from luminal tumours that undergo apocrine metaplasia, a common process in normal breast tissue in which cells lose their ER expression through an unknown biological mechanism, leading to their ancestral androgen-driven, apocrine fate (Iggo, Richard 2018). Almost 12 % of all breast cancer subtypes are considered as MA tumours, characterised by high AR and a gene expression profile akin to that of ER+ luminal subtypes despite lack of ER (Doane et al. 2006; Farmer et al. 2005; Iggo, RD 2011). In order to further understand the molecular basis of TNBC, Lehmann *et al.*, assessed 21 publicly available data sets of primary human breast cancers and classified TNBC into 3 main groups: 1) Basal-like (including BL-1 and BL-2), 2) Mesenchymal-like (including mesenchymal (M) and mesenchymal stem-like (MSL), and 3) Luminal androgen receptor (LAR) (Lehmann et al. 2011). Interestingly, the LAR subgroup is heavily enriched in hormonally regulated pathways including steroid synthesis and androgen/estrogen metabolism, thought to be due to relatively higher expression of AR (at both mRNA and protein levels) in this TNBC subgroup compared to the others, and expression of downstream AR targets and coactivators (e.g. *DHCR24*, *ALCAM*, *FASN*, *FKBP5*, *APOD*, *PIP*, *SPDEF*, and *CLDN8*) (Lehmann et al. 2011). The MDA-MB-453 breast cancer cell line has been widely used as a model of the MA BC subtype due to high levels of AR and HER2 expression, however, some studies classified MDA-MB-453 as LAR because it does not have amplification of the HER2 gene (Espinosa Fernandez et al. 2020; Tseng et al. 2017). The expression of AR in MA tumours is associated with histological evidence of apocrine differentiation (Niemeier et al. 2010). The correlation between AR and patient

outcome is still unclear in ER- breast cancers. Some studies have shown that AR is correlated with improved disease-free survival (DFS) and overall survival (OS) (Luo et al. 2010; Qu et al. 2013; Vera-Badillo et al. 2014) in ER- breast cancer patients. However, others claimed that there is no clear association between AR expression and outcomes of ER- breast cancer (Hu, R et al. 2011; Park et al. 2011; Peters et al. 2009; Vera-Badillo et al. 2014). High heterogeneity within the ER- breast subtype could be an explanation for these contradictory results (McNamara, KM et al. 2014).

#### 1.2.5.4. Androgen receptor signalling in ER+ breast cancer

Androgens have been shown to inhibit both basal and estrogen-stimulated proliferation of ER+/AR+ breast cancer cell lines such as MCF-7 (Hickey, T et al. 2012; Macedo et al. 2006; Ortmann et al. 2002), ZR-75-1 (Birrell et al. 1995; Poulin, R, Baker & Labrie 1988) and T-47D (Birrell et al. 1995; Cops et al. 2008; Greeve et al. 2004; Ortmann et al. 2002). This growth inhibition can be reversed by anti-androgens or silencing AR, indicating a protective role for androgens against the development and progression of ER+ disease (Birrell et al. 1995; Cops et al. 2008; Dimitrakakis & Bondy 2009; Macedo et al. 2006). As expected, the inhibitory effect of DHT is not observed in cell lines that do not express AR, including the MDA-MB-231 and BT-20 TNBC cell lines, which indicates the essential requirement of AR expression to mediate the growth inhibitory effects of androgens on breast cancer cells (Birrell et al. 1995). Also, it has been demonstrated that AR can inhibit the transactivation activity of ER in human breast cancer cells through competitive binding of the AR to EREs associated with estrogen-regulated genes (Peters et al. 2009). Historically, non-aromatizable AR agonists such as Fluoymesterone have been used as a hormonal therapy for advanced breast cancer with an efficacy comparable with Tamoxifen (Tormey et al. 1983). Combinational therapy of Fluoymesterone and Tamoxifen was more effective than Tamoxifen alone in cases of advanced breast cancer (Ingle et al. 1991). Despite having clinical benefits, therapeutic use of androgenic agents to treat breast cancer fell from favour due to their side effects, which included increased aggressive behaviour, hirsutism and masculinization (McNamara, KM et al. 2014). Enobosarm, which is a novel well-tolerated, non-steroidal, selective AR modulator (SARM) with an agonistic effect on AR in ER+ cells, has shown promise in a phase II trial of women with advanced, hormone-sensitive breast cancer (Overmoyer et al. 2014; Yuan et al. 2020).

Taken together, these studies highlight the growth inhibitory effects of androgens through the AR signalling axis in malignant ER+ breast tissues, resulting in protection against cancer progression due to the reduction of cell proliferation and tumour growth.

#### **1.2.5.5. Androgen receptor signalling in ER- breast cancer**

Contrary to growth-inhibitory effects in the ER+ context, androgens have been shown to promote the proliferation of some ER- breast cancer cell lines, including the MDA-MB-453 model of MA breast cancer (Birrell et al. 1995; Doane et al. 2006; Ni et al. 2011). AR antagonism or AR knock down inhibits growth of this cell line, suggesting AR as having oncogenic activity (Birrell et al. 1995; Lehmann et al. 2011; Robinson, J et al. 2011). Interestingly, the global AR DNA binding events in MDA-MB-453 cells have been shown to be more similar to the ER cistrome in MCF-7 cells, compared to the AR cistrome in LN-CaP prostate cancer cells, where AR stimulates cell growth (Robinson, J et al. 2011). This suggests that when ER is absent, AR binds to and regulates ER *cis*-regulatory elements, resulting in a transcriptional programme reminiscent of ER-mediated transcription in luminal breast cancers (Robinson, J et al. 2011). This may potentially explain the oncogenic role of AR in the MA breast cancer subtype. Other proposed mechanisms for the growth-stimulatory effects of androgens in MA BC include 1) androgen-induction of *WNT7B* and *HER3* expression levels leads to activation of Wnt/ $\beta$ -catenin and HER2 signalling pathways, which are required for androgen-induced proliferation of these cells (Ni et al. 2011); 2) androgen-induced activation of PI3K/AKT pathway through activation of HER2/HER3 signalling, leads to enhanced phosphorylation of MAD1 and subsequently increases the oncogenic function of MYC (Ni et al. 2013); 3) androgen-induction of ERK phosphorylation

through activation of HER2 signalling, increases cell proliferation (Naderi & Hughes-Davies 2008). The AR-ERK feedback loop in MA tumours has been investigated as a potential therapeutic option using the AR inhibitor Flutamide and the MEK inhibitor CI-1040 *in vitro* and *in vivo* (Naderi, Chia & Liu 2011). The combination therapy showed a synergic inhibitory effect on the viability of MDA-MB-453, HCC-1954 and HCC-202 MA cells using the MTT metabolic assay (a proxy for proliferation), and tumour growth of MDA-MB-453 xenografts *in vivo* (Naderi, Chia & Liu 2011). The MFM-223 cells line is classified as a model of the LAR subtype (due to being negative for HER2 expression), and unlike MDA-MB-453 cells, their growth is suppressed by androgens (Hackenberg et al. 1991; Lehmann et al. 2011). Importantly, the divergent proliferative effects of androgens in AR+/ER- breast cancer models highlight the fact that inhibiting AR will likely not benefit all ER- patients and that careful investigation into the mechanisms of context-specific AR signalling is needed. This is supported by the underwhelming clinical benefit rates (19 %) reported for the AR antagonist Bicalutamide in TNBC (Gucalp et al. 2013).

### **1.2.6. Nuclear receptor cross-talk in cancer**

While it is becoming very apparent that steroid receptors play a crucial role in breast cancer alone, what has not been clearly investigated in depth is how they act together. Nuclear receptor cross-talk studies reveal the fact that co-activation of steroid receptors can affect each other's function and change the cellular response. Steroid hormones regulate a wide range of physiological processes by activating multiple receptors simultaneously inside a cell, where their combined action modulates the development and differentiation of the mammary gland. Consistent with this pivotal role, there is evidence

of nuclear receptors cross-talk is also linked to breast cancer (De Bosscher et al. 2020; Swinstead, Paakinaho & Hager 2018). Genome-wide studies have provided important information on nuclear receptor genomic cross-talk. NR cell-based assays are often configured to monitor ligand-modulated reporter gene expression. As TFs, NRs directly control gene expression by binding to HREs. The steroid NRs can interact with HREs in mouse mammary tumor virus (MMTV) promoter sequences that has been used in cross-talk studies. MMTV is a retrovirus that is transmitted either endogenously through the germ line or exogenously as infectious viral particles in the mother's milk. Although the virus is latently transforming, MMTV does not encode an oncogene (reviewed in (Grimm & Nordeen 1998)). MMTV is almost exclusively expressed in the mammary gland (reviewed in (Grimm & Nordeen 1998)). The activity of the MMTV promoter is greatly enhanced by steroid hormones, most notably glucocorticoids and progestins (reviewed in (Grimm & Nordeen 1998)). The ER agonist E2, but not ER antagonists, significantly suppress GR-mediated transcription in MCF-7 cells co-transfected with a MMTV-Luc (mouse mammary tumor virus promoter luciferase) reporter and GR reporter constructs using luciferase reporter assays (Kinyamu & Archer 2003). Since GR-mediated chromatin remodelling has been shown to be facilitated by the BRG1 chromatin-remodelling complex, the effect of estrogens on GR-dependent recruitment of BRG1 to the MMTV promoter was also investigated using CHIP-PCR. Interestingly, BRG1 association with the MMTV promoter was inhibited by E2 in the MCF-7-MMTV-GR cells (Kinyamu & Archer 2003). More recently, simultaneous co-activation of ER and GR was shown to reprogram the binding landscape of each other at specific recognition sites in an engineered murine breast cancer cell line (Miranda et al. 2013). GR reprograms chromatin accessibility for ER, through recruiting chromatin remodellers (SWI/SNF complexes) and assisting ER to bind to chromatin upon



co-treatment with E2 and a GR agonist (Dexamethasone; (DEX)) compared with E2 treatment only (Miranda et al. 2013). At the same time, ER activation causes changes in the chromatin structure at glucocorticoid receptor elements (GREs) allowing for GR recruitment at those sites upon DEX and E2 co-treatment compared with DEX-treatment only (Miranda et al. 2013). This highlights that induction of each steroid receptor, during mammary gland differentiation or initiation and development of breast cancer, can be critical for shaping the binding profiles of the other and subsequent cellular responses to hormone stimulation (Miranda et al. 2013).

The PR has also been shown to modulate ER genomic activity in BC. In support of earlier findings (Ballaré et al. 2003), the novel RIME technique revealed that activation of PR under estrogenic conditions increased the physical protein interaction of PR with ER and other known ER cofactors such as NR1P1, GATA3, and TLE3 in T-47D and MCF-7 ER+ BC cells (Mohammed et al. 2015). Consequently, activated PR re-directed ER chromatin binding in T-47D and MCF-7 cells, resulting in a unique transcriptome correlated with a good clinical outcome of the disease (Mohammed et al. 2015). Progesterone significantly inhibited E2-induced growth of ER+ MCF-7 and T-47D cell line xenografts in NSG mice and *ex vivo* primary ER+ breast tumour explants (Mohammed et al. 2015). A combination of progesterone treatment with the ER antagonist Tamoxifen in estrogenic conditions showed an increased anti-growth effect compared with progesterone or Tamoxifen alone in both MCF-7 and T-47D xenografts (Mohammed et al. 2015). Therefore, the functional significance of ER and PR cross-talk is regulation of a gene expression program associated with low tumorigenicity; hence, better disease outcome.

Finally, ER and AR cross-talk has been indicated in multiple ER+ breast cancer studies. AR activation with its agonists (e.g. DHT, Testosterone) inhibits the growth

stimulatory effect of estrogens on ER+ breast cancer cells including ZR-75-1, T-47D, and MCF-7 (Ando et al. 2002; Lapointe et al. 1999; Need et al. 2012; Panet-Raymond et al. 2000; Peters et al. 2009; Poulin, R. et al. 1989). A direct interaction between the LBD of ER $\alpha$  (not ER $\beta$ ) and the NTD domain of AR has been shown in yeast and mammalian two-hybrid systems but not using more contemporary techniques and physiologically relevant models (Panet-Raymond et al. 2000). Also, co-transfected AR and ER $\alpha$  expression vectors in CV-1 cells (well-used models for transfection studies) revealed the impact of AR and ER $\alpha$  in modulating the transactivation activity of the other (Panet-Raymond et al. 2000). The ER $\alpha$  effect on AR activity was shown to be E2-dependent, where E2 had no effect on the AR system unless ER $\alpha$  was co-transfected (Panet-Raymond et al. 2000). Similarly, E2-induced ER $\alpha$  activity was dramatically suppressed by AR co-expression in the presence of androgen using the double reporter construct in the same cells (Panet-Raymond et al. 2000). These findings suggest another explanation for the physiological interplay among androgens and estrogens where direct interaction of receptors may allow for an additional level of NR cross-talk that increases the complexity of steroid signalling pathways (Panet-Raymond et al. 2000). Although the direct interaction between ER and AR has been shown in two-hybrid systems, Peters *et al.*, could not detect a strong interaction between endogenous AR and ER proteins in T-47D breast cancer cells (Peters et al. 2009). They suggest another contributing mechanism of ER and AR cross-talk in these cells, where AR inhibits the E2-stimulated growth (Peters et al. 2009). In the ER+ breast cancer context, AR competes with ER for binding to regulatory regions of ER target genes (Need et al. 2012; Peters et al. 2009), which highlights the importance of the ratio of AR:ER expression in the regulation of breast cancer cell proliferation and predicting clinical outcome (Ando et al. 2002; Cochrane et al. 2014; Peters et al. 2009; Qu et al. 2013; Vera-Badillo et al. 2014).

Interestingly, it has been shown that the AR:ER expression ratio influences the activity of steroid receptor co-activator ARA70 (NCOA4) in ER+ breast cancer cells (Lanzino, M et al. 2005). Although ARA70 increased E2-dependent gene expression in MCF-7 cells in a low AR:ER ratio condition, induction of a high AR:ER ratio caused ARA70 to interact with AR to antagonise ER signalling (Lanzino, M et al. 2005). Also, androgen-induced reduction in expression levels of ER and MYC have been reported as a potential mechanism of action of androgens in prohibiting breast cancer cell growth (Dimitrakakis et al. 2003). Ligand-dependent AR-mediated inhibition of Cyclin-D1 (CCND1) at the mRNA and protein levels reduced the proliferation of MCF-7 breast cancer cells, depicting another mechanism of growth-inhibitory effects of AR by modulating cell cycle pathways (Lanzino, M. et al. 2010). Apart from ARA70, cofactors common to AR and ER and their roles in AR/ER cross-talk have not been widely investigated. One aim of this thesis was to reveal novel co-factors that may influence AR/ER transcriptional activity.

### **1.2.7. NR co-regulators and cancer**

It has been shown that NR co-regulators are not exclusively “bridging” molecules that simply link NRs and the RNA pol II transcriptional machinery, but have vast enzymatic processes, which modulate transcription, including chromatin modification and remodelling, initiation, elongation and termination of RNA pol II-mediated transcription, mRNA splicing and translation, and post translational modifications (PTMs) of NRs, other co-regulators and/or histones (Lonard, David M & O'Malley 2007). Co-regulators act as partner molecules that help in forming large protein complexes to modulating NR target gene activity (Manavathi, Samanthapudi & Gajulapalli 2014). Co-regulators can be other

TFs that bind DNA and may positively or negatively influence the transcriptional activity of NRs. Co-regulators that increase the transcriptional activity of NRs through their intrinsic histone acetyltransferase (HAT) activity or by recruiting HATs to the chromatin are referred to co-activators, while those that reduce transcription through recruiting the HDACs and compacting the chromatin structure are called co-repressors (Lonard, David M & O'Malley 2007). By modulating gene expression regulated by hormones, growth factors and cytokines, co-regulators can promote pathological processes associated with cancer, including cell proliferation, differentiation, carcinogenesis and metastasis (Lonard, David M & O'Malley 2007). Pioneer factors are an example of co-regulators that can associate with compacted chromatin to facilitate the binding of additional TFs. NRs can also tether to these factors (Jozwik & Carroll 2012). For instance, FOXA1 and GATA2/3 are examples of pioneer factors that open the chromatin to modulate NR DNA binding (Jozwik & Carroll 2012). Although FOXA proteins usually open the compacted chromatin structure, FOXA1 can also recruit additional factors, such as TLE proteins, to promote chromatin inaccessibility (Sekiya & Zaret 2007). This unique property of FOXA1 results from their structural similarity to linker histone proteins (Clark et al. 1993). FOXA1 is an ER co-regulator that its presence is essential for almost all ER-binding events in the genome in ER+ breast cancer context (Carroll et al. 2005). Also, FOXA1 acts as an AR co-regulator in prostate cancer, where its knock-down results in a reprogramming of AR and in an altered gene expression program (Sahu et al. 2011). The GATA family of NR co-regulators prominently contribute to the development of many cancer types in human patients. For instance, GATA2 is an AR co-regulator, playing key role in prostate cancer progression (He et al. 2014). GATA3 is an ER co-regulator that plays a critical role in cell fate decisions during mammary gland development and cancer progression (Theodorou et al. 2013). However,

whether GATA3 impacts AR signalling in breast cancer is unknown. Therefore, this thesis aimed to investigate the role of GATA3 on AR signalling in breast cancer.

### **1.3. Role of the GATA3 transcription factor in the mammary glands**

#### **1.3.1. GATA family members, their similar protein structure and DNA binding features**

Specification and maintenance of differentiated cell types arising from multipotent progenitor cells is a fundamental aspect of development. The specification of cell fate is mediated in part by hierarchical networks of TFs and cis-regulatory elements that control their genomic activity. TFs are often organized into multi-gene families and play essential roles in activating and/or repressing target genes that dictate specific cell fates (reviewed in (Lentjes et al. 2016)). The GATA family of TFs, of which there are six (GATA1 to GATA6), in mammals, are such master regulators. GATA family members are located on six distinct chromosomal regions in the human genome ([Figure 14](#)) and share common structural features (reviewed in (Lentjes et al. 2016)). Lineage-specific GATA factors play complex and widespread roles in cell fate decisions and tissue morphogenesis. GATA family members are classified into two subfamilies of hematopoietic (GATA1, 2, 3) and cardiac (GATA4, 5, 6) factors based on their tissue-specific expression patterns (Lentjes et al. 2016; Tremblay, Sanchez-Ferras & Bouchard 2018). GATA1 and GATA2 play key roles in erythroid and myeloid lineages, whilst GATA4, GATA5, and GATA6 are found in mesoderm and endoderm-derived tissues and organs including liver, heart, and intestines where they are involved in smooth muscle differentiation and endoderm formation (Patient & McGhee 2002). Conversely, GATA3 is expressed in both hematopoietic (e.g. T cells (Frelin et al.

2013; Tindemans et al. 2014)) and non-hematopoietic tissues such as the kidney (Grote et al. 2008), central nervous system (van Doorninck et al. 1999), skin (Kaufman et al. 2003) and mammary glands (Kouros-Mehr et al. 2006), where it plays a key regulatory role in developmental pathways. Importantly, GATA family members share different degrees of homology at the amino acid level. Although the zinc finger motifs are more than 70 % homologous among all the six GATA members in humans, the total protein homology between different family members is variable. For instance, there is an approximate 55 % homology between GATA2 and GATA3 and only 20 % homology between GATA3 and GATA4 (reviewed in (Zaidan & Ottersbach 2018)).

GATA family members have two transactivation domains (TAD) at their amino terminus (N-terminal domain), which diverge considerably among GATA factors, and two highly conserved zinc-finger DNA binding domains at their carboxyl terminus (C-terminal domain) called ZnFn1 (N-finger) and ZnFn2 (C-finger) ([Figure 15](#)) (Morrisey et al. 1997). All GATA family members recognize a hexa-nucleotide consensus recognition element of 5'-(A/T) GATA (A/G)-3' on DNA, which is centred on the GATA motif in the regulatory regions of target genes. ZnFn2 is necessary and sufficient for GATA proteins to recognise and bind to the consensus sequence (Yang, HY & Evans 1992). ZnFn1 binds the GATA recognition sequence independently of ZnFn2, especially for GATA2 and GATA3, but with a slightly different sequence preference (5'-GATC-3'). ZnFn1 chromatin binding stabilizes the GATA-DNA interactions (reviewed in (Bates et al. 2008)). In addition, both zinc fingers can contribute to self-association of GATA proteins and interaction with other protein partners, including other TFs or other GATA family members (Bates et al. 2008; Crispino et al. 2001; Trainor, Ghirlando & Simpson 2000; Wilkinson-White et al. 2015). It has been shown that GATA factors can form homo or heterodimers *in vivo* and *in vitro* (reviewed in

(Bates et al. 2008)). This is thought to have key roles in mediating combinatorial and synergistic transcriptional regulation by GATA factors, and also in the assembly of high-order protein DNA complexes (reviewed in (Bates et al. 2008)).

GATA factors can modulate the chromatin structure as pioneer factors, which open compacted chromatin, implicating GATA factors in the epigenetic control of chromosome structure during progressive stages of developmental pathways, including erythroid or mammary epithelial cell differentiation (Chen et al. 2012). For instance, GATA3 functions as a pioneer TF by inducing *de novo* formation of enhancers at previously closed chromatin, probably by recruiting a nucleosome remodeller (Takaku et al. 2016). Another mechanism by which GATA proteins contribute to transcriptional regulation is by facilitating chromosomal looping to mediate long-range control of gene expression. For example, structural studies using the chromosome conformation capture (3C) technique demonstrated an important role for GATA3 and STAT6 in the formation and maintenance of long-range chromatin interactions at the T helper type 2 (T<sub>H</sub>2) locus control region (LCR) (Spilianakis et al. 2005). The genomic region of the T<sub>H</sub>2 cytokine locus covers more than 120 kilobases (kb) and encodes genes for the *Il4*, *Il5* and *Il13* cytokines (Spilianakis et al. 2005). Also, it has been shown that GATA1 and its co-factor Friend-of-GATA proteins 1 (FOG1) can directly occupy looped enhancers and target gene promoters at the  $\beta$ -globin locus in fetal liver erythroblasts (Vakoc et al. 2005). GATA1 has been shown to affect higher order chromatin organization during suppression of the *Kit* gene in erythroid cells (Jing et al. 2008). *Kit* encodes a receptor tyrosine kinase that is essential for normal haematopoiesis, and is expressed in haematopoietic stem cells and lineage progenitor cells during early erythropoiesis (Broudy 1997; Munugalavadla et al. 2005). In immature erythroid cells, a distal enhancer bound by GATA2 is in physical proximity with the active

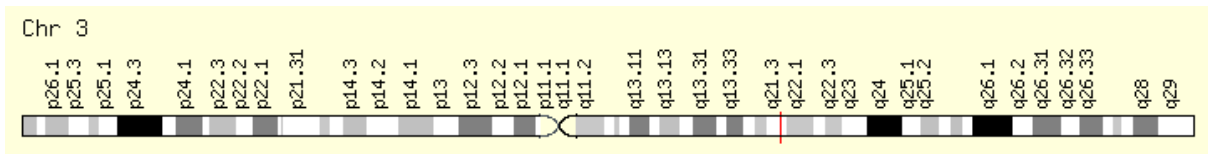
*Kit* promoter. Upon cell maturation, GATA1 displaces GATA2 and triggers a loss of the enhancer/promoter interaction (Jing et al. 2008). This exchange of GATA1 and GATA2 on the chromatin mediates a transition in looped chromatin organization of the *Kit* gene (Jing et al. 2008). Downregulation of *Kit* upon terminal erythroid differentiation occurs in association with complete exchange of GATA1 for GATA2 (Jing et al. 2008). Since this thesis specifically involves GATA3 in breast cancer, I will now focus on the role of GATA3 in mammary gland development and breast cancer progression.



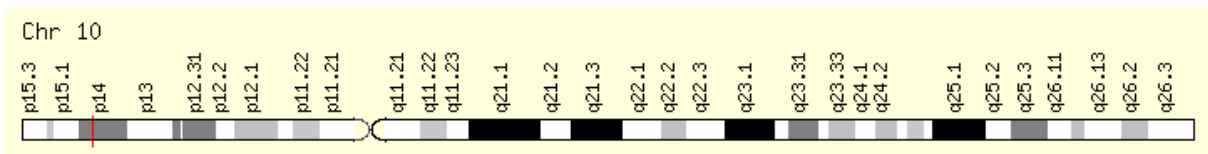
**GATA1 (Xp11.23)** \_ chrX:48,644,962-48,652,718 (GRCh37/hg19) \_ size: 7,757 bases



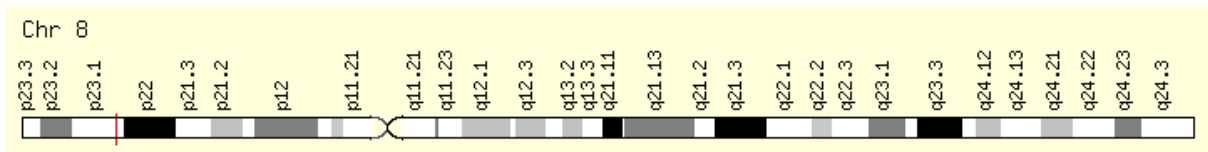
**GATA2 (3q21.3)** \_ chr3:128,198,265-128,212,030 (GRCh37/hg19) \_ size: 13,766 bases



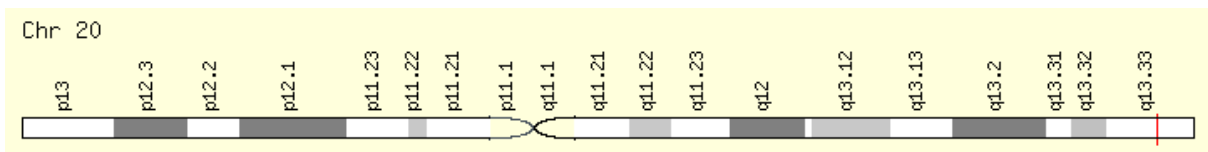
**GATA3 (10p14)** \_ chr10:8,095,567-8,117,164 (GRCh37/hg19) \_ size: 21,598 bases



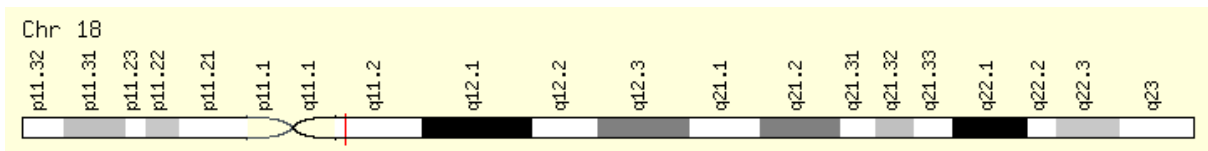
**GATA4 (8p23.1)** \_ chr8:11,534,468-11,617,511 (GRCh37/hg19) \_ size: 83,044 bases



**GATA5 (14q11.2)** \_ chr20:61,038,553-61,051,026 (GRCh37/hg19) \_ size: 12,474 bases

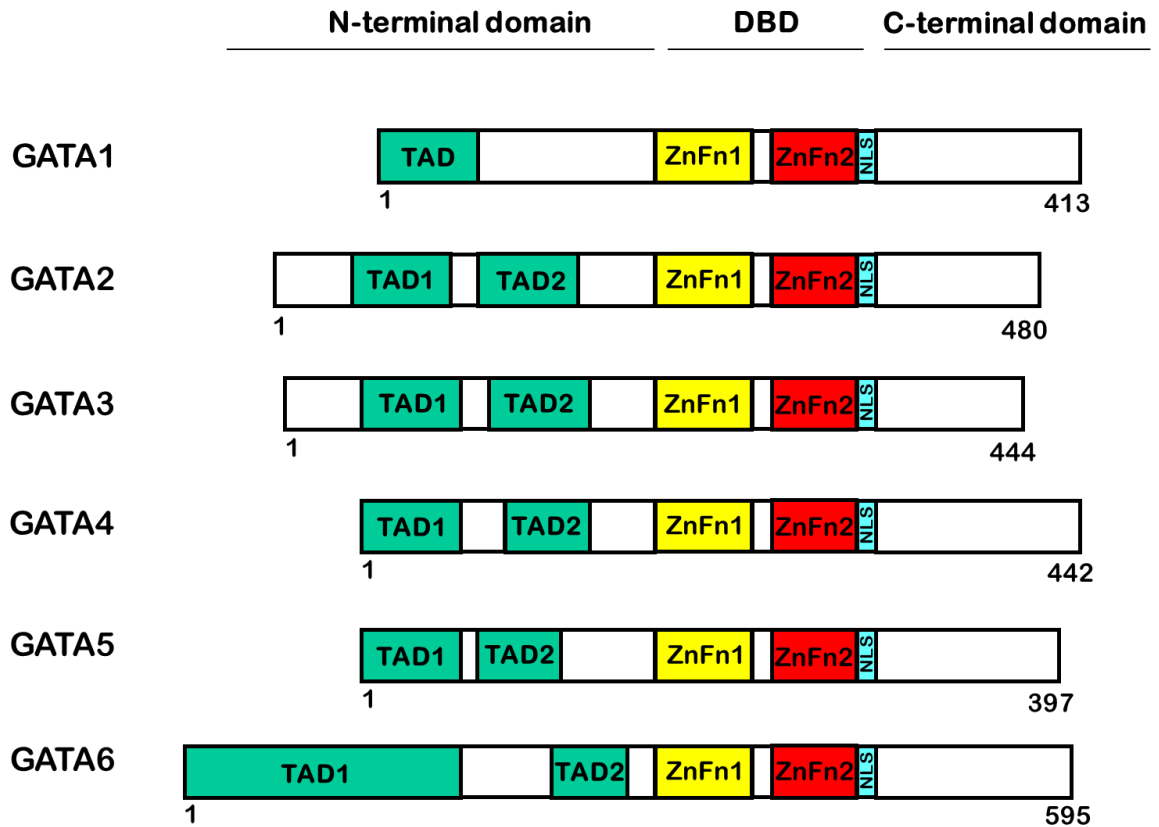


**GATA6 (18q11.2)** \_ chr18:19,749,404-19,782,491 (GRCh37/hg19) \_ Size: 33,088 bases



**Figure 14: Chromosomal location of GATA family members**

GATA family members are in different chromosomal locations in the genome ([www.GeneCards.org](http://www.GeneCards.org)).



**Figure 15: The GATA family of proteins**

Members of the GATA family are sub-classified as GATA1, 2, 3 and GATA4, 5, 6 based on their tissue-specific expression patterns and their structural similarity. They contain two highly conserved zinc finger motifs (ZnFn) in the DBD domain, a less conserved transactivation domain (TAD) in the N-terminal domain, and a conserved nuclear localization sequence (NLS) in the C-terminal domain.

### 1.3.2. The GATA3 transcription factor

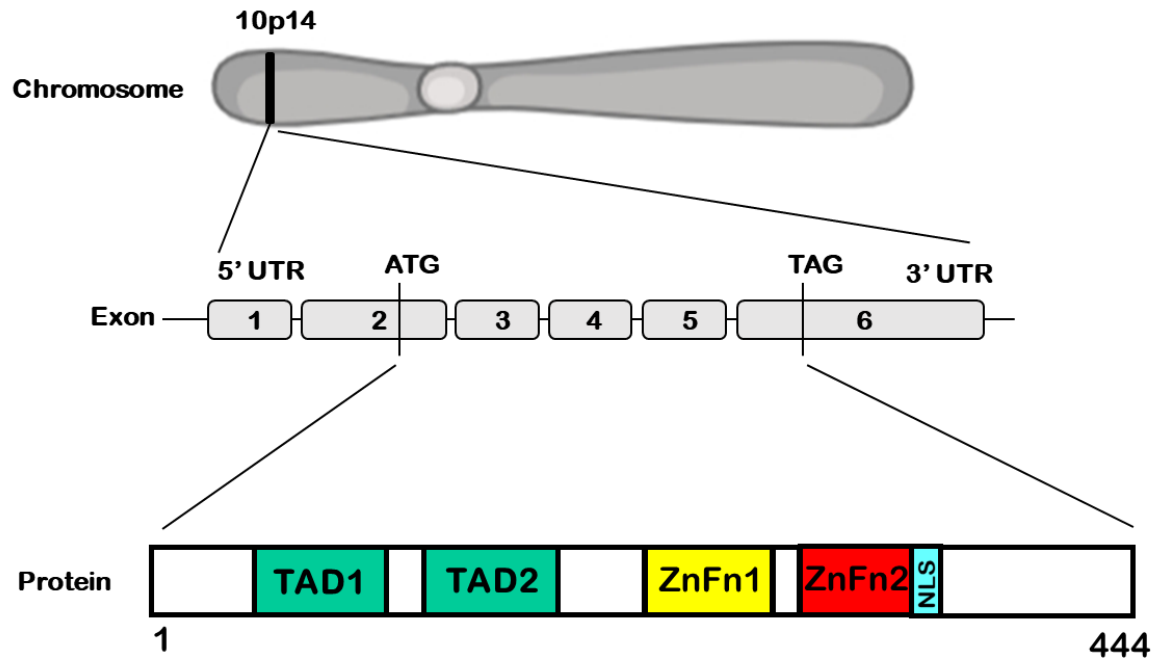
The *GATA3* gene (located on chromosome 10p14) consists of six exons along 20 kb of genomic DNA and encodes a 444 amino acid protein that must translocate into the nucleus to exert its transcriptional activity ([Figure 16](#)). Like all GATA factors, GATA3 contains an NLS to direct the protein into the nucleus (Chook & Blobel 2001). Nuclear transportation of GATA3 from the cytoplasm is mediated by p38 MAPK activity through phosphorylation of serine residues on GATA3, leading to a protein-protein interaction with the nuclear transporter protein importin- $\alpha$  at the NLS site (Maneechotesuwan et al. 2007). Importin- $\alpha$  is one of the nuclear import proteins that plays an important role in transporting large proteins from the cytoplasm into the nucleus (Conti et al. 1998; Goldfarb et al. 2004). It has been previously shown that deletion of the NLS in GATA3 (aa 249–311) prevents its nuclear localization (Yang, Z et al. 1994). Furthermore, the affinity of the interaction between importin- $\alpha$  and the GATA3 NLS is a key parameter in determining nuclear transport efficiency, which may be directly enhanced by GATA3 phosphorylation (Chook & Blobel 2001; Goldfarb et al. 2004; Stochaj & Silver 1992). Structural studies and further mutational analyses revealed that the two ZnFns of GATA3 bind to two proximal 5'-GAT-3' sequences that are located in the same major groove of the DNA in different positions (ZnFn1 binds to the SHL5.5 position and ZnFn2 binds to SHL6.5 position in the nucleosome) (Tanaka et al. 2020). Also, the strand orientation of the sequence between ZnFn1 and ZnFn2 is not random, suggesting the importance of specific engagement of GATA3 zinc fingers with DNA in the outcome of its chromatin binding (Chen et al. 2012; Tanaka et al. 2020). Mutation of one of the two 5'-GAT-3' sequences substantially reduced the ZnFn binding affinity of GATA3, indicating that both 5'-GAT-3' sequences are required for high affinity ZnFn binding to the nucleosome (Tanaka et al. 2020). Preferential

positioning of GATA3 has been observed near the periphery of the nucleosome, which is considered to be more accessible for TF binding, compared to central positions in the nucleosome structure (Tanaka et al. 2020). These data indicate that the rotational position of the nucleosomal 5'-GAT-3' sequences is an essential factor for stable GATA3 binding to the nucleosome periphery.

Genome-based technologies like ChIP-seq and assay for transposase-accessible chromatin using sequencing (ATAC-seq) have been used in a recent study to explore the mechanistic basis by which GATA3 functions as a pioneer TF in a cellular reprogramming event relevant to BC, the mesenchymal to epithelial transition (MET), following ectopic expression of GATA3 in a typically GATA3-negative breast cancer cell line (MDA-MB-231) (Takaku et al. 2016). It has been previously demonstrated that ectopic expression of GATA3 promotes MET in MDA-MB-231 cells, with a consequent reduction in metastatic capacity (Chou et al. 2013). At the individual locus level, many GATA3 binding sites, including MET-associated gene loci, exhibited increased accessibility in GATA3 over-expressing cells (Takaku et al. 2016). In some instances, GATA3 pioneered new binding sites in inaccessible chromatin and reprogrammed the local chromatin structure to be accessible after its binding (~1/4 loci). However, mostly GATA3 bound to pre-accessible chromatin and modestly increased the width of the accessible region (~1/2 loci). At the rest of the target loci, GATA3 bound inaccessible sites and failed to remodel local structures (~1/4 loci). Therefore, the chromatin structure at these loci remained refractory to transposition following GATA3 ectopic-expression (Takaku et al. 2016). Interestingly, GATA3-mediated alterations in local chromatin accessibility was dependent on presence of a functional TAD1 domain. Deletion of the TAD1 did not alter the chromatin binding ability of GATA3, but crippled chromatin reprogramming ability, resulting in failure to induce MET (Takaku

et al. 2016). The data presented in this study provided key mechanistic insights into GATA3-mediated chromatin reprogramming during MET, and suggested unexpected complexity to TF pioneering.

Multiple proteins are known to mediate GATA3 cell-specific transcriptional activity. In  $T_H2$  cells, GATA3 directly interacts with Smad3, an intracellular signal transducer of TGF $\beta$ , and mediates the recruitment of Smad3 to GATA3 binding sites independently of Smad3 binding to DNA (Blokzijl, ten Dijke & Ibanez 2002). This synergistic cooperation of GATA3-Smad3 complexes promoted transcription of *IL-5* and *IL-10* cytokines (Blokzijl, ten Dijke & Ibanez 2002). It has been evident in human mammary tumours that, as tumour grade increases, GATA3 expression is silenced through various mechanisms such as DNA methylation (Carr et al. 2012). FOXM1 is a negative regulator of GATA3 and has been shown to repress transcription of *GATA3* and subsequent protein expression *in vivo* through induction of hypermethylation of the CpG islands at the *GATA3* promoter, in association with DNA methyl transferase 3b (DNMT3b) (Carr et al. 2012). Acute loss of FOXM1 resulted in remarkable enhancement in GATA3-mediated differentiation of luminal epithelial cells in mouse mammary glands (Carr et al. 2012). Also, in the breast cancer context, at loci where GATA3 pioneered the opening of inaccessible chromatin, co-factors such as the SWI/SNF complex can be recruited, leading to nucleosome eviction and opening of the local chromatin structure (Takaku et al. 2016).



**Figure 16: GATA3 Functional domains and protein structure**

The *GATA3* gene is located on chromosome 10p14, contains 6 exons and its protein is composed of 444 amino acids harbouring two amino terminal transactivation domains (TAD1 and TAD2), and two zinc-finger motifs (ZnFn1 and ZnFn2). The ZnFn2 binds to DNA containing the canonical GATA motif, (A/T) GATA (A/G). However, ZnFn1 seems to have broader specificity. The ATG is the translation start codon located in exon 2 and the TAG is the translation stop codon located in exon 6. The nuclear localization sequence, or NLS, is the sequence that interacts with import proteins to drive the active process of nuclear import.

### **1.3.3. Role of GATA3 in normal mammary gland development and differentiation**

GATA3 plays an integral role in luminal cell differentiation in the mammary glands (Asselin-Labat et al. 2007). However, little is known about GATA3-mediated transcriptional regulation of luminal lineage genes. Efforts to understand luminal cell differentiation in mouse mammary glands have mostly focused on ER, which is only expressed in approximately half of the mature luminal cells and is present in fibroblasts and other stromal cells (Cheng et al. 2004). GATA3 has been shown to have a key role in the development and differentiation of mammary ducts (Kouros-Mehr & Werb 2006). It has been identified as the most highly enriched TF in microarray screening of mammary epithelial cells in both the TEB and the mature duct microenvironment of pubertal mice (Kouros-Mehr & Werb 2006). In human breast tissue, normal ducts and lobules were focally positive for GATA3 expression, with generally increased expression in malignant cells (Yoon et al. 2010). During mouse embryonic development, GATA3 expression is detected in primordial mammary buds and is limited only to the luminal epithelial layer, not the myoepithelium (Shackleton et al. 2006; Stingl et al. 2006). Similar to what has been seen in ERKO mice (Feng et al. 2007), targeted loss of GATA3 in mouse mammary glands through MMTV-Cre mediated deletion of floxed-GATA3 led to defects in mammary branching morphogenesis (Kouros-Mehr et al. 2006). GATA3KO mice failed to form TEBs during puberty, showed irregular luminal diameters, deficiencies in side branching and invasion of the epithelium into the stroma (Asselin-Labat et al. 2007; Kouros-Mehr et al. 2006). In mice, GATA3 expression has also been shown to be necessary to maintain the differentiated state of luminal epithelial cells in the adult mammary glands after completion of mammary development (Kouros-Mehr et al. 2006). GATA3 plays a significant

role in the lineage determination and maturation of mammary epithelial cells by pushing mammary cells into the luminal cell fate (Asselin-Labat et al. 2007; Kouros-Mehr et al. 2006; Naylor & Ormandy 2007). Targeted deletion of GATA3 from mammary progenitor cells (MMTV-Cre; GATA3<sup>flox/flox</sup>) blocks luminal cell differentiation, whereas forced expression of GATA3 in mammary stem-cell-enriched populations promotes differentiation into luminal cells (Kouros-Mehr et al. 2006). During pregnancy in mice, GATA3 participates in luminal epithelial differentiation required for lobuloalveolar development (Asselin-Labat et al. 2007). It also has been shown that lack of GATA3 perturbed differentiation and proliferation of luminal cells, which caused severely impaired lactogenesis, yielded smaller lobuloalveolar units, and decreased the characteristic markers of alveolar differentiation (Asselin-Labat et al. 2007). GATA3-depleted adult mammary glands have demonstrated severe luminal cellular defects such as disorganization of the duct and decreased cell–cell adhesion, which are important properties for invasion and metastasis (Kouros-Mehr et al. 2006). Interestingly, detachment from the basement membrane and caspase-mediated luminal cell death results from long term GATA3 deletion in adult mammary epithelial cells (Kouros-Mehr et al. 2006).

Taken together, these findings indicate that GATA3 is necessary for normal mammary gland development, ductal elongation and branching, luminal epithelial differentiation, and maintenance of the differentiated state of the mammary epithelial cells.



#### 1.3.4. GATA3 and breast cancer progression

Given the essential role of GATA3 in determining the epithelial cell lineage in the mammary gland, it is not surprising that GATA3 plays a role in breast cancer biology and prognosis. GATA3 is a prominent marker of the luminal pattern of gene expression, however, it is expressed in all BC subtypes to some extent. The overall positivity of GATA3 is 99.51 % in luminal A-like, 97.70 % in luminal B-like, 68.50 % in HER2 overexpressing and 20.16 % in TNBC tumours (Shaoxian et al. 2017). Importantly, several microarray studies found expression of GATA3 as one of the best predictors of ER positive status among primary breast cancers (Hoch et al. 1999; Kouros-Mehr, Kim, et al. 2008). Accordingly, GATA3 has been shown to be expressed strongly in ER+ cell lines (MCF-7, T-47D) but not in ER-negative lines (MDA-MB-231, HBL-100) through cDNA microarray analysis (Hoch et al. 1999). Meta-analysis using the OncoPrint Research Platform confirmed increased GATA3 expression in ER+ cancers compared with ER- cancers throughout 17 studies (Fang, Chen & Weigel 2009). As expected from its normal function, GATA3 is a strong predictor of tumour differentiation, ER status and clinical outcome (low GATA3 expression was a predictive of poor clinical outcome in all patients as well as ER+ subtype) in breast tumours, with independent prognostic significance above conventional variables (reviewed at (Kouros-Mehr, Kim, et al. 2008)). Absence of GATA3 expression is associated with tumour types with a propensity for invasive growth and a poor prognosis, larger tumor size, positive lymph node status, higher histology grade, ER-negative status, HER2 overexpression as well as increased risk for recurrence and metastasis (Mehra et al. 2005). Interestingly, a positive regulatory loop has been found between ER and GATA3 in T-47D breast cancer cells (Eeckhoute et al. 2007). ER binds to a regulatory element 10 kb downstream of the GATA3 transcription start site and regulates the expression of GATA3

(Eeckhoute et al. 2007). GATA3 binds to two enhancer sites within the ER gene which are required for RNA pol II recruitment at the ER promoter and regulates ER expression (Eeckhoute et al. 2007).

The direct role of GATA3 in breast cancer pathogenesis and progression has been investigated using the MMTV-PyMT mouse model of breast cancer that develops spontaneous mammary tumours that progress from ER+ hyperplasia to an ER-, poorly differentiated adenocarcinoma (Kouros-Mehr, Bechis, et al. 2008). Therefore, PyMT mice serves as an excellent model to understand the tumour progression process, and are very comparable to human breast carcinomas (Lin et al. 2003). Microarray analysis revealed a high correlation between downregulation of luminal differentiation genes and tumour progression in this model. Partial loss of GATA3 expression occurred during the transition from adenoma (5-week tumour) to early carcinoma (8-week tumour), however, the tumour cells in late carcinomas (18-week tumour) and metastatic cells in the lungs were consistently GATA3 negative (Kouros-Mehr, Bechis, et al. 2008). Moreover, restoration of GATA3 expression in late- stage carcinomas was sufficient to induce luminal differentiation markers (e.g. E-cadherin and  $\beta$ -casein), as well as a significant reduction in dissemination to distant sites and metastatic seeding (Kouros-Mehr, Bechis, et al. 2008). These data indicate that GATA3 negatively impact the breast cancer progression.

### **1.3.5. Role of GATA3 in tumour invasion and metastasis**

The established hallmarks of cancer include sustaining of proliferative signalling, escaping from growth suppressors and resisting programmed cell death, angiogenesis, immortality and the ability of the cells to invade and metastasise (Hanahan & Weinberg

2011). Metastasis is a multistage biological process consisting of extracellular matrix remodelling, dissemination from the original organ and spreading of the tumour cells through the circulation, then finally reaching and surviving in distant organs. During metastasis, the cancer cells lose their epithelial features and differentiated status and gain stem-like and mesenchymal properties leading to increased motility and other aggressive behaviours (Valastyan & Weinberg 2011). This transformation of epithelium-derived tumour cells from epithelial to a mesenchymal phenotype is called the epithelial to mesenchymal transition (EMT) (Valastyan & Weinberg 2011). GATA3, as a critical negative regulator of tumour features, suppresses the expression of key EMT proteins and reduces invasion of breast cancer cells *in vitro* and *in vivo* (Chou et al. 2013; Dydensborg et al. 2009; Yan et al. 2010). It has been shown using microarray analysis of primary tumours that GATA3 negatively regulates the expression of several genes associated with breast cancer lung metastasis such as *ID1*, *ID3*, *KRTHB1*, *LY6E* and *RARRES3* (Dydensborg et al. 2009). Conversely, GATA3 up-regulated the expression of genes encoding known inhibitors of lung metastasis including *DLC1* and *PAEP* in tumours developing from GATA3-overexpressing LM2 cells, which are derived from the highly aggressive MDA-MB-231 model (Dydensborg et al. 2009). Of note, the LM2 cells express significantly lower levels of GATA3 compared with the parental MDA-MB-231 strain. These data are consistent with microarray data from human breast cancer patients, indicating a strong positive correlation between high expression of GATA3 and absence of metastases, specifically to the lungs (Dydensborg et al. 2009). Similarly, another study showed that tumours derived from basal BC cell lines, 4T1 (mouse) and MDA-MB-231 (human), that over-expressed GATA3 decreased lung metastases and altered the tumour microenvironment through induction of microRNA (miR)-29b (anti-metastatic microRNA) expression *in vivo* (Chou et

al. 2013). Increased expression of miR-29b resulted in down-regulation of pro-metastatic genes through repression of targets in the tumour microenvironment such as *ANGPTL4*, *LOX*, *MMP9*, *VEGF-A*, *ITGA6* and *ITGB1*, thereby suppressing lung metastasis (Chou et al. 2013). Forced expression of GATA3 in MDA-MB-231 cells promotes the cells to undergo MET transition and reduces cell invasiveness through reduction of vimentin, N-cadherin, and MMP-9 expression (mesenchymal markers) and an increase of E-cadherin expression (an epithelial marker) (Yan et al. 2010). Consistently, GATA3 expressing MDA-MB-231 cells displayed a cuboidal-like epithelial phenotype identical to MCF-7 cells (Yan et al. 2010). Also, injecting GATA3 expressing MDA-MB-231 cells into the mammary fat pad of immunocompromised mice resulted in development of smaller tumours without distal metastasis, contrary to mice bearing empty vector controls, supporting a tumour suppressive role for GATA3 in blocking tumour development and metastasis independent of ER signalling (Yan et al. 2010). Further investigations in the same study has shown that GATA3 knockdown in the non-invasive ER+ breast cancer cell line MCF-7, reduced the expression level of E-cadherin and triggered fibroblastic transformation. Also, tumours derived from MCF-7 GATA3-depleted cells developed quicker and the mice showed shorter survival rate and distant metastasis to lung and liver (Yan et al. 2010). Taken together, these findings suggest a critical role for GATA3 in inhibiting the invasive behaviour and metastatic features of ER- breast tumours via inhibition of EMT.

### **1.3.6. GATA3 cooperates with ER and FOXA1 as a definitive luminal complex in ER+ breast cancer**

ER drives growth in the majority of human breast cancers through regulation of genes associated with tumour growth. ER occupies distal enhancers (Carroll et al. 2005)

and brings them to the proximity of promoters of the regulated target genes by estrogen-induced chromosomal loops (Pan et al. 2008). GATA3 is an important ER co-regulator, playing a pivotal role in mediating enhancer accessibility and transcriptional potential at the ER regulatory regions in ER+ breast cancer cells (Theodorou et al. 2013). GATA3 has been shown to be an ER interacting protein in MCF-7 cells (Mohammed et al. 2013), human PDX tumours and primary human breast cancer clinical tissues using the powerful approach of RIME and qPLEX-RIME, respectively (Papachristou et al. 2018). The qPLEX-RIME technique is an unbiased proteomic technique that enables comprehensive mapping of endogenous protein interactomes with higher sensitivity and statistical robustness compared to traditional RIME (Papachristou et al. 2018). Serandour *et al.*, developed a new genomic approach termed ChIP-exonuclease (ChIP-exo), which uses an exonuclease to trim the ChIP'd DNA to a precise distance from the crosslinking site (Rhee & Pugh 2011, 2012), to gain clearer insights into the spatial relationship between GATA3, FOXA1 (as two key ER-associated TFs), and ER (Serandour et al. 2013). In the latter study, GATA3 motifs were shown to be adjacent to the central ERE, however, the Forkhead motifs were further away from the ERE motifs. This suggested structure shows how these three key factors form to define a transcriptionally active cis-regulatory element in luminal ER+ breast cancers (Lacroix & Leclercq 2004). GATA3 and FOXA1 are both estrogen-regulated genes (Eeckhoute et al. 2006; Eeckhoute et al. 2007) that are highly correlated with ER expression in ER+ breast cancers (Lacroix & Leclercq 2004). ER, FOXA1, and GATA3 are definitive genes for ER+ luminal breast cancer tumours (Perou et al. 2000; Sorlie et al. 2003) and their expression is required for establishment of an E2-responsive ER complex (Kong et al. 2011). Ectopic expression of GATA3, FOXA1 and ER in ER-, estrogen-unresponsive MDA-MB-231 cells partially restored ER binding capacity and transcriptional activity (Kong et al. 2011).

Furthermore, E2-stimulation of ER+ MCF-7 breast cancer cells resulted in the formation of a tripartite enhanceosome complex of ER, FOXA1, and GATA3 which co-localizes at common binding sites across the genome to induce ER target gene transcription through recruitment of RNA Pol II and p300 to the chromatin (Kong et al. 2011). Both FOXA1 and GATA3 are known ER pioneer factors, where they independently associate with compacted chromatin and directly facilitate chromatin accessibility for ER (Carroll et al. 2006; Theodorou et al. 2013). FOXA1 is required for ER to bind to DNA and exert its estrogen-induced transcriptional activity (Carroll et al. 2005). In contrast, GATA3 is not critical for ER binding but loss of GATA3 induces redistribution of about a third of all E2-stimulated ER binding events (Theodorou et al. 2013). ER binding events gained upon GATA3 depletion had a low dependency on GATA3 presence, however, the lost ER binding events were GATA3 dependent. Also, it has been shown that sites that lose their ER binding affinity upon GATA3 depletion are also depleted of active histone marks (H3K4me1 and H3K27ac), suggesting that GATA3 mediates these enhancers directly, likely as a pioneer factor (Theodorou et al. 2013). These findings also suggest that GATA3 is required for ER to bind to chromatin lacking active histones. The GATA3 independent ER binding events occurred at sites exhibiting elevated occupancy of FOXA1 and active histone marks (Theodorou et al. 2013). The remaining ER sites that were not changed upon GATA3 loss, occurred at enhancers with active histone marks that are occupied with both GATA3 and FOXA1, indicating less dependency of these ER sites to the presence of GATA3, but more dependence on FOXA1. GATA3-mediated ER redistribution was associated with changes in gene expression (Theodorou et al. 2013). The up-regulated genes were significantly correlated with the “stronger” ER binding events that occurred in the absence of GATA3 (e.g. *TRAK1*, *TGFB1*, and *FGFR3*) whereas the down-regulated genes were significantly

correlated with the “weaker” ER binding events upon silencing GATA3 (e.g. *CA12*, *CAV1*, and *METRNL*) (Theodorou et al. 2013). GATA3 has been shown to directly bind to the FOXA1 promoter in primary mammary epithelial cultures of adult mice (Kouros-Mehr et al. 2006). This study has identified FOXA1 as a putative member of the GATA3 genetic regulatory network. Expression of GATA3 and FOXA1 is strongly correlated in mammary and breast cancer microarray datasets (Usary, Jerry et al. 2004). Moreover, the overexpression of GATA3 was shown to upregulate FOXA1 expression, suggesting that GATA3 activates gene expression (Usary, Jerry et al. 2004). Furthermore, genome-wide mapping of FOXA1 after GATA3 depletion in MCF-7 cells revealed that GATA3 also modulates the FOXA1 chromatin binding profile, suggesting that GATA3 acts upstream of FOXA1 in determining the ER binding profile (Theodorou et al. 2013). Collectively, these data suggest an important role for GATA3 in ER signalling by facilitating ER DNA binding to a subset of regulatory sites at compacted chromatin, and by mediating FOXA1 as its required co-regulator.

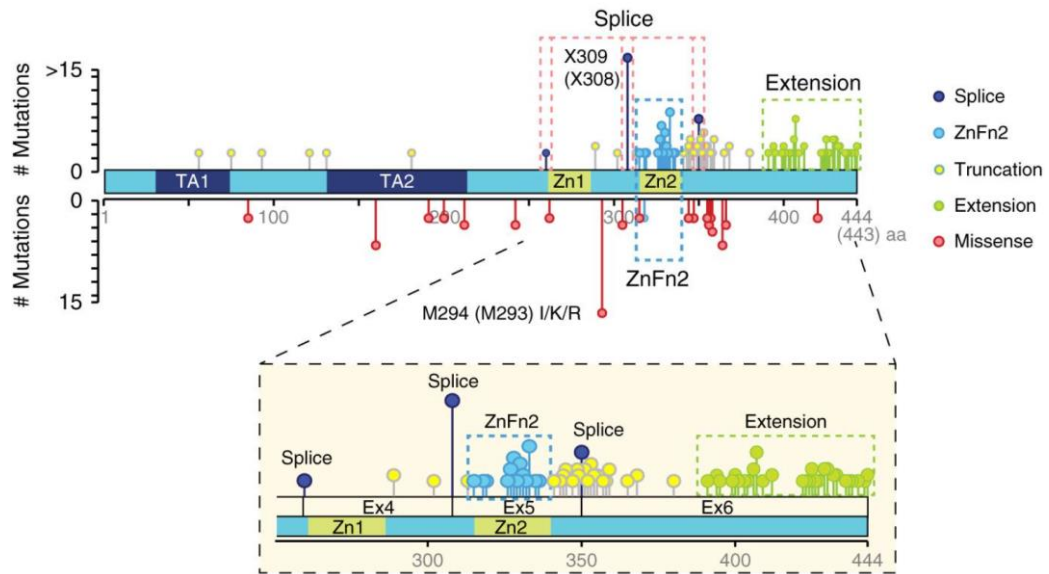
### **1.3.7. GATA3 mutations in breast cancer**

*GATA3* mutations were reported in humans for the first time by Van Esch *et al.*, in 2000, identified in a family with a rare disease called HDR syndrome characterised by hypoparathyroidism, sensorineural deafness and renal disease (Van Esch et al. 2000). Following that, in 2004, six different *GATA3* heterozygous mutations were identified in breast cancer tissues and breast tumour-derived cell lines (Usary, J. et al. 2004). Very quickly after that, *GATA3* was shown to be one of the most frequently mutated genes in primary breast cancers, with approximately 10 % incidence rates (Banerji et al. 2012;

Chanock et al. 2007; Stephens et al. 2012). Now there are more than 70 *GATA3* somatic mutations reported in breast cancers (Gaynor et al. 2013). The majority of these mutations (>95 %) are found in ER+ luminal A or B breast cancer subtypes (Gaynor et al. 2013). Most *GATA3* mutations are rare or unique frameshift indels distributed along the 3' end of the gene structure ([Figure 17](#)) (Pereira et al. 2016). Four groups of frameshift mutations have been identified for *GATA3* including, 1) ZnFn2 mutations, which occur within the C terminal zinc finger; 2) truncating mutations, located downstream of the C terminal zinc finger; 3) extension mutations, which occur in exon 6 and disrupt the stop codon, and 4) splice mutations, which mainly occur between intron 4 and exon 5 (Takaku et al. 2018). Depending on the location of the mutations in the *GATA3* gene, they are predicted to differentially influence protein production, such as truncated or extended proteins (Usary, J. et al. 2004). Due to the relatively high frequency of *GATA3* mutations in primary breast cancers, they were first considered driver mutations (reviewed in (Takaku, Grimm & Wade 2015)). Although some show tumour-promoting functions (Cohen et al. 2014; Gustin et al. 2017; Takaku et al. 2018), *GATA3* mutations are paradoxically associated with longer patient survival (Pereira et al. 2016) and improved response to endocrine therapy (Ellis et al. 2012). Jiang *et al.*, collated the clinicopathologic characteristics of patients with breast cancers harbouring *GATA3* mutations from two separate cohorts (The Cancer Genome Atlas (TCGA), n=934) and the Fudan University Shanghai Cancer Centre (FUSCC), n=308) and examined the association of *GATA3* somatic mutations with patient survival rate (Jiang et al. 2014). They detected *GATA3* mutations in 8.8 % and 14.9 % of patients with luminal-like breast cancer in TCGA and FUSCC cohorts, respectively, and noted that they were significantly associated with improved overall survival and favourable disease-free outcome. This was also observed in patients with luminal disease who had received



adjuvant endocrine therapy, indicating a potential predictive value for endocrine therapy in this population (Jiang et al. 2014).



**Figure 17: Distribution of GATA3 mutations in breast tumors (Takaku et al. 2018)**

Takaku *et al.*, classified GATA3 mutations into 5 groups as depicted in the figure, including 1) Splice site mutations; 2) ZnFn2 mutations; 3) Truncation mutations; 4) Extension mutations; and 5) Missense mutations. The mutation data shown in this figure was obtained from the METABRIC cohort via the cBio Cancer Genomics Portal. The missense and truncation mutations are dispersed along the GATA3 protein sequence. However, most GATA3 mutations are localized in exon 5 and 6. ZnFn2 mutations and extension mutations exclusively occur in the second zinc-finger motifs and far end of CTD domain, respectively. Infrequent splice site mutations are shown in dark blue localized in exon 4 and 5 (Takaku et al. 2018).

The most common GATA3 mutation, X308 splice, has recently been investigated in a study involving carrier patients (66/231 tumours) of a Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort (Hruschka et al. 2020). These patients were significantly associated with signatures of good prognosis after endocrine therapy, smaller tumour size, lower tumour stage and grade, PR expression, and improved outcomes compared to patients without GATA3 X308 splice mutations or those carrying other GATA3 mutations (Hruschka et al. 2020). Importantly, this mutation was exclusively observed in patients with ER+ tumours (Hruschka et al. 2020). Also, investigating differentially expressed genes in tumours harbouring X308 splice mutations compared with all the other tumours revealed a negative correlation with gene signatures of bad prognosis and strong down-regulation of cell cycle and inflammation-associated genes (e.g. *E2F2*, *E2F4*, *PCNA* and *MKI67*), which is consistent with the favourable prognosis of these patients (Hruschka et al. 2020). The X308 splice mutation of GATA3 reduced ER enrichment to chromatin, especially at enhancer regions in mutant T-47D cells upon E2-stimulation compared with the hormone-starved control cells, suggesting that this mutation interferes mainly with the fine-tuning of ER-dependent transcription, which is in consistent with the pioneering role of GATA3 (Hruschka et al. 2020). In addition, the X308 splice GATA3 mutation interfered with the progesterone-induced antiproliferative function of PR in T-47D cells expressing mutated GATA3 compared to the control cells (Hruschka et al. 2020). These data showed that the GATA3 X308 splice mutation interferes with both ER and PR signalling, suggesting a crucial role of GATA3 as a co-factor for both ER and PR.

The D336fs frameshift mutation is another common GATA3 mutation in primary human breast cancers, having been reported 16 times in TCGA and METABRIC datasets

(Gustin et al. 2017). The MCF-7 model is the best example of an ER+ BC line carrying a heterozygous frameshift mutation (D336fs) in the second zinc finger of GATA3, which makes it a useful model to study this clinically relevant phenomenon. As a result of this mutation, MCF-7 cells express both the full length GATA3 protein, as well as a truncated protein (approximately 37 kDa) which has significantly higher steady state levels compared to the full-length variant (Adomas et al. 2014). The D336fs mutation does not affect nuclear localization of GATA3 in MCF-7 cells, or in the T-47D ER+ BC cell line in which mutant GATA3 is induced (Adomas et al. 2014), however, the mutant form of GATA3 was present in the cytoplasm of the cells. This suggests that this truncation mutation has impaired interaction with chromatin and is readily translocate from the nucleus to the cytoplasm (Adomas et al. 2014). It has been shown that the GATA3 cis-regulatory profile in MCF-7 cells bearing the mutation is bigger than in T-47D cells (wild-type GATA3), however, no major functional differences among genes associated with the GATA3 cistrome was identified. This may suggest that excess binding events in MCF-7 cells are not transcriptionally functional (Adomas et al. 2014). It has been suggested that the bigger GATA3 cistrome in MCF-7 cells could be due to a compromised ability of the mutant protein to recognise the specific GATA3 motif (Adomas et al. 2014). However, the proportion of GATA3 peaks containing the GATA3 motif was almost identical between the two lines, suggesting that the mutation did not influence chromatin binding specificity in MCF-7 cells (Adomas et al. 2014).

Although GATA3 mutations are widely thought to be LOF, Takaku *et al.*, demonstrated for the first time that not all mutations in the GATA3 TF are equivalent to each other in terms of patient prognosis (Takaku et al. 2018). ZnFn2 mutations impact breast cancer through both gain (GOF) and loss of function. Patients with a mutation in the GATA3 ZnFn2 region had significantly worse survival (10-year survival rate) compared to

patients carrying the wild-type GATA3 allele or other GATA3 mutations such as splice site mutations, suggesting that ZnFn2 mutations might be more damaging than complete loss of the binding domain (Takaku et al. 2018). Systematic gene expression analysis of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-ed T-47D cells harbouring a ZnFn2 2-nucleotide deletion mutation at arginine 330 (R330) position of GATA3 gene, indicated that a more aggressive gene signature was found in ZnFn2 mutant cells compared to the parental T-47D cells (Takaku et al. 2018). R330fs mutant T-47D tumours in nude mice had a higher growth rate, determined by a significantly higher luminescent signal compared to control mice (Takaku et al. 2018). Also, CHIP-seq analysis has shown that the GATA3 R330fs mutation induces redistribution of GATA3 binding events on the chromatin, resulting in upregulation of genes including *TWIST1* and *SNAIL2* (consistent with up-regulation of cell movement- and invasion-related pathways) and downregulation of genes such as PR, which is consistent with reduced cell differentiation-related and development-related pathways (Takaku et al. 2018). Since expression of PR is a key prognostic marker in breast cancer, its down-regulation is associated with a weaker anti-proliferative effect of progesterone and worse prognosis of the disease (Purdie et al. 2014). Additionally, activated-PR has been shown to direct ER cistrome within breast cancer cells, resulting in a unique gene expression programme that is associated with good clinical outcome (Mohammed et al. 2015). This indicates that the GATA3 ZnFn2 mutations play a role in modulating PR expression and progesterone sensitivity (Takaku et al. 2018). Although there are accumulating studies revealing the functional effect of the GATA3 mutations, the mechanism by which various types of mutations contribute to clinical prognosis and outcomes of patients with breast cancer remain to be elucidated.

## 1.4. Summary

GATA3 is a regulator of ER signaling in ER+ BC (Eeckhoute et al. 2007) and plays important roles in development and differentiation of the mammary gland (Asselin-Labat et al. 2007; Kouros-Mehr et al. 2006). Most ER+ BCs also express the AR (Ricciardelli et al. 2018), which has been shown to inhibit estrogen-induced growth of this disease subtype (Hickey, TE et al. 2021). Given the importance of AR in ER+ BC, there has been increasing interest in targeting the AR to treat ER+ BC. In ER- BC patients, AR positivity is also associated with improved disease-free survival and clinical and pathologic prognostic factors (e.g. lower tumour grade, smaller tumour, and increasing age) (Agoff et al. 2003), but its role in this disease context is controversial (Hickey, T et al. 2012). One key TF that has been strongly correlated with AR expression in BCs (both ER+ (Boto & Harigopal 2017) and TNBC (Kim et al. 2016)) is GATA3. To elicit its transcriptional activity, AR does not function alone, but within a transcriptional complex along with several co-regulatory proteins that are implicated in a diverse number of cellular functions. However, no study has broadly investigated the AR interacting proteins in different BC subtypes, to date, and no study has specifically examined the influence of GATA3 on AR signalling in different breast cancer contexts.

## 1.5. Hypothesis and Aims

I hypothesise that, 1) GATA3 is an important AR co-regulatory protein that contributes to AR-mediated growth inhibition in ER+ breast cancer and 2) GATA3 promotes AR-mediated cellular reprogramming associated with luminal lineage identity in a breast cancer context. These hypotheses are investigated using a variety of cell-based and PDX models of breast cancer, using modern next-generation sequencing techniques and different proteomic approaches including RIME.

**Aim 1:** Determine whether activation of AR and/or ER alters the GATA3 cistrome in breast cancer cells;

**Aim 2:** Examine whether GATA3 impacts the growth-inhibitory effect of AR in ER+ breast cancer;

**Aim 3:** Investigate the significance of GATA3 and AR interaction in breast cancer

**Aim 4:** Assess the pre-clinical significance of AR and GATA3 interactions in ER+ *in vivo* breast cancer models.

## **Chapter-2**

**AR interacts with GATA3 to induce a luminal epithelial phenotype in breast cancer through regulation of lineage-restricted genes**



The following chapter includes a manuscript formatted for submission to *Genome Biology* journal, followed by supplementary figures and extended data. This chapter makes up a significant proportion completed as a part of this PhD. A general discussion of this chapter had been included in Chapter 4.

## Statement of Authorship

Title of Paper	AR interacts with GATA3 to induce a luminal epithelial phenotype in breast cancer through regulation of lineage-restricted genes
Publication Status	<input type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input checked="" type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	This chapter is formatted for submission to Genome Biology journal

### Principal Author

Name of Principal Author (Candidate)	Lella Hosseinzadeh		
Contribution to the Paper	Performed experiments (Cell culture, qPCR, Western blotting, IF, PLA, ChIP-seq and ChIP-PCR, CO-IP, siRNA transfection, confocal imaging and slide scanning), Data mining and analysing, interpretation of data Experimental design, Manuscript writing, drafting and editing Preparation of figures		
Overall percentage (%)	75%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature	<table border="1"> <tr> <td>Date</td> <td>03.03.2021</td> </tr> </table>	Date	03.03.2021
Date	03.03.2021		

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Amy R. Dwyer		
Contribution to the Paper	Technical assistance, supervision, data interpretation, editing manuscript.		
Signature	<table border="1"> <tr> <td>Date</td> <td>03/03/21</td> </tr> </table>	Date	03/03/21
Date	03/03/21		

Name of Co-Author	Stephen M. Pederson		
Contribution to the Paper	Bioinformatic analysis		
Signature	<table border="1"> <tr> <td>Date</td> <td>3/3/21</td> </tr> </table>	Date	3/3/21
Date	3/3/21		

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Geraldine Laven-Law		
Contribution to the Paper	Technical assistance, bioinformatic analysis		
Signature		Date	3/3/21

Name of Co-Author	Zoya Kikhtyak		
Contribution to the Paper	Technical assistance		
Signature		Date	4/3/2021

Name of Co-Author	Wayne D. Tilley		
Contribution to the Paper	Funding, supervision, and data interpretation		
Signature		Date	05/07/2021

Name of Co-Author	Theresa E. Hickey		
Contribution to the Paper	Funding, supervision, data interpretation, study design and editing manuscript		
Signature		Date	3/3/21

## **Chapter-2: AR interacts with GATA3 to induce a luminal epithelial phenotype in breast cancer through regulation of lineage-restricted genes**

Leila Hosseinzadeh<sup>1</sup>, Amy R. Dwyer<sup>1</sup>, Stephen M. Pederson<sup>1</sup>, Geraldine Laven-Law<sup>1</sup>, Zoya Kikhtyak<sup>1</sup>, Wayne D. Tilley<sup>1</sup>, Theresa E. Hickey<sup>1</sup>.

<sup>1</sup> Dame Roma Mitchel Cancer Research Laboratories, Adelaide Medical School, University of Adelaide, Adelaide, Australia

## Abstract

**Background:** GATA3 transcription factor is considered critical for luminal development and differentiation in normal and malignant mammary epithelial cells (MECs). The androgen receptor (AR) has also been associated with luminal gene expression profiles in breast cancer, independent of the estrogen receptor (ER), and has been shown to promote a basal to luminal phenotype transition in mouse MECs. To date, the potential interaction of GATA3 and AR in transcriptional regulation of lineage driver genes in breast cancer has never been investigated.

**Results:** Our unbiased proteomic analysis identified GATA3 as a novel AR interacting protein in a variety of breast cancer cell types regardless of ER expression. We showed that AR and GATA3 interact in the cytoplasm and nucleus of normal and malignant breast epithelia, an interaction increased by androgen treatment. Androgen stimulation of breast cancer cells also induced nuclear translocation of AR and GATA3 and resulted in enrichment of co-localized AR and GATA3 chromatin binding events. Using ER<sup>+</sup> and/or ER<sup>-</sup> breast cancer cell lines and ER<sup>+</sup> patient-derived xenograft (PDX) models of breast cancer, we identified a conserved subset of AR agonist-induced AR and GATA3 co-occupied cis-regulatory elements across all models. Knockdown experiments indicated that GATA3 acts as an AR co-regulator to upregulate transcription of known luminal-lineage genes (e.g. *EHF*, *AQP3*, and *KDM4B*). Also, we showed an induction in the chromatin accessibility at those subsets of common AR-GATA3 cis-regulatory elements, which drive the luminal lineage identity in breast cancer upon androgen stimulation.

**Conclusions:** Our findings show a cooperative role for AR and GATA3 in driving the luminal-lineage identity in mammary epithelial cells. Also, our data explains the association between AR and luminal gene profiles in breast cancer.

## Introduction

Breast cancer (BC) is the most common malignancy worldwide (Bray et al. 2018) and approximately 80 % of cases are considered to be driven by oncogenic estrogen receptor alpha (ER) signalling. Breast tumors expressing ER, termed as ER+, can be molecularly characterised into Luminal A and B subtypes. Since the luminal A subtype does not express the human epidermal growth factor 2 (HER2) and has a lower level of ki67 expression compared to the luminal B subtype, it is associated with a slightly better prognosis (Reis-Filho & Pusztai 2011). The remaining 20-25 % of breast cancer cases are negative for the expression of ER (termed as ER-) and are divided into HER2+, Basal-like, and normal-like molecular subtypes (Parker et al. 2009; Sørlie et al. 2001). Interestingly, the majority of ER+ breast cancers (more than 90 %) (Ricciardelli et al. 2018) and about 50 % of all ER- breast cancers (Niemeier et al. 2010; Park et al. 2010) express AR. There is currently a lot of interest in the role of AR in breast cancer, but its actions may differ among breast cancer subtypes (Hickey et al. 2012).

The AR has been shown by numerous clinical studies to be an independent predictor of breast cancer survival and outcome of ER+ breast cancer (Ricciardelli et al. 2018). In ER+ breast cancers, AR was able to interfere with ER-dependent transcriptional activity through competitive binding to select estrogen response elements (EREs) (Peters et al. 2009). However, using genome-wide methodologies, it has very recently been shown

that competitive binding to EREs is not the key mechanism by which AR inhibits the stimulatory effect of ER in breast cancer cells (Hickey et al. 2021). In the latter study, AR was shown to be a tumour-suppressor in ER+ BC, where it inhibits estrogen-stimulated growth of ER+ BC by reducing recruitment of ER, and its co-activators p300, and SRC3, to chromatin at key cell cycle genes, leading to growth arrest (Hickey et al. 2021).

The role of AR in ER- breast cancers is still equivocal. Despite some trials reporting that AR positivity in ER- breast cancer cases is associated with poor prognosis (Choi et al. 2015; Jiang et al. 2016), there are several other studies that showed AR expression either is not correlated with prognosis (Jongen et al. 2019; Zaborowski et al. 2019) or is associated with a better prognosis (Guiu et al. 2018; Kucukzeybek et al. 2018). AR positivity has also been shown to be associated with improved disease-free survival and more benign clinical and pathologic factors (e.g. lower tumor grade and smaller tumor) in ER- BC (Hu, XQ et al. 2017; Luo et al. 2010). While AR activity has been increasingly investigated in AR+/ER- breast cancers (Farmer et al. 2005; Iggo, R 2018; Iggo, RD 2011), the specific role of AR signaling in this disease context is not well understood.

Since AR is a transcription factor that interacts with multiple co-regulatory proteins that control its transcriptional activity, we undertook the first unbiased proteomic analysis of AR in ER+ and ER- BC contexts in order to better understand AR function in this disease.

## Results

### GATA3 is a novel AR interacting protein in breast cancer

To identify AR interacting proteins in breast cancer cells, we used an unbiased proteomic approach called RIME (Rapid Immunoprecipitation Mass spectrometry of

Endogenous proteins) (Mohammed et al. 2016). Briefly, asynchronous AR+/ER+ (ZR-75-1 and T-47D) and AR+/ER- (MFM-223 and MDA-MB-453) breast cancer cell lines were cross-linked and AR protein complexes immunoprecipitated from purified chromatin prior to peptide digestion and identification by mass spectrometry. To identify high confidence interactors, three independent replicates representing consecutive passages of cells were performed for each cell line. Identification of AR and several established AR interacting proteins (e.g. different members of the SWI/SNF chromatin remodelling complex and heat shock proteins) highlights the accuracy and robustness of our RIME datasets across all *in vitro* models ([Supplementary 1-A, Extended data 1-1](#)), and alludes to conservation of some features of AR signalling across different breast cancer contexts. We identified **110**, **59**, **130** and **119** AR interacting proteins in ZR-75-1, T-47D, MFM-223, and MDA-MB-453 cells, respectively, grouped into unique and common AR interacting proteins between 2 or more breast cancer models using an up-set plot ([Figure 1-A](#)). Interestingly, while FOXA1 is a known AR interactor in prostate cancer (PCa) (Paltoglou et al. 2017; Robinson et al. 2014; Stelloo et al. 2018), and has been shown to influence AR activity in MDA-MB-453 breast cancer cells (Robinson et al. 2011), it was not detected in all breast cancer models (FOXA1 was only detected in MDA-MB-453 3 replicates and 1 replicate of MFM-223 cells). A total of eleven proteins were identified that interacted with AR in all breast cancer cell lines investigated ([Figure 1-A,B](#)). We further classified these interacting proteins by their molecular function into DNA-binding transcription factors, RNA binding proteins and proteins with either catalytic, transferase, or SNARE (Soluble NSF attachment protein receptor) activities ([Figure 1-B](#)). Interestingly, GATA3 was identified as one of the DNA-binding transcription factors that interacted with AR in all models ([Figure 1-A-B](#)). Four unique GATA3 peptides were consistently detected across all the cell lines and exclusively



belong to the *homo sapiens* GATA3 protein, confirming the authenticity of GATA3 as a candidate AR interacting protein in breast cancer cells regardless of ER status (Supplementary 1-B,C). Since GATA3 is a critical regulator of luminal epithelial differentiation in mammary glands (Kouros-Mehr et al. 2008) and AR has also been implicated in promotion of a luminal mammary epithelial cell phenotype in mice (Tarulli et al. 2019), and is associated with a luminal gene signature in human ER- breast cancers (Lehmann et al. 2011; Robinson et al. 2011), we focused the current study on the interaction between AR and GATA3 in breast cancer.

Interaction between AR and GATA3 was confirmed using the Proximity Ligation Assay (PLA) in all four cell lines (Figure 1-C, Supplementary 2-A). Activation of AR by the most potent natural ligand, 5- $\alpha$ -dihydrotestosterone (DHT), significantly increased the number of AR-GATA3 interactions (Figure 1-D, Supplementary 2-B,C), which further increased with time (Supplementary 2-D). Surprisingly, some AR-GATA3 interactions were also detected under hormone-deprived conditions (Figure 1-C, Supplementary 2-D), suggesting that hormone activation of AR enhanced but was not required for interaction with GATA3. The AR-GATA3 interaction was further validated by co-immunoprecipitation (Co-IP) assays (Figure 1-E,F). Pull down with AR was associated with detection of GATA3 in all four breast cancer cell lines, and the GATA3 signal was increased upon treatment with DHT (Figure 1-E), consistent with the PLA results. Reciprocal pull down of GATA3 was associated with detection of AR in a DHT-dependent manner in the ER- breast cancer cell lines (Figure 1-F), but AR was not detected in the ER+ breast cancer cell lines regardless of DHT treatment (Supplementary 2-E). The latter observation is likely due to substantially lower AR protein levels in ER+ compared to the ER- breast cancer lines (Supplementary 2-F) and the fact that Co-IP methodology is less sensitive than RIME and PLA methodologies.

To show that AR interacts with GATA3 in primary tissues, we conducted PLA for clinical ER+/AR+ (n=2) and ER-/AR+ (n=2) malignant breast tissues, and observed AR-GATA3 interactions in all cases (Figure 1-G). More interactions were evident in ER+ tissues compared with the ER- tissues (Figure 1-G, Supplementary 2-G), which is consistent with the fact that GATA3 expression is higher in ER+ compared to ER- breast cancers (Fararjeh et al. 2018). Interestingly, we also observed AR-GATA3 interactions in the epithelial cells of non-malignant breast tissues from reduction mammoplasties (n=2). Although observational, these data indicate that an interaction between AR and GATA3 occurs in normal and malignant clinical contexts.

### GATA3 nuclear translocation is induced upon hormone stimulation in breast cancer

In PLA experiments, we observed both nuclear and cytoplasmic interaction between AR and GATA3 *in vitro* (Supplementary 3-A) and *in vivo* (Supplementary 3-B), so we assessed whether AR activation would induce nuclear translocation of AR and GATA3 via dual label immunofluorescence (IF). Androgen (DHT) stimulation caused translocation of AR in all cell line models, an expected feature indicative of AR activation (Figure 2-A-D). Notably, DHT also induced nuclear translocation of GATA3 with strong co-localization between the two factors (Figure 2-A-D), suggesting that AR and GATA3 are translocated as a complex. Since GATA3 is an important ER co-regulator in the ER+ breast cancer context, we investigated whether treatment with estrogen would have a similar effect on sub-cellular localization of GATA3. Our findings showed that the most potent natural ER ligand, 17- $\beta$ -estradiol (E2) had a moderate effect on AR and GATA3 nuclear translocation (Supplementary 3-C,D). Strong co-localization of AR and GATA3 was sustained upon E2 treatment, suggesting that this is an AR-mediated phenomenon, consistent with a previous

report of E2-induced AR nuclear translocation in ER+ but not ER- breast cancer cell lines (D'Amato et al. 2016). Indeed, when cells were treated with a combination of E2+DHT, AR and GATA3 co-localize in the nucleus of nearly all cells ([Supplementary 3-C,D](#)), similar to treatment with DHT alone ([Figure 2-A-B](#)). Collectively, this data suggests that AR and GATA3 interact in the cytoplasm of breast cancer cells in the absence of hormone, but activation of AR increases interaction and induces nuclear translocation of both factors into the nucleus as a complex.

### AR agonist-induced GATA3 cis-regulatory elements are co-occupied by AR at active chromatin loci

Given that activated AR increased interaction with GATA3 (described in [Figure 1-C-F](#)) and induced nuclear translocation of both factors (described in [Figure 2-A-D](#)), we next assessed the effect of AR activation in shaping the genome-wide chromatin binding profile (cistrome) of GATA3 in breast cancer. We performed GATA3 ChIP-seq in our four breast cancer cell lines following a 4-hour treatment with vehicle or DHT. Each experiment was performed using 3 independent biological replicates from consecutive passages of cells to generate consensus cistromes representing reproducible binding events. We also performed GATA3 ChIP-seq in tumours from two ER+/AR+ patient-derived xenograft models (PDXs) of endocrine resistant disease treated for 5 days *in vivo* with a vehicle control or an AR agonist (DHT or the selective AR modulator (SARM; Enobosarm), as previously reported (Hickey et al. 2021). We confirmed GATA3 expression in both PDX models ([Supplementary 4-A-B](#)) before performing ChIP-seq experiments. Principle component analysis (PCA) of the GATA3 ChIP-seq data revealed that treatment with AR agonist was the major source of variation between replicate experiments for each model ([Supplementary 5-A-F](#)), indicating a major impact of AR activation on the GATA3 cistrome

in breast cancer cells. DiffBind analysis revealed significant ( $FDR \leq 0.05$ ) alteration of GATA3 binding sites upon androgen stimulation in all our *in vitro* and *in vivo* breast cancer models ([Supplementary 6-A-F](#)). Although androgen stimulation mainly resulted in enrichment of new GATA3 binding sites, consistent with more GATA3 entering the nucleus, the HCI-005 PDX model showed a comparable number of significantly gained and lost GATA3 binding events ([Supplementary 6-A-F](#)). Notably, we observed a similar pattern of differentiated peaks across all the cell line models. Also, the GAR15-13 PDX model show similar patterns to cell lines models, however, HCI-005 show differences, possibly, due to various phenotypes of endocrine resistance. In order to assess the genome-wide relationship between GATA3 and AR chromatin occupancy following AR activation, we performed AR ChIP-seq co-incident with the GATA3 ChIP-seq in the MDA-MB-453 and MFM-223 ER- breast cancer cell lines and used our publicly available AR ChIP-seq data ((Hickey et al. 2021), ([GSE123770](#)) for the ER+ breast cancer models (ZR-75-1, T-47D, Gar15-13, and HCI-005). Two-factor MA plots were generated to show the relationship between AR and GATA3 binding following AR activation in each model, revealing 5 different sub-groups representing shared or unique AR and GATA3 binding events ([Figure 3-A-C, Supplementary 7-A-C](#)). Notably, the majority of GATA3 binding events induced by AR activation were co-occupied by AR (shown in red dots in MA plots) representing 84 % of total androgen-induced GATA3 binding sites in ZR-75-1, 94 % in T-47D, 82 % in MFM-223, 86 % in MDA-MB-453, 80 % in Gar15-13 PDX, and 50 % in HCI-005 PDX models ([Supplementary 7-D-I](#)). These observations are consistent with the effect of AR activation on interaction and nuclear translocation of GATA3 and suggest that AR and GATA3 are either binding to their respective motifs in close proximity to each other or that one factor is tethering another. Discriminative DNA motif analysis of AR agonist-induced GATA3

binding sites identified AR response elements (AREs) among the most highly enriched motifs across all the breast cancer models (Figure 3-D, Supplementary 7-J, Extended-data-2), although GATA3 motifs were also significantly enriched. Interestingly, loci corresponding to AR agonist-induced AR and GATA3 recruitment were located near known AR target genes in breast cancer cells (e.g. *SEC14L2*, *C1orf116*, *ZBTB16*, and *RANBP3L*) (Hickey et al. 2021), indicating that GATA3 may play a key role in AR-mediated transcription (Figure 3-A-C, Supplementary 7-A-C). To determine changes to the transcriptional activation status of chromatin associated with AR agonist-induced AR-GATA3 binding events, we performed histone H3 lysine 27 acetylase (H3K27ac) ChIP-seq in the ER- breast cancer models (Supplementary 8-A-C, Extended-data-3) and utilized our publicly available H3K27ac ChIP-seq data for the other models (Hickey et al. 2021). Heatmaps show an increased H3K27ac signal at the AR agonist-induced AR-GATA3 binding events across all breast cancer models, especially at stronger peaks (Figure 3-E-G, Supplementary 9-A-C). This suggests that in addition to an induction of AR-GATA3 binding events, androgen stimulation remodels the chromatin landscape at these newly gained AR-GATA3 binding sites. Examples of H3K27ac enrichment at loci co-occupied by AR and GATA3 at AR target genes is shown in Figure 3 (Figure 3-H). Taken with Figure 1 and Figure 2, our ChIP-seq data support the concept that activation of AR induces nuclear translocation of both AR and GATA3, and this results in new, transcriptionally active GATA3 DNA binding events that are closely associated with AR DNA binding sites.

## AR agonist-induced AR-GATA3 cis-regulatory elements are associated with gene regulatory profiles involved in development and differentiation of the mammary epithelium

To understand the biological significance of transcriptionally active AR agonist-induced AR-GATA3 binding events at the transcriptomic level, we first annotated these binding events to potential targets by filtering for genes within a 100 kb distance from the transcription start sites (TSS) in each of our breast cancer models. We found thousands of candidate target genes associated with AR and GATA3 common bindings (Figure 4-A, Supplementary 10-A). Then, we conducted RNA-seq assays in all our *in vitro* ER+ and ER- breast cancer models for DHT-treated cells vs the unstimulated control cells. We also utilized *in vivo* RNA-seq data from ((Hickey et al. 2021), [GSE123770](#)) for the PDX models (Enobosarm or DHT vs vehicle datasets). After performing differential expression analysis for each model (Supplementary 10-B-G), we integrated the annotated genes from AR-GATA3 common regions with the differentially expressed genes (DEGs) from RNA-seq data, narrowing down the list of candidates AR-GATA3 targets (Figure 4-A, Supplementary 10-A). We then applied gene set enrichment analysis (GSEA) to these candidate AR-GATA3 target genes to identify associated biological processes. This analysis uncovered Gene Ontology (GO) terms pertaining to development and differentiation of mammary epithelial tissue (such as mammary gland development, mammary gland duct morphogenesis, epithelial cell differentiation, cell-cell adhesion, and apical-junction assembly) (Figure 4-B-D, Supplementary 11-A-C). Also, GSEA revealed positive enrichment of hallmark gene sets associated with canonical AR signalling and luminal differentiation (e.g., Androgen-response, and apical-junction, early and late estrogen response, and IL2-STAT5-signalling), and suppression of gene sets related to proliferation and cell cycle (e.g., G2/M-checkpoint, E2F-targets, mitotic-spindle, and MYC-targets-V1) *in vitro* and/or *in vivo* (Figure 4-E,F,

[Supplementary 12-A-C](#)). Collectively, these data suggest that AR and GATA3 form a complex in which they coregulate genes involved in mammary gland development and differentiation. This led us to explore whether GATA3 is required for AR-mediated regulation of these target genes.

### GATA3 acts as an AR co-regulator in breast cancer cells

To investigate whether DHT-induced GATA3 binding sites co-occupied by AR (shown in red in Figure 3) are AR-dependent events, we performed GATA3 ChIP-PCR in the presence or absence of siRNA-mediated AR knock-down in T-47D (as an ER+ model) and MDA-MB-453 (as an ER- model) cells at representative loci (e.g. *SEC14L2*, *ZBTB16*, and *C1orf116*) highlighted in Figure 5-A ([Figure 5-A](#)). Cells were treated with and without DHT. As a negative control, we also tested a GATA3 locus associated with the *c-FOC* gene that was not influenced by AR activation in any of the models. Silencing AR abolished DHT-induced enrichment of GATA3 at representative loci in both models and, as expected, GATA3 enrichment was not altered by AR knockdown at the *c-FOC* locus ([Figure 5-B](#)). Furthermore, abolition of AR prevented DHT-induced expression of genes associated with these loci in both ER+ and ER- models ([Supplementary 13-A](#)). The efficacy of AR knock-down and its effect on GATA3 protein expression in T-47D and MDA-MB-453 breast cancer cells was checked through western blotting ([Supplementary 13-B,C](#)). No significant effect on GATA3 protein expression was evident after silencing AR in either model, indicating that loss of GATA3 at AR-GATA3 co-occupied loci was not due to reduced levels of GATA3 ([Supplementary 13-B,C](#)). Taken together, these results indicate that GATA3 binding at AR target genes is dependent on AR.

To determine whether GATA3 functions as an AR co-regulator at these loci, we assessed the effect of GATA3 silencing on the ability of AR to bind DNA and induce transcription at the same representative loci highlighted in Figure 5-A. As a negative control, we assessed AR binding at a locus associated with the *FKBP5* AR target gene that is not co-occupied by GATA3. While silencing GATA3 significantly reduced DHT-induced enrichment of AR to representative loci, AR binding was not abolished (Figure 5-C). As expected, there was no significant change in AR enrichment at the *FKBP5* GATA3-independent locus after GATA3 knock-down (Figure 5-C). These data indicate that GATA3 facilitates but is not critical for AR binding at DHT-induced co-occupied loci. Importantly, GATA3 knock-down significantly reduced DHT-induced transcript levels of the representative AR target genes (Supplementary 14-A), indicating that the reduction of AR binding at associated loci had a functional impact on AR transactivation capacity. The efficacy of GATA3 knock-down and its effect on AR protein expression level were checked by western blotting in the T-47D and MDA-MB-453 breast cancer cells (Supplementary 14-B,C). Silencing GATA3 did not change the AR protein level in the T-47D and MDA-MB-453 cells. This suggests that the reduction of AR binding in the absence of GATA3 is not due to reduced AR protein expression (Supplementary 14-B,C).

### AR agonist-induced modifications of the chromatin landscape at AR-GATA3 binding events associates with regulation of lineage-restricted genes

Since studies have shown a key role for GATA3 in promoting and maintaining the luminal epithelial phenotype in mammary glands (Kouros-Mehr et al. 2008; Kouros-Mehr et al. 2006), and AR has also been implicated in promotion of the luminal phenotype in mice (Tarulli et al. 2019), we specifically examined the relationship of AR and GATA3 at genes known to be important for establishment of the luminal lineage. Overlapping the



candidate AR-GATA3 DEG targets with GSEA luminal and basal signature gene sets ([Extended-data-4](#)), we identified tens of up-regulated luminal epithelial markers indicative of luminal phenotype promotion in all our luminal breast cancer models ([Figure 6-A-C](#), [Supplementary 15-A-C](#)). Furthermore, among those basal marker genes that were down-regulated upon androgen stimulation in each of our models, we found tens of basal genes that were direct targets of AR and GATA3 in each of our breast cancer models ([Figure 6-A-C](#), [Supplementary 15-A-C](#)). AR-GATA3 target genes that are located within transcriptionally active regions of chromatin include 33 up-regulated luminal marker genes (e.g., *EHF*, *CNTNAP2*, *FGFR2*, *AQP3*, *SPDEF*, *KDM4B*, and *CLDN8*) ([Supplementary 15-D](#)) and 8 down-regulated basal marker genes (e.g. *PALLD*, *ZNF519*, *SMAD3*, and *ELK3*) that were commonly found across at least 3 out of 6 of our breast cancer models ([Supplementary 15-E](#)). The effect of AR and GATA3 knock-down on the DHT-induced binding enrichment of *EHF*, *AQP3*, *KDM4B* (luminal markers) ([Supplementary 16-A-C](#)) was assessed through ChIP-PCR experiments in T-47D and MDA-MB-453 cell lines. Silencing AR significantly abolished DHT-induced enrichment of GATA3 at luminal representative loci in both ER+ and ER- breast cancer models, and silencing GATA3 negatively impacted the DHT-induced enrichment of AR binding at the selected luminal representative loci ([Supplementary 16](#)). This shows, for the first time, that AR and GATA3 co-operate to promote a luminal phenotype in both ER+ and ER- contexts. Also, using qPCR, we assessed the effect of AR and GATA3 knock-down on the mRNA expression level of selected luminal targets in both of the *in vitro* models (T-47D and MDA-MB-453) in DHT-treated cells compared with the vehicle-treated control cells ([Figure 6-D,F](#)). Knock-down of AR or GATA3 significantly reduced the DHT-induced mRNA expression level of luminal genes *in vitro* ([Figure 6-D-F](#)).

Collectively, our findings revealed that androgen stimulation induces AR-mediated translocation of GATA3 into the nucleus, resulting in a gain of new GATA3 binding sites that are predominantly associated with AR target genes, where GATA3 co-regulates AR function in promoting the expression of luminal epithelial driver genes and repressing the expression of the basal marker genes in breast cancer luminal models.

## Discussion

Herein, we identified GATA3 as a novel AR interacting protein in breast cancer *in vitro* and *in vivo* models irrespective of ER expression. We showed that AR agonist stimulated breast cancer cells induced nuclear translocation of AR and GATA3, which suggests that AR and GATA3 are translocated as a complex. The knockdown experiments reveal that GATA3 acts as AR co-regulator in breast cancer, facilitates AR binding at DHT-induced AR-GATA3 co-occupied loci and impacts on AR transactivation capacity. Exploring the relationship of AR and GATA3 at genes associated with the luminal phenotype revealed that the importance of AR-GATA3 interaction and function is in the promotion of luminal and repression of basal lineage-restricted genes, in breast cancer models *in vitro* and *in vivo*.

Despite the fact that there are many known AR-associated factors in normal and malignant prostate epithelial cells that can shape AR function (Chmelar et al. 2007), including FOXA1 (Sahu et al. 2011) and GATA2 (He et al. 2014; Wu et al. 2014), the critical regulators are still being identified. Concerning the breast cancer context, there is huge gap in our knowledge regards lack of a comprehensive catalogue of AR interacting factors in different subtypes of the disease. Like ER, to elicit its transcriptional activity, AR does not function alone, but within a transcriptional complex. Although, they are some

literatures that investigated the ER-associated factors (e.g. FOXA1 (Carroll et al. 2005), PBX-1 (Magnani et al. 2011), AP-2 $\gamma$  (Tan et al. 2011) and GREB1 (Mohammed et al. 2013)) in breast cancer, possibly because ER is the driver transcription factor in majority of the breast tumours, investigation of AR interacting factors in breast cancer has never been studied before. Therefore, we undertook the first comprehensive unbiased proteomic analysis of AR using RIME in both ER+ and ER- breast cancer models to better understand AR function in this disease. We identified GATA3 as a novel AR interacting protein among all ER+ and ER- models. Although, FOXA1 has been shown to be required for AR DNA-binding and functionality at a subset of bindings in molecular apocrine breast cancer cells (Robinson et al. 2011) we did not detect FOXA1 as a common AR interacting protein across all the models. In addition to GATA3, the other two transcription factors that interacted with AR in all models were JUNB and ERF, both previously reported as AR interacting proteins in prostate cancer (Paltoglou et al. 2017). However, due to the well-established role of GATA3 in mammary gland development and differentiation (Asselin-Labat et al. 2007) we focused to reveal the significant of AR-GATA3 interactions within breast cancer. It has been previously shown that in the mammary gland, GATA3 is exclusively expressed in luminal epithelial cells (both ductal and alveolar) (Asselin-Labat et al. 2007), and regulates the differentiation of mouse mammary stem cells into mature luminal epithelial cells (Kouros-Mehr et al. 2008). GATA3 has also been shown to maintain the differentiated state of the mature luminal epithelium in the murine mammary gland (Kouros-Mehr et al. 2006). Ectopic overexpression of GATA3, ER and FOXA1 in basal-like breast cancer cells has been shown to induce a luminal phenotype at the cellular and molecular levels (Kong et al. 2011). Interestingly, in a study by Tarulli *et al.*, 2019 AR has been shown to implicate in promotion of the luminal phenotype in mouse mammary epithelial cells *in vivo*, and its

inhibition increased basal cell activity *in vitro* (Tarulli et al. 2019). Interestingly, GATA3 expression is more strongly correlated with AR than with ER in all clinical breast cancer subtypes (Boto & Harigopal 2017; Kim et al. 2016). Therefore, we hypothesized that GATA3 may interact with AR to induce a luminal phenotype. Our data revealed a key role for AR-GATA3 in promoting luminal lineage features in breast cancer *in vitro* and *in vivo*. We showed that GATA3 acts as an AR co-regulator, impacting the AR transcriptional activities at a subset of AR agonist-induced loci associated with altered expression of lineage-restricted genes. For instance, EHF, KDM4B and AQP3 were identified as AR-GATA3 luminal target genes that were up-regulated upon AR agonist stimulation across all the *in vitro* and *in vivo* breast cancer models.

EHF is one of the epithelium-specific ETS (ESE) transcription factors that is highly expressed in the breast epithelium and contribute to the development of luminal progenitors from bipotent MaSCs during the luminal differentiation (Pellacani et al. 2016), however, *GATA3* is critical for further differentiation into mature luminal cells (Kouros-Mehr et al. 2006; Pellacani et al. 2016). Bioinformatic analysis of a mammary epithelial microarray dataset reveals EHF as an epithelial-specific transcription factor that may cooperate with GATA3 in its gene regulatory network (Kouros-Mehr & Werb 2006). EHF functions as a tumour-suppressor in prostate cancer through repressing expression of *EZH2* and promoting expression of the tumour suppressor *Nkx3.1* (Kunderfranco et al. 2010). Also, EHF ectopic expression in prostate cancer cells inhibited stem like properties and promoted epithelial differentiation through repressing EMT drivers, such as *TWIST1*, *ZEB2*, *NANOG* and *POU5F1* (Albino et al. 2012).

KDM4B is a H3K9me3 / H3K9me2 histone demethylase that associates with activation or maintenance of gene expression (Gaughan et al. 2013). KDM4B is an ER co-

regulator that interacts and forms a complex with GATA3 at the ER regulatory enhancers to regulate the ER expression in breast cancer cells (Gaughan et al. 2013). Also, KDM4B has been also showed to interact AR and regulates its signalling in endometrial (Qiu et al. 2015) and prostate cancer (Coffey et al. 2013). Mammary epithelial identity is partially shaped by the DNA methylation landscape, which varies between cell types (Huh et al. 2015). Genes that are highly expressed in mammary stem/progenitor cells (e.g. HOXA1 and TCF7L1) are hypomethylated (Bloushtain-Qimron et al. 2008), conversely, genes express in mature luminal cells including luminal-driver GATA3 transcription factor show more transcriptional activation, implying that DNA methylation is important in regulating the expression of lineage-specific transcription factors (Maruyama et al. 2011). It has been shown that KDM4B is essential for estrogen-dependent gene expression (Gaughan et al. 2013) and also, it is an important regulator of mammary gland development and differentiation, involving in differentiation direction of mammary stem cells to luminal stem cells and mature luminal cells (Holliday et al. 2018).

AQP3 is one of the aquaporin family members that express in mice and human MECs (Kaihoko et al. 2020). AQP3 has diffuse localization in the cytoplasm of ductal MECs and concentrated localization in the basolateral membrane of alveolar MECs during the late pregnancy and lactation periods (Kaihoko et al. 2020). AQP3 has been shown to be required for FGF2-induced migration of MDA-MB-231 cells (Cao et al. 2013). However, the role of AQP3 in breast cancer and in association with GATA3, AR has never been studied.

Collectively, our findings suggest a novel aspect of transcription factor cooperation in lineage determination of normal mammary and breast cancer cells. however, further investigations are needed to reveal the details of AR-GATA3 regulation of luminal target genes in the context of breast cancer progression.

## Material and Methods

### Cell culture

ZR-75-1 (ATCC®CRL-1500TM), T-47D (ATCC®HTB-133TM), MFM-223 (DMSZ, ACC 422), and MDA-MB-453 (HTB-131TM) breast cancer cell lines were obtained from American Type Cell Culture Collection (ATCC) (Manassas, VA, USA) or DSMZ-German Collection of Microorganisms and Cell Cultures GmbH (Leibniz-institute, DSMZ, Germany). All cell lines were routinely tested for mycoplasma infection and authenticity confirmed by short tandem repeat profiling (Cell Bank Australia). ZR-75-1 and T-47D were maintained in RPMI-1640 medium (Invitrogen) containing 10 % Foetal Bovine Serum (FBS) and 2 nM L-Glutamine (Sigma). MFM-223 were cultured in EMEM (Invitrogen) containing 10 % FBS, 2 nM L-Glutamine (Sigma), 1 x Non-essential Amino Acids (Sigma) and 1 x Insulin-Transferrin-Sodium Selenite (Sigma). MDA-MB-453 cells were maintained in DMEM (Invitrogen) medium containing 10 % FBS, 2 nM L-Glutamine (Sigma) and 1 x Sodium Pyruvate (Sigma). All lines were incubated at 37 °C and 5 % CO<sub>2</sub>. For hormone stimulation experiments, cells were starved for 72 hours in 5 % dextran-coated charcoal stripped (DCC) FBS medium.

### **Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME)**

The RIME technique was performed as described previously (Mohammed et al. 2016). Briefly, cells (MDA-MB-453, MFM-223, ZR-75-1, and T-47D) were seeded at approximately 80 % confluence in their appropriate growth medium and cultured for 48 hours cross-linked in 1 % formaldehyde for 7 minutes, quenched with 0.2 M Glycine, chromatin isolated then subjected to immunoprecipitation using magnetic beads pre-bound with 10 µg of AR antibody (Santa Cruz Biotechnology, AR N20; sc-816). An on-bead

peptide digestion was performed and a 2-5  $\mu$ l aliquot of diluted peptide mixture was analysed by Nano-LC-MS/MS. Peptides were identified using MS. Only those interacting proteins that were identified in 3 of 3 independent biological replicates were considered for further analysis. Additional filtering was achieved by excluding non-specific interactions that appeared in >1 of the 3 replicates of matching IgG negative controls.

### **Proximity ligation assay (PLA)**

The ZR-75-1, T-47D, MFM-223 and MDA-MB-453 breast cancer cells were cultured in hormone deprived condition for 72 hours before seeding on top of sterilized coverslips in 6-well culture plates. Cells were treated with Ethanol (vehicle control), 10 nM E2 and/or 10 nM DHT accordingly for 4 hours before fixation with 4 % Paraformaldehyde for 10 minutes in room temperature. Cells were then permeabilized with 0.1 % Triton X-100 for an hour at room temperature and stained with AR (LS-Bio; LS-B3326) and GATA3 (Invitrogen; 1A12-1D9) antibodies diluted in 10 % Donkey serum (in PBS) overnight at 4 °C. PLA probes were mixed and diluted 1/5 in Duolink In Situ Antibody diluent for 1 hour at 37 °C in a humid chamber. Ligation and amplification steps were conducted for 30 minutes and 100 minutes, respectively, at 37 °C according to the manufacturer's instructions (Duolink<sup>®</sup> PLA kit; Sigma-Aldrich). Cell nuclei were stained with DAPI (Invitrogen) before mounting with Duolink In Situ Mounting Medium.

### **Confocal imaging**

Images were sequentially acquired on an Olympus FV3000 confocal microscope. 9 sections that were systematically sampled at 20 x magnification were selected from 5 random spots for each slide as technical replicates. Level adjustments were applied across

entire images. Quantification of the total number of the cells and the cytoplasmic and nuclear foci inside was performed using Fiji software (ImageJ).

### **Immunofluorescent staining (IF)**

ZR-75-1, T-47D, MFM-223 and MDA-MB-453 breast cancer cells were cultured in DCC FBS media for 72 hours and then seeded on top of sterilized coverslips in 6-well plates and treated with Ethanol (vehicle control), 10 nM E2 and/or 10 nM DHT accordingly for 4 hours before fixation with 4 % Paraformaldehyde for 10 minutes at room temperature. Permeabilization was performed with 0.05 % Triton X-100 (in PBS). Dual immunofluorescence staining with AR (LS-Bio; LS-B3326) and GATA3 (Invitrogen; 1A12-1D9) antibodies was performed at 4 °C, overnight. Cells were incubated for 30 minutes at room temperature with secondary antibodies of Goat-anti-Rabbit (life Technologies; Alexa Fluor 568) and Goat-anti-mouse (life Technologies; Alexa Fluor 488) diluted in 10 % goat serum with the dilution of 1 in 400 before DAPI staining (Invitrogen) and mounting with Dako fluorescent mounting medium. Mounting media was air-dried for 24 hours before imaging.

### **Slide scanning**

Immunofluorescent stained slides were scanned through the AxioScan.Z1 (ZEISS). Representative images were processed using ZEN 3.0 (blue edition) software (ZEISS). All representative images were taken with the scale of 10 um maintaining channel intensity range of DAPI in blue (black(250); white(2000)), GATA3 in green (black(400); white(1600)), AR in red (black(130); white(260)).



## **Co-immunoprecipitation assay (Co-IP)**

ZR-75-1, T-47D, MFM-223, and MA-MB-453 cells were seeded at  $9 \times 10^6$  cells/plate,  $9 \times 10^6$  cells/plate,  $11 \times 10^6$  cells/plate, and  $10 \times 10^6$  cells/plate, respectively, and treated with ethanol or 10 nM DHT accordingly for 4 hours before harvest. Cells were cross-linked with 1 % formaldehyde, quenched with 2 M Glycine pH 7.5, and collected by scraping. The cells were suspended in lysis buffer (10 mM Tris-HCl (pH 8), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % Na-Deoxycholate, and 0.5 % N-lauryl sarcosine) in the presence of protease inhibitors (Complete(R), Roche) and sonicated for 10 cycles of '30 seconds on, 30 seconds off', using a Diagenode Bioruptor. After centrifugation, the supernatant was collected and immunoprecipitated with protein A magnetic beads (Dynabeads®, Invitrogen) pre-bound with 5 ug/IP of GATA3 (Abcam; ab199428) or AR (Abcam; ab108341) at 4 °C overnight excluding the Input samples. The following day, the beads were washed 4 times with the wash buffer (20 mM Tris HCL Ph 8, 50 mM NaCl, 1 mM EDTA, 0.1 % Tween 20, and 2 mM DTT (MW 154.25)) at room temperature, eluted with 30 µL of 0.2 M Glycine pH 2.6, neutralized with 1 M Tris-HCL pH 8, and boiled at 95 °C for 10 minutes to elute associated proteins, prior to analysis by Western blotting.

## **Western blotting**

Cells were harvested by scraping and lysed into Radioimmunoprecipitation assay (RIPA) buffer. Protein concentration was quantified by BCA protein assay (Thermo Fisher Scientific). Proteins were separated by SDS-PAGE on 10 % Criterion TGX Stain Free-gels (BIO-RAD) and then transferred to Amersham nitrocellulose blotting membranes (GE Healthcare). Blocking was carried out in 5 % skim milk in TBST for 2 hours. Immunoblotting for AR (Santa Cruz Biotechnology; N20; sc-816); ER (Santa Cruz Biotechnology; F-10; sc-

8002); GATA3 (Abcam; ab199428); B-actin (Abcam; ab6376) and GAPDH (Merck Millipore; mab374) was performed overnight, then membranes were washed with PBS + 0.1 % Tween-20 for three 10-minutes rounds. HRP-coupled secondary antibodies (Goat anti-Rabbit (Dako; 1:2000) or Rabbit anti-Mouse (Dako; 1:2000) were detected with Clarity Western ECL Substrate (BIO-RAD) and visualised using a BIO-RAD ChemiDoc (MP) imaging system. Veriblot (Abcam; ab131367) was used for the detection of the primary antibodies for Co-IP samples before ECL detection.

### **siRNA knockdown experiments**

Pre-designed siRNAs against GATA3, AR and a negative (non-targeting) control siRNA (Allstar-Neg. Control siRNA) were purchased commercially ([Extended-data-5](#)). siGATA3-1 and -2 (10 nM each) were used in T-47D cells, siGATA3-3 and -4 (5 nM each) were used in MDA-MB-453 cells. Two siARs were used with the concentration of 10 nM for both cell line models, where appropriate. 5 and 10 nM of siControl was used T-47D and MDA-MB-453 cells, respectively. Cells were transfected with siRNAs by reverse transfection using RNAiMAX (Invitrogen) (0.5 ul/cm<sup>2</sup>) at the time of seeding according to the manufacturer's protocols (Thermo Fisher Scientific, Waltham, MA, USA). For CHIP-PCR experiments, cells were cultured in full media supplemented with siRNA transfection mix for 24 hours, then hormone deprived for 48 hours before treatment. However, cells were transfected with siRNA for 48 hours for qPCR experiments.

### **RNA extraction and quantitative real-time PCR (qPCR)**

T-47D and MDA-MB-453 cells were seeded at 0.5 X 10<sup>6</sup> cells/well in 6-well plates and simultaneously transfected with siRNAs against GATA3 or AR. The medium and transfection mix were changed the following day to a hormone-depleted medium

supplemented with 5 % DCC-FBS. After 48 hours (media refreshed each day), both cell models were treated for 6 hours with Ethanol or 10 nM DHT, before collection with TriReagent (Sigma). Total RNA was extracted using Trizol (Sigma) followed by DNase treatment using the TURBO DNase Kit (Invitrogen) according to manufacturer's protocol. Reverse transcription was performed with 500 ng of total RNA using the iScript cDNA Synthesis Kit (BIO-RAD). The resulting cDNA was diluted 1:10 and used for qRT-PCR. mRNA levels were normalised to *GAPDH* using the  $\Delta\Delta C_t$  method in BIO-RAD CFX-manager software, as previously described (Hu, DG et al. 2016).

### **ChIP-sequencing (ChIP-seq)**

ZR-75-1, T-47D, MFM-223 and MDA-MB-453 cells were seeded in 15 cm plates at  $9 \times 10^6$ ,  $9 \times 10^6$ ,  $11 \times 10^6$ , and  $10 \times 10^6$  cells/plate, respectively, in their normal growth medium. Media was changed to phenol-red-free medium supplemented with 5 % DCC - stripped FBS after 24 hours and cells were allowed to grow for 2 days prior to treatment with daily media changes. Cells were then treated with either Vehicle (Ethanol) or 10 nM DHT for 4 hours prior to fixation and harvest. Each experiment was done in three independent biological replicates representing consecutive passages of cells. Additionally, frozen tumour tissues left over from experiments with the endocrine-resistant ER+ PDX models of breast cancer (Gar15-13 and HCI-005) were used for ChIP-seq representing (Gar15-13; Enobosarm vs vehicle) and (HCI-005; DHT vs vehicle) arms as described in (Hickey et al. 2021). In the latter experiments, tumours were harvested 5 days post treatment and cryo-sectioned before cross-linking. Tumors from vehicle (GAR15-13,  $n = 3$ ; HCI-005,  $n = 5$ ), Enobosarm (GAR15-13,  $n = 3$ ) and DHT (HCI-005,  $n = 5$ ) were used for PDX ChIP-seq. Cross-linking and sonication were carried out as mentioned above for Co-IP experiments for both *in vitro* and *in vivo* models. Immunoprecipitations were performed

using 5 ug/IP of GATA3 (Abcam; ab199428) (for all *in vitro* and *in vivo* models) and AR (Abcam; ab108341), or 2 ug/IP of H3K27ac (Abcam; ab4729) (only for ER- *in vitro* models) antibodies. Before sequencing the DNA samples of all *in vitro* and *in vivo* experiments, ChIP-PCR reactions were prepared using iQ SYBR Green Supermix (BIO-RAD) using primers listed in [Extended-data-5](#). PCR was performed with the CFX384 Real Time PCR Detection System (BIO-RAD) and standard cycling conditions. ChIP-PCR data were analysed by the percentage input method and further analysed as fold enrichment over negative control. DNA was sequenced using an Illumina NextSeq 500 (High Output) with 75 bp single-end reads. Raw data was processed in Galaxy. Briefly, trimmed FASTQ files were aligned to the hg19 genome assembly using Bowtie2 (version: 2.3.4.3, default parameters); mapped reads with a minimum MAPQ <10 and duplicate reads were removed using SAMtools; peaks were called using MACS2 callpeak (version: 2.1.1, default settings), with a pooled input sample as the control. Only peaks found in 2 out of the 3 replicates were kept for the consensus peak-set. For figures, ChIP-seq data was visualized using the Integrative Genomics Viewer (<https://igv.org/app/>). Heatmaps (Galaxy Version 3.3.2.0.1) and PCAs (Galaxy Version 3.3.2.0.0) were generated using Deeptools. Peak annotations were performed using Cisgenome (v2.0).

## **ChIP-PCR**

T-47D and MDA-MB-453 cells were seeded ( $9 \times 10^6$ ,  $10 \times 10^6$  cells/plate, respectively) in their normal growth media and transfected with siRNAs against either AR (10 nM siAR-1 and siAR-2), GATA3 (10 nM of siGATA3-1 and -2 in T-47D, and 5 nM of siGATA3-3 and -4 in MDA-MB-453), or siControl (5 or 10 nM accordingly) at time of seeding. Media was changed to appropriate hormone-depleted media containing 5 % DCC-FBS on the following day. Treating (10 nM DHT vs Ethanol in both cell lines for 4 hours),

cross-linking, harvesting, ChIP processing and PCR reactions were performed as previously mentioned in the ChIP-seq section.

## **RNA-seq**

ZR-75-1, T-47D, MFM-223 and MDA-MB-453 *in vitro* cells were seeded and treated with 10 nM DHT (vs vehicle) for 6 hours prior to collection with TriReagent (Sigma). Three independent replicate experiments representing consecutive passages of cells were used to generate samples for RNA-seq. RNA was extracted from cells using the Direct-Zol RNA kit (Zymo Research). RNA integrity was assessed using the Experion RNA StdSens Analysis kit (700-7111, BIO-RAD) on the Experion Automated Electrophoresis System (BIO-RAD) and quantified by Nanodrop 2000 (Thermo Fisher Scientific). Total RNA was supplied to the Cancer Research UK Cambridge Institute (CRUK-CI) Genomics Core Facility for library preparation and high throughput sequencing. Conversion of the RNA into sequencing libraries was performed using the TruSeq® Total RNA HT kit (Illumina). Sequencing was performed using the Illumina HiSeq 2500 with single-end 40 bp reads. Raw data was processed by the CRUK-CI Bioinformatic Core. Reads were aligned to hg19 (TopHat) and differential expression analysis was performed using DESeq.

RNA from the *in vivo* models (PDXs) was extracted using the RNeasy kit (QIAGEN). Extracted RNA from PDXs was sequenced using the Illumina NextSeq 500 and aligned to hg19 using CLC Genomics Workbench 11 (QIAGEN) as described previously (Hickey et al. 2021).

For RNA-seq validation, total RNA was DNase-treated using the TURBO DNase kit (Invitrogen) and reverse transcribed using the iScript cDNA Synthesis kit (BIO-RAD). PCR with reverse transcription was performed as per ChIP-PCR but using primers outlined in [Extended-data-5](#). Gene expression was calculated by the  $2^{-\Delta\Delta Ct}$  method and normalized to

the expression of *GAPDH* (*in vitro* models) or *IPO8* and *PUM1* (tumors), using the CFX Manager Software (BIO-RAD).

### **Immunohistochemistry (IHC)**

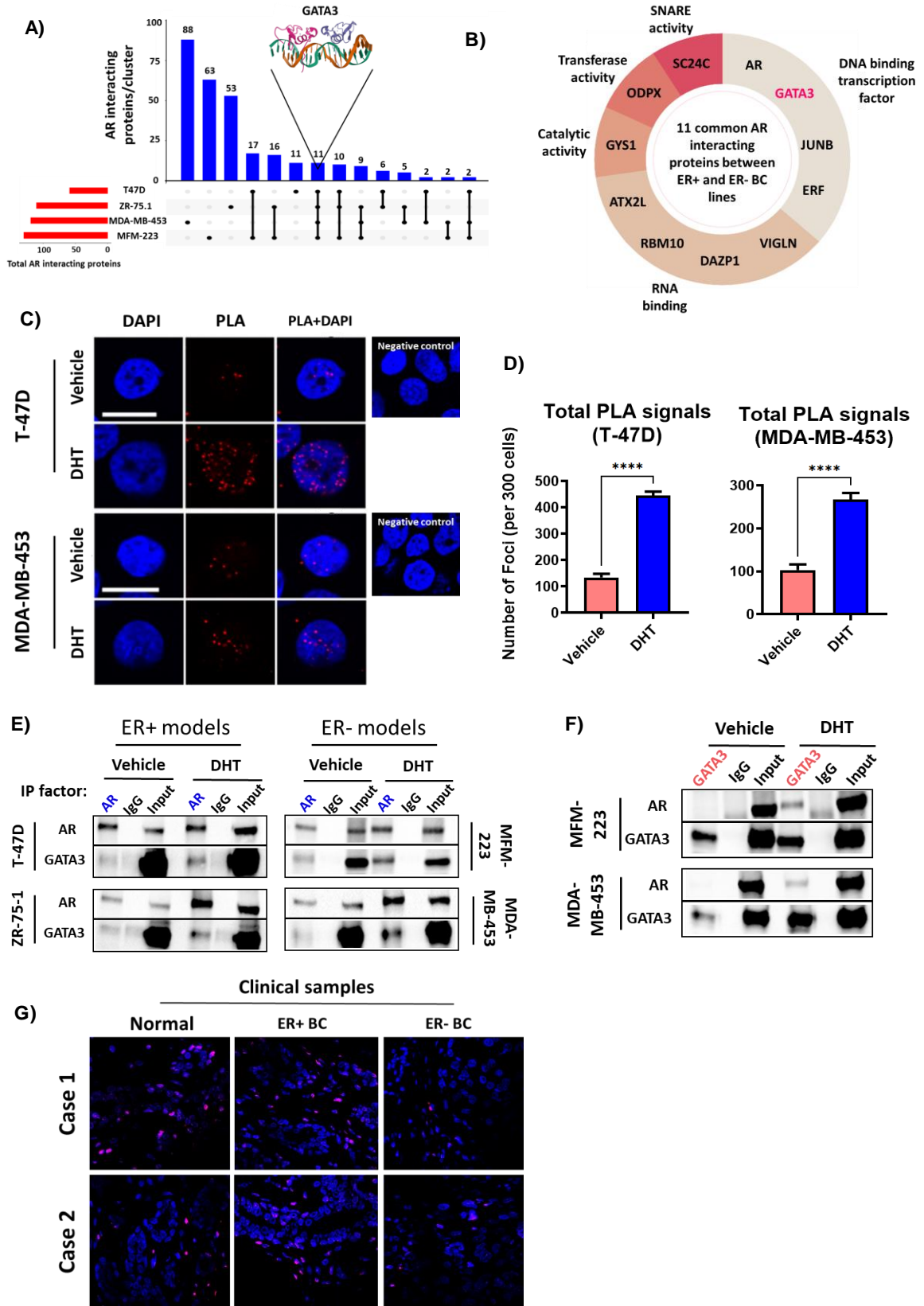
Standard immunohistochemical techniques were performed as previously described (Hickey et al. 2015) and employed using the following primary antibodies: AR (N-20, Santa Cruz SC-816) with the applied dilution of 1:1000; and GATA3 (ab199428, Abcam) with the applied dilution of 1:500. Appropriate positive and negative controls were included in all experiments. Slides were incubated overnight at 4 °C. Slides were scanned through Nanozoomer slide scanner (Hamamatsu #C9600).

### **Statistical analyses**

Statistical calculations were performed using GraphPad Prism (GraphPad Software). Normality was assumed for all statistical tests unless otherwise stated. Multiple comparisons were adjusted for Tukey's test, Student's t-test and ANOVA tests, where appropriate. All tests were two-sided with a 95 % confidence interval and a P value <0.05 was indicative of statistical significance. All ChIP-PCR experiments for AR, GATA3, and H3K27ac and qPCRs in all the models analysed using a two-way ANOVA, followed by Tukey's multiple comparisons test. Rest of the experiments were analysed using a two-way ANOVA, followed by Student's t-test.

# Main figures

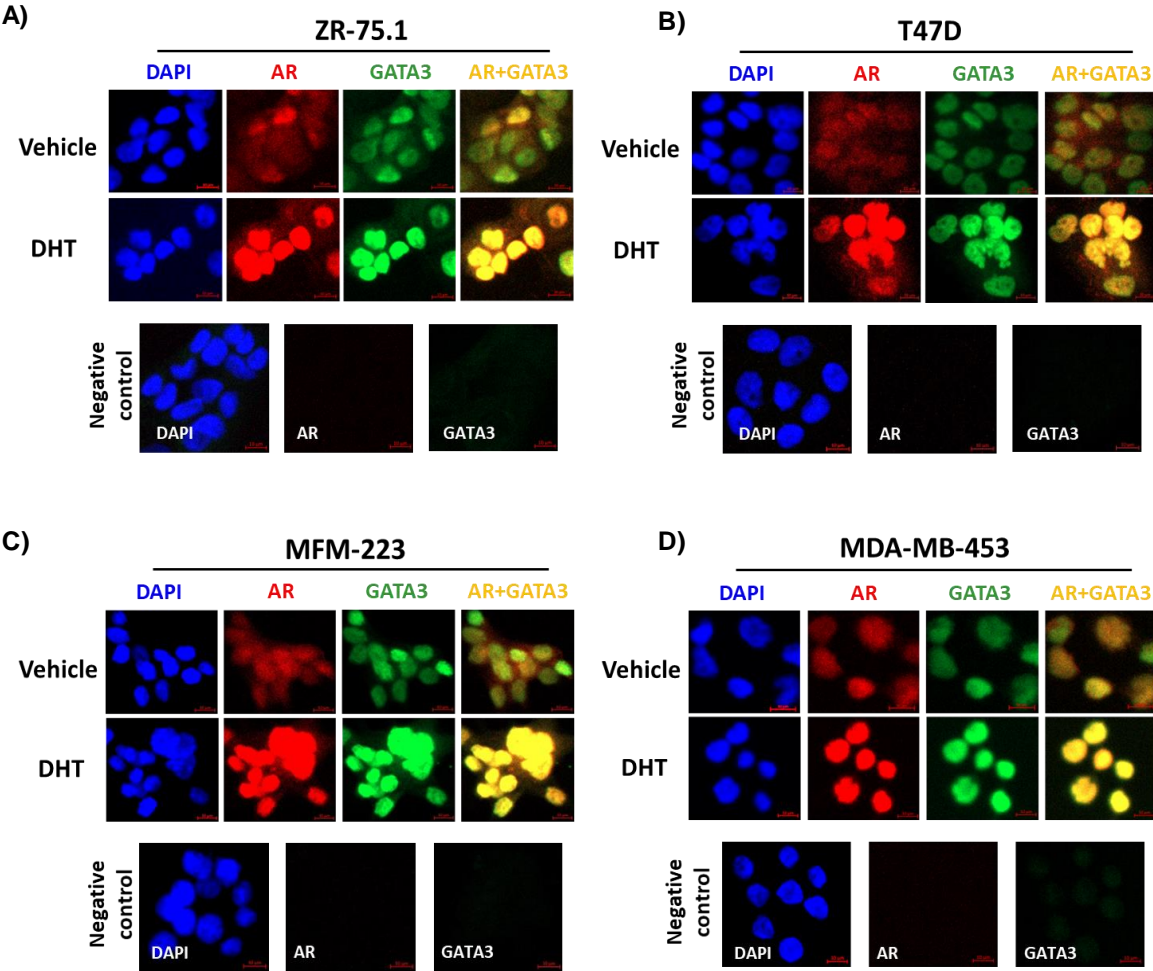
Figure 1: GATA3 is a novel AR interacting protein in breast cancer



**Figure-1: GATA3 is a novel AR interacting protein in breast cancer.** (A) Up-set plot representing common and unique AR interacting proteins among ER+ and ER- breast cancer cell lines, identifying GATA3 as one of the common interactors. (B) Pie chart showing the 11 AR interacting proteins common to ER+ and ER- *in vitro* breast cancer models, classified based on their molecular functions. (C) Proximity ligation assays (PLA) showing AR-GATA3 interactions in T-47D (ER+) and MDA-MB-453 (ER-) breast cancer lines comparing DHT-treated cells to vehicle-treated cells. Negative controls represent an assay performed without the secondary antibody. Images were captured using confocal-microscopy (scale bar: 20um). (D) Quantification of total PLA foci (average number of spots per 300 counted nuclei) in T-47D (ER+) and MDA-MB-453 (ER-) breast cancer lines after 4 hours of treatment with DHT. Two-way ANOVA with Student's t-test was used to determine statistically significant differences. Data shown as mean  $\pm$  SEM of five replicates and are representative of one independent experiment; \*\*\*\* p<0.0001. (E) Western blot showing AR-GATA3 interaction in 4 ER+ and ER- breast cancer lines. AR was immunoprecipitated from breast cancer cells and GATA3 was detected in association with AR protein in all models. IgG served as a negative control in all assays. (F) Western blot showing AR-GATA3 interactions through immunoprecipitating GATA3 in ER- models. (G) PLA images representing AR-GATA3 interactions in normal and malignant (ER+ and ER-) clinical breast tissues.

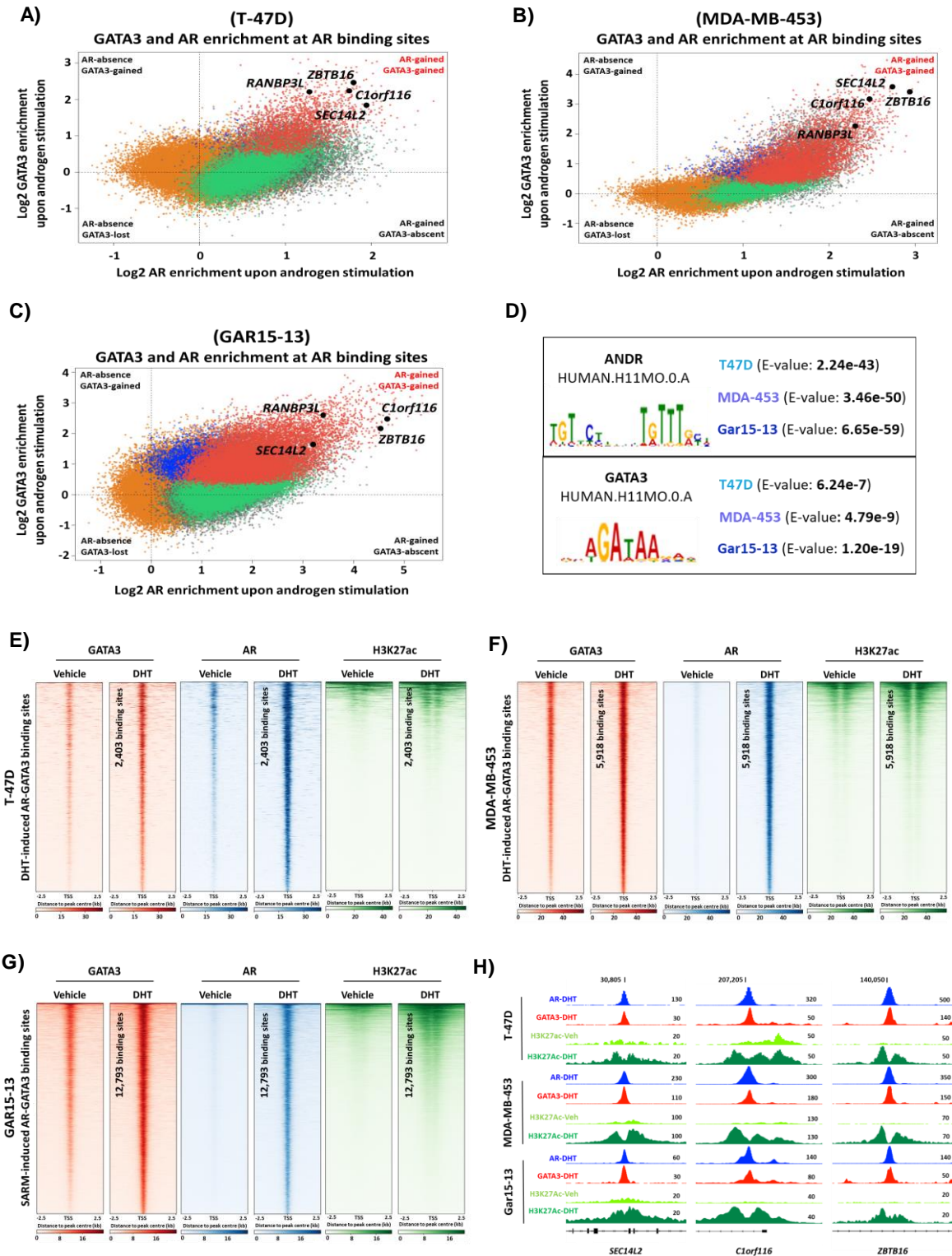


Figure 2: Activated AR induces AR and GATA3 nuclear translocation



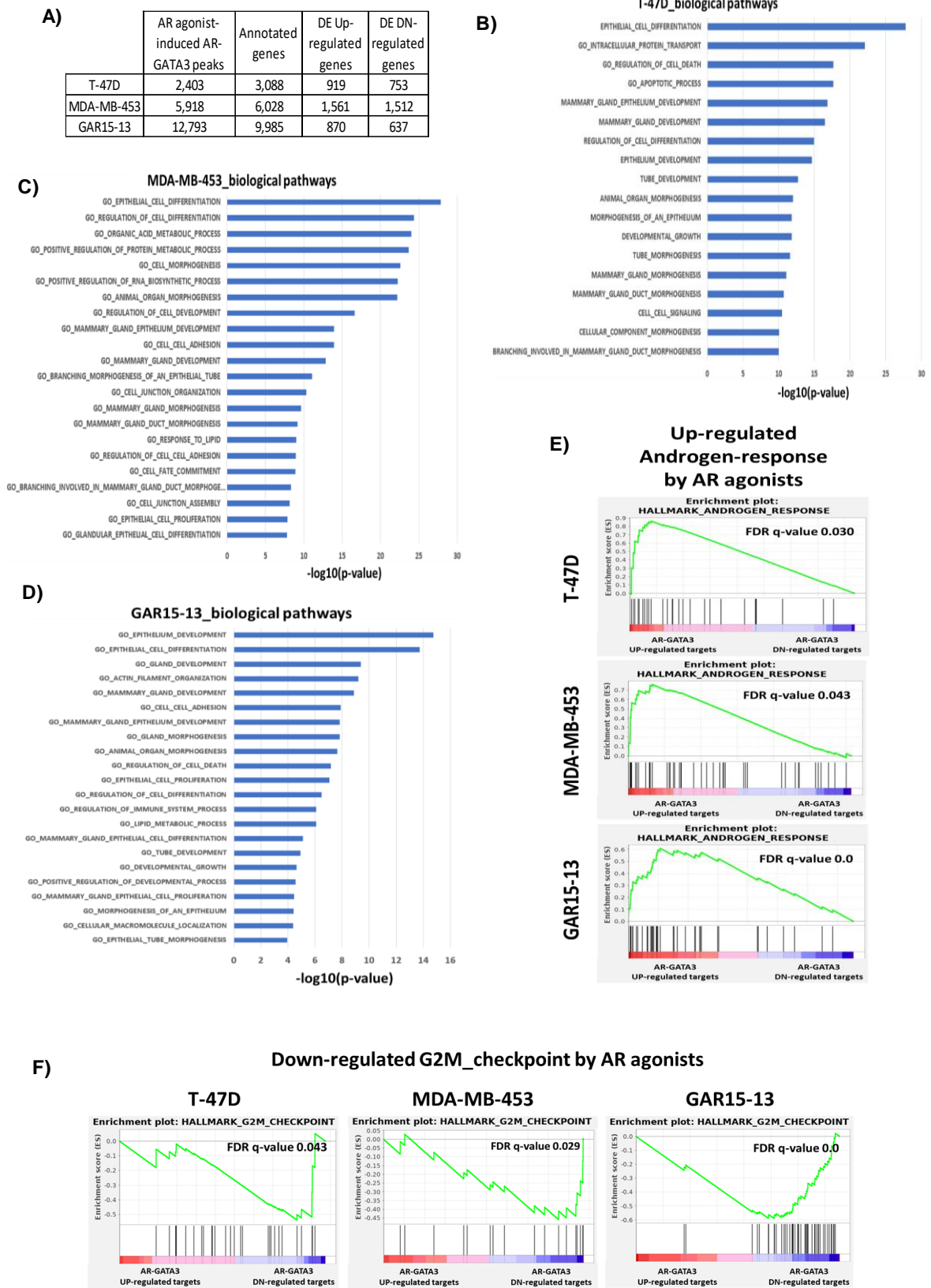
**Figure-2: Activated AR induces AR and GATA3 nuclear translocation.** Representative images showing the co-localization of AR and GATA3 in **(A)** ZR-75-1, **(B)** T-47D, **(C)** MFM-223, and **(D)** MDA-MB-453 breast cancer cells. Images were captured using a slide scanner (Zeiss Axio Scan.Z1) and processed using ZEN 2 (blue edition) software. Negative controls for all the *in vitro* models include an assay without the secondary antibody. All representative images were taken with the scale of 10 um and same intensity range for each channel. Cells were treated for 4 hours with 10 nM DHT (vs vehicle-treated control cells) prior to fixation and permeabilization.

**Figure 3: AR agonist-induced GATA3 cis-regulatory elements are co-occupied by AR at transcriptionally active chromatin loci**



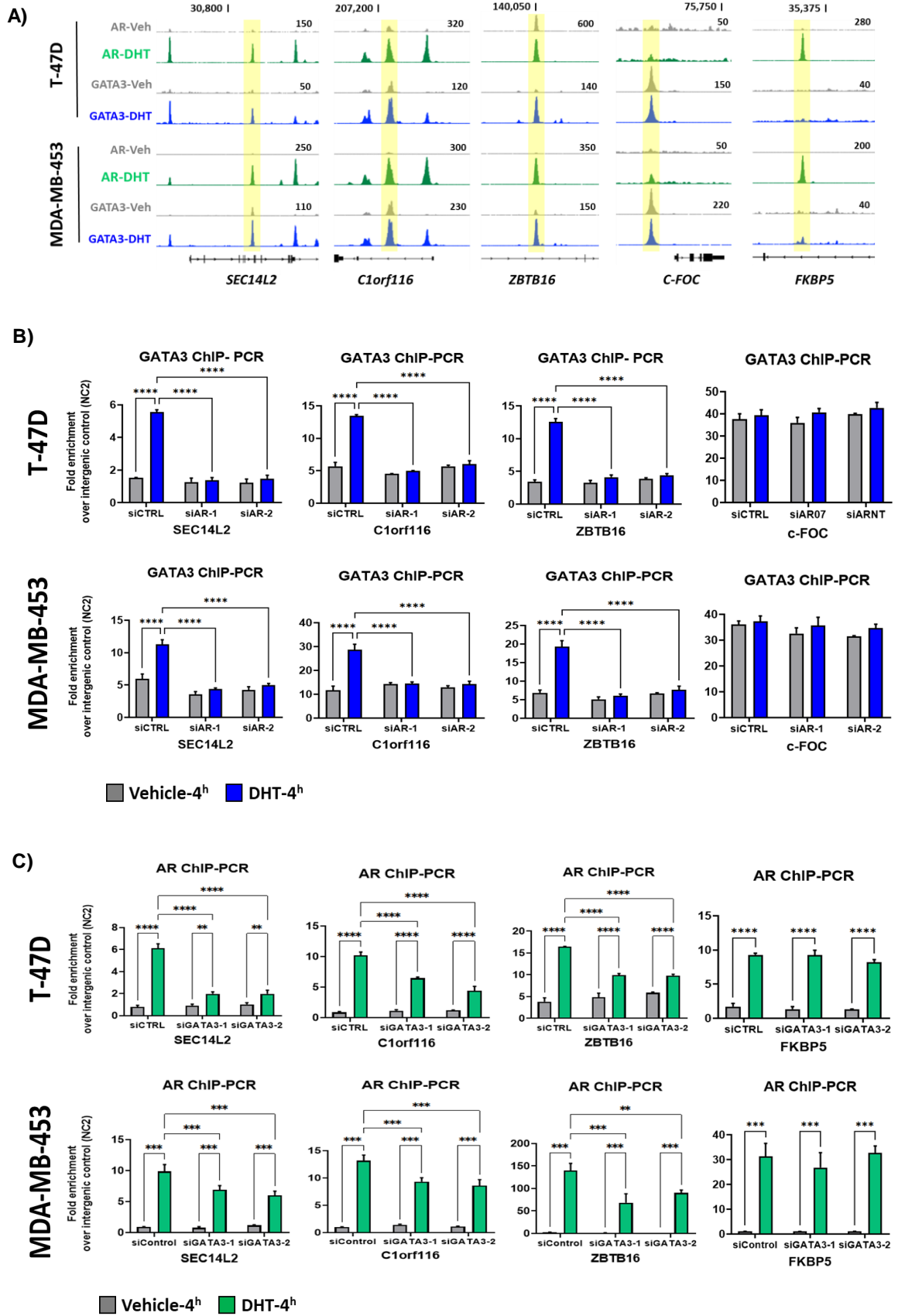
**Figure-3: AR agonist-induced GATA3 cis-regulatory elements are co-occupied by AR at transcriptionally active chromatin loci.** Distribution of the AR and GATA3 cistrome upon androgen stimulation is shown using two-factor log-ratio (M) plots for **(A)** T-47D and **(B)** MDA-MB-453 *in vitro* models and **(C)** GAR15-13 *in vivo* PDX model. Peaks are represented by coloured dots and classified into 5 different sub-groups: Orange and grey dots represent GATA3 and AR binding sites that do not overlap with the other factor, respectively; blue dots represent AR agonist-induced GATA3 peaks that are not shared with AR; green dots represent AR-GATA3 overlapping peaks that GATA3 peaks are not gained with AR agonist; and red dots represent AR-GATA3 peaks that are significantly gained upon AR activation. Example Loci associated with AR target genes (*ZBTB16*, *C1orf116*, *SE3C14L2*, *RANBP3L*) are highlighted in black dots and selected from the AR-GATA3 peaks that are induced by AR agonist in each model. **(D)** Motif analysis of AR agonist-induced GATA3 binding sites revealed AR as the most highly enriched motif across all *in vitro* and *in vivo* breast cancer models. Heatmaps showing AR agonist-induced AR-GATA3 binding sites and the associated H3K27ac binding events in **(E)** T-47D, **(F)** MDA-MB-453, and **(G)** GAR15-13 breast cancer models. **(H)** Genome browser image displaying averaged AR, GATA3, and H3K27ac ChIP-seq signals at binding sites associated with AR target genes (*SE3C14L2*, *C1orf116*, and *ZBTB16*) in three replicates of T-47D (top), MDA-MB-453 (middle), and GAR15-13 (bottom) breast cancer models.

**Figure 4: AR agonist-induced AR-GATA3 binding sites are implicated in development and differentiation of mammary epithelium**



**Figure-4: AR agonist-induced AR-GATA3 binding sites are implicated in development and differentiation of mammary epithelium. (A)** Table showing the number of AR agonist-induced AR-GATA3 binding sites and their associated genes that are differentially up- or down-regulated in both *in vitro* and *in vivo* models. Pathway analysis representing significant enrichment of pathways including mammary gland epithelium development and differentiation in **(B)** T-47D, **(C)** MDA-MB-453, and **(D)** GAR15-13 breast cancer models. **(E,F)** Gene set enrichment plots representing up- and down-regulated hallmark Androgen\_response and hallmark G2M\_checkpoint, respectively for *in vitro* and *in vivo* models.

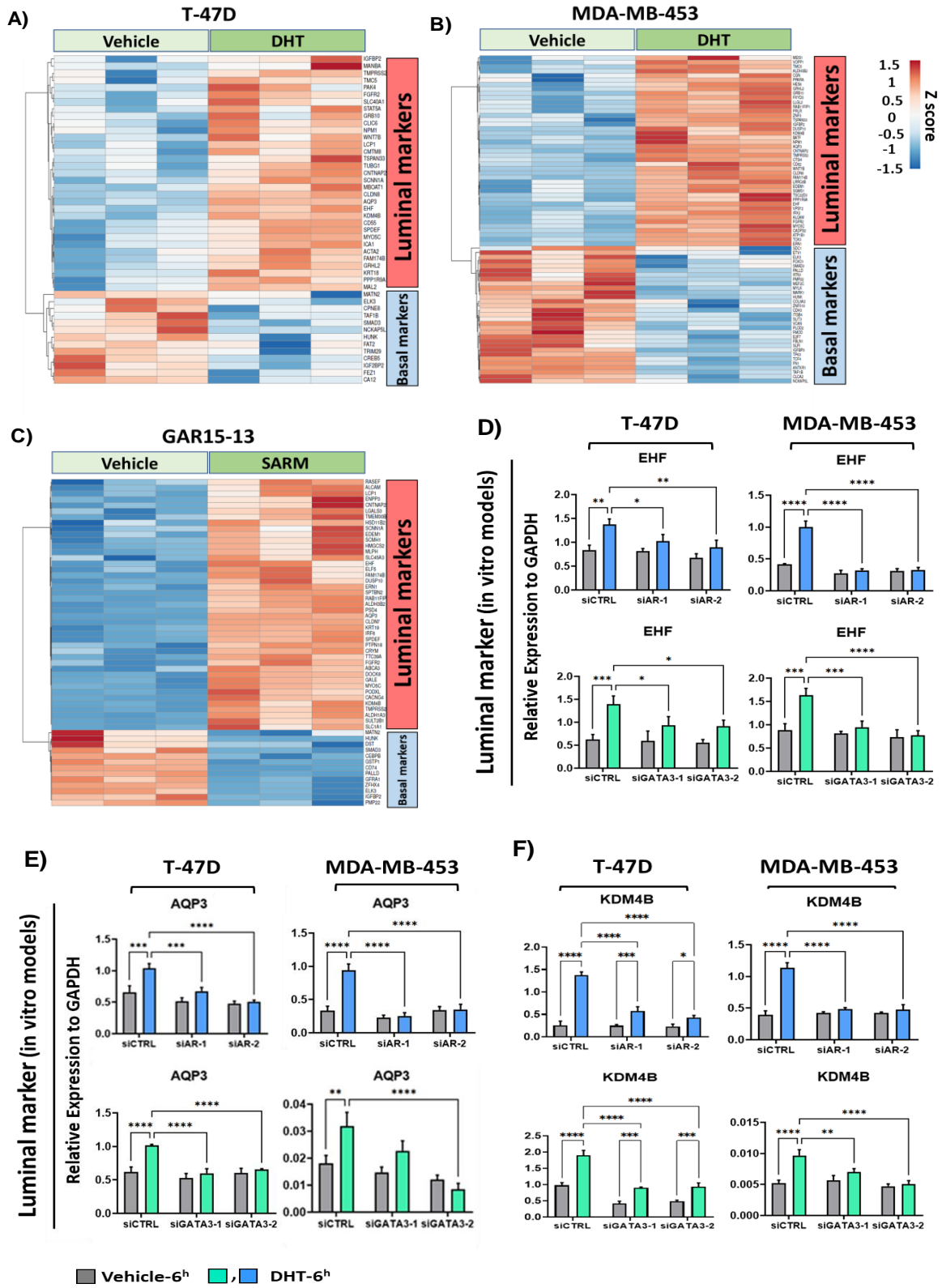
Figure 5: GATA3 acts as an AR co-regulator in breast cancer cells



**Figure-5: GATA3 acts as an AR co-regulator in breast cancer cells. (A)** Genome browser images showing averaged AR and GATA3 ChIP-seq signals at binding sites associated with AR target genes (*SEC14L2*, *C1orf116*, and *ZBTB16*) and genes used as negative controls for knock-down experiments (*c-FOC*, and *FKBP5*). Peaks represent the average signal of three independent replicates of T-47D (top) and MDA-MB-453 (bottom) cells. **(B)** Bar graphs showing the GATA3 ChIP-PCR results following AR knock-down at loci (represented in **A**) associated with AR target genes *SEC14L2*, *C1orf116*, *ZBTB16*, and *c-FOC* (negative control) in T-47D and MDA-MB-453 breast cancer cells transfected with siControl and two different siRNAs against AR. **(C)** Bar graphs representing the AR ChIP-PCR results following GATA3 knock-down at loci (represented in **A**) associated with AR target genes *SEC14L2*, *C1orf116*, *ZBTB16*, and *FKBP5* (negative control) in T-47D and MDA-MB-453 breast cancer cells transfected with siControl and two different siRNAs against GATA3. Cells in **B** and **C** were treated with DHT vs vehicle for 4 hours before harvesting. Two-way ANOVA with Tukey's multiple comparisons test was used to determine statistically significant differences in both **B** and **C**. Data shown as mean  $\pm$  SD of 3 replicates and are representative of two independent experiment; \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$ .



Figure 6: AR agonist-induced AR-GATA3 loci are associated with lineage-restricted genes in breast cancer *in vitro* and *in vivo* models



**Figure-6: AR agonist-induced AR-GATA3 loci are associated with lineage-restricted genes in breast cancer *in vitro* and *in vivo* models.** Heatmaps representing up- and down-regulated luminal and basal AR-GATA3 targets upon AR activation in **(A)** T-47D, **(B)** MDA-MB-453, and **(C)** GAR15-13 *in vitro* and *in vivo* breast cancer models, respectively. **(D-F)** Bar graphs representing the effect of AR knock-down and GATA3 knock-down on mRNA expression level of the *EHF*, *AQP3*, and *KDM4B* luminal gene markers in DHT-treated cells compared to control cells in T-47D, MDA-MB-453. Two-way ANOVA with Tukey's multiple comparisons test was used to determine statistically significant differences in **D**, **E** and **F**. Data shown as mean  $\pm$  SD of 3 replicates and are representative of two independent experiment; \*  $p=0.01$ , \*\*  $p<0.01$ , \*\*\*  $p<0.001$ , and \*\*\*\*  $p<0.0001$ .

## References

Albino, D, Longoni, N, Curti, L, Mello-Grand, M, Pinton, S, Civenni, G, Thalmann, G, D'Ambrosio, G, Sarti, M, Sessa, F, Chiorino, G, Catapano, CV & Carbone, GM 2012, 'ESE3/EHF controls epithelial cell differentiation and its loss leads to prostate tumors with mesenchymal and stem-like features', *Cancer Res*, vol. 72, no. 11, Jun 1, pp. 2889-2900.

Asselin-Labat, ML, Sutherland, KD, Barker, H, Thomas, R, Shackleton, M, Forrest, NC, Hartley, L, Robb, L, Grosveld, FG, van der Wees, J, Lindeman, GJ & Visvader, JE 2007, 'Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation', *Nat Cell Biol*, vol. 9, no. 2, Feb, pp. 201-209.

Bloushtain-Qimron, N, Yao, J, Snyder, EL, Shipitsin, M, Campbell, LL, Mani, SA, Hu, M, Chen, H, Ustyansky, V, Antosiewicz, JE, Argani, P, Halushka, MK, Thomson, JA, Pharoah, P, Porgador, A, Sukumar, S, Parsons, R, Richardson, AL, Stampfer, MR, Gelman, RS, Nikolskaya, T, Nikolsky, Y & Polyak, K 2008, 'Cell type-specific DNA methylation patterns in the human breast', *Proc Natl Acad Sci U S A*, vol. 105, no. 37, Sep 16, pp. 14076-14081.

Boto, A & Harigopal, M 2017, 'GATA3 Expression in Breast Cancers is Strongly Associated with AR Expression in a Large TMA Study', *IJCP*, vol. 103.

Bray, F, Ferlay, J, Soerjomataram, I, Siegel, RL, Torre, LA & Jemal, A 2018, 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA Cancer J Clin*, vol. 68, no. 6, Nov, pp. 394-424.

Cao, XC, Zhang, WR, Cao, WF, Liu, BW, Zhang, F, Zhao, HM, Meng, R, Zhang, L, Niu, RF, Hao, XS & Zhang, B 2013, 'Aquaporin3 is required for FGF-2-induced migration of human breast cancers', *PLoS One*, vol. 8, no. 2, p. e56735.

Carroll, JS, Liu, XS, Brodsky, AS, Li, W, Meyer, CA, Szary, AJ, Eeckhoute, J, Shao, W, Hestermann, EV, Geistlinger, TR, Fox, EA, Silver, PA & Brown, M 2005, 'Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1', *Cell*, vol. 122, no. 1, Jul 15, pp. 33-43.

Chmelar, R, Buchanan, G, Need, EF, Tilley, W & Greenberg, NM 2007, 'Androgen receptor coregulators and their involvement in the development and progression of prostate cancer', *Int J Cancer*, vol. 120, no. 4, Feb 15, pp. 719-733.

Choi, JE, Kang, SH, Lee, SJ & Bae, YK 2015, 'Androgen receptor expression predicts decreased survival in early stage triple-negative breast cancer', *Ann Surg Oncol*, vol. 22, no. 1, Jan, pp. 82-89.

Coffey, K, Rogerson, L, Ryan-Munden, C, Alkharaif, D, Stockley, J, Heer, R, Sahadevan, K, O'Neill, D, Jones, D, Darby, S, Staller, P, Mantilla, A, Gaughan, L & Robson, CN 2013, 'The lysine demethylase, KDM4B, is a key molecule in androgen receptor signalling and turnover', *Nucleic Acids Res*, vol. 41, no. 8, Apr, pp. 4433-4446.

D'Amato, NC, Gordon, MA, Babbs, B, Spoelstra, NS, Carson Butterfield, KT, Torkko, KC, Phan, VT, Barton, VN, Rogers, TJ, Sartorius, CA, Elias, A, Gertz, J, Jacobsen, BM & Richer, JK

2016, 'Cooperative Dynamics of AR and ER Activity in Breast Cancer', *Mol Cancer Res*, vol. 14, no. 11, Nov, pp. 1054-1067.

Fararjeh, AS, Tu, SH, Chen, LC, Liu, YR, Lin, YK, Chang, HL, Chang, HW, Wu, CH, Hwang-Verslues, WW & Ho, YS 2018, 'The impact of the effectiveness of GATA3 as a prognostic factor in breast cancer', *Hum Pathol*, vol. 80, Oct, pp. 219-230.

Farmer, P, Bonnefoi, H, Becette, V, Tubiana-Hulin, M, Fumoleau, P, Larsimont, D, Macgrogan, G, Bergh, J, Cameron, D, Goldstein, D, Duss, S, Nicoulaz, AL, Brisken, C, Fiche, M, Delorenzi, M & Iggo, R 2005, 'Identification of molecular apocrine breast tumours by microarray analysis', *Oncogene*, vol. 24, no. 29, Jul 7, pp. 4660-4671.

Gaughan, L, Stockley, J, Coffey, K, O'Neill, D, Jones, DL, Wade, M, Wright, J, Moore, M, Tse, S, Rogerson, L & Robson, CN 2013, 'KDM4B is a master regulator of the estrogen receptor signalling cascade', *Nucleic Acids Res*, vol. 41, no. 14, Aug, pp. 6892-6904.

Guiu, S, Mollevi, C, Charon-Barra, C, Boissière, F, Crapez, E, Chartron, E, Lamy, PJ, Gutowski, M, Bourgier, C, Romieu, G, Simony-Lafontaine, J & Jacot, W 2018, 'Prognostic value of androgen receptor and FOXA1 co-expression in non-metastatic triple negative breast cancer and correlation with other biomarkers', *Br J Cancer*, vol. 119, no. 1, Jul, pp. 76-79.

He, B, Lanz, RB, Fiskus, W, Geng, C, Yi, P, Hartig, SM, Rajapakshe, K, Shou, J, Wei, L, Shah, SS, Foley, C, Chew, SA, Eedunuri, VK, Bedoya, DJ, Feng, Q, Minami, T, Mitsiades, CS, Frolov, A, Weigel, NL, Hilsenbeck, SG, Rosen, DG, Palzkill, T, Ittmann, MM, Song, Y, Coarfa, C,

O'Malley, BW & Mitsiades, N 2014, 'GATA2 facilitates steroid receptor coactivator recruitment to the androgen receptor complex', *Proc Natl Acad Sci U S A*, vol. 111, no. 51, Dec 23, pp. 18261-18266.

Hickey, TE, Irvine, CM, Dvinge, H, Tarulli, GA, Hanson, AR, Ryan, NK, Pickering, MA, Birrell, SN, Hu, DG, Mackenzie, PI, Russell, R, Caldas, C, Raj, GV, Dehm, SM, Plymate, SR, Bradley, RK, Tilley, WD & Selth, LA 2015, 'Expression of androgen receptor splice variants in clinical breast cancers', *Oncotarget*, vol. 6, no. 42, Dec 29, pp. 44728-44744.

Hickey, TE, Robinson, JLL, Carroll, JS & Tilley, WD 2012, 'Minireview: The Androgen Receptor in Breast Tissues: Growth Inhibitor, Tumor Suppressor, Oncogene?', *Molecular Endocrinology*, vol. 26, no. 8, pp. 1252-1267.

Hickey, TE, Selth, LA, Chia, KM, Laven-Law, G, Milioli, HH, Roden, D, Jindal, S, Hui, M, Finlay-Schultz, J, Ebrahimie, E, Birrell, SN, Stelloo, S, Iggo, R, Alexandrou, S, Caldon, CE, Abdel-Fatah, TM, Ellis, IO, Zwart, W, Palmieri, C, Sartorius, CA, Swarbrick, A, Lim, E, Carroll, JS & Tilley, WD 2021, 'The androgen receptor is a tumor suppressor in estrogen receptor-positive breast cancer', *Nat Med*, Jan 18.

Holliday, H, Baker, LA, Junankar, SR, Clark, SJ & Swarbrick, A 2018, 'Epigenomics of mammary gland development', *Breast Cancer Res*, vol. 20, no. 1, Sep 3, p. 100.

Hu, DG, Selth, LA, Tarulli, GA, Meech, R, Wijayakumara, D, Chanawong, A, Russell, R, Caldas, C, Robinson, JL, Carroll, JS, Tilley, WD, Mackenzie, PI & Hickey, TE 2016, 'Androgen

and Estrogen Receptors in Breast Cancer Coregulate Human UDP-Glucuronosyltransferases 2B15 and 2B17', *Cancer Res*, vol. 76, no. 19, Oct 1, pp. 5881-5893.

Hu, XQ, Chen, WL, Ma, HG & Jiang, K 2017, 'Androgen receptor expression identifies patient with favorable outcome in operable triple negative breast cancer', *Oncotarget*, vol. 8, no. 34, Aug 22, pp. 56364-56374.

Huh, SJ, Clement, K, Jee, D, Merlini, A, Choudhury, S, Maruyama, R, Yoo, R, Chytil, A, Boyle, P, Ran, FA, Moses, HL, Barcellos-Hoff, MH, Jackson-Grusby, L, Meissner, A & Polyak, K 2015, 'Age- and pregnancy-associated DNA methylation changes in mammary epithelial cells', *Stem Cell Reports*, vol. 4, no. 2, Feb 10, pp. 297-311.

Iggo, R 2018, 'Classification of breast tumours into molecular apocrine, luminal and basal groups based on an explicit biological model', *bioRxiv*, p. 270975.

Iggo, RD 2011, 'New insights into the role of androgen and oestrogen receptors in molecular apocrine breast tumours', *Breast Cancer Res*, vol. 13, no. 6, p. 318.

Jiang, HS, Kuang, XY, Sun, WL, Xu, Y, Zheng, YZ, Liu, YR, Lang, GT, Qiao, F, Hu, X & Shao, ZM 2016, 'Androgen receptor expression predicts different clinical outcomes for breast cancer patients stratified by hormone receptor status', *Oncotarget*, vol. 7, no. 27, Jul 5, pp. 41285-41293.

Jongen, L, Floris, G, Wildiers, H, Claessens, F, Richard, F, Laenen, A, Desmedt, C, Ardui, J, Punie, K, Smeets, A, Berteloot, P, Vergote, I & Neven, P 2019, 'Tumor characteristics and outcome by androgen receptor expression in triple-negative breast cancer patients treated with neo-adjuvant chemotherapy', *Breast Cancer Res Treat*, vol. 176, no. 3, Aug, pp. 699-708.

Kaihoko, Y, Tsugami, Y, Suzuki, N, Suzuki, T, Nishimura, T & Kobayashi, K 2020, 'Distinct expression patterns of aquaporin 3 and 5 in ductal and alveolar epithelial cells in mouse mammary glands before and after parturition', *Cell Tissue Res*, vol. 380, no. 3, Jun, pp. 513-526.

Kim, S, Moon, BI, Lim, W, Park, S, Cho, MS & Sung, SH 2016, 'Expression patterns of GATA3 and the androgen receptor are strongly correlated in patients with triple-negative breast cancer', *Hum Pathol*, vol. 55, Sep, pp. 190-195.

Kong, SL, Li, G, Loh, SL, Sung, WK & Liu, ET 2011, 'Cellular reprogramming by the conjoint action of ER $\alpha$ , FOXA1, and GATA3 to a ligand-inducible growth state', *Mol Syst Biol*, vol. 7, Aug 30, p. 526.

Kouros-Mehr, H, Kim, JW, Bechis, SK & Werb, Z 2008, 'GATA-3 and the regulation of the mammary luminal cell fate', *Curr Opin Cell Biol*, vol. 20, no. 2, Apr, pp. 164-170.



Kouros-Mehr, H, Slorach, EM, Sternlicht, MD & Werb, Z 2006, 'GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland', *Cell*, vol. 127, no. 5, Dec 1, pp. 1041-1055.

Kouros-Mehr, H & Werb, Z 2006, 'Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis', *Dev Dyn*, vol. 235, no. 12, Dec, pp. 3404-3412.

Kucukzeybek, BB, Bayoglu, IV, Kucukzeybek, Y, Yıldız, Y, Oflazoglu, U, Atahan, MK, Taskaynatan, H, Alacacioglu, A, Yigit, S & Tarhan, MO 2018, 'Prognostic significance of androgen receptor expression in HER2-positive and triple-negative breast cancer', *Pol J Pathol*, vol. 69, no. 2, pp. 157-168.

Kunderfranco, P, Mello-Grand, M, Cangemi, R, Pellini, S, Mensah, A, Albertini, V, Malek, A, Chiorino, G, Catapano, CV & Carbone, GM 2010, 'ETS transcription factors control transcription of EZH2 and epigenetic silencing of the tumor suppressor gene Nkx3.1 in prostate cancer', *PLoS One*, vol. 5, no. 5, May 10, p. e10547.

Lehmann, BD, Bauer, JA, Chen, X, Sanders, ME, Chakravarthy, AB, Shyr, Y & Pietenpol, JA 2011, 'Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies', *J Clin Invest*, vol. 121, no. 7, Jul, pp. 2750-2767.

Luo, X, Shi, YX, Li, ZM & Jiang, WQ 2010, 'Expression and clinical significance of androgen receptor in triple negative breast cancer', *Chin J Cancer*, vol. 29, no. 6, Jun, pp. 585-590.

Magnani, L, Ballantyne, EB, Zhang, X & Lupien, M 2011, 'PBX1 genomic pioneer function drives ER $\alpha$  signaling underlying progression in breast cancer', *PLoS Genet*, vol. 7, no. 11, Nov, p. e1002368.

Maruyama, R, Choudhury, S, Kowalczyk, A, Bessarabova, M, Beresford-Smith, B, Conway, T, Kaspi, A, Wu, Z, Nikolskaya, T, Merino, VF, Lo, PK, Liu, XS, Nikolsky, Y, Sukumar, S, Haviv, I & Polyak, K 2011, 'Epigenetic regulation of cell type-specific expression patterns in the human mammary epithelium', *PLoS Genet*, vol. 7, no. 4, Apr, p. e1001369.

Mohammed, H, D'Santos, C, Serandour, AA, Ali, HR, Brown, GD, Atkins, A, Rueda, OM, Holmes, KA, Theodorou, V, Robinson, JL, Zwart, W, Saadi, A, Ross-Innes, CS, Chin, SF, Menon, S, Stingl, J, Palmieri, C, Caldas, C & Carroll, JS 2013, 'Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor', *Cell Rep*, vol. 3, no. 2, Feb 21, pp. 342-349.

Mohammed, H, Taylor, C, Brown, GD, Papachristou, EK, Carroll, JS & D'Santos, CS 2016, 'Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes', *Nat Protoc*, vol. 11, no. 2, Feb, pp. 316-326.

Niemeier, LA, Dabbs, DJ, Beriwal, S, Striebel, JM & Bhargava, R 2010, 'Androgen receptor in breast cancer: expression in estrogen receptor-positive tumors and in estrogen receptor-negative tumors with apocrine differentiation', *Mod Pathol*, vol. 23, no. 2, Feb, pp. 205-212.

Paltoglou, S, Das, R, Townley, SL, Hickey, TE, Tarulli, GA, Coutinho, I, Fernandes, R, Hanson, AR, Denis, I, Carroll, JS, Dehm, SM, Raj, GV, Plymate, SR, Tilley, WD & Selth, LA 2017, 'Novel Androgen Receptor Coregulator GRHL2 Exerts Both Oncogenic and Antimetastatic Functions in Prostate Cancer', *Cancer Res*, vol. 77, no. 13, Jul 1, pp. 3417-3430.

Park, S, Koo, J, Park, HS, Kim, JH, Choi, SY, Lee, JH, Park, BW & Lee, KS 2010, 'Expression of androgen receptors in primary breast cancer', *Ann Oncol*, vol. 21, no. 3, Mar, pp. 488-492.

Parker, JS, Mullins, M, Cheang, MC, Leung, S, Voduc, D, Vickery, T, Davies, S, Fauron, C, He, X, Hu, Z, Quackenbush, JF, Stijleman, IJ, Palazzo, J, Marron, JS, Nobel, AB, Mardis, E, Nielsen, TO, Ellis, MJ, Perou, CM & Bernard, PS 2009, 'Supervised risk predictor of breast cancer based on intrinsic subtypes', *J Clin Oncol*, vol. 27, no. 8, Mar 10, pp. 1160-1167.

Pellacani, D, Bilenky, M, Kannan, N, Heravi-Moussavi, A, Knapp, D, Gakkhar, S, Moksa, M, Carles, A, Moore, R, Mungall, AJ, Marra, MA, Jones, SJM, Aparicio, S, Hirst, M & Eaves, CJ 2016, 'Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks', *Cell Rep*, vol. 17, no. 8, Nov 15, pp. 2060-2074.

Peters, AA, Buchanan, G, Ricciardelli, C, Bianco-Miotto, T, Centenera, MM, Harris, JM, Jindal, S, Segara, D, Jia, L, Moore, NL, Henshall, SM, Birrell, SN, Coetzee, GA, Sutherland, RL, Butler, LM & Tilley, WD 2009, 'Androgen receptor inhibits estrogen receptor-alpha

activity and is prognostic in breast cancer', *Cancer Res*, vol. 69, no. 15, Aug 1, pp. 6131-6140.

Qiu, MT, Fan, Q, Zhu, Z, Kwan, SY, Chen, L, Chen, JH, Ying, ZL, Zhou, Y, Gu, W, Wang, LH, Cheng, WW, Zeng, J, Wan, XP, Mok, SC, Wong, KK & Bao, W 2015, 'KDM4B and KDM4A promote endometrial cancer progression by regulating androgen receptor, c-myc, and p27kip1', *Oncotarget*, vol. 6, no. 31, Oct 13, pp. 31702-31720.

Reis-Filho, JS & Puztai, L 2011, 'Gene expression profiling in breast cancer: classification, prognostication, and prediction', *Lancet*, vol. 378, no. 9805, Nov 19, pp. 1812-1823.

Ricciardelli, C, Bianco-Miotto, T, Jindal, S, Butler, LM, Leung, S, McNeil, CM, O'Toole, SA, Ebrahimie, E, Millar, EKA, Sakko, AJ, Ruiz, AI, Vowler, SL, Huntsman, DG, Birrell, SN, Sutherland, RL, Palmieri, C, Hickey, TE & Tilley, WD 2018, 'The Magnitude of Androgen Receptor Positivity in Breast Cancer Is Critical for Reliable Prediction of Disease Outcome', *Clin Cancer Res*, vol. 24, no. 10, May 15, pp. 2328-2341.

Robinson, JL, Hickey, TE, Warren, AY, Vowler, SL, Carroll, T, Lamb, AD, Papoutsoglou, N, Neal, DE, Tilley, WD & Carroll, JS 2014, 'Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype', *Oncogene*, vol. 33, no. 50, Dec 11, pp. 5666-5674.

Robinson, JL, Macarthur, S, Ross-Innes, CS, Tilley, WD, Neal, DE, Mills, IG & Carroll, JS 2011, 'Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1', *EMBO J*, vol. 30, no. 15, Jun 24, pp. 3019-3027.

Sahu, B, Laakso, M, Ovaska, K, Mirtti, T, Lundin, J, Rannikko, A, Sankila, A, Turunen, JP, Lundin, M, Konsti, J, Vesterinen, T, Nordling, S, Kallioniemi, O, Hautaniemi, S & Jänne, OA 2011, 'Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer', *EMBO J*, vol. 30, no. 19, Sep 13, pp. 3962-3976.

Sørli, T, Perou, CM, Tibshirani, R, Aas, T, Geisler, S, Johnsen, H, Hastie, T, Eisen, MB, van de Rijn, M, Jeffrey, SS, Thorsen, T, Quist, H, Matese, JC, Brown, PO, Botstein, D, Lønning, PE & Børresen-Dale, AL 2001, 'Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications', *Proc Natl Acad Sci U S A*, vol. 98, no. 19, Sep 11, pp. 10869-10874.

Stelloo, S, Nevedomskaya, E, Kim, Y, Hoekman, L, Bleijerveld, OB, Mirza, T, Wessels, LFA, van Weerden, WM, Altelaar, AFM, Bergman, AM & Zwart, W 2018, 'Endogenous androgen receptor proteomic profiling reveals genomic subcomplex involved in prostate tumorigenesis', *Oncogene*, vol. 37, no. 3, Jan 18, pp. 313-322.

Tan, SK, Lin, ZH, Chang, CW, Varang, V, Chng, KR, Pan, YF, Yong, EL, Sung, WK & Cheung, E 2011, 'AP-2 $\gamma$  regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription', *EMBO J*, vol. 30, no. 13, May 13, pp. 2569-2581.

Tarulli, GA, Laven-Law, G, Shehata, M, Walters, KA, Denis, IM, Rahman, MM, Handelsman, DJ, Dean, NR, Tilley, WD & Hickey, TE 2019, 'Androgen Receptor Signalling Promotes a Luminal Phenotype in Mammary Epithelial Cells', *J Mammary Gland Biol Neoplasia*, vol. 24, no. 1, Mar, pp. 99-108.

Wu, D, Sunkel, B, Chen, Z, Liu, X, Ye, Z, Li, Q, Grenade, C, Ke, J, Zhang, C, Chen, H, Nephew, KP, Huang, TH, Liu, Z, Jin, VX & Wang, Q 2014, 'Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer', *Nucleic Acids Res*, vol. 42, no. 6, Apr, pp. 3607-3622.

Zaborowski, M, Pearson, A, Sioson, L, Gill, AJ & Ahadi, MS 2019, 'Androgen receptor immunoexpression in triple-negative breast cancers: is it a prognostic factor?', *Pathology*, vol. 51, no. 3, Apr, pp. 327-329.

## Supplementary figures

### Supplementary 1: GATA3 is a novel AR interacting protein in breast cancer cells

A)

AR RIME	Description	Cell line	Replicates	ΣCoverage	Σ# Proteins	Σ# Unique Peptides	Σ# Peptides	Σ# PSMs
	[ANDR_HUMAN] Accession number of P10275 MW [kDa]: 99.1	ZR-75-1	rep1	31.52	1	19	19	142
			rep2	34.13	1	20	20	162
			rep3	29.35	1	18	18	115
		T47D	rep1	20.54	1	9	9	32
			rep2	15.00	1	9	9	34
			rep3	16.74	1	11	11	41
		MFM-223	rep1	34.67	1	21	21	187
			rep2	2.50	1	1	1	2
			rep3	31.30	1	19	19	157
MDA-MB-453	rep1	28.48	1	18	18	115		
	rep2	32.07	1	20	20	156		
	rep3	26.52	1	18	18	87		

B)

4 unique peptides for GATA3	Query Cover	E value	Per. Ident	Organism
<b>DVSPDPSLSTPGSAGSAR</b>	100%	2.00E-11	100.00%	Homo sapiens(taxid:9606)
<b>NSSFNPAALSR</b>	100%	3.00E-05	100.00%	Homo sapiens(taxid:9606)
<b>KVHDSLEDFPK</b>	100%	4.00E-06	100.00%	Homo sapiens(taxid:9606)
<b>ALGSHHTASPWNLSPFSK</b>	100%	4.00E-13	100.00%	Homo sapiens(taxid:9606)

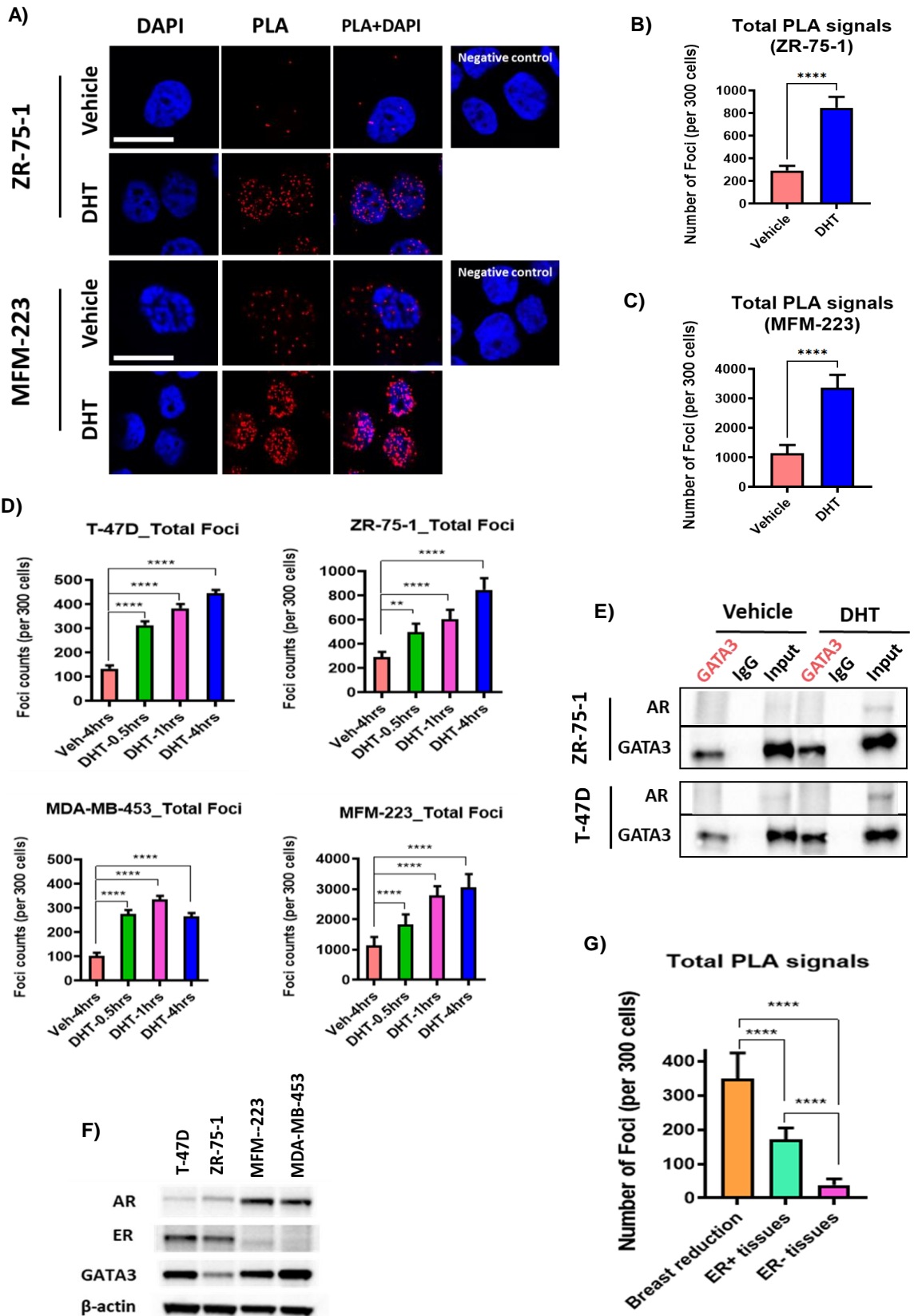
C)

GATA3 human protein sequence				
10	20	30	40	50
MEVTADQPRW	VSHHHPAVLN	GQHPDTHHPG	LSHSYMDAAQ	YPLPEEVDVL
60	70	80	90	100
FNIDGQGNHV	PPYYGNSVRA	TVQRYPPTHH	GSQVCRPPLL	HGSLPWLDGG
110	120	130	140	150
<b>KALGSHHTAS</b>	<b>PWNLSPFSKT</b>	SIHHGSPGPL	SVYPPASSSS	LSGGHASPFL
160	170	180	190	200
FTFPPTPPKD	<b>VSPDPSLSTP</b>	<b>GSAGSAR</b> QDE	KECLKYQVPL	PDSMKLESSH
210	220	230	240	250
SRGSMTALGG	ASSSTHHPIT	TYPYVPEYS	SGLFPPSSLL	GGSPTFGCK
260	270	280	290	300
SRPKARSSTG	RECVNCGATS	TPLWRRDGTG	HLYCNACGLY	HKMNGQNRPL
310	320	330	340	350
IKPKRRLSAA	RRAGTSCANC	QTTTTLWRR	NANGDPVCNA	CGLYYKLHNI
360	370	380	390	400
NRPLTMKKEG	IQTRNRKMSS	KSKKCKVHD	<b>SLEDFPKNSS</b>	<b>FNPAALSRHM</b>
410	420	430	440	
SSLSHISPF	HSSHMLTPT	PMHPPSSLSF	GPHHPSSMVT	AMG

**Supplementary-1: GATA3 is a novel AR interacting protein in breast cancer cells. (A)** AR was identified in 3/3 independent biological replicates in the AR RIME results of all the *in vitro* breast cancer models, highlighting the specificity of the RIME experiments. **(B)** Four unique GATA3 peptides were consistently identified in AR RIME data across all the breast cancer cell lines. Each unique sequence was blasted for the whole protein dataset of UniProtKB/Swiss (swissprot) database. Only peptides with 90-100 % identity, 90-100 % coverage and the E-value of 0-1e-2 were included as a confirmed unique peptide for the *homo sapiens* GATA3 protein. **(C)** The FASTA format of amino-acid sequence of GATA3 protein is shown according to the UniProtKB/Swiss(swissprot) database. Image shows the location of each of the 4 unique peptides listed in **B** along the GATA3 protein sequence, highlighted by corresponding colours.



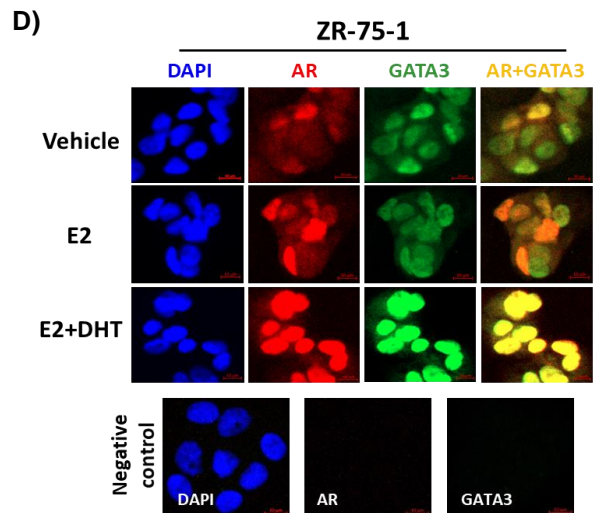
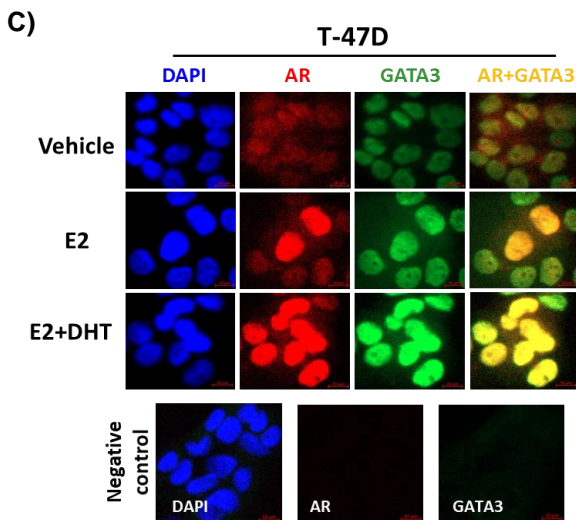
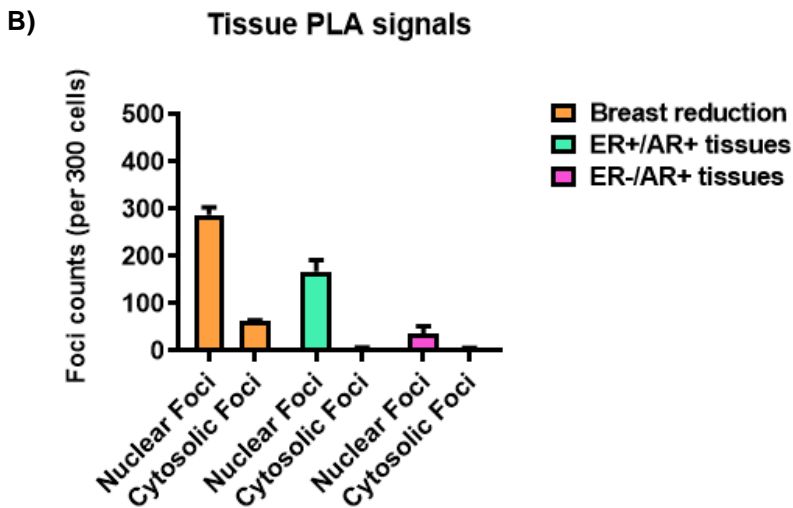
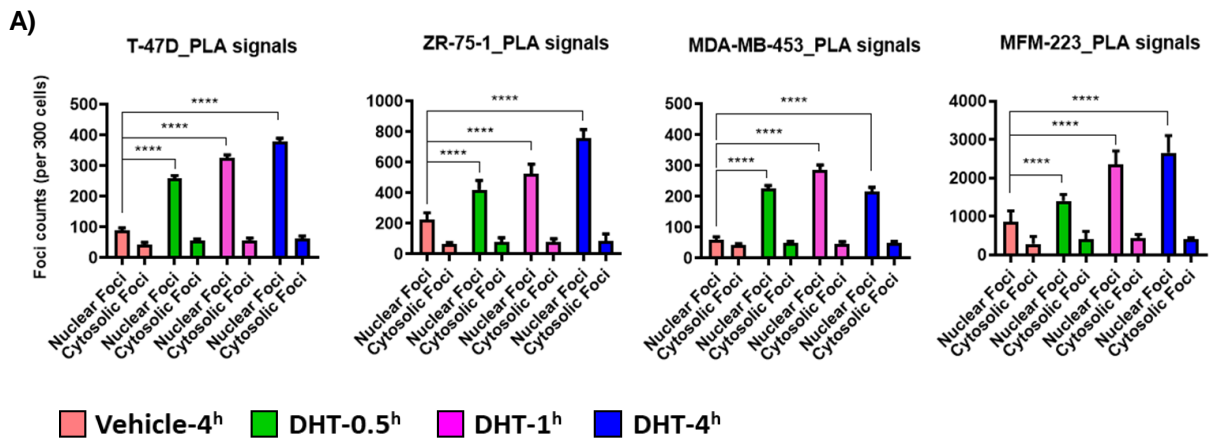
Supplementary 2: Validation of the AR and GATA3 interactions in breast cancer



**Supplementary-2: Validation of the AR and GATA3 interactions in breast cancer. (A)**

Representative confocal images showing AR-GATA3 interactions in the presence or absence of DHT treatment in ZR-75-1 and MFM-223 breast cancer cells. The negative PLA control represents an assay performed without the secondary antibody. Bar graphs showing the total number of foci per 300 counted nuclei in ZR-75-1 **(B)** and MFM-223 **(C)** DHT-treated cells compared to the control cells. **(D)** Bar graphs representing AR-GATA3 interactions in DHT treated cells in different time-points compared to the unstimulated cells. Two-way ANOVA with Student's t-test was used to determine statistically significant differences in **B**, **C** and **D**. Data shown as mean  $\pm$  SEM of 5 replicates and are representative of two independent experiment; \*\*  $p=0.001$ , and \*\*\*\*  $p<0.0001$ . **(E)** Western blot showing GATA3 immunoprecipitation in ER+ breast cancer cells, without AR detection in association with GATA3 protein. IgG served as a negative control in both assays. **(F)** Western blotting displaying the protein expression levels of AR, ER, and GATA3 in ER+ and ER- breast cancer cell lines. **(G)** Bar chart showing the quantification of total foci in malignant (ER+  $n=2$  and ER-  $n=2$ ) and non-malignant ( $n=2$ ) breast cancer tissues. Two-way ANOVA with Tukey's multiple comparisons test was used to determine statistically significant differences in **G**. Data shown as mean  $\pm$  SEM of 3 replicates and are representative of one independent experiment; \*\*\*\*  $p<0.0001$ .

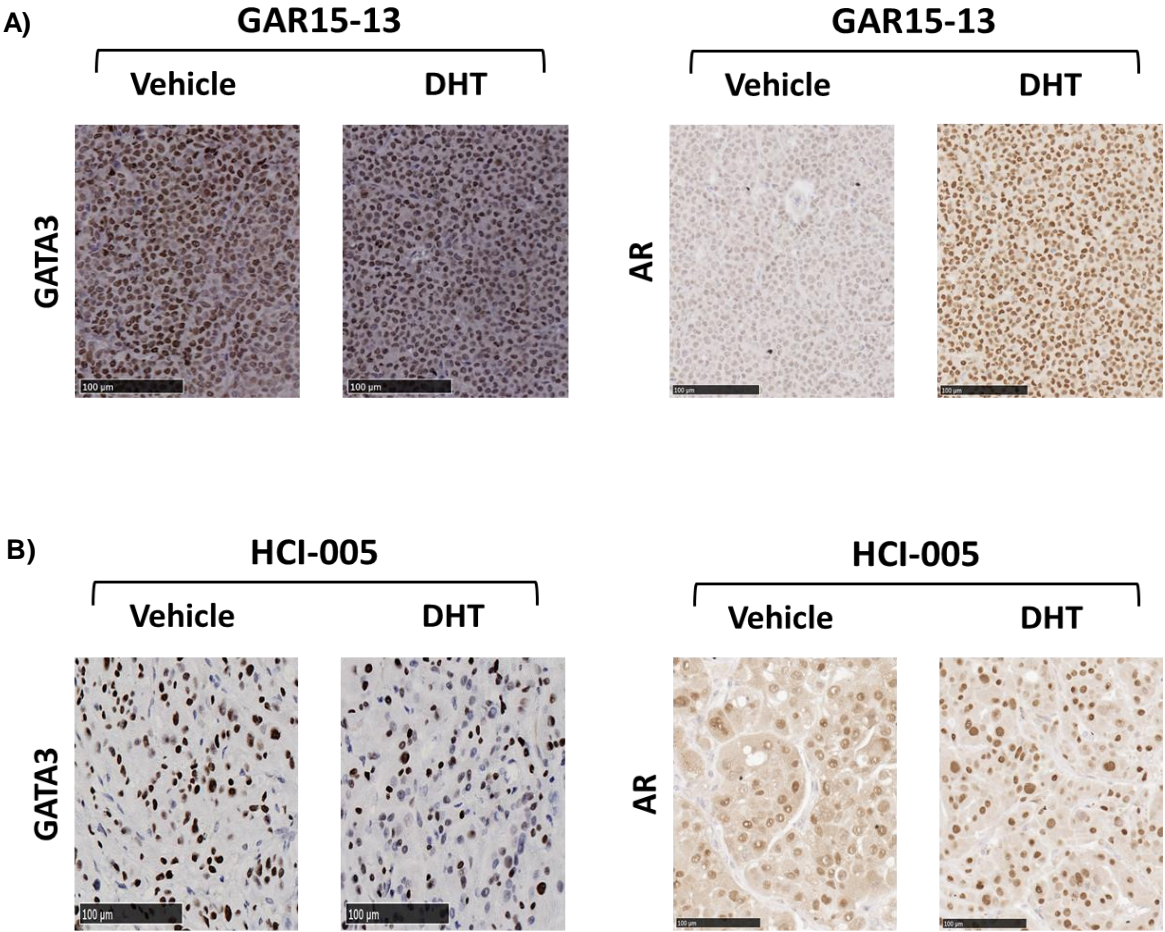
**Supplementary 3: AR and GATA3 interactions occur in both nucleus and cytoplasm of breast cancer cells, and estrogen treatment caused moderate GATA3 nuclear translocation**



**Supplementary-3: AR and GATA3 interactions occur in both nucleus and cytoplasm of the breast cells, and estrogen treatment caused moderate GATA3 nuclear translocation.**

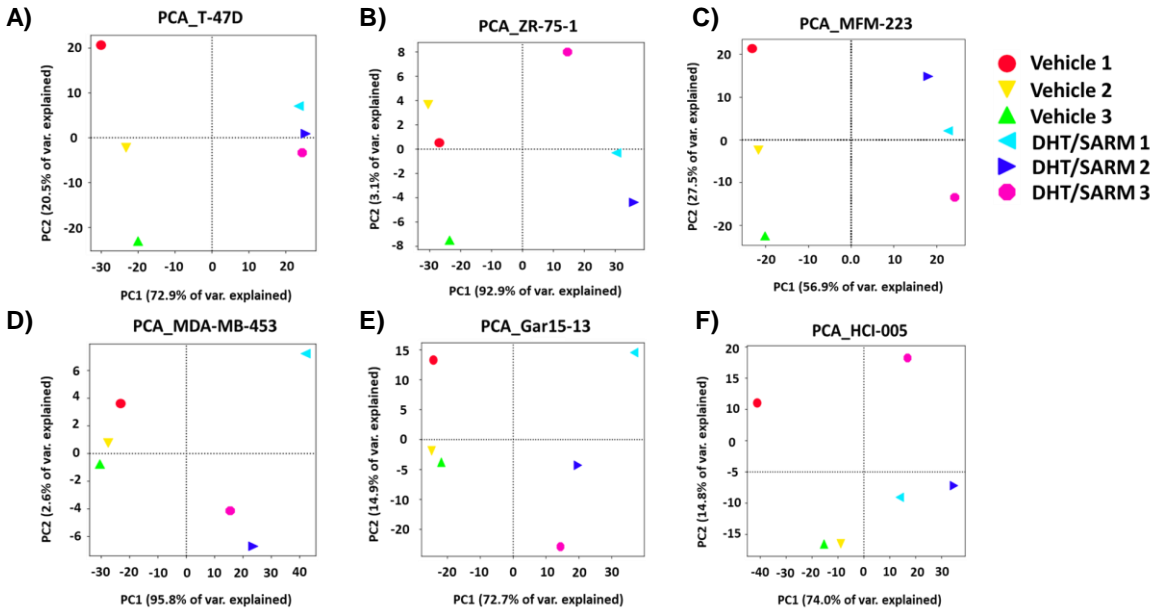
**(A)** Bar charts showing the quantification of nuclear and cytosolic interactions of AR and GATA3 in DHT-treated breast cancer cells compared with vehicle treated cells across 4 breast cancer cell lines. Two-way ANOVA with Student's t-test was conducted to determine statistically significant differences in **A**. Data shown as mean  $\pm$  SEM of 5 replicates and are representative of two independent experiment; \*\*\*\*  $p < 0.0001$ . **(B)** Nuclear and cytosolic foci are detected in both normal and malignant ER+ and ER- breast tissues through the PLA technique. Representative images of dual IF experiment showing a moderate effect of E2 on AR and GATA3 nuclear translocation in **(C)** T-47D and **(D)** ZR-75-1 breast cancer cells. Negative controls shown in this figure are cells without the secondary antibody staining.

Supplementary 4: Validation of AR and GATA3 expression in GAR15-13 and HCI-005 PDX models



**Supplementary-4: Validation of AR and GATA3 expression in GAR15-13 and HCI-005 PDX models. (A,B)** Immunohistochemistry images representing the positivity of AR and GATA3 expression in GAR15-13 and **(C,D)** HCI-005 PDXs in vehicle and AR agonist-treated tissues. Images are representative of n=3 independent tissues of AR agonist-treated and n=3 vehicle-treated tumours from each PDX tissue. Scale bars, 100  $\mu$ m.

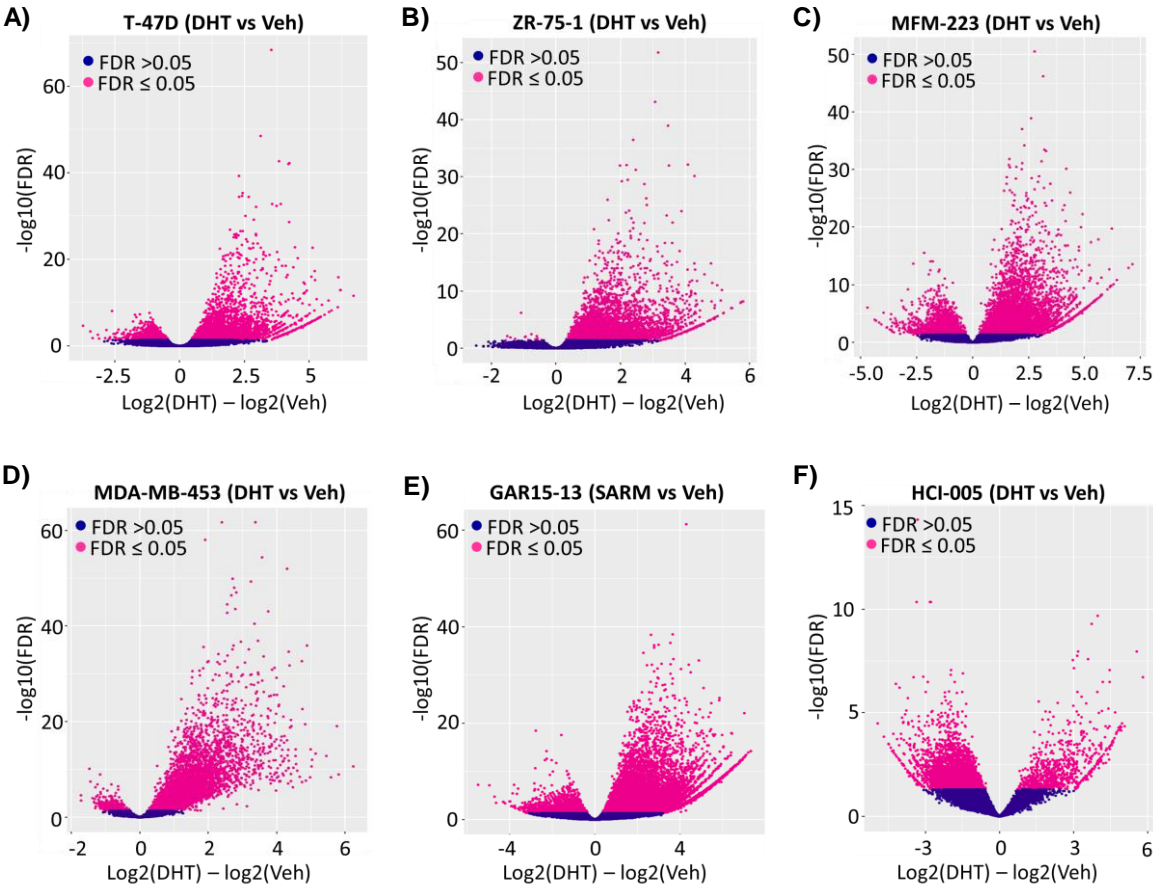
**Supplementary 5: AR activation is the major source of variation in GATA3 ChIP-seq datasets**



**Supplementary-5: AR activation is the major source of variation in GATA3 ChIP-seq datasets. (A-F)** Principle component analysis (PCA) clustering the AR agonist-treated cells away from the vehicle-treated cells across all the breast cancer models, characterizing the trends exhibited by the AR agonists compared with the vehicle treated cells in each model. Red, yellow, and green colours representing three replicates of vehicle-treated samples, however, light and dark blue and purple representing 3 replicates of DHT/SARM-treated samples in each model.

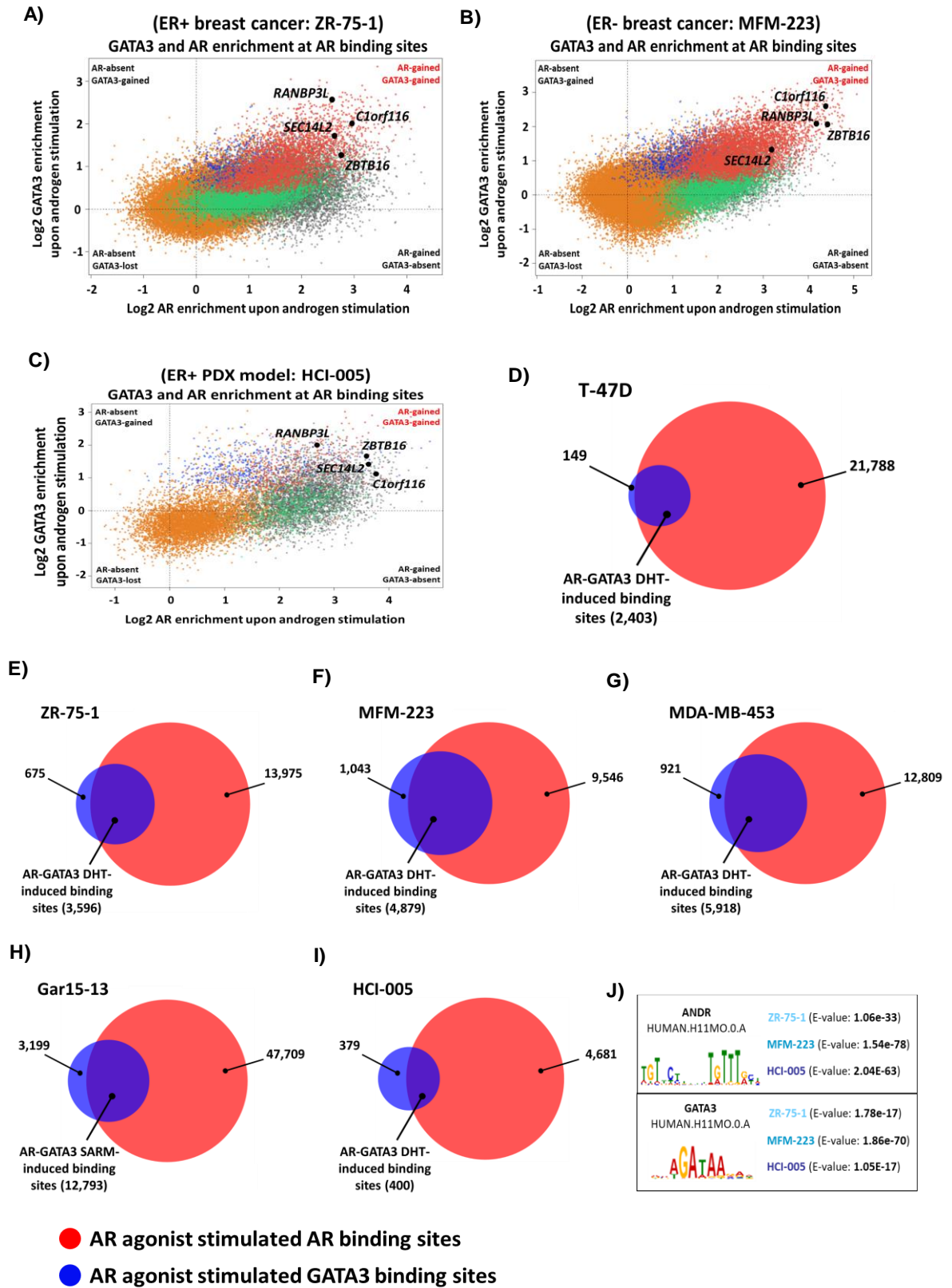


**Supplementary 6: AR agonism caused significant enrichment of new GATA3 binding sites across all the *in vitro* and *in vivo* breast cancer models**



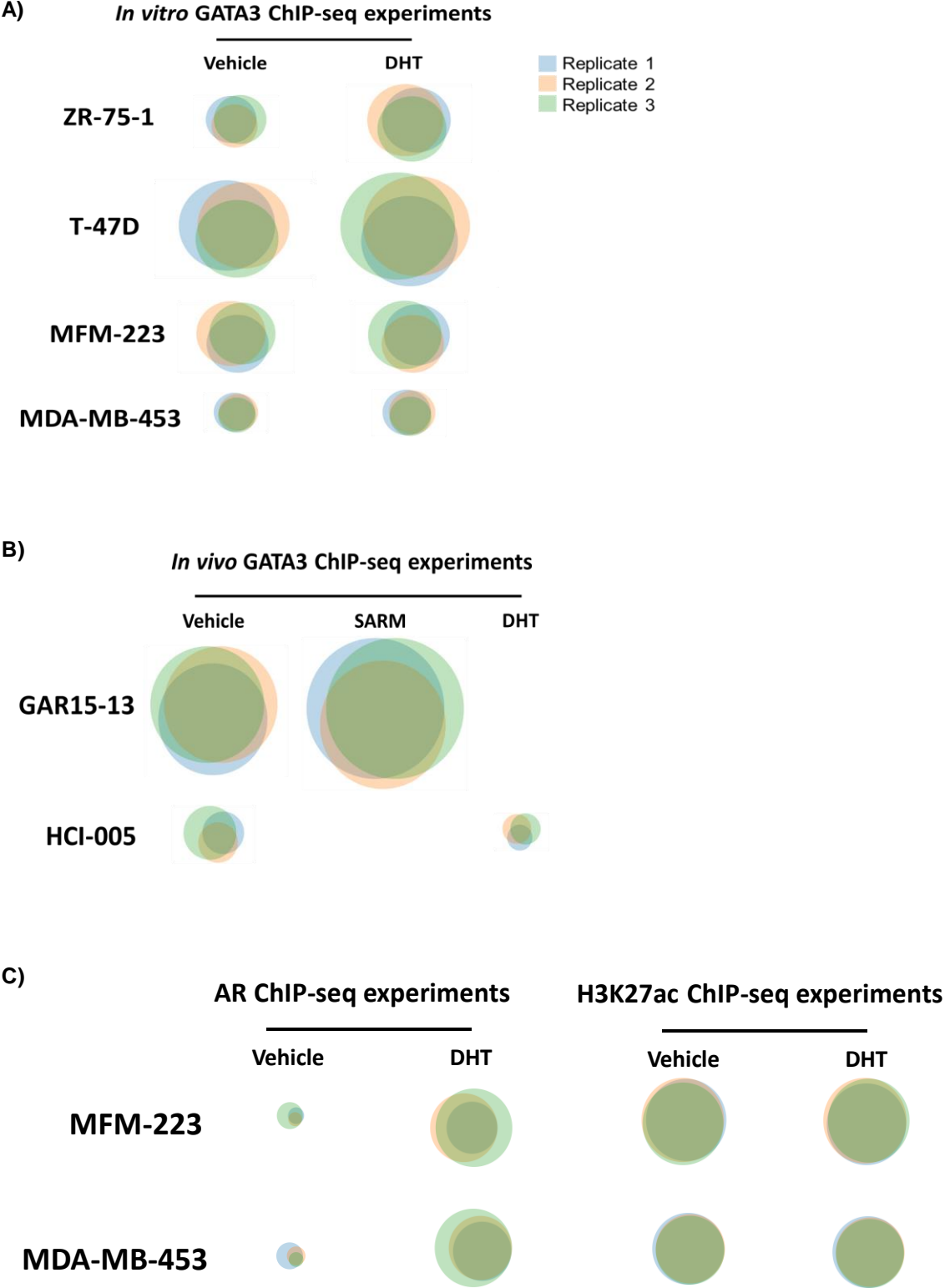
**Supplementary-6: AR agonism caused significant enrichment of new GATA3 binding sites across all the *in vitro* and *in vivo* breast cancer models.** Volcano plots (A-F) showing the differentially enriched and depleted GATA3 binding sites ( $FDR \leq 0.05$ ) upon AR activation across all the *in vitro* and *in vivo* breast cancer models. The purple dots represent peaks that did not undergo a significant change ( $FDR \geq 0.05$ ) with AR activation. The pink dots represent significant ( $FDR \leq 0.05$ ) changes to GATA3 peaks in each model.

## Supplementary 7: Distribution of AR and GATA3 ChIP-seq data upon AR activation in breast cancer cells



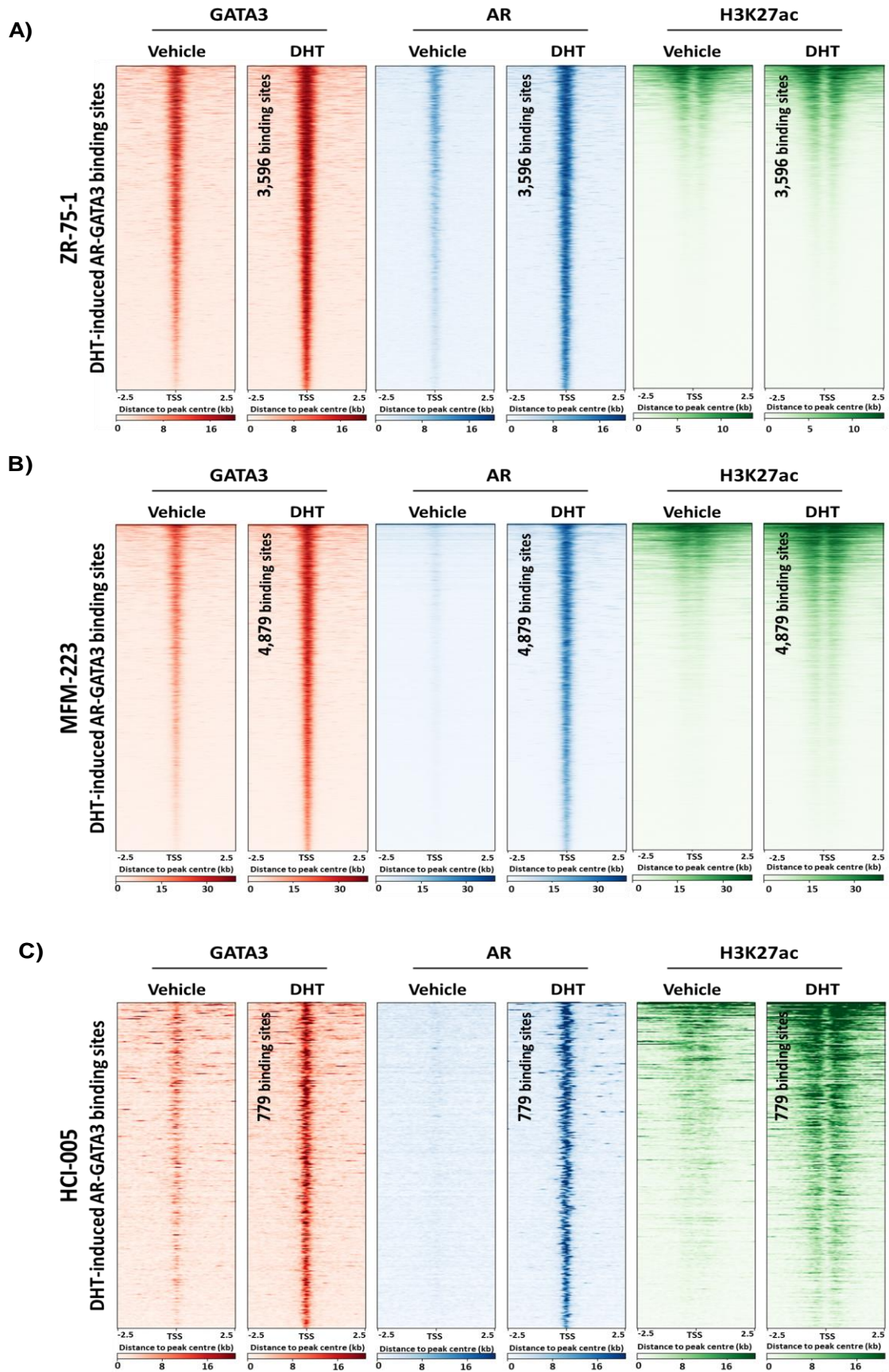
**Supplementary-7: Distribution of AR and GATA3 ChIP-seq data upon AR activation in breast cancer cells.** Two-factor MA plots representing distribution of the AR and GATA3 cistromes upon AR activation in **(A)** ZR-75-1 and **(B)** MFM-223 *in vitro* models and the **(C)** HCI-005 *in vivo* model. Peaks are represented by coloured dots and classified into 5 different sub-groups: Orange and grey dots represent GATA3 and AR binding sites that do not overlap with the other factor, respectively; blue dots represent AR agonist-induced GATA3 peaks that are not shared with AR; green dots represent AR-GATA3 overlapping peaks that GATA3 peaks are not gained with AR agonist; and red dots represent AR-GATA3 peaks that are significantly gained upon AR activation. Example loci associated with AR target genes (*ZBTB16*, *C1orf116*, *SE3C14L2*, *RANBP3L*) are highlighted in black dots and selected from the AR-GATA3 peaks that are induced by AR agonist in each model. Venn diagrams represent the overlap of AR agonist-induced binding sites of AR with the differentially enriched GATA3 binding events upon AR activation across the *in vitro* **(D-G)** and *in vivo* **(H and I)** breast cancer models. **(J)** Motif analysis of AR agonist-induced GATA3 binding sites representing AR as the most highly enriched motif across all models.

Supplementary 8: Replicate concordance for GATA3, AR, and H3K27ac CHIP-seq experiments



**Supplementary-8: Replicate concordance for GATA3, AR, and H3K27ac ChIP-seq experiments.** Venn diagrams represent the concordance of the 3 biological replicates of each vehicle and DHT treated GATA3 ChIP-seq experiment in two ER+ (ZR-75-1 and T-47D) and two ER- (MFM-223 and MDA-MB-453) breast cancer *in vitro* cells **(A)**, as well as 2 ER+/AR+/GATA3+ GAR15-13, and HCI-005 *in vivo* PDX models **(B)**. All the *in vitro* models were treated (DHT vs Vehicle) for 4 hours before cross-linking and harvest. The PDX models were treated for 5 days before getting processed for ChIP-seq experiments. **(C)** Venn diagrams show the concordance of 3 biological replicates of vehicle and DHT-treated cells (4 hours) for AR and H3K27ac ChIP-seq experiments in two ER- (MFM-223 and MDA-MB-453) breast cancer cell line models. Three replicates of each ChIP-seq experiment are shown in 3 different colours.

Supplementary 9: H3K27ac enrichment at AR-GATA3 binding events upon AR activation



**Supplementary-9: H3K27ac enrichment at AR-GATA3 binding events upon AR activation.**

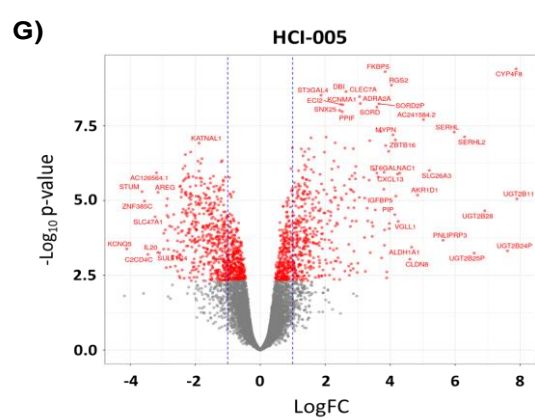
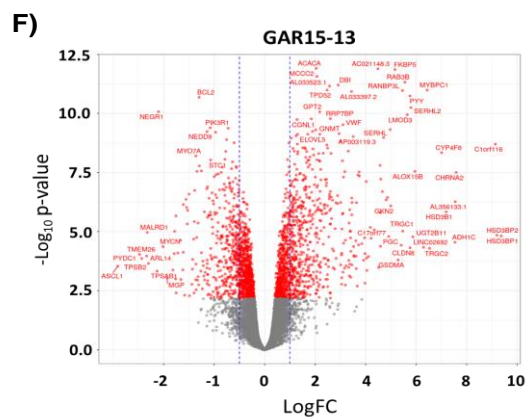
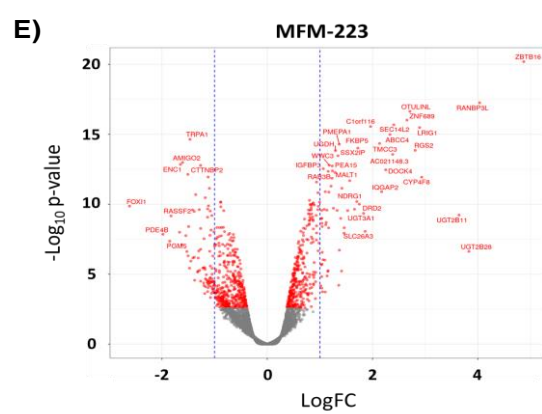
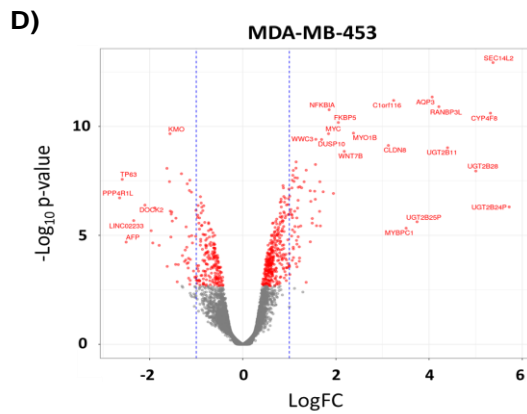
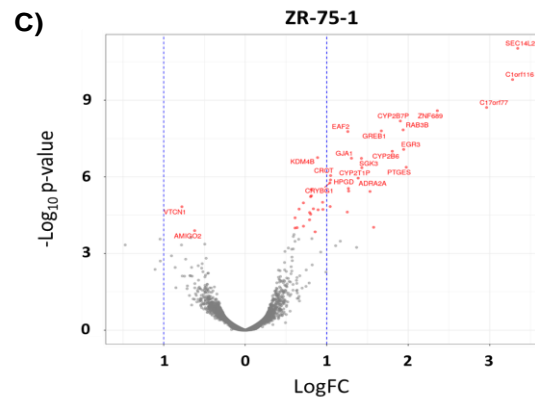
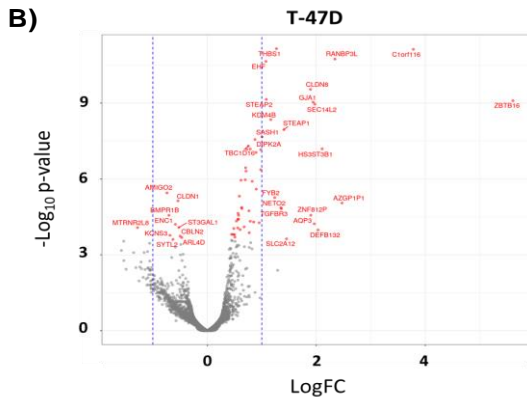
Heatmaps depict increased H3K27ac ChIP-seq signals at the AR agonist-induced AR-GATA3 cis-regulatory elements in **(A)** ZR-75-1 (ER+ breast cancer cell line), **(B)** MFM-223 (ER- breast cancer cell line), and the **(C)** HCI-005 ER+/AR+/GATA+ PDX model. Indicative of the transcriptionally active state of the chromatin at those specific DHT-induced AR-GATA3 loci in each of the models.



**Supplementary 10: Differentially up- and down-regulated genes associated with AR agonist-induced AR-GATA3 peaks in breast cancer models**

**A)**

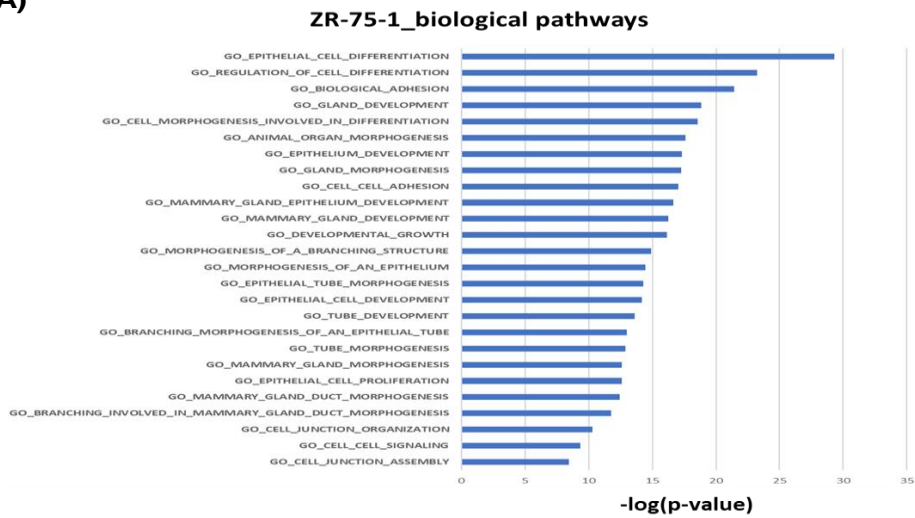
	AR agonist-induced AR-GATA3 peaks	Annotated genes	DE Up-regulated genes	DE DN-regulated genes
ZR-75-1	3,596	4,038	1,217	932
MFM-223	4,879	5,418	1,645	1,409
HCI-005	400	1,219	174	63



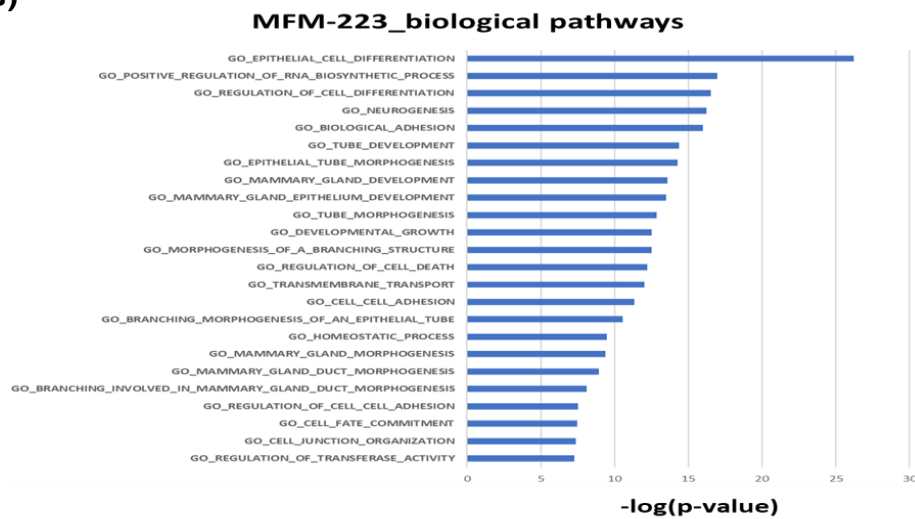
**Supplementary-10: Differentially up- and down-regulated genes associated with AR agonist-induced AR-GATA3 peaks in breast cancer models. (A)** Table showing the number of AR agonist-induced AR-GATA3 binding sites and their associated genes within a 100 kb window from the TSS that are differentially up- or down-regulated in ZR-75-1, MFM-223 cells and HCI-005 PDX breast cancer models. **(B-G)** Volcano plots showing genes that are up- or down-regulated in response to an AR agonist in each of *in vitro* and *in vivo* breast cancer models. Red dots represent genes that are significantly ( $FDR \leq 0.05$ ) up- or down-regulated after AR activation, and grey dots represent genes not significantly changed upon AR activation in each model.

**Supplementary 11: Biological pathways associated with the differentially expressed AR-GATA3 gene targets in breast cancer in vitro and in vivo models**

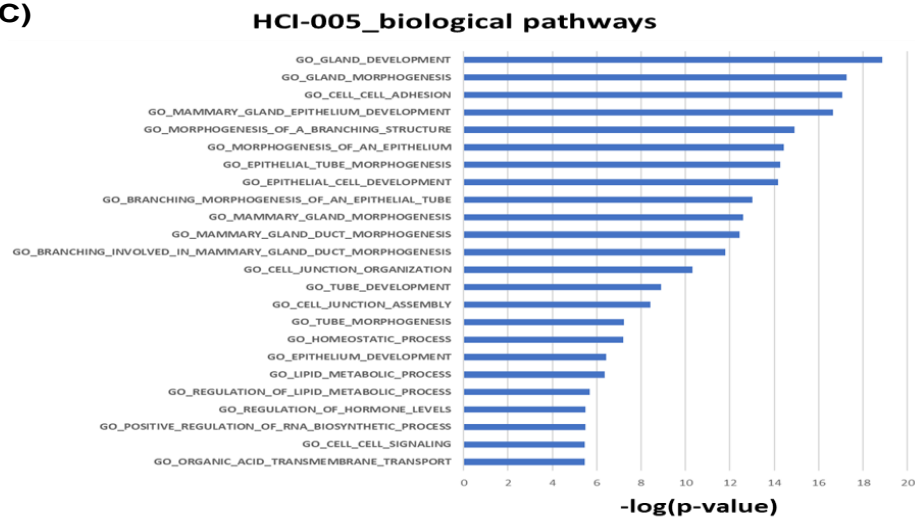
**A)**



**B)**



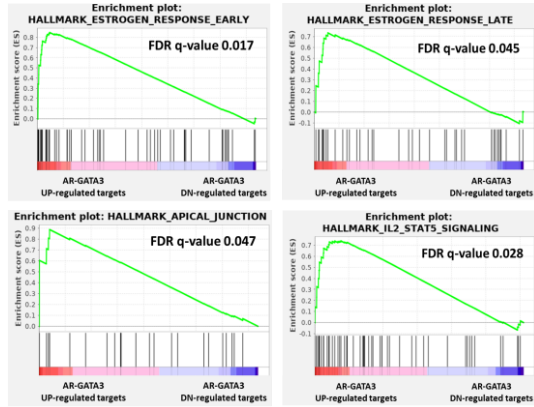
**C)**



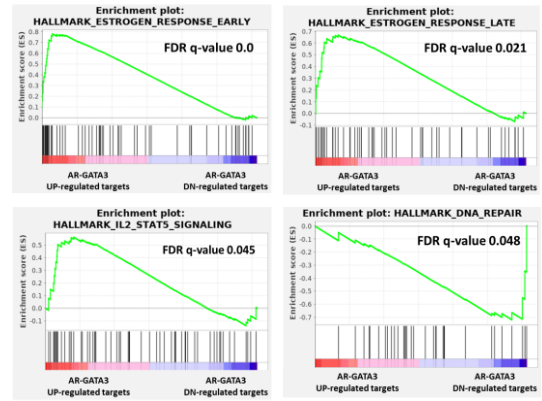
**Supplementary-11: Biological pathways associated with differentially expressed AR-GATA3 gene targets in breast cancer *in vitro* and *in vivo* models.** Bar graph showing GO (gene ontology) pathway analysis showing significantly enriched biological pathways associated with AR-GATA3 differentially expressed gene targets in **(A)** ZR-75-1, **(B)** MFM-223, and **(C)** HCI-005 breast cancer models.

## Supplementary 12: Gene set enrichment analysis (GSEA) of AR-GATA3 gene targets in breast cancer

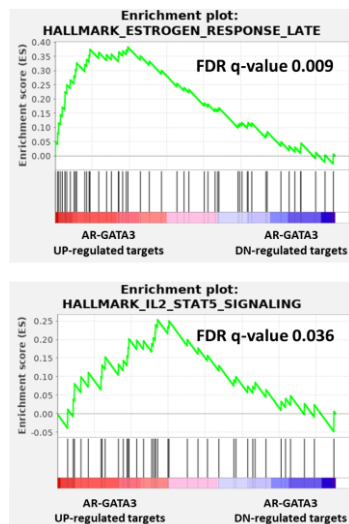
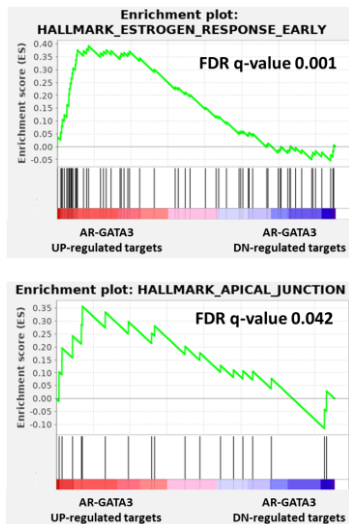
A) Up-regulated pathways by DHT in T-47D cells



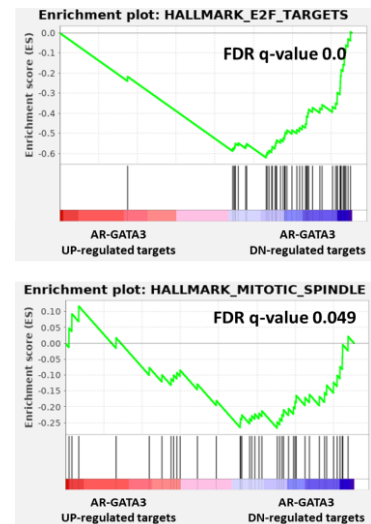
B) Up and down-regulated pathways by DHT in MDA-MB-453 cells



C) Up-regulated pathways by Enobosarm in GAR15-13 PDX

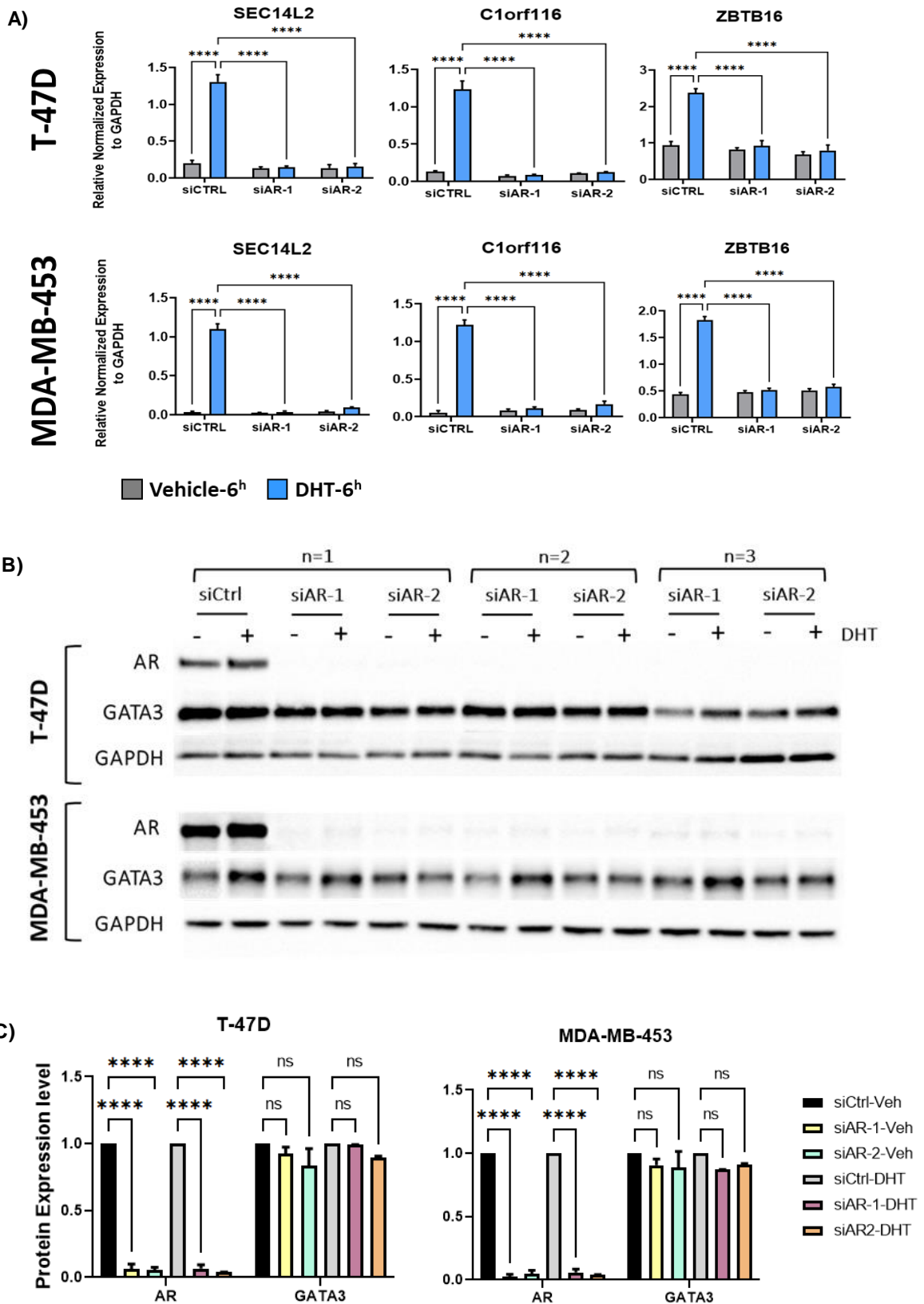


Down-regulated pathways by Enobosarm in GAR15-13 PDX



**Supplementary-12: Gene set enrichment analysis (GSEA) of AR-GATA3 targets in breast cancer.** GSEA plots representing the up-regulated and down-regulated hallmarks associated with the AR-GATA3 up- and down-regulated targets in **(A)** T-47D, **(B)** MDA-MB-453, and **(C)** GAR15-13 breast cancer models.

**Supplementary 13: AR knock-down significantly suppressed the expression of AR target genes but did not affect the GATA3 protein expression level in breast cancer cells**

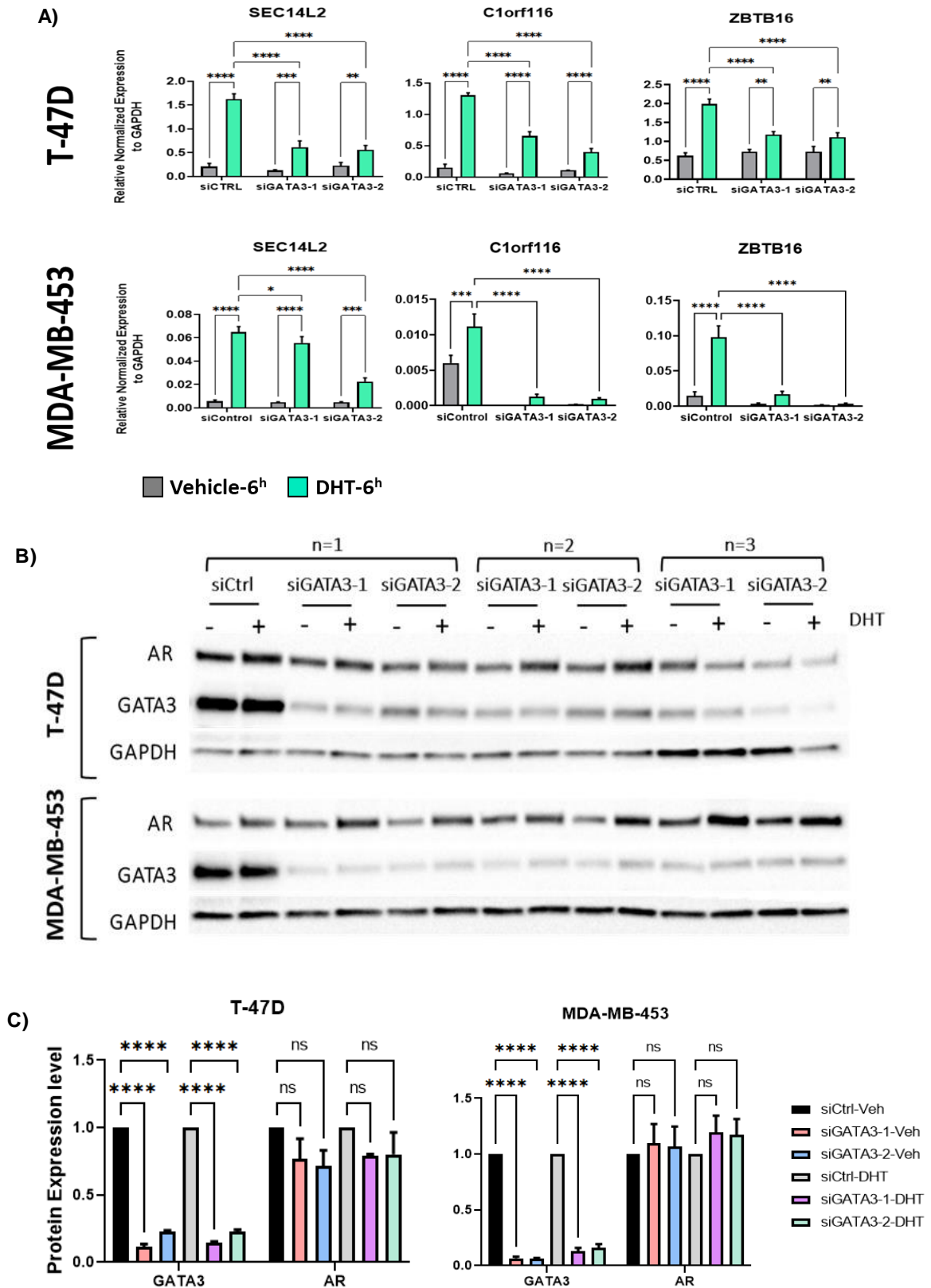


**Supplementary-13: AR knock-down significantly ablated DHT-induced expression of AR target genes but did not affect level of GATA3 protein expression in breast cancer cells.**

**(A)** Bar graphs show the effect of AR silencing on mRNA expression levels of *SEC14L2*, *C1orf116*, and *ZBTB16* genes in DHT-treated cells (6 hours) compared with vehicle-treated T-47D and MDA-MB-453 breast cancer cell line models. Two-way ANOVA with Tukey's multiple comparisons test was conducted to determine statistically significant differences in **A**. Data shown as mean  $\pm$  SD of 3 replicates and are representative of three independent experiments; \*\*\*\*  $p < 0.0001$ . **(B)** Immunoblotting images show the efficacy of AR knock-down in three independent biological replicate experiments in which cells were transfected with one of two different siRNAs specific for AR compared with the non-targeting siControl in T-47D (top) and MDA-MB-453 (bottom) breast cancer cell line models. After 2 days of siRNA treatment, cells were treated with DHT or vehicle for 4 hours prior to harvest. Both blots were stained for GAPDH as a control for protein loading. **(C)** Bar graphs showing quantification of AR and GATA3 specific bands in Western blots shown in (B). Each band is normalized to its respective GAPDH band as well as its corresponding treatment control. Two-way ANOVA with Student's t-test was conducted to determine statistically significant differences in **C**. Data shown as mean  $\pm$  SD of 2 replicates (3<sup>rd</sup> replicate is removed from the stats of both blots due to the unusual variation compared with the first two replicates) and are representative of one independent experiment; \*\*\*\*  $p < 0.0001$ .

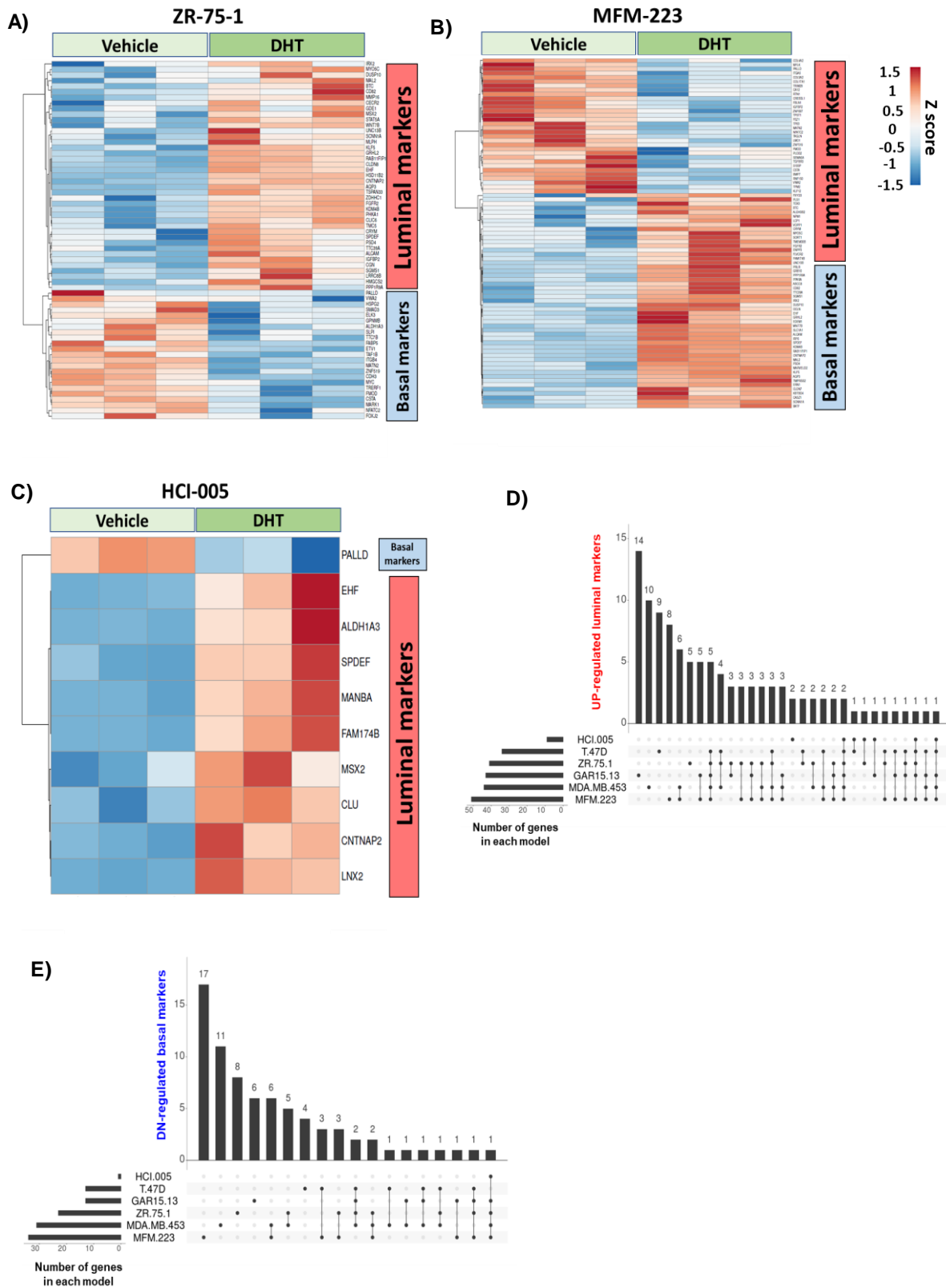


**Supplementary 14: GATA3 knock-down significantly reduced the expression of AR target genes but did not affect AR protein expression level in breast cancer cells**



**Supplementary-14: GATA3 knock-down significantly reduced expression of AR target genes but did not affect AR protein expression level in breast cancer cells. (A)** Bar graphs show the effect of GATA3 silencing on the mRNA expression level of *SEC14L2*, *C1orf116*, and *ZBTB16* AR target genes in DHT-treated cells (6 hours) compared with vehicle-treated cells of T-47D and MDA-MB-453 breast cancer models. Two-way ANOVA with Tukey's multiple comparisons test was conducted to determine statistically significant differences in **A**. Data shown as mean  $\pm$  SD of 3 replicates and are representative of three independent experiments; \*  $p=0.001$ , \*\*  $p<0.01$ , \*\*\*  $p<0.001$ , and \*\*\*\*  $p<0.0001$ . **(B)** Immunoblotting images show the efficacy of GATA3 knock-down in three independent biological replicate experiments in which cells were transfected with one of two different siRNAs against GATA3 compared with the siControl transfected cells in T-47D and MDA-MB-453 breast cancer models. Cells were transfected for 2 days then treated with DHT vs Vehicle for 4 hours prior to harvest. Both blots were stained for GAPDH as a control for protein loading. **(C)** Bar graphs represent quantification of western blot results shown in (B). Each band of interest is normalized to its corresponding GAPDH band and also the relative treatment control. Two-way ANOVA with Student's t-test was conducted to determine statistically significant differences in **C**. Data shown as mean  $\pm$  SD of 2 replicates (3<sup>rd</sup> replicate is removed from the stats of both blots due to the un-usual variation compared with the first two replicates) and are representative of one independent experiment; \*\*\*\*  $p<0.0001$ .

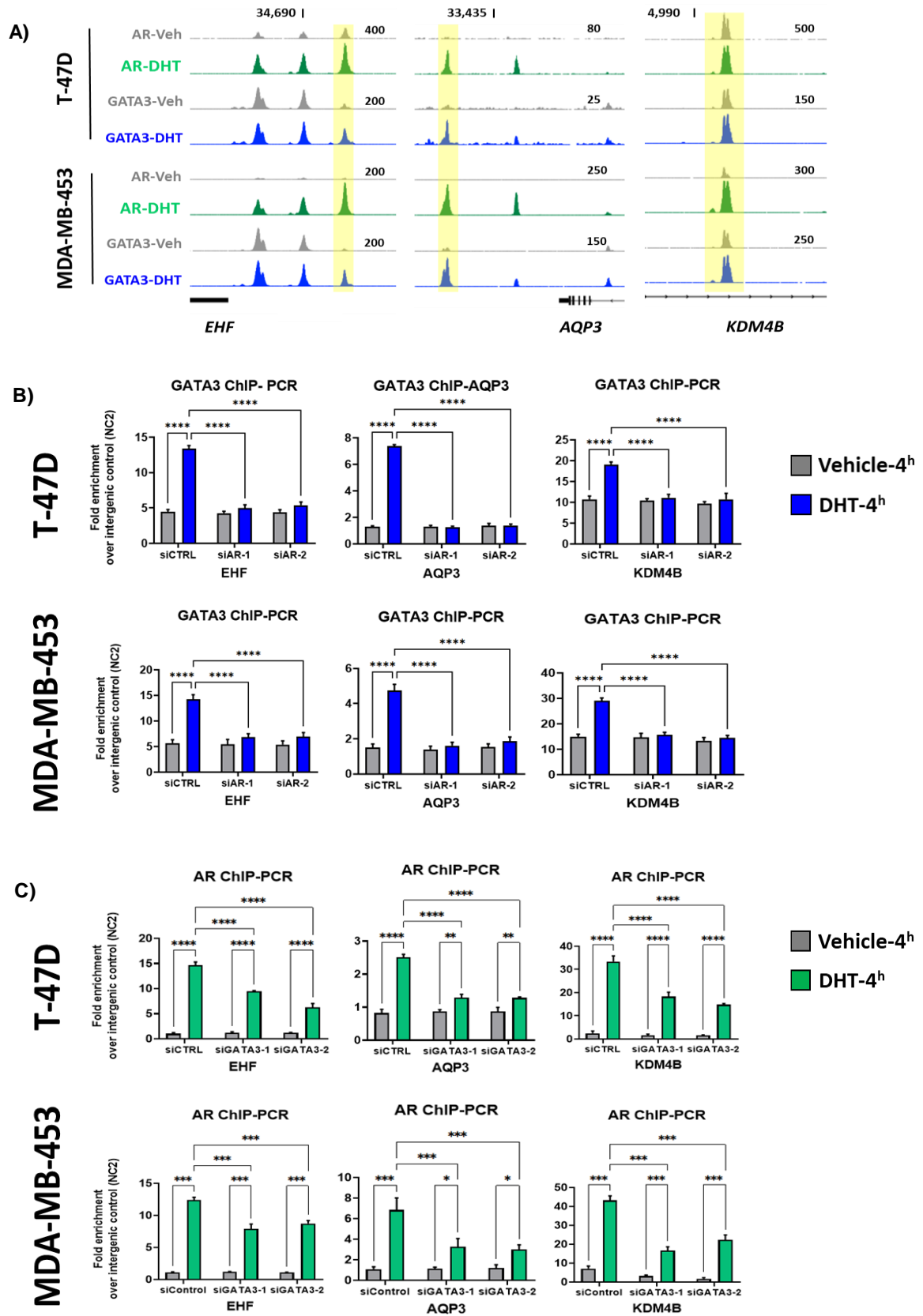
Supplementary 15: AR agonist-induced AR-GATA3 loci are associated with lineage-restricted genes in breast cancer *in vitro* and *in vivo* models



**Supplementary-15: AR agonist-induced AR-GATA3 co-occupied genomic loci are associated with lineage-restricted genes in breast cancer *in vitro* and *in vivo* models.**

Heatmaps representing up- and down-regulated luminal and basal AR-GATA3 gene targets upon AR activation in **(A)** ZR-75-1, **(B)** MFM-223, and **(C)** HCl-005 *in vitro* and *in vivo* breast cancer models, respectively. Up-set plots showing the total number of **(D)** up-regulated luminal genes and **(E)** down-regulated basal genes in each of *in vitro* and *in vivo* models, highlighting the common and unique genes in each group.

Supplementary 16: GATA3 and AR binding at loci associated with regulation of luminal epithelial genes in breast cancer cells



**Supplementary-16: GATA3 and AR binding at loci associated with regulation of luminal epithelial genes in breast cancer cells. (A)** Genome browser images showing averaged AR and GATA3 ChIP-seq signals at binding sites associated with luminal epithelial genes (*EHF*, *AQP3*, and *KDM4B*). Peaks represent an average signal from three independent ChIP-seq replicates of T-47D (top) and MDA-MB-453 (bottom) cells. **(B)** Bar charts showing GATA3 ChIP-PCR results following AR knock-down at loci (represented in **A**) associated with regulation of luminal markers *EHF*, *AQP3*, and *KDM4B* in T-47D and MDA-MB-453 breast cancer cells. **(C)** Bar graphs representing the AR ChIP-PCR results following GATA3 knock-down at loci (represented in **A**) associated with regulation of luminal markers *EHF*, *AQP3*, and *KDM4B* in T-47D and MDA-MB-453 breast cancer cells. Cells in **B** and **C** were transfected for 2 days before being treated with DHT or vehicle for 4 hours before harvesting. Two-way ANOVA with Tukey's multiple comparisons test was used to determine statistically significant differences in both **B** and **C**. Data shown as mean  $\pm$  SD of 3 replicates and are representative of two independent experiments; \*  $p=0.001$ , \*\*  $p<0.01$ , \*\*\*  $p<0.001$ , and \*\*\*\*  $p<0.0001$ .

## Extended data

### Extended data 1: AR RIME interactome in different breast cancer cell lines

AR interacting proteins identified by RIME in 3/3 replicates excluding IgG:	
<b>ZR-75.1</b>	
<b>Accession</b>	<b>Description</b>
P50548	ETS domain-containing transcription factor ERF OS=Homo sapiens GN=ERF PE=1 SV=2 - [ERF_HUMAN]
P13807	Glycogen [starch] synthase, muscle OS=Homo sapiens GN=GYS1 PE=1 SV=2 - [GYS1_HUMAN]
P63241	Eukaryotic translation initiation factor 5A-1 OS=Homo sapiens GN=EIF5A PE=1 SV=2 - [IF5A1_HUMAN]
Q02790	Peptidyl-prolyl cis-trans isomerase FKBP4 OS=Homo sapiens GN=FKBP4 PE=1 SV=3 - [FKBP4_HUMAN]
P17275	Transcription factor jun-B OS=Homo sapiens GN=JUNB PE=1 SV=1 - [JUNB_HUMAN]
P10275	Androgen receptor OS=Homo sapiens GN=AR PE=1 SV=3 - [ANDR_HUMAN]
P08779	Keratin, type I cytoskeletal 16 OS=Homo sapiens GN=KRT16 PE=1 SV=4 - [K1C16_HUMAN]
P18754	Regulator of chromosome condensation OS=Homo sapiens GN=RCC1 PE=1 SV=1 - [RCC1_HUMAN]
P04075	Fructose-bisphosphate aldolase A OS=Homo sapiens GN=ALDOA PE=1 SV=2 - [ALDOA_HUMAN]
Q8WWM7	Ataxin-2-like protein OS=Homo sapiens GN=ATXN2L PE=1 SV=2 - [ATX2L_HUMAN]
P11940	Polyadenylate-binding protein 1 OS=Homo sapiens GN=PABPC1 PE=1 SV=2 - [PABP1_HUMAN]
P02538	Keratin, type II cytoskeletal 6A OS=Homo sapiens GN=KRT6A PE=1 SV=3 - [K2C6A_HUMAN]
P60866	40S ribosomal protein S20 OS=Homo sapiens GN=RPS20 PE=1 SV=1 - [RS20_HUMAN]
P98179	RNA-binding protein 3 OS=Homo sapiens GN=RBM3 PE=1 SV=1 - [RBM3_HUMAN]
P84103	Serine/arginine-rich splicing factor 3 OS=Homo sapiens GN=SRSF3 PE=1 SV=1 - [SRSF3_HUMAN]
Q9Y5S9	RNA-binding protein 8A OS=Homo sapiens GN=RBM8A PE=1 SV=1 - [RBM8A_HUMAN]
P13010	X-ray repair cross-complementing protein 5 OS=Homo sapiens GN=XRCC5 PE=1 SV=3 - [XRCC5_HUMAN]
Q15366	Poly(rC)-binding protein 2 OS=Homo sapiens GN=PCBP2 PE=1 SV=1 - [PCBP2_HUMAN]
P60981	Dextrin OS=Homo sapiens GN=DSTN PE=1 SV=3 - [DEST_HUMAN]
P13639	Elongation factor 2 OS=Homo sapiens GN=EEF2 PE=1 SV=4 - [EF2_HUMAN]
P38919	Eukaryotic initiation factor 4A-III OS=Homo sapiens GN=EIF4A3 PE=1 SV=4 - [IF4A3_HUMAN]
P62306	Small nuclear ribonucleoprotein F OS=Homo sapiens GN=SNRPF PE=1 SV=1 - [RUXF_HUMAN]
O14745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1 OS=Homo sapiens GN=SLC9A3R1 PE=1 SV=4 - [NHRF1_HUMAN]

P23526	Adenosylhomocysteinase OS=Homo sapiens GN=AHCY PE=1 SV=4 - [SAHH_HUMAN]
P17931	Galectin-3 OS=Homo sapiens GN=LGALS3 PE=1 SV=5 - [LEG3_HUMAN]
Q04695	Keratin, type I cytoskeletal 17 OS=Homo sapiens GN=KRT17 PE=1 SV=2 - [K1C17_HUMAN]
P12956	X-ray repair cross-complementing protein 6 OS=Homo sapiens GN=XRCC6 PE=1 SV=2 - [XRCC6_HUMAN]
Q15717	ELAV-like protein 1 OS=Homo sapiens GN=ELAVL1 PE=1 SV=2 - [ELAV1_HUMAN]
Q12905	Interleukin enhancer-binding factor 2 OS=Homo sapiens GN=ILF2 PE=1 SV=2 - [ILF2_HUMAN]
P60842	Eukaryotic initiation factor 4A-I OS=Homo sapiens GN=EIF4A1 PE=1 SV=1 - [IF4A1_HUMAN]
O00330	Pyruvate dehydrogenase protein X component, mitochondrial OS=Homo sapiens GN=PDHX PE=1 SV=3 - [ODPX_HUMAN]
P62906	60S ribosomal protein L10a OS=Homo sapiens GN=RPL10A PE=1 SV=2 - [RL10A_HUMAN]
P62244	40S ribosomal protein S15a OS=Homo sapiens GN=RPS15A PE=1 SV=2 - [RS15A_HUMAN]
P10599	Thioredoxin OS=Homo sapiens GN=TXN PE=1 SV=3 - [THIO_HUMAN]
P63104	14-3-3 protein zeta/delta OS=Homo sapiens GN=YWHAZ PE=1 SV=1 - [1433Z_HUMAN]
Q15056	Eukaryotic translation initiation factor 4H OS=Homo sapiens GN=EIF4H PE=1 SV=5 - [IF4H_HUMAN]
Q99471	Prefoldin subunit 5 OS=Homo sapiens GN=PFDN5 PE=1 SV=2 - [PFD5_HUMAN]
P21291	Cysteine and glycine-rich protein 1 OS=Homo sapiens GN=CSRP1 PE=1 SV=3 - [CSRP1_HUMAN]
Q8N163	Cell cycle and apoptosis regulator protein 2 OS=Homo sapiens GN=CCAR2 PE=1 SV=2 - [CCAR2_HUMAN]
Q9BRA2	Thioredoxin domain-containing protein 17 OS=Homo sapiens GN=TXNDC17 PE=1 SV=1 - [TXD17_HUMAN]
O14497	AT-rich interactive domain-containing protein 1A OS=Homo sapiens GN=ARID1A PE=1 SV=3 - [ARI1A_HUMAN]
P06753	Tropomyosin alpha-3 chain OS=Homo sapiens GN=TPM3 PE=1 SV=2 - [TPM3_HUMAN]
Q92925	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 2 OS=Homo sapiens GN=SMARCD2 PE=1 SV=3 - [SMRD2_HUMAN]
Q92922	SWI/SNF complex subunit SMARCC1 OS=Homo sapiens GN=SMARCC1 PE=1 SV=3 - [SMRC1_HUMAN]
Q13435	Splicing factor 3B subunit 2 OS=Homo sapiens GN=SF3B2 PE=1 SV=2 - [SF3B2_HUMAN]
P52943	Cysteine-rich protein 2 OS=Homo sapiens GN=CRIP2 PE=1 SV=1 - [CRIP2_HUMAN]
Q8TAQ2	SWI/SNF complex subunit SMARCC2 OS=Homo sapiens GN=SMARCC2 PE=1 SV=1 - [SMRC2_HUMAN]
Q15427	Splicing factor 3B subunit 4 OS=Homo sapiens GN=SF3B4 PE=1 SV=1 - [SF3B4_HUMAN]
Q15365	Poly(rC)-binding protein 1 OS=Homo sapiens GN=PCBP1 PE=1 SV=2 - [PCBP1_HUMAN]



P31946	14-3-3 protein beta/alpha OS=Homo sapiens GN=YWHAB PE=1 SV=3 - [1433B_HUMAN]
P62316	Small nuclear ribonucleoprotein Sm D2 OS=Homo sapiens GN=SNRPD2 PE=1 SV=1 - [SMD2_HUMAN]
Q00341	Vigilin OS=Homo sapiens GN=HDLBP PE=1 SV=2 - [VIGLN_HUMAN]
Q14498	RNA-binding protein 39 OS=Homo sapiens GN=RBM39 PE=1 SV=2 - [RBM39_HUMAN]
P42766	60S ribosomal protein L35 OS=Homo sapiens GN=RPL35 PE=1 SV=2 - [RL35_HUMAN]
O43390	Heterogeneous nuclear ribonucleoprotein R OS=Homo sapiens GN=HNRNPR PE=1 SV=1 - [HNRPR_HUMAN]
Q9BVJ7	Dual specificity protein phosphatase 23 OS=Homo sapiens GN=DUSP23 PE=1 SV=1 - [DUS23_HUMAN]
O60336	Mitogen-activated protein kinase-binding protein 1 OS=Homo sapiens GN=MAPKBP1 PE=1 SV=4 - [MABP1_HUMAN]
Q96EP5	DAZ-associated protein 1 OS=Homo sapiens GN=DAZAP1 PE=1 SV=1 - [DAZP1_HUMAN]
P02747	Complement C1q subcomponent subunit C OS=Homo sapiens GN=C1QC PE=1 SV=3 - [C1QC_HUMAN]
P32969	60S ribosomal protein L9 OS=Homo sapiens GN=RPL9 PE=1 SV=1 - [RL9_HUMAN]
Q09666	Neuroblast differentiation-associated protein AHNAK OS=Homo sapiens GN=AHNAK PE=1 SV=2 - [AHNK_HUMAN]
P62899	60S ribosomal protein L31 OS=Homo sapiens GN=RPL31 PE=1 SV=1 - [RL31_HUMAN]
Q16630	Cleavage and polyadenylation specificity factor subunit 6 OS=Homo sapiens GN=CPSF6 PE=1 SV=2 - [CPSF6_HUMAN]
O00512	B-cell CLL/lymphoma 9 protein OS=Homo sapiens GN=BCL9 PE=1 SV=4 - [BCL9_HUMAN]
P49207	60S ribosomal protein L34 OS=Homo sapiens GN=RPL34 PE=1 SV=3 - [RL34_HUMAN]
P26583	High mobility group protein B2 OS=Homo sapiens GN=HMGB2 PE=1 SV=2 - [HMGB2_HUMAN]
Q3MHD2	Protein LSM12 homolog OS=Homo sapiens GN=LSM12 PE=1 SV=2 - [LSM12_HUMAN]
P98175	RNA-binding protein 10 OS=Homo sapiens GN=RBM10 PE=1 SV=3 - [RBM10_HUMAN]
P14625	Endoplasmin OS=Homo sapiens GN=HSP90B1 PE=1 SV=1 - [ENPL_HUMAN]
Q969G3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1 OS=Homo sapiens GN=SMARCE1 PE=1 SV=2 - [SMCE1_HUMAN]
P50914	60S ribosomal protein L14 OS=Homo sapiens GN=RPL14 PE=1 SV=4 - [RL14_HUMAN]
P46783	40S ribosomal protein S10 OS=Homo sapiens GN=RPS10 PE=1 SV=1 - [RS10_HUMAN]
Q15637	Splicing factor 1 OS=Homo sapiens GN=SF1 PE=1 SV=4 - [SF01_HUMAN]
Q8N684	Cleavage and polyadenylation specificity factor subunit 7 OS=Homo sapiens GN=CPSF7 PE=1 SV=1 - [CPSF7_HUMAN]
Q13247	Serine/arginine-rich splicing factor 6 OS=Homo sapiens GN=SRSF6 PE=1 SV=2 - [SRSF6_HUMAN]

Q9BTT0	Acidic leucine-rich nuclear phosphoprotein 32 family member E OS=Homo sapiens GN=ANP32E PE=1 SV=1 - [AN32E_HUMAN]
P78527	DNA-dependent protein kinase catalytic subunit OS=Homo sapiens GN=PRKDC PE=1 SV=3 - [PRKDC_HUMAN]
P62913	60S ribosomal protein L11 OS=Homo sapiens GN=RPL11 PE=1 SV=2 - [RL11_HUMAN]
P04792	Heat shock protein beta-1 OS=Homo sapiens GN=HSPB1 PE=1 SV=2 - [HSPB1_HUMAN]
Q9UHF7	Zinc finger transcription factor Trps1 OS=Homo sapiens GN=TRPS1 PE=1 SV=2 - [TRPS1_HUMAN]
Q99426	Tubulin-folding cofactor B OS=Homo sapiens GN=TBCB PE=1 SV=2 - [TBCB_HUMAN]
Q99496	E3 ubiquitin-protein ligase RING2 OS=Homo sapiens GN=RNFB PE=1 SV=1 - [RING2_HUMAN]
P21333	Filamin-A OS=Homo sapiens GN=FLNA PE=1 SV=4 - [FLNA_HUMAN]
P23771	Trans-acting T-cell-specific transcription factor GATA-3 OS=Homo sapiens GN=GATA3 PE=1 SV=1 - [GATA3_HUMAN]
O15504	Nucleoporin-like protein 2 OS=Homo sapiens GN=NUPL2 PE=1 SV=1 - [NUPL2_HUMAN]
P53992	Protein transport protein Sec24C OS=Homo sapiens GN=SEC24C PE=1 SV=3 - [SC24C_HUMAN]
Q8N5F7	NF-kappa-B-activating protein OS=Homo sapiens GN=NKAP PE=1 SV=1 - [NKAP_HUMAN]
Q66PJ3	ADP-ribosylation factor-like protein 6-interacting protein 4 OS=Homo sapiens GN=ARL6IP4 PE=1 SV=2 - [AR6P4_HUMAN]
P26368	Splicing factor U2AF 65 kDa subunit OS=Homo sapiens GN=U2AF2 PE=1 SV=4 - [U2AF2_HUMAN]
P55072	Transitional endoplasmic reticulum ATPase OS=Homo sapiens GN=VCP PE=1 SV=4 - [TERA_HUMAN]
Q7Z5L9	Interferon regulatory factor 2-binding protein 2 OS=Homo sapiens GN=IRF2BP2 PE=1 SV=2 - [I2BP2_HUMAN]
P40763	Signal transducer and activator of transcription 3 OS=Homo sapiens GN=STAT3 PE=1 SV=2 - [STAT3_HUMAN]
O75131	Copine-3 OS=Homo sapiens GN=CPNE3 PE=1 SV=1 - [CPNE3_HUMAN]
Q00796	Sorbitol dehydrogenase OS=Homo sapiens GN=SORD PE=1 SV=4 - [DHSO_HUMAN]
P02452	Collagen alpha-1(I) chain OS=Homo sapiens GN=COL1A1 PE=1 SV=5 - [CO1A1_HUMAN]
Q00535	Cyclin-dependent-like kinase 5 OS=Homo sapiens GN=CDK5 PE=1 SV=3 - [CDK5_HUMAN]
O43143	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 OS=Homo sapiens GN=DHX15 PE=1 SV=2 - [DHX15_HUMAN]
Q9H6T0	Epithelial splicing regulatory protein 2 OS=Homo sapiens GN=ESRP2 PE=1 SV=1 - [ESRP2_HUMAN]
O75643	U5 small nuclear ribonucleoprotein 200 kDa helicase OS=Homo sapiens GN=SNRNP200 PE=1 SV=2 - [U520_HUMAN]
O00567	Nucleolar protein 56 OS=Homo sapiens GN=NOP56 PE=1 SV=4 - [NOP56_HUMAN]
Q96HC4	PDZ and LIM domain protein 5 OS=Homo sapiens GN=PDLIM5 PE=1 SV=5 - [PDLI5_HUMAN]

P09622	Dihydrolipoyl dehydrogenase, mitochondrial OS=Homo sapiens GN=DLD PE=1 SV=2 - [DLDH_HUMAN]
P12270	Nucleoprotein TPR OS=Homo sapiens GN=TPR PE=1 SV=3 - [TPR_HUMAN]
Q9BUJ2	Heterogeneous nuclear ribonucleoprotein U-like protein 1 OS=Homo sapiens GN=HNRNPUL1 PE=1 SV=2 - [HNRL1_HUMAN]
P49588	Alanine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=AARS PE=1 SV=2 - [SYAC_HUMAN]
Q9NTZ6	RNA-binding protein 12 OS=Homo sapiens GN=RBM12 PE=1 SV=1 - [RBM12_HUMAN]
Q6P2Q9	Pre-mRNA-processing-splicing factor 8 OS=Homo sapiens GN=PRPF8 PE=1 SV=2 - [PRP8_HUMAN]
Q14152	Eukaryotic translation initiation factor 3 subunit A OS=Homo sapiens GN=EIF3A PE=1 SV=1 - [EIF3A_HUMAN]
P78347	General transcription factor II-I OS=Homo sapiens GN=GTF2I PE=1 SV=2 - [GTF2I_HUMAN]
P02461	Collagen alpha-1(III) chain OS=Homo sapiens GN=COL3A1 PE=1 SV=4 - [CO3A1_HUMAN]

AR interacting proteins identified by RIME in 3/3 replicates excluding IgG:	
T-47D	
Accession	Description
P26599	Polypyrimidine tract-binding protein 1 OS=Homo sapiens GN=PTBP1 PE=1 SV=1 - [PTBP1_HUMAN]
P13807	Glycogen [starch] synthase, muscle OS=Homo sapiens GN=GYS1 PE=1 SV=2 - [GYS1_HUMAN]
Q8N163	Cell cycle and apoptosis regulator protein 2 OS=Homo sapiens GN=CCAR2 PE=1 SV=2 - [CCAR2_HUMAN]
Q12905	Interleukin enhancer-binding factor 2 OS=Homo sapiens GN=ILF2 PE=1 SV=2 - [ILF2_HUMAN]
P11142	Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1 - [HSP7C_HUMAN]
P60866	40S ribosomal protein S20 OS=Homo sapiens GN=RPS20 PE=1 SV=1 - [RS20_HUMAN]
P10275	Androgen receptor OS=Homo sapiens GN=AR PE=1 SV=3 - [ANDR_HUMAN]
P17275	Transcription factor jun-B OS=Homo sapiens GN=JUNB PE=1 SV=1 - [JUNB_HUMAN]
O00330	Pyruvate dehydrogenase protein X component, mitochondrial OS=Homo sapiens GN=PDHX PE=1 SV=3 - [ODPX_HUMAN]
P50548	ETS domain-containing transcription factor ERF OS=Homo sapiens GN=ERF PE=1 SV=2 - [ERF_HUMAN]
P62306	Small nuclear ribonucleoprotein F OS=Homo sapiens GN=SNRPF PE=1 SV=1 - [RUXF_HUMAN]
O14745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1 OS=Homo sapiens GN=SLC9A3R1 PE=1 SV=4 - [NHRF1_HUMAN]
P14678	Small nuclear ribonucleoprotein-associated proteins B and B' OS=Homo sapiens GN=SNRPB PE=1 SV=2 - [RSMB_HUMAN]
Q8WXF1	Paraspeckle component 1 OS=Homo sapiens GN=PSPC1 PE=1 SV=1 - [PSPC1_HUMAN]
P09429	High mobility group protein B1 OS=Homo sapiens GN=HMGB1 PE=1 SV=3 - [HMGB1_HUMAN]
Q16629	Serine/arginine-rich splicing factor 7 OS=Homo sapiens GN=SRSF7 PE=1 SV=1 - [SRSF7_HUMAN]
Q15637	Splicing factor 1 OS=Homo sapiens GN=SF1 PE=1 SV=4 - [SF01_HUMAN]
O60506	Heterogeneous nuclear ribonucleoprotein Q OS=Homo sapiens GN=SYNCRIP PE=1 SV=2 - [HNRPQ_HUMAN]
P13010	X-ray repair cross-complementing protein 5 OS=Homo sapiens GN=XRCC5 PE=1 SV=3 - [XRCC5_HUMAN]
P42166	Lamina-associated polypeptide 2, isoform alpha OS=Homo sapiens GN=TMPO PE=1 SV=2 - [LAP2A_HUMAN]
Q96PM9	Zinc finger protein 385A OS=Homo sapiens GN=ZNF385A PE=1 SV=2 - [Z385A_HUMAN]
P98175	RNA-binding protein 10 OS=Homo sapiens GN=RBM10 PE=1 SV=3 - [RBM10_HUMAN]
P02533	Keratin, type I cytoskeletal 14 OS=Homo sapiens GN=KRT14 PE=1 SV=4 - [K1C14_HUMAN]
Q8WWM7	Ataxin-2-like protein OS=Homo sapiens GN=ATXN2L PE=1 SV=2 - [ATX2L_HUMAN]

P0CG12	Chromosome transmission fidelity protein 8 homolog isoform 2 OS=Homo sapiens GN=CTHF8 PE=1 SV=1 - [CTF8A_HUMAN]
Q6STE5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 3 OS=Homo sapiens GN=SMARCD3 PE=1 SV=1 - [SMRD3_HUMAN]
P53992	Protein transport protein Sec24C OS=Homo sapiens GN=SEC24C PE=1 SV=3 - [SC24C_HUMAN]
Q9BUJ2	Heterogeneous nuclear ribonucleoprotein U-like protein 1 OS=Homo sapiens GN=HNRNPUL1 PE=1 SV=2 - [HNRL1_HUMAN]
Q00341	Vigilin OS=Homo sapiens GN=HDLBP PE=1 SV=2 - [VIGLN_HUMAN]
Q96EP5	DAZ-associated protein 1 OS=Homo sapiens GN=DAZAP1 PE=1 SV=1 - [DAZP1_HUMAN]
Q14011	Cold-inducible RNA-binding protein OS=Homo sapiens GN=CIRBP PE=1 SV=1 - [CIRBP_HUMAN]
O60336	Mitogen-activated protein kinase-binding protein 1 OS=Homo sapiens GN=MAPKBP1 PE=1 SV=4 - [MABP1_HUMAN]
P09874	Poly [ADP-ribose] polymerase 1 OS=Homo sapiens GN=PARP1 PE=1 SV=4 - [PARP1_HUMAN]
P23771	Trans-acting T-cell-specific transcription factor GATA-3 OS=Homo sapiens GN=GATA3 PE=1 SV=1 - [GATA3_HUMAN]
O15504	Nucleoporin-like protein 2 OS=Homo sapiens GN=NUPL2 PE=1 SV=1 - [NUPL2_HUMAN]
P49750	YLP motif-containing protein 1 OS=Homo sapiens GN=YLPM1 PE=1 SV=3 - [YLPM1_HUMAN]
P22087	rRNA 2'-O-methyltransferase fibrillarin OS=Homo sapiens GN=FBL PE=1 SV=2 - [FBRL_HUMAN]
P12956	X-ray repair cross-complementing protein 6 OS=Homo sapiens GN=XRCC6 PE=1 SV=2 - [XRCC6_HUMAN]
O00148	ATP-dependent RNA helicase DDX39A OS=Homo sapiens GN=DDX39A PE=1 SV=2 - [DX39A_HUMAN]
Q01085	Nucleolysin TIAR OS=Homo sapiens GN=TIAL1 PE=1 SV=1 - [TIAR_HUMAN]
O43809	Cleavage and polyadenylation specificity factor subunit 5 OS=Homo sapiens GN=NUDT21 PE=1 SV=1 - [CPSF5_HUMAN]
P83731	60S ribosomal protein L24 OS=Homo sapiens GN=RPL24 PE=1 SV=1 - [RL24_HUMAN]
Q15287	RNA-binding protein with serine-rich domain 1 OS=Homo sapiens GN=RNPS1 PE=1 SV=1 - [RNPS1_HUMAN]
P18754	Regulator of chromosome condensation OS=Homo sapiens GN=RCC1 PE=1 SV=1 - [RCC1_HUMAN]
Q15942	Zyxin OS=Homo sapiens GN=ZYX PE=1 SV=1 - [ZYX_HUMAN]
Q8TD19	Serine/threonine-protein kinase Nek9 OS=Homo sapiens GN=NEK9 PE=1 SV=2 - [NEK9_HUMAN]
O00512	B-cell CLL/lymphoma 9 protein OS=Homo sapiens GN=BCL9 PE=1 SV=4 - [BCL9_HUMAN]
P20073	Annexin A7 OS=Homo sapiens GN=ANXA7 PE=1 SV=3 - [ANXA7_HUMAN]
O43143	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 OS=Homo sapiens GN=DHX15 PE=1 SV=2 - [DHX15_HUMAN]
Q14498	RNA-binding protein 39 OS=Homo sapiens GN=RBM39 PE=1 SV=2 - [RBM39_HUMAN]

Q9UBV2	Protein sel-1 homolog 1 OS=Homo sapiens GN=SEL1L PE=1 SV=3 - [SE1L1_HUMAN]
Q9Y2X3	Nucleolar protein 58 OS=Homo sapiens GN=NOP58 PE=1 SV=1 - [NOP58_HUMAN]
Q96T37	Putative RNA-binding protein 15 OS=Homo sapiens GN=RBM15 PE=1 SV=2 - [RBM15_HUMAN]
Q9ULV4	Coronin-1C OS=Homo sapiens GN=CORO1C PE=1 SV=1 - [COR1C_HUMAN]
Q9H0D6	5'-3' exoribonuclease 2 OS=Homo sapiens GN=XRN2 PE=1 SV=1 - [XRN2_HUMAN]
O75369	Filamin-B OS=Homo sapiens GN=FLNB PE=1 SV=2 - [FLNB_HUMAN]
P11387	DNA topoisomerase 1 OS=Homo sapiens GN=TOP1 PE=1 SV=2 - [TOP1_HUMAN]
Q5M775	Cytospin-B OS=Homo sapiens GN=SPECC1 PE=1 SV=1 - [CYTSB_HUMAN]
P02452	Collagen alpha-1(I) chain OS=Homo sapiens GN=COL1A1 PE=1 SV=5 - [CO1A1_HUMAN]

AR interacting proteins identified by RIME in 3/3 replicates excluding IgG:	
MFM-223	
Accession	Description
P17275	Transcription factor jun-B OS=Homo sapiens GN=JUNB PE=1 SV=1 - [JUNB_HUMAN]
P98179	RNA-binding protein 3 OS=Homo sapiens GN=RBM3 PE=1 SV=1 - [RBM3_HUMAN]
Q12905	Interleukin enhancer-binding factor 2 OS=Homo sapiens GN=ILF2 PE=1 SV=2 - [ILF2_HUMAN]
O00567	Nucleolar protein 56 OS=Homo sapiens GN=NOP56 PE=1 SV=4 - [NOP56_HUMAN]
P07900	Heat shock protein HSP 90-alpha OS=Homo sapiens GN=HSP90AA1 PE=1 SV=5 - [HS90A_HUMAN]
P62937	Peptidyl-prolyl cis-trans isomerase A OS=Homo sapiens GN=PPIA PE=1 SV=2 - [PPIA_HUMAN]
P10275	Androgen receptor OS=Homo sapiens GN=AR PE=1 SV=3 - [ANDR_HUMAN]
P63241	Eukaryotic translation initiation factor 5A-1 OS=Homo sapiens GN=EIF5A PE=1 SV=2 - [IF5A1_HUMAN]
Q9UKM9	RNA-binding protein Raly OS=Homo sapiens GN=RALY PE=1 SV=1 - [RALY_HUMAN]
Q15717	ELAV-like protein 1 OS=Homo sapiens GN=ELAVL1 PE=1 SV=2 - [ELAV1_HUMAN]
P09874	Poly [ADP-ribose] polymerase 1 OS=Homo sapiens GN=PARP1 PE=1 SV=4 - [PARP1_HUMAN]
P08238	Heat shock protein HSP 90-beta OS=Homo sapiens GN=HSP90AB1 PE=1 SV=4 - [HS90B_HUMAN]
O43809	Cleavage and polyadenylation specificity factor subunit 5 OS=Homo sapiens GN=NUDT21 PE=1 SV=1 - [CPSF5_HUMAN]
P12956	X-ray repair cross-complementing protein 6 OS=Homo sapiens GN=XRCC6 PE=1 SV=2 - [XRCC6_HUMAN]
Q8WXF1	Paraspeckle component 1 OS=Homo sapiens GN=PSPC1 PE=1 SV=1 - [PSPC1_HUMAN]
P50548	ETS domain-containing transcription factor ERF OS=Homo sapiens GN=ERF PE=1 SV=2 - [ERF_HUMAN]
P61956	Small ubiquitin-related modifier 2 OS=Homo sapiens GN=SUMO2 PE=1 SV=3 - [SUMO2_HUMAN]
Q92804	TATA-binding protein-associated factor 2N OS=Homo sapiens GN=TAF15 PE=1 SV=1 - [RBP56_HUMAN]
Q13263	Transcription intermediary factor 1-beta OS=Homo sapiens GN=TRIM28 PE=1 SV=5 - [TIF1B_HUMAN]
P43243	Matrin-3 OS=Homo sapiens GN=MATR3 PE=1 SV=2 - [MATR3_HUMAN]
Q92925	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 2 OS=Homo sapiens GN=SMARCD2 PE=1 SV=3 - [SMRD2_HUMAN]
Q8N163	Cell cycle and apoptosis regulator protein 2 OS=Homo sapiens GN=CCAR2 PE=1 SV=2 - [CCAR2_HUMAN]
P13807	Glycogen [starch] synthase, muscle OS=Homo sapiens GN=GYS1 PE=1 SV=2 - [GYS1_HUMAN]

P08621	U1 small nuclear ribonucleoprotein 70 kDa OS=Homo sapiens GN=SNRNP70 PE=1 SV=2 - [RU17_HUMAN]
P09429	High mobility group protein B1 OS=Homo sapiens GN=HMGB1 PE=1 SV=3 - [HMGB1_HUMAN]
P98175	RNA-binding protein 10 OS=Homo sapiens GN=RBM10 PE=1 SV=3 - [RBM10_HUMAN]
P62158	Calmodulin OS=Homo sapiens GN=CALM1 PE=1 SV=2 - [CALM_HUMAN]
P07737	Profilin-1 OS=Homo sapiens GN=PFN1 PE=1 SV=2 - [PROF1_HUMAN]
P0CG48	Polyubiquitin-C OS=Homo sapiens GN=UBC PE=1 SV=3 - [UBC_HUMAN]
P0C0S5	Histone H2A.Z OS=Homo sapiens GN=H2AFZ PE=1 SV=2 - [H2AZ_HUMAN]
P26368	Splicing factor U2AF 65 kDa subunit OS=Homo sapiens GN=U2AF2 PE=1 SV=4 - [U2AF2_HUMAN]
P18754	Regulator of chromosome condensation OS=Homo sapiens GN=RCC1 PE=1 SV=1 - [RCC1_HUMAN]
P55795	Heterogeneous nuclear ribonucleoprotein H2 OS=Homo sapiens GN=HNRNPH2 PE=1 SV=1 - [HNRH2_HUMAN]
O43684	Mitotic checkpoint protein BUB3 OS=Homo sapiens GN=BUB3 PE=1 SV=1 - [BUB3_HUMAN]
P60660	Myosin light polypeptide 6 OS=Homo sapiens GN=MYL6 PE=1 SV=2 - [MYL6_HUMAN]
Q16629	Serine/arginine-rich splicing factor 7 OS=Homo sapiens GN=SRSF7 PE=1 SV=1 - [SRSF7_HUMAN]
P32119	Peroxiredoxin-2 OS=Homo sapiens GN=PRDX2 PE=1 SV=5 - [PRDX2_HUMAN]
O14979	Heterogeneous nuclear ribonucleoprotein D-like OS=Homo sapiens GN=HNRNPD L PE=1 SV=3 - [HNRDL_HUMAN]
Q15942	Zyxin OS=Homo sapiens GN=ZYX PE=1 SV=1 - [ZYX_HUMAN]
P25815	Protein S100-P OS=Homo sapiens GN=S100P PE=1 SV=2 - [S100P_HUMAN]
Q8WWM7	Ataxin-2-like protein OS=Homo sapiens GN=ATXN2L PE=1 SV=2 - [ATX2L_HUMAN]
Q9Y5S9	RNA-binding protein 8A OS=Homo sapiens GN=RBM8A PE=1 SV=1 - [RBM8A_HUMAN]
P02533	Keratin, type I cytoskeletal 14 OS=Homo sapiens GN=KRT14 PE=1 SV=4 - [K1C14_HUMAN]
P68032	Actin, alpha cardiac muscle 1 OS=Homo sapiens GN=ACTC1 PE=1 SV=1 - [ACTC_HUMAN]
Q96124	Far upstream element-binding protein 3 OS=Homo sapiens GN=FUBP3 PE=1 SV=2 - [FUBP3_HUMAN]
P62995	Transformer-2 protein homolog beta OS=Homo sapiens GN=TRA2B PE=1 SV=1 - [TRA2B_HUMAN]
P23526	Adenosylhomocysteinase OS=Homo sapiens GN=AHCY PE=1 SV=4 - [SAHH_HUMAN]
Q9UBV2	Protein sel-1 homolog 1 OS=Homo sapiens GN=SEL1L PE=1 SV=3 - [SE1L1_HUMAN]
Q99829	Copine-1 OS=Homo sapiens GN=CPNE1 PE=1 SV=1 - [CPNE1_HUMAN]
O75494	Serine/arginine-rich splicing factor 10 OS=Homo sapiens GN=SRSF10 PE=1 SV=1 - [SRS10_HUMAN]
P62306	Small nuclear ribonucleoprotein F OS=Homo sapiens GN=SNRPF PE=1 SV=1 - [RUXF_HUMAN]



P63104	14-3-3 protein zeta/delta OS=Homo sapiens GN=YWHAZ PE=1 SV=1 - [1433Z_HUMAN]
P30050	60S ribosomal protein L12 OS=Homo sapiens GN=RPL12 PE=1 SV=1 - [RL12_HUMAN]
P62316	Small nuclear ribonucleoprotein Sm D2 OS=Homo sapiens GN=SNRPD2 PE=1 SV=1 - [SMD2_HUMAN]
P62249	40S ribosomal protein S16 OS=Homo sapiens GN=RPS16 PE=1 SV=2 - [RS16_HUMAN]
Q13162	Peroxiredoxin-4 OS=Homo sapiens GN=PRDX4 PE=1 SV=1 - [PRDX4_HUMAN]
Q8N684	Cleavage and polyadenylation specificity factor subunit 7 OS=Homo sapiens GN=CPSF7 PE=1 SV=1 - [CPSF7_HUMAN]
Q969G3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1 OS=Homo sapiens GN=SMARCE1 PE=1 SV=2 - [SMCE1_HUMAN]
O00512	B-cell CLL/lymphoma 9 protein OS=Homo sapiens GN=BCL9 PE=1 SV=4 - [BCL9_HUMAN]
O14497	AT-rich interactive domain-containing protein 1A OS=Homo sapiens GN=ARID1A PE=1 SV=3 - [ARI1A_HUMAN]
P42167	Lamina-associated polypeptide 2, isoforms beta/gamma OS=Homo sapiens GN=TMPO PE=1 SV=2 - [LAP2B_HUMAN]
Q8WVV9	Heterogeneous nuclear ribonucleoprotein L-like OS=Homo sapiens GN=HNRNPLL PE=1 SV=1 - [HNRL_L_HUMAN]
P53992	Protein transport protein Sec24C OS=Homo sapiens GN=SEC24C PE=1 SV=3 - [SC24C_HUMAN]
Q6NXG1	Epithelial splicing regulatory protein 1 OS=Homo sapiens GN=ESRP1 PE=1 SV=2 - [ESRP1_HUMAN]
P11940	Polyadenylate-binding protein 1 OS=Homo sapiens GN=PABPC1 PE=1 SV=2 - [PABP1_HUMAN]
Q9ULV4	Coronin-1C OS=Homo sapiens GN=CORO1C PE=1 SV=1 - [COR1C_HUMAN]
P14678	Small nuclear ribonucleoprotein-associated proteins B and B' OS=Homo sapiens GN=SNRPB PE=1 SV=2 - [RSMB_HUMAN]
Q16630	Cleavage and polyadenylation specificity factor subunit 6 OS=Homo sapiens GN=CPSF6 PE=1 SV=2 - [CPSF6_HUMAN]
P62826	GTP-binding nuclear protein Ran OS=Homo sapiens GN=RAN PE=1 SV=3 - [RAN_HUMAN]
O60506	Heterogeneous nuclear ribonucleoprotein Q OS=Homo sapiens GN=SYNCRIP PE=1 SV=2 - [HNRPQ_HUMAN]
Q13151	Heterogeneous nuclear ribonucleoprotein A0 OS=Homo sapiens GN=HNRNPA0 PE=1 SV=1 - [ROA0_HUMAN]
Q6NVV1	Putative 60S ribosomal protein L13a protein RPL13AP3 OS=Homo sapiens GN=RPL13AP3 PE=5 SV=1 - [R13P3_HUMAN]
P09012	U1 small nuclear ribonucleoprotein A OS=Homo sapiens GN=SNRPA PE=1 SV=3 - [SNRPA_HUMAN]
Q8NAV1	Pre-mRNA-splicing factor 38A OS=Homo sapiens GN=PRPF38A PE=1 SV=1 - [PR38A_HUMAN]
Q96EP5	DAZ-associated protein 1 OS=Homo sapiens GN=DAZAP1 PE=1 SV=1 - [DAZP1_HUMAN]
P61326	Protein mago nashi homolog OS=Homo sapiens GN=MAGOH PE=1 SV=1 - [MGN_HUMAN]

P20073	Annexin A7 OS=Homo sapiens GN=ANXA7 PE=1 SV=3 - [ANXA7_HUMAN]
P60866	40S ribosomal protein S20 OS=Homo sapiens GN=RPS20 PE=1 SV=1 - [RS20_HUMAN]
Q00341	Vigilin OS=Homo sapiens GN=HDLBP PE=1 SV=2 - [VIGLN_HUMAN]
O75340	Programmed cell death protein 6 OS=Homo sapiens GN=PDCD6 PE=1 SV=1 - [PDCD6_HUMAN]
P46734	Dual specificity mitogen-activated protein kinase kinase 3 OS=Homo sapiens GN=MAP2K3 PE=1 SV=2 - [MP2K3_HUMAN]
P49750	YLP motif-containing protein 1 OS=Homo sapiens GN=YLPM1 PE=1 SV=3 - [YLPM1_HUMAN]
P18206	Vinculin OS=Homo sapiens GN=VCL PE=1 SV=4 - [VINC_HUMAN]
Q06830	Peroxiredoxin-1 OS=Homo sapiens GN=PRDX1 PE=1 SV=1 - [PRDX1_HUMAN]
P36578	60S ribosomal protein L4 OS=Homo sapiens GN=RPL4 PE=1 SV=5 - [RL4_HUMAN]
Q15427	Splicing factor 3B subunit 4 OS=Homo sapiens GN=SF3B4 PE=1 SV=1 - [SF3B4_HUMAN]
P12532	Creatine kinase U-type, mitochondrial OS=Homo sapiens GN=CKMT1A PE=1 SV=1 - [KCRU_HUMAN]
PODMV9;P08107	Heat shock 70 kDa protein 1B OS=Homo sapiens GN=HSPA1B PE=1 SV=1 - [HS71B_HUMAN]
P46781	40S ribosomal protein S9 OS=Homo sapiens GN=RPS9 PE=1 SV=3 - [RS9_HUMAN]
P23771	Trans-acting T-cell-specific transcription factor GATA-3 OS=Homo sapiens GN=GATA3 PE=1 SV=1 - [GATA3_HUMAN]
P62318	Small nuclear ribonucleoprotein Sm D3 OS=Homo sapiens GN=SNRPD3 PE=1 SV=1 - [SMD3_HUMAN]
Q15287	RNA-binding protein with serine-rich domain 1 OS=Homo sapiens GN=RNPS1 PE=1 SV=1 - [RNPS1_HUMAN]
P62899	60S ribosomal protein L31 OS=Homo sapiens GN=RPL31 PE=1 SV=1 - [RL31_HUMAN]
P62244	40S ribosomal protein S15a OS=Homo sapiens GN=RPS15A PE=1 SV=2 - [RS15A_HUMAN]
P49207	60S ribosomal protein L34 OS=Homo sapiens GN=RPL34 PE=1 SV=3 - [RL34_HUMAN]
Q99986	Serine/threonine-protein kinase VRK1 OS=Homo sapiens GN=VRK1 PE=1 SV=1 - [VRK1_HUMAN]
P45973	Chromobox protein homolog 5 OS=Homo sapiens GN=CBX5 PE=1 SV=1 - [CBX5_HUMAN]
Q14011	Cold-inducible RNA-binding protein OS=Homo sapiens GN=CIRBP PE=1 SV=1 - [CIRBP_HUMAN]
P26373	60S ribosomal protein L13 OS=Homo sapiens GN=RPL13 PE=1 SV=4 - [RL13_HUMAN]
P62829	60S ribosomal protein L23 OS=Homo sapiens GN=RPL23 PE=1 SV=1 - [RL23_HUMAN]
P13489	Ribonuclease inhibitor OS=Homo sapiens GN=RNH1 PE=1 SV=2 - [RINI_HUMAN]
Q15637	Splicing factor 1 OS=Homo sapiens GN=SF1 PE=1 SV=4 - [SF01_HUMAN]
P46783	40S ribosomal protein S10 OS=Homo sapiens GN=RPS10 PE=1 SV=1 - [RS10_HUMAN]

Q8TAQ2	SWI/SNF complex subunit SMARCC2 OS=Homo sapiens GN=SMARCC2 PE=1 SV=1 - [SMRC2_HUMAN]
Q01085	Nucleolysin TIAR OS=Homo sapiens GN=TIAL1 PE=1 SV=1 - [TIAR_HUMAN]
O60336	Mitogen-activated protein kinase-binding protein 1 OS=Homo sapiens GN=MAPKBP1 PE=1 SV=4 - [MABP1_HUMAN]
O15504	Nucleoporin-like protein 2 OS=Homo sapiens GN=NUPL2 PE=1 SV=1 - [NUPL2_HUMAN]
P83731	60S ribosomal protein L24 OS=Homo sapiens GN=RPL24 PE=1 SV=1 - [RL24_HUMAN]
Q5JTV8	Torsin-1A-interacting protein 1 OS=Homo sapiens GN=TOR1AIP1 PE=1 SV=2 - [TOIP1_HUMAN]
Q9UHF7	Zinc finger transcription factor Trps1 OS=Homo sapiens GN=TRPS1 PE=1 SV=2 - [TRPS1_HUMAN]
Q9NY12	H/ACA ribonucleoprotein complex subunit 1 OS=Homo sapiens GN=GAR1 PE=1 SV=1 - [GAR1_HUMAN]
Q14978	Nucleolar and coiled-body phosphoprotein 1 OS=Homo sapiens GN=NOLC1 PE=1 SV=2 - [NOLC1_HUMAN]
Q01130	Serine/arginine-rich splicing factor 2 OS=Homo sapiens GN=SRSF2 PE=1 SV=4 - [SRSF2_HUMAN]
Q09666	Neuroblast differentiation-associated protein AHNAK OS=Homo sapiens GN=AHNAK PE=1 SV=2 - [AHNK_HUMAN]
O00330	Pyruvate dehydrogenase protein X component, mitochondrial OS=Homo sapiens GN=PDHX PE=1 SV=3 - [ODPX_HUMAN]
Q96KR1	Zinc finger RNA-binding protein OS=Homo sapiens GN=ZFR PE=1 SV=2 - [ZFR_HUMAN]
O15042	U2 snRNP-associated SURP motif-containing protein OS=Homo sapiens GN=U2SURP PE=1 SV=2 - [SR140_HUMAN]
O43823	A-kinase anchor protein 8 OS=Homo sapiens GN=AKAP8 PE=1 SV=1 - [AKAP8_HUMAN]
Q92499	ATP-dependent RNA helicase DDX1 OS=Homo sapiens GN=DDX1 PE=1 SV=2 - [DDX1_HUMAN]
Q14152	Eukaryotic translation initiation factor 3 subunit A OS=Homo sapiens GN=EIF3A PE=1 SV=1 - [EIF3A_HUMAN]
Q9NR30	Nucleolar RNA helicase 2 OS=Homo sapiens GN=DDX21 PE=1 SV=5 - [DDX21_HUMAN]
O76021	Ribosomal L1 domain-containing protein 1 OS=Homo sapiens GN=RSL1D1 PE=1 SV=3 - [RL1D1_HUMAN]
Q9NWB6	Arginine and glutamate-rich protein 1 OS=Homo sapiens GN=ARGLU1 PE=1 SV=1 - [ARGL1_HUMAN]
Q9ULJ6	Zinc finger MIZ domain-containing protein 1 OS=Homo sapiens GN=ZMIZ1 PE=1 SV=3 - [ZMIZ1_HUMAN]
Q5BKZ1	DBIRD complex subunit ZNF326 OS=Homo sapiens GN=ZNF326 PE=1 SV=2 - [ZN326_HUMAN]
POCG12	Chromosome transmission fidelity protein 8 homolog isoform 2 OS=Homo sapiens GN=CTHF8 PE=1 SV=1 - [CTF8A_HUMAN]
P04196	Histidine-rich glycoprotein OS=Homo sapiens GN=HRG PE=1 SV=1 - [HRG_HUMAN]
Q6P2Q9	Pre-mRNA-processing-splicing factor 8 OS=Homo sapiens GN=PRPF8 PE=1 SV=2 - [PRP8_HUMAN]

P43246	DNA mismatch repair protein Msh2 OS=Homo sapiens GN=MSH2 PE=1 SV=1 - [MSH2_HUMAN]
Q5M775	Cytospin-B OS=Homo sapiens GN=SPECC1 PE=1 SV=1 - [CYTSB_HUMAN]

AR interacting proteins identified by RIME in 3/3 replicates excluding IgG:	
MDA-MB-453	
Accession	Description
P07355	Annexin A2 OS=Homo sapiens GN=ANXA2 PE=1 SV=2 - [ANXA2_HUMAN]
P32322	Pyrraline-5-carboxylate reductase 1, mitochondrial OS=Homo sapiens GN=PYCR1 PE=1 SV=2 - [P5CR1_HUMAN]
Q92925	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 2 OS=Homo sapiens GN=SMARCD2 PE=1 SV=3 - [SMRD2_HUMAN]
P17275	Transcription factor jun-B OS=Homo sapiens GN=JUNB PE=1 SV=1 - [JUNB_HUMAN]
P10275	Androgen receptor OS=Homo sapiens GN=AR PE=1 SV=3 - [ANDR_HUMAN]
P05204	Non-histone chromosomal protein HMG-17 OS=Homo sapiens GN=HMGN2 PE=1 SV=3 - [HMGN2_HUMAN]
Q15717	ELAV-like protein 1 OS=Homo sapiens GN=ELAVL1 PE=1 SV=2 - [ELAV1_HUMAN]
Q969G3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1 OS=Homo sapiens GN=SMARCE1 PE=1 SV=2 - [SMCE1_HUMAN]
P63241	Eukaryotic translation initiation factor 5A-1 OS=Homo sapiens GN=EIF5A PE=1 SV=2 - [IF5A1_HUMAN]
P10515	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial OS=Homo sapiens GN=DLAT PE=1 SV=3 - [ODP2_HUMAN]
P11177	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial OS=Homo sapiens GN=PDHB PE=1 SV=3 - [ODPB_HUMAN]
P39019	40S ribosomal protein S19 OS=Homo sapiens GN=RPS19 PE=1 SV=2 - [RS19_HUMAN]
P49773	Histidine triad nucleotide-binding protein 1 OS=Homo sapiens GN=HINT1 PE=1 SV=2 - [HINT1_HUMAN]
P08559	Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial OS=Homo sapiens GN=PDHA1 PE=1 SV=3 - [ODPA_HUMAN]
Q9H3K6	Bola-like protein 2 OS=Homo sapiens GN=BOLA2 PE=1 SV=1 - [BOLA2_HUMAN]
P50548	ETS domain-containing transcription factor ERF OS=Homo sapiens GN=ERF PE=1 SV=2 - [ERF_HUMAN]
P13807	Glycogen [starch] synthase, muscle OS=Homo sapiens GN=GYS1 PE=1 SV=2 - [GYS1_HUMAN]
P62857	40S ribosomal protein S28 OS=Homo sapiens GN=RPS28 PE=1 SV=1 - [RS28_HUMAN]
O00330	Pyruvate dehydrogenase protein X component, mitochondrial OS=Homo sapiens GN=PDHX PE=1 SV=3 - [ODPX_HUMAN]
Q02790	Peptidyl-prolyl cis-trans isomerase FKBP4 OS=Homo sapiens GN=FKBP4 PE=1 SV=3 - [FKBP4_HUMAN]
Q8TAQ2	SWI/SNF complex subunit SMARCC2 OS=Homo sapiens GN=SMARCC2 PE=1 SV=1 - [SMRC2_HUMAN]
Q16401	26S proteasome non-ATPase regulatory subunit 5 OS=Homo sapiens GN=PSMD5 PE=1 SV=3 - [PSMD5_HUMAN]
O96019	Actin-like protein 6A OS=Homo sapiens GN=ACTL6A PE=1 SV=1 - [ACL6A_HUMAN]
Q14166	Tubulin--tyrosine ligase-like protein 12 OS=Homo sapiens GN=TTL12 PE=1 SV=2 - [TTL12_HUMAN]

Q96AE4	Far upstream element-binding protein 1 OS=Homo sapiens GN=FUBP1 PE=1 SV=3 - [FUBP1_HUMAN]
Q92481	Transcription factor AP-2-beta OS=Homo sapiens GN=TFAP2B PE=1 SV=2 - [AP2B_HUMAN]
P63165	Small ubiquitin-related modifier 1 OS=Homo sapiens GN=SUMO1 PE=1 SV=1 - [SUMO1_HUMAN]
P10768	S-formylglutathione hydrolase OS=Homo sapiens GN=ESD PE=1 SV=2 - [ESTD_HUMAN]
Q96EP5	DAZ-associated protein 1 OS=Homo sapiens GN=DAZAP1 PE=1 SV=1 - [DAZP1_HUMAN]
P48047	ATP synthase subunit O, mitochondrial OS=Homo sapiens GN=ATP5O PE=1 SV=1 - [ATPO_HUMAN]
P36776	Lon protease homolog, mitochondrial OS=Homo sapiens GN=LONP1 PE=1 SV=2 - [LONM_HUMAN]
Q96DG6	Carboxymethylenebutenolidase homolog OS=Homo sapiens GN=CMBL PE=1 SV=1 - [CMBL_HUMAN]
O14979	Heterogeneous nuclear ribonucleoprotein D-like OS=Homo sapiens GN=HNRNPD PE=1 SV=3 - [HNRDL_HUMAN]
P62888	60S ribosomal protein L30 OS=Homo sapiens GN=RPL30 PE=1 SV=2 - [RL30_HUMAN]
P09622	Dihydrolipoyl dehydrogenase, mitochondrial OS=Homo sapiens GN=DLD PE=1 SV=2 - [DLDH_HUMAN]
Q92785	Zinc finger protein ubi-d4 OS=Homo sapiens GN=DPF2 PE=1 SV=2 - [REQU_HUMAN]
Q9H0D6	5'-3' exoribonuclease 2 OS=Homo sapiens GN=XRN2 PE=1 SV=1 - [XRN2_HUMAN]
P23193	Transcription elongation factor A protein 1 OS=Homo sapiens GN=TCEA1 PE=1 SV=2 - [TCEA1_HUMAN]
P42166	Lamina-associated polypeptide 2, isoform alpha OS=Homo sapiens GN=TMPO PE=1 SV=2 - [LAP2A_HUMAN]
Q96C36	Pyrroline-5-carboxylate reductase 2 OS=Homo sapiens GN=PYCR2 PE=1 SV=1 - [P5CR2_HUMAN]
P62136	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit OS=Homo sapiens GN=PPP1CA PE=1 SV=1 - [PP1A_HUMAN]
Q1KMD3	Heterogeneous nuclear ribonucleoprotein U-like protein 2 OS=Homo sapiens GN=HNRNPUL2 PE=1 SV=1 - [HNRL2_HUMAN]
P11940	Polyadenylate-binding protein 1 OS=Homo sapiens GN=PABPC1 PE=1 SV=2 - [PABP1_HUMAN]
Q99873	Protein arginine N-methyltransferase 1 OS=Homo sapiens GN=PRMT1 PE=1 SV=2 - [ANM1_HUMAN]
Q13363	C-terminal-binding protein 1 OS=Homo sapiens GN=CTBP1 PE=1 SV=2 - [CTBP1_HUMAN]
O14497	AT-rich interactive domain-containing protein 1A OS=Homo sapiens GN=ARID1A PE=1 SV=3 - [ARI1A_HUMAN]
Q9BZK7	F-box-like/WD repeat-containing protein TBL1XR1 OS=Homo sapiens GN=TBL1XR1 PE=1 SV=1 - [TBL1R_HUMAN]
P00568	Adenylate kinase isoenzyme 1 OS=Homo sapiens GN=AK1 PE=1 SV=3 - [KAD1_HUMAN]
Q13765	Nascent polypeptide-associated complex subunit alpha OS=Homo sapiens GN=NACA PE=1 SV=1 - [NACA_HUMAN]

















P23771	Trans-acting T-cell-specific transcription factor GATA-3 OS=Homo sapiens GN=GATA3 PE=1 SV=1 - [GATA3_HUMAN]
P27635	60S ribosomal protein L10 OS=Homo sapiens GN=RPL10 PE=1 SV=4 - [RL10_HUMAN]
Q92922	SWI/SNF complex subunit SMARCC1 OS=Homo sapiens GN=SMARCC1 PE=1 SV=3 - [SMRC1_HUMAN]
O15347	High mobility group protein B3 OS=Homo sapiens GN=HMGB3 PE=1 SV=4 - [HMGB3_HUMAN]
P51531	Probable global transcription activator SNF2L2 OS=Homo sapiens GN=SMARCA2 PE=1 SV=2 - [SMCA2_HUMAN]
Q13310	Polyadenylate-binding protein 4 OS=Homo sapiens GN=PABPC4 PE=1 SV=1 - [PABP4_HUMAN]
Q9UHF7	Zinc finger transcription factor Trps1 OS=Homo sapiens GN=TRPS1 PE=1 SV=2 - [TRPS1_HUMAN]
P51532	Transcription activator BRG1 OS=Homo sapiens GN=SMARCA4 PE=1 SV=2 - [SMCA4_HUMAN]
Q99829	Copine-1 OS=Homo sapiens GN=CPNE1 PE=1 SV=1 - [CPNE1_HUMAN]
Q9BVJ7	Dual specificity protein phosphatase 23 OS=Homo sapiens GN=DUSP23 PE=1 SV=1 - [DUS23_HUMAN]
Q16543	Hsp90 co-chaperone Cdc37 OS=Homo sapiens GN=CDC37 PE=1 SV=1 - [CDC37_HUMAN]
O43447	Peptidyl-prolyl cis-trans isomerase H OS=Homo sapiens GN=PPIH PE=1 SV=1 - [PPIH_HUMAN]
P61081	NEDD8-conjugating enzyme Ubc12 OS=Homo sapiens GN=UBE2M PE=1 SV=1 - [UBC12_HUMAN]
P18615	Negative elongation factor E OS=Homo sapiens GN=NELFE PE=1 SV=3 - [NELFE_HUMAN]
Q01105	Protein SET OS=Homo sapiens GN=SET PE=1 SV=3 - [SET_HUMAN]
O60884	DnaJ homolog subfamily A member 2 OS=Homo sapiens GN=DNAJA2 PE=1 SV=1 - [DNJA2_HUMAN]
P05388	60S acidic ribosomal protein P0 OS=Homo sapiens GN=RPLP0 PE=1 SV=1 - [RLA0_HUMAN]
Q8WWM7	Ataxin-2-like protein OS=Homo sapiens GN=ATXN2L PE=1 SV=2 - [ATX2L_HUMAN]
O94776	Metastasis-associated protein MTA2 OS=Homo sapiens GN=MTA2 PE=1 SV=1 - [MTA2_HUMAN]
P63244	Receptor of activated protein C kinase 1 OS=Homo sapiens GN=RACK1 PE=1 SV=3 - [RACK1_HUMAN]
Q9NTK5	Obg-like ATPase 1 OS=Homo sapiens GN=OLA1 PE=1 SV=2 - [OLA1_HUMAN]
P55317	Hepatocyte nuclear factor 3-alpha OS=Homo sapiens GN=FOXA1 PE=1 SV=2 - [FOXA1_HUMAN]
O95861	3'(2'),5'-bisphosphate nucleotidase 1 OS=Homo sapiens GN=BPNT1 PE=1 SV=1 - [BPNT1_HUMAN]
Q8N684	Cleavage and polyadenylation specificity factor subunit 7 OS=Homo sapiens GN=CPSF7 PE=1 SV=1 - [CPSF7_HUMAN]
Q12824	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1 OS=Homo sapiens GN=SMARCB1 PE=1 SV=2 - [SNF5_HUMAN]
Q99623	Prohibitin-2 OS=Homo sapiens GN=PHB2 PE=1 SV=2 - [PHB2_HUMAN]











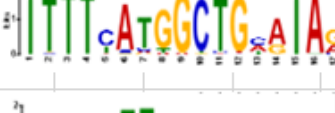

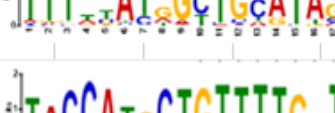


Q9Y266	Nuclear migration protein nudC OS=Homo sapiens GN=NUDC PE=1 SV=1 - [NUDC_HUMAN]
O95232	Luc7-like protein 3 OS=Homo sapiens GN=LUC7L3 PE=1 SV=2 - [LC7L3_HUMAN]
Q9UHD8	Septin-9 OS=Homo sapiens GN=SEPT9 PE=1 SV=2 - [SEPT9_HUMAN]
P49750	YLP motif-containing protein 1 OS=Homo sapiens GN=YLPM1 PE=1 SV=3 - [YLPM1_HUMAN]
Q9Y678	Coatomer subunit gamma-1 OS=Homo sapiens GN=COPG1 PE=1 SV=1 - [COPG1_HUMAN]
Q96JK9	Mastermind-like protein 3 OS=Homo sapiens GN=MAML3 PE=1 SV=4 - [MAML3_HUMAN]
O96013	Serine/threonine-protein kinase PAK 4 OS=Homo sapiens GN=PAK4 PE=1 SV=1 - [PAK4_HUMAN]
Q96FW1	Ubiquitin thioesterase OTUB1 OS=Homo sapiens GN=OTUB1 PE=1 SV=2 - [OTUB1_HUMAN]
P27694	Replication protein A 70 kDa DNA-binding subunit OS=Homo sapiens GN=RPA1 PE=1 SV=2 - [RFA1_HUMAN]
P78347	General transcription factor II-I OS=Homo sapiens GN=GTF2I PE=1 SV=2 - [GTF2I_HUMAN]
P98175	RNA-binding protein 10 OS=Homo sapiens GN=RBM10 PE=1 SV=3 - [RBM10_HUMAN]
Q92879	CUGBP Elav-like family member 1 OS=Homo sapiens GN=CELF1 PE=1 SV=2 - [CELF1_HUMAN]
O43681	ATPase ASNA1 OS=Homo sapiens GN=ASNA1 PE=1 SV=2 - [ASNA_HUMAN]
P29401	Transketolase OS=Homo sapiens GN=TKT PE=1 SV=3 - [TKT_HUMAN]
Q8NFD5	AT-rich interactive domain-containing protein 1B OS=Homo sapiens GN=ARID1B PE=1 SV=2 - [ARI1B_HUMAN]
O60716	Catenin delta-1 OS=Homo sapiens GN=CTNND1 PE=1 SV=1 - [CTND1_HUMAN]
Q96T60	Bifunctional polynucleotide phosphatase/kinase OS=Homo sapiens GN=PNKP PE=1 SV=1 - [PNKP_HUMAN]
P18583	Protein SON OS=Homo sapiens GN=SON PE=1 SV=4 - [SON_HUMAN]
P05455	Lupus La protein OS=Homo sapiens GN=SSB PE=1 SV=2 - [LA_HUMAN]
O60264	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 OS=Homo sapiens GN=SMARCA5 PE=1 SV=1 - [SMCA5_HUMAN]
Q13283	Ras GTPase-activating protein-binding protein 1 OS=Homo sapiens GN=G3BP1 PE=1 SV=1 - [G3BP1_HUMAN]
Q9UNE7	E3 ubiquitin-protein ligase CHIP OS=Homo sapiens GN=STUB1 PE=1 SV=2 - [CHIP_HUMAN]
Q14839	Chromodomain-helicase-DNA-binding protein 4 OS=Homo sapiens GN=CHD4 PE=1 SV=2 - [CHD4_HUMAN]
P56545	C-terminal-binding protein 2 OS=Homo sapiens GN=CTBP2 PE=1 SV=1 - [CTBP2_HUMAN]
O43719	HIV Tat-specific factor 1 OS=Homo sapiens GN=HTATSF1 PE=1 SV=1 - [HTSF1_HUMAN]
Q96I25	Splicing factor 45 OS=Homo sapiens GN=RBM17 PE=1 SV=1 - [SPF45_HUMAN]
Q16539	Mitogen-activated protein kinase 14 OS=Homo sapiens GN=MAPK14 PE=1 SV=3 - [MK14_HUMAN]
P22314	Ubiquitin-like modifier-activating enzyme 1 OS=Homo sapiens GN=UBA1 PE=1 SV=3 - [UBA1_HUMAN]



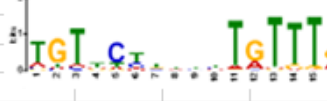
















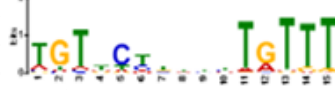












Q03252	Lamin-B2 OS=Homo sapiens GN=LMNB2 PE=1 SV=4 - [LMNB2_HUMAN]
Q05048	Cleavage stimulation factor subunit 1 OS=Homo sapiens GN=CSTF1 PE=1 SV=1 - [CSTF1_HUMAN]
P53992	Protein transport protein Sec24C OS=Homo sapiens GN=SEC24C PE=1 SV=3 - [SC24C_HUMAN]
Q9BZJ0	Crooked neck-like protein 1 OS=Homo sapiens GN=CRNKL1 PE=1 SV=4 - [CRNL1_HUMAN]
Q8IZ83	Aldehyde dehydrogenase family 16 member A1 OS=Homo sapiens GN=ALDH16A1 PE=1 SV=2 - [A16A1_HUMAN]
Q6IN85	Serine/threonine-protein phosphatase 4 regulatory subunit 3A OS=Homo sapiens GN=PPP4R3A PE=1 SV=1 - [P4R3A_HUMAN]
Q8WYA6	Beta-catenin-like protein 1 OS=Homo sapiens GN=CTNBL1 PE=1 SV=1 - [CTBL1_HUMAN]
Q13523	Serine/threonine-protein kinase PRP4 homolog OS=Homo sapiens GN=PRPF4B PE=1 SV=3 - [PRP4B_HUMAN]
Q8NE71	ATP-binding cassette sub-family F member 1 OS=Homo sapiens GN=ABCF1 PE=1 SV=2 - [ABCF1_HUMAN]
O94906	Pre-mRNA-processing factor 6 OS=Homo sapiens GN=PRPF6 PE=1 SV=1 - [PRP6_HUMAN]
Q5M775	Cytospin-B OS=Homo sapiens GN=SPECC1 PE=1 SV=1 - [CYTSB_HUMAN]
Q14683	Structural maintenance of chromosomes protein 1A OS=Homo sapiens GN=SMC1A PE=1 SV=2 - [SMC1A_HUMAN]
Q96QC0	Serine/threonine-protein phosphatase 1 regulatory subunit 10 OS=Homo sapiens GN=PPP1R10 PE=1 SV=1 - [PP1RA_HUMAN]
Q00341	Vigilin OS=Homo sapiens GN=HDLBP PE=1 SV=2 - [VIGLN_HUMAN]
Q9H2P0	Activity-dependent neuroprotector homeobox protein OS=Homo sapiens GN=ADNP PE=1 SV=1 - [ADNP_HUMAN]
P51610	Host cell factor 1 OS=Homo sapiens GN=HCFC1 PE=1 SV=2 - [HCFC1_HUMAN]
















Extended data 2: Motifs enriched upon AR agonist stimulation in different breast cancer models



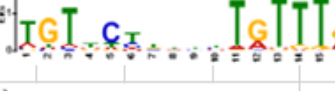










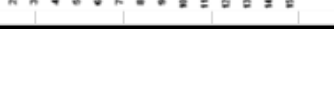

ZR-75.1		
ID	Motif logo	E-value
GCR		4.79E-50
FOXA1		3.83E-47
PRGR		5.81E-45
FOXA2		2.35E-42
FOXM1		6.54E-37
ANDR		1.06E-33
FOXJ2		1.12E-28
FOXA3		2.69E-26
FO XK1		5.88E-18
GATA3		1.78E-17
GATA6		1.36E-15
FOXP2		2.45E-14
FOXC1		1.37E-13
SRY		1.30E-10
SOX5		7.22E-08
FOXO4		1.09E-07

T-47D		
ID	Motif logo	E-value
GCR		5.60E-65
PRGR		2.94E-56
ANDR		2.24E-43
GATA4		1.78E-17
GATA1		4.79E-09
GATA2		5.30E-09
GATA6		1.86E-07
GATA3		6.24E-07
LEF1		6.07E-03
LYL1		2.42E-02
TAL1		3.81E-02
OZF		5.99E-02
SOX2		6.69E-02
ZN260		4.28E-01
ZN250		5.96E-01

MFM-223		
ID	Motif logo	E-value
GCR		7.41e-312
PRGR		2.48E-276
ANDR		1.54E-78
GATA3		1.86E-70
GATA1		3.17E-53
LYL1		1.54E-22
RXRA		3.05E-21
RARG		6.87E-21
NR1D1		2.59E-20
FEV		7.85-18
RORA		1.36E-15
ZN341		1.99E-15
SMAD3		1.32E-14
ESR2		2.00E-14
ITF2		1.39E-13

MDA-MB-453		
ID	Motif logo	E-value
GCR		8.51E-119
PRGR		1.16E-95
ANDR		3.46E-50
GATA4		2.54E-10
GATA3		4.79E-09
GATA2		1.72E-08
AP2C		8.96E-07
AP2A		1.25E-06
LYL1		8.43E-06
RARG		1.71E-05
ZN322		3.86E-05
SOX4		5.85E-04
ZBT48		6.57E-04
FOXA3		8.47E-04
FEV		9.40E-04

GAR15-13		
ID	Motif logo	E-value
GCR		2.39E-317
PRGR		2.29E-282
ANDR		6.65E-59
GATA3		1.20E-19
LYL1		9.23E-18
GATA6		4.63E-13
RORA		5.13E-17
NR1D1		3.35E-16
RARG		2.17E-14
RXRA		3.34E-13
SMAD3		4.06E-12
CLOCK		8.62E-11
ZN322		1.93E-10
TGIF1		1.27E-09
STF1		3.76E-09
NR4A2		5.83E-09

HCI-005		
ID	Motif logo	E-value
GCR		7.13E-113
PRGR		1.07E-104
ANDR		2.04E-63
GATA4		6.34E-21
GATA6		5.67E-17
GATA3		1.05E-17
ELF3		6.01E-03
SMCA1		5.18E-02
ZN260		1.42E-01
SOX4		1.45E-01
SOX9		1.96E-01
SRY		6.06E-01
SOX2		8.52E-01
SOX3		1.45E+00
EHF		3.05E+00

**Extended data 3: ChIP-seq data analysis information in different breast cancer models**

<b>ZR-75-1 (ER+ BC cell line): GATA3 ChIP-seq</b>				
<b>Treatment</b>	<b>Replicate</b>	<b>Mapped reads (post-filter)</b>	<b>Peaks</b>	<b>Consensus peaks</b>
Vehicle	1	23,585,902	53,568	48,807
	2	22,801,542	44,043	
	3	17,348,458	57,348	
E2	1	25,500,639	56,569	47,290
	2	25,168,979	39,456	
	3	22,833,295	52,207	
DHT	1	22,986,082	56,152	58,037
	2	24,189,405	69,127	
	3	24,960,138	57,024	
E2+DHT	1	18,301,158	65,496	48,864
	2	19,463,879	39,674	
	3	22,528,153	50,768	
Pooled input		22,362,699		

<b>T-47D (ER+ BC cell line): GATA3 ChIP-seq</b>				
<b>Treatment</b>	<b>Replicate</b>	<b>Mapped reads (post-filter)</b>	<b>Peaks</b>	<b>Consensus peaks</b>
Vehicle	1	24,571,905	92,192	77,109
	2	25,795,187	83,858	
	3	26,279,645	68,720	
E2	1	25,088,842	83,590	81,964
	2	25,819,969	84,440	
	3	26,890,871	90,547	
DHT	1	29,135,527	71,311	81,567
	2	26,353,541	86,591	
	3	24,936,265	100,000	
E2+DHT	1	26,883,074	77,791	79,162
	2	26,185,310	83,174	
	3	25,931,067	87,711	
Pooled input		25,761,594		



MFM-223 (ER- BC cell line): GATA3 ChIP-seq				
Treatment	Replicate	Mapped reads (post-filter)	Peaks	Consensus peaks
Vehicle	1	15,535,070	50,592	54,092
	2	17,141,923	63,258	
	3	17,964,809	57,463	
DHT	1	18,091,254	57,426	56,671
	2	18,614,483	51,920	
	3	17,831,070	72,231	
Pooled input		17,863,999		

MDA-MB-453 (ER- BC cell line): GATA3 ChIP-seq				
Treatment	Replicate	Mapped reads (post-filter)	Peaks	Consensus peaks
Vehicle	1	19,384,810	28,969	25,810
	2	19,346,890	26,015	
	3	19,667,556	23,881	
DHT	1	19,364,625	34,665	29,842
	2	18,930,930	31,747	
	3	20,061,070	25,580	
Pooled input		4,539,847		

GAR15-13 (ER+ BC PDX model): GATA3 ChIP-seq				
Treatment	Tumour	Mapped reads (post-filter)	Peaks	Consensus peaks
Vehicle	1	17,254,744	91,976	97,817
	2	18,436,329	103,653	
	3	19,221,721	109,041	
SARM	1	15,025,034	122,410	110,487
	2	17,492,949	101,687	
	3	20,107,820	118,853	
Pooled input		26,127,246		

HCI-005 (ER+ BC PDX model): GATA3 ChIP-seq				
Treatment	Tumour	Mapped reads (post-filter)	Peaks	Consensus peaks
Vehicle	1	17,325,890	14,050	15,651
	2	20,048,228	13,666	
	3	16,686,047	26,630	
DHT	1	17,748,150	8,837	7,173
	2	19,208,412	5,916	
	3	18,957,335	9,559	
Pooled input		17,745,689		

MFM-223 (ER- BC cell line): AR ChIP-seq				
Treatment	Replicate	Mapped reads (post-filter)	Peaks	Consensus peaks
Vehicle	1	18,864,107	159	168
	2	18,467,549	113	
	3	19,087,134	418	
DHT	1	20,187,809	8,951	14,493
	2	18,428,943	15,512	
	3	16,832,755	20,224	
Pooled input		17,863,999		

MDA-MB-453 (ER- BC cell line): AR ChIP-seq				
Treatment	Replicate	Mapped reads (post-filter)	Peaks	Consensus peaks
Vehicle	1	20,889,107	67	30
	2	19,178,841	33	
	3	19,583,539	20	
DHT	1	19,790,092	15,612	18,778
	2	20,112,325	17,942	
	3	19,410,899	26,162	
Pooled input		4,539,847		

MFM-223 (ER- BC cell line): H3K27ac ChIP-seq				
Treatment	Replicate	Mapped reads (post-filter)	Peaks	Consensus peaks
Vehicle	1	14,529,360	46,803	47,555
	2	15,977,139	49,413	
	3	16,805,995	47,198	
DHT	1	18,316,621	43,723	45,633
	2	15,839,748	48,170	
	3	14,144,462	46,154	
Pooled input		17,863,999		

MDA-MB-453 (ER- BC cell line): H3K27ac ChIP-seq				
Treatment	Replicate	Mapped reads (post-filter)	Peaks	Consensus peaks
Vehicle	1	13,791,813	28,969	25,810
	2	15,630,653	26,015	
	3	16,390,426	23,881	
DHT	1	13,703,643	34,665	29,842
	2	16,288,903	31,747	
	3	16,321,087	25,580	
Pooled input		4,539,847		

Extended data 4: List of Luminal and Basal genes

Differentially expressed genes in Luminal vs Basal adult cells															
KRT18	SHANK2	PLAC8	KLK10	KIT	GDE1	CLCN3	FOXA1	CASZ1	LMNTD2	TBX3	AGR2	CEBPA			
ESR1	FGB	BCL2L14	R3HDM1	ABCC8	MUC15	LALBA	NRAS	CCDC92	LNX2	TGM2	LLGL2	IRX2			
STC2	PIEZO2	AQP5	MSX2	PDZK1IP1	FCGBP	NUP210	GRB10	DNAAF3	MEIS1	TMCO3	CDH1	BARX2			
HP	TSPAN1	PTN	NUPR1	CMTM8	LTK	PLB1	GRB14	DRC3	MEIS3	TMPPRS6	CKMT1B	DLX3			
CITED1	SLC5A1	TNFAIP2	ATP1B1	BTC	WAP	NCALD	INHBA	DUSP10	MINDY1	TNFSF11	CDH3	GATA4			
RASEF	MBOAT1	CSN1S2A	TMPPRS13	TMEM30B	OGFRL1	ARHGEF38	GSK3B	EDEM1	MYB	TOX3	CRIP1	GATA3			
RHOV	PTPN18	THSD4	CAPN8	MANBA	AIFM3	HMGCS2	CHD3	EEF1A2	NSD3	TP53INP2	SPINT1	HLF			
PRLR	SMIM22	2210407C18RIK	CRYM	MYO5C	SYNE4	RSP01	PLP1	EPS8L1	PAK4	TSPAN13	SH3YL1	GSC			
FOXA1	ZFP750	ESRRB	ANO3	BIK	MUC1	FAM174B	FOS	ERN1	PGAP6	TUBG1	PRSS8	SOX2			
LY6D	MPLH	WFDC3	EHF	DNAIC6	SLC8B1	TRIM6	MAFF	FBXO36	PHKA1	VOPP1	GRHL2	ZIC3			
MUC4	DKKL1	TMPPRS4	ATP6V1B1	LRRC8B	DNAJC12	ADRA1A	PPARA	FGF13	PLEKHG3	VPS33B	FXYD3	ARX			
OCLN	TMPPRS2	RAB11FIP1	ERBB3	LGALS3	RAB17	REG3G	CYP24A1	FGL1	PON3	WNK4	RAB25	FOXP3			
ARG1	PODXL	SPTBN2	CLDN6	CEACAM1	IGFALS	ALDH1A3	CSN1S2B	FLVCR2	PROM2	WNT5A	INHBB	MESP2			
2010300C02RIK	TPH1	BSPRY	SPNS2	UNC13B	FOLR1	SLC45A3	CXCL15	FYCO1	PSD4	WNT7B	CRABP2	HESX1			
SLC28A3	CSN3	TTC39A	ITIH2	CECR2	CADM4	SRCIN1	ACTA2	G6PD	PTPN6	YIPF6	IRF6	GAS7			
MARVELD2	CADPS2	TJP3	CLU	ENPP3	STRA6L	PLS1	ABCA7	GADD45G	PVALB	ZDHC1	IRX4	GATA5			
PKP2	NECTIN4	TMC5	CLDN7	CD55	MMP16	ABCA3	ACOT11	GALE	PXYLP1	ZFHX2	ASCL2	IRX5			
NAV3	ESM1	MUC20	AA986860	BDH1	FAAH	SPTLC3	ALCAM	GMPR	RABL3	ZSCAN18	XBP1	AQP3			
CLDN3	CEL	CAPN5	SLC22A18	TRF	UGT8A	2610528J11RIK	ALDH3B1	GPRC5C	REEP6	CEBPB	TLX1	PRICKLE2			
BGLAP3	SLC44A4	CEACAM10	SCNN1A	CLDN1	GLYCAM1	TMEM125	ALDH3B2	H4C3	SCMH1	STAT5A	PBX1	PARD6B			
CGN	RASSF6	TSPAN33	CLDN8	NXF7	HEPACAM2	PREX1	ANKMY2	H4C9	SGMS1	KRT7	SOX13	KDM4B			
BTN1A1	GIPC2	BASP1	TRP53INP1	IGSF5	CXCL15	UPK3A	AQP11	HDAC11	SLC16A5	LTF	DLX4				
SPP1	FAM25C	LIPH	KLHDC7A	FER1L4	S100A8	NPM1	ARFGEF3	HES6	SLC40A1	MFGE8	TSC22D3				
VILL	WFDC18	GDPD1	FAM234B	SLC46A3	CD14	SUMO1	ATP6V0E2	HID1	SLC7A2	CD44	ZNF3				
KRT19	SULF2	AREG	CD82	CYSTM1	GABRP	KRT8	BATF	HOXB2	SLC7A4	S100A14	VPS72				
GCA	CTSH	SLC1A1	ALDOC	CSN2	AGTR1A	E-CAD	BBOF1	HSD11B2	SPDEF	PPP2R2C	EDF1				
ELF5	PPP1R9A	CSN1S1	RASGRF1	TMC4	TACSTD2	PGR	BTRC	IL13RA1	SPINK1	ICA1	E4F1				
CLIC6	DOCK8	FOX11	WNT4	UPK2	HHIPL2	FGFR2	C1ORF210	KBTBD4	SPRR1A	ST14	FOXD1				
SORT1	LCP1	GPD1	ACSBG1	RFTN1	CNTNAP2	KLF5	CACNB3	KLHL5	SULT2B1	AP1M2	MYNN				
F11R	TMEM56	CLDN4	MAL2	MMP15	KCNN4	ELF3	CACNG4	LAMA5	TANGO2	IGFBP2	ZXDC				

Differentially expressed genes in Basal vs Luminal adult cells																
WIF1	VCAN	CACNA1C	RTN1	NTNG2	AOC3	C1S1	ZEB2	MFAP5	MT2	NR1D2	ELK3	WBP5				
ANTXR1	TSHZ2	KRT17	WNT6	SPOCK2	CDH13	FAT2	FOS	MMP10	GPNUMB	AFF4	FOXA2	S100A4				
CPXM2	COL4A2	SLC2A4	ASB2	DLK2	CLMP	SEMA6A	EGR1	NDRG1	CAV1	ST18	NFE2	CTGF				
CACNA2D1	TRP63	FAM65B	SOBP	AATK	NGFR	EYA1	SMAD3	NSG1	CD74	KLF15	KLF12	COL6A2				
KRT5	NKD1	IRX4	CKB	SH3RF3	PPIC	HUNK	CD55	P3H2	CAR9	NFE2L2	HNF4G	MAP1B				
CDH3	FSTL1	COL7A1	MARK1	TRIL	COL5A2	ADAMTS2	ALDH1A3	PI3	PMP22	YBX1	NKX2-5	EMP1				
COL14A1	COL18A1	BZRAP1	SPON2	KRT14	FBN1	GM17019	CXCL14	PKP1	CYP2C55	GMEB1	FOXC2	LOXL2				
DLL1	LMO1	PCDH7	COL4A1	RELN	NIPAL4	TTC7B	CCND2	PTGES	MYL6A	MEF2A	TBX18	PLAU				
COL17A1	ACTG2	RNASE4	MSRB3	DKK3	NEBL	SERPINB10	AKR1C1	PTHLH	CALD1	CBL	MSC	BCAT1				
SLIT3	NEXN	TMEM200A	CXCL14	SYNPO2	KLHL29	HSPG2	AMIGO2	PTPRZ1	MBD1	ZNF358	LHX9	TFPI				
FAM129A	MYL9	ITGA9	EML1	GM21149	TPST1	KIRREL	CA12	RHCG	ZSCAN21	TAF1B	ETV4	CAV2				
AXL	CLIP3	RNF150	ISM1	MAPT	TRIM29	ADAMTS18	CLCA2	S100A2	CEBPB	TAF4B	PLAG1	SRGN				
COL5A1	ADAMTS20	ACTA2	NRG1	SERPINB11	ADAMTS1	FKBP10	CSTA	S100A7	FOXJ2	TRERF1	ZNF655	CSF1				
TGFB11	PLOD2	BGN	NECTIN3	LGALS1	SERPINF1	THEM5	DSG1	S100A8	KLF9	MAFG	SHOX2	EMP3				
ELOVL4	LAMA1	KCNMB1	GJA1	EPHB1	CNTN2	PTPRM	DST	S100A9	SNAPC2	PGBD1	PPARG	PHLDA1				
CHST15	KCNQ5	ID4	GAS1	DACT3	IGSF10	EOGT	FABP5	S100P	HSF2	HES7	FLI1	IGF2BP2				
TPM2	HS6ST2	RCN3	MEF2C	CACNA1G	PLA2G7	SPEER4D	FEZ1	SDC1	PER1	TWIST1	ETV5	PTRF				
POSTN	CAMK4	MRV1	JPH2	WNT10A	ASXL3	RGL1	FLRT3	SERPINE2	ZNF519	GATA6	ETV1	FN1				
HMCN1	TAGLN	HDGFRP3	HTR1D	GALNT16	KIF1A	ZFHX4	GJB5	SFN	TSHZ1	EID2B	GLIS3	IGFBP7				
IGFBP3	RAI2	TGFBR3	SEMA6D	ADCY7	AFF3	BMP7	GPX2	SLC39A6	ZNF238	ZNF775	LBH	MMP14				
VWA1	POPODC2	CHST11	FBLN1	NRG2	CPNE8	VWA2	HTRA1	SLPI	ZNF529	CEBPD	FOSL1	MICAL2				
AEBP1	NPTX1	1500015O10RIK	PALLD	BEGAIN	GM21190	VCAM1	IL1A	SPRR1A	FOXK1	TCF4	FOXF2	F2R				
MYLK	GFRA1	FMOD	NT5E	NCKAP5L	PKNOX2	ITGB4	IL1B	SPRR1B	ZNF318	ZNF175	ZEB1	F3				
ANGPTL2	LOXL3	SCN7A	HGF	SCN5A	P3H1	CAMK2D	ITGA6	TP63	MYC	ZNF697	ETS1	AKR1B1				
SDK2	LHFP	SPARC	DRP2	ABCB1B	PTCH2	MAP3K3	IVL	VIM	SIX5	E2F7	PTX3					
IGFBP2	MYH11	CTTNBP2	SOGA1	DSC3	PHYHIP	GSK3B	JAG1	CALLA	NR2C2	SMAD3	MSX1					
CNN1	TSPAN18	PSD2	ABAT	WNT11	EFCAB1	MATN2	KRT16	COL6A1	CREB5	FOXO1	MSX2.1					
JAG2	ZFP365	TENM2	TENM3	PALM	NKD2	NFATC2	KRT6	OLFML2B	MAFF	RELB	HMGNS5					
MOXD1	PDPN	DPYSL3	DCHS1	LRP1	FIX1	NTRK3	KRT75	SF3A3	EN1	ZBTB16	FOXO1					
SNAI2	GSTP1	LRRRC8	ST3GAL2	TCEAL3	MATN4	SOX2	LGALS7	HSPA4L	ZFP37	CREB3L1	RAB34					

## Extended data 5: List of siRNAs, antibodies, RT-PCT primers, and ChIP-PCR primers

### List of siRNAs:

	Scene	Anti-scene	Company
siAR-1	CAAGGGAGGUUACACAAA	UUUGGUGUAACCUCCUUG	Sigma
siAR-2	CGAGAGAGCUGCAUCAGUUUU	AACUGAUGCAGCUCUCUCUGUU	GenePharma
siGATA3-JC-1	AAACUAGGUCUGAUUUUCAUU	UGAAUAUCAGACCUAGUUUU	Sigma
siGATA3-JC-2	CUUUUUUGCAUCUGGGUAG	CUACCCAGAUGCAAUAAAG	Sigma
siGATA3-JC3-(exo-5)	GAACUGUCAGACCACCACAUU	UGUGGUGGUCUGACAGUUCUU	Sigma
siGATA3-JC4-(exo-6)	GACCAGAAACCGAAAAUGUU	CAUUUUUCGGUUUCUGGUCUU	Sigma
siControl	Cat No: 1027281	Lot No: 284526154	Allstar-Neg. Control siRNA

### List of antibodies that have been used in each specific experiment:

	RIME			
	Company	Cat numbers	Spices	Volume
AR	Santa Cruz Biotechnology	sc-816	Rabbit	10ug / experiment

	PLA / IF			
	Company	Cat numbers	Spices	Dilluton factor
AR	LS-Bio	LS-B3326	Rabbit	1 in 400
GATA3	Invitrogen	1A12-1D9	Mouse	1 in 100
Alexa Fluor-568 Goat-anti-Rabbit IgG	life technology	A11036	Goat	1 in 400
Alexa Fluor-488 Goat-anti-Mouse IgG	life technology	A11029	Goat	1 in 400

	ChIP-seq/ ChIP-PCR/Co-IP			
	Company	Cat numbers	Spices	Volume
AR	Abcam	ab108341	Rabbit	5ug / IP
GATA3	Abcam	ab199428	Rabbit	5ug / IP
H3K27AC	Abcam	ab4729	Rabbit	2ug / IP

	Western blotting			
	Company	Cat numbers	Spices	Dilluton factor
AR	Santa Cruz Biotechnology	sc-816	Rabbit	1 in 1000
GATA3	Abcam	ab199428	Rabbit	1 in 500
ER	Santa Cruz Biotechnology	sc-8002	Mouse	1 in 1000
B-actin	Abcam	ab6376	Mouse	1 in 2000
GAPDH	Merck Millipore	mab374	Mouse	1 in 2000
Polyclonal goat-anti-Rabbit / HRP	Dako	PO448	—	1 in 2000
Polyclonal goat-anti-Mouse	Dako	PO161	—	1 in 2000
Veriblot	Abcam	ab131367	—	1 in 1000

	IHC			
	Company	Cat numbers	spices	Dilluton factor
AR	Santa Cruz Biotechnology	sc-816	Rabbit	1 in 1000
GATA3	Abcam	ab199428	Rabbit	1 in 500

**List of CHIP-PCR primers:**

	<b>Forward</b>	<b>Reverse</b>	<b>Melting TM</b>	<b>Company</b>
<b>SEC14L2</b>	GGCCAAACTGCAGAACCAAA	TCCTGCCTAGGTCTCTGTG	55	Sigma
<b>C1orf116</b>	ATCACAACCACCGCTACGAT	AATGTGTGCAACAACAGAGGC	55	Sigma
<b>ZBTB16</b>	CCCTGCGTCTGTACTCATTG	CTGCACGCCTGCTTACATA	55	Sigma
<b>c-FOC</b>	AGCCGCTTCCCTGTTACTA	TTCCCTTCTCCACTCTTCGG	60	Sigma
<b>FKBP5</b>	GCTCTGACTTATTGTTCTTACTGCC	TTGCTGTCAGCACATCGAGTTCA	55	Sigma
<b>EHF</b>	AGGCACTGGGGCCATTTATT	TGCATACAGGGCTTCGTTCA	55	Sigma
<b>AQP3</b>	GAGCCATTTACCACCGTTGC	CGACCTCCAAGCACCTATT	55	Sigma
<b>KDM4B</b>	CACTGTGGCTTCAGACCTTC	TCGCAAAGACTCTACCAAAGA	55	Sigma
<b>NC2</b>	GTGAGTGCCAGTTAGAGCATCTA	GGAACCAGTGGGTCTTGAAGTG	55	Sigma

**List of RT-PCR primers:**

	<b>Forward</b>	<b>Reverse</b>	<b>Melting TM</b>	<b>Company</b>
<b>SEC14L2</b>	GCCGAATCCAGATGACTATTTTCT	GATGTTGTCAATGTCCTTTTGCTT	55	Sigma
<b>C1orf116</b>	CATCTCTGCCCTTTGAAACAAAA	GGGCATCACCCGAAACAAG	55	Sigma
<b>ZBTB16</b>	GAGATCCTCTTCCACCGCAAT	CCGCATACAGCAGGTCATC	55	Sigma
<b>EHF</b>	ACAACCACCAGTCACCTTCC	CCATCCACACGCTCCAGAAT	60	Sigma
<b>AQP3</b>	AGATGCTCCACATCCGCTAC	AGGGAGGCTGTGCCTATGAA	55	Sigma
<b>KDM4B</b>	ACTTCAACAAATACGTGGCCTAC	CGATGTCATCATACGTCTGCC	55	Sigma
<b>GAPDH</b>	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG	55	Sigma

## **Chapter-3**

# **GATA3 facilitates AR tumour-suppressive function in ER+ breast cancer**

The following chapter includes a manuscript formatted for submission to *Cancer Research* journal, followed by supplementary figures. A general discussion of this chapter had been included in Chapter 4.



## Statement of Authorship

Title of Paper	GATA3 facilitates AR tumour-suppressive function in ER+ breast cancer
Publication Status	<input type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input checked="" type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	This chapter is formatted for submission to Cancer Research journal

### Principal Author

Name of Principal Author (Candidate)	Leila Hosseinzadeh
Contribution to the Paper	Performed experiments (Cell culture, qPCR, Western blotting, ChIP-seq and ChIP-PCR, siRNA transfection, generating tagged cells, cell growth assays through IncuCyte), Data mining and analysing, data interpretation, Experimental design, Manuscript writing, drafting, editing, and figure preparation
Overall percentage (%)	75%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 03.03.2021

### Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Amy R. Dwyer
Contribution to the Paper	Technical assistance, supervision, data interpretation, editing manuscript
Signature	Date 03/03/21

Name of Co-Author	Geraldine Laven-Law
Contribution to the Paper	Technical assistance, bioinformatic analysis
Signature	Date 3/3/21

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Zoya Kikhtyak
Contribution to the Paper	Technical assistance
Signature	Date 4/3/2021

Name of Co-Author	Wayne D. Tilley
Contribution to the Paper	Funding, supervision, and data interpretation
Signature	Date 05/03/2021

Name of Co-Author	Theresa E. Hickey
Contribution to the Paper	Funding, supervision, data interpretation, study design and editing manuscript
Signature	Date 3/3/21

## **Chapter-3: GATA3 facilitates AR tumour-suppressive function in ER+ breast cancer**

Leila Hosseinzadeh<sup>1</sup>, Amy R. Dwyer<sup>1</sup>, Geraldine Laven-Law<sup>1</sup>, Zoya Kikhtyak<sup>1</sup>, Wayne D. Tilley<sup>1</sup>, Theresa E. Hickey<sup>1</sup>.

1 Dame Roma Mitchel Cancer Research Laboratories, Adelaide Medical School, University of Adelaide, Adelaide, Australia

## Abstract

**Background:** The GATA3 transcription factor is an established regulator of estrogen receptor-alpha (ER) signalling in ER+ breast cancer and plays important roles in development and differentiation of the mammary gland. Most ER+ breast cancers also express the androgen receptor (AR), which has been shown to inhibit ER-driven growth of this disease subtype. Recently, GATA3 has been identified as a novel AR interacting protein in ER+ breast cancer cells. Therefore, we aimed to investigate how GATA3 influences AR signalling in ER+ breast cancer and determine whether it is important for AR-mediated growth inhibition.

**Results:** In order to assess the effect of hormones in shaping the GATA3 cistrome in ER+ breast cancer context, we investigated the GATA3 cistrome in vehicle, E2, DHT, and E2+DHT hormone treatment conditions in two ER+ breast cancer cell lines (T-47D and ZR-75-1). We found that both estrogen and androgen treatment rearranged the GATA3 cistrome in these models. Pathway analysis showed that E2-induced GATA3 binding events mediates ER signalling to induce growth and cell cycle progression in both models. However, AR-activation hugely impacts the E2-induced GATA3 binding profile. Importantly, we identified that androgen-induced GATA3 cistrome involved in AR-mediated ER reprogramming and anti-tumor function of AR in ER+ breast cancer cells.

**Conclusions:** Our findings show that the GATA3 cistrome is regulated by estrogen and androgen hormones in ER+ breast cancer cells. Also, our data identified that androgen induction (in the background of estrogens) competitively take GATA3 away from the ER target genes to a subset of newly gained androgen-induced sites involved in AR growth-inhibitory function.

## Introduction

Breast cancer is the most common malignancy among women worldwide (Bray et al. 2018). The majority of breast cancers (>80 %) express the estrogen receptor-alpha (ER), the main nuclear transcription factor that mediates the oncogenic effects of estrogens (Perou et al. 2000). ER+ tumours are the most differentiated breast cancer subtypes, with a better outcome in comparison to the remaining 20-25 % of breast cancer subtypes, collectively referred to as ER- tumours (Fragomeni, Sciallis & Jeruss 2018; Reis-Filho & Pusztai 2011). The dependency of breast tumours on ER to sustain their growth underpins the use of systemic endocrine treatments that target ER to inhibit its activity (Castrellon 2017). Therapeutic agents that either inhibit estrogen production (aromatase inhibitors; AIs) (Thürlimann et al. 2005), block the binding of estrogen to the ER $\alpha$  (selective estrogen receptor modulators (SERMs, e.g. tamoxifen) (Fisher et al. 2005), or degrade ER (e.g. Fulvestrant) (Howell 2005; Vergote & Abram 2006) are used as adjuvant therapy for ER+ breast cancers and have resulted in significant improvement in breast cancer patient survival rates (Howell et al. 2004). Despite this, development of resistance is a major clinical issue and patients can relapse months or years after the initial treatment. Resistance to adjuvant endocrine therapy occurs in almost 10-15 % of ER+ breast cancer patients with early stage disease within 5 years (Dowsett et al. 2010), 30 % within 15 years (EBCTCG 2005), and ultimately most patients with advanced or metastatic ER+ breast cancers become resistant to endocrine treatments within 2-3 years of commencing endocrine therapy (Dixon 2014). Therefore, there is an urgent need to discover new therapeutic strategies to treat ER+ breast cancer to circumvent the emergence of resistance to current endocrine treatments.

Most (>90 %) ER+ breast cancers also express the androgen receptor (AR) (Peters et al. 2009). AR has been introduced as an independent predictor of breast cancer survival (reviewed at (Cops et al. 2008; Hickey et al. 2012)) with a favourable prognostic indication (reviewed at (Hickey et al. 2012; Ricciardelli et al. 2018)), which is significantly associated with reduced risk of relapse, longer survival (Yu et al. 2011) and better outcomes among patients with ER+ tumours (Park et al. 2011). It has been very recently shown that ligand activated AR acts as a tumour suppressor in ER+ breast cancer at all stages of disease including endocrine-resistant tumours, which strongly supports an AR agonist strategy for treatment (Hickey et al. 2021). AR suppresses endogenous ER activity in ER+ breast cancers by impeding the recruitment of ER or its co-activators, SRC-3 and p300, to chromatin, therefore suppressing the expression of critical cell cycle associated ER target genes (Hickey et al. 2021). AR also up-regulates target genes that may contribute to growth inhibition, via sequestration of p300 and SRC-3 away from ER target genes (Hickey et al. 2021). The latter study focussed on p300 and SRC-3 as co-factors known to be important for ER and AR signalling, but other co-factors common to both receptors are likely involved in cross-talk between these pathways.

To elicit transcriptional activity, steroid receptors like ER and AR do not function alone, but within a large, multi-protein transcriptional complex. In breast cancer, much is known about the ER complex, but little about the AR complex. Among several co-regulatory proteins, one key transcription factor that has been positively implicated in mediating estrogen-induced ER signalling in breast cancer is GATA3, which belongs to the GATA family of transcription factors (Eeckhoute et al. 2007; Kong et al. 2011). GATA3 is required for morphogenesis of luminal epithelial cells in the normal mammary gland, and promotes differentiation (Asselin-Labat et al. 2007). GATA3 has been shown to function in

collaboration with the ER transcriptional complex at multiple loci of the genome to regulate ER activity (Takaku, Grimm & Wade 2015; Theodorou et al. 2013). Interestingly, we recently identified GATA3 as a novel AR interacting protein in breast cancer cells regardless of ER status, and showed that it facilitates AR binding and transcriptional activity at AR target genes, and co-operates with AR to up-regulate luminal lineage genes (Chapter-2). Therefore, GATA3 is an important co-regulator of ER and AR signalling in breast cancer. Herein, we aimed to specifically examine the role of GATA3 in mediating cross-talk between these two receptors.

## Results

### Estrogen stimulation of ER+ breast cancer cells rearranges the GATA3 cistrome

In order to generate the first comprehensive and reproducible GATA3 cistrome and to investigate the effect of estrogens on shaping the GATA3 binding profile in ER+ breast cancer context, we conducted GATA3 ChIP-seq experiments using two estrogen (E2)-dependent ER+ breast cancer cell lines (T-47D and ZR-75-1). T-47D and ZR-75-1 breast cancer cell lines were used in this study since they endogenously and exclusively express wild-type forms of ER and GATA3. Cells were treated with vehicle or E2 (10 nM) for 4 hours following 3 days of hormone deprivation to match conditions previously utilized to generate ER consensus cistromes (Hickey *et al.*, 2021). We generated consensus GATA3 cistromes for both cell lines (E2 vs vehicle) using three consecutive passages of cells representing three independent biological replicates, for reproducibility. Peaks identified in at least 2 of 3 replicates were included in the consensus cistromes (Supplementary 1-a-d). We discovered that estrogen stimulation reproducibly shifts the GATA3 cistrome towards a new subset of binding sites compared with vehicle-treated cells in both ER+ breast cancer cell models (Figure 1-a, Supplementary 1-e), which has never been reported previously. Principle component analysis (PCA) confirmed that estrogen treatment was the main variable in clustering the stimulated cells away from the controls in both cell line models (Figure 1-b, Supplementary 1-f). Integrating the GATA3 data with our publicly available E2-treated ER ChIP-seq data from Hickey *et al.*, 2021, revealed that more than 60 % of ER binding sites are shared with GATA3 cis-regulatory elements in T-47D cells (Figure 1-c). We also observed about 50 % overlap between ER and GATA3 cistromes for the ZR-75-1 cell line under estrogenic conditions (Supplementary 2-a). DiffBind analysis revealed



406 and 219 GATA3 peaks that were significantly ( $FDR \leq 0.05$ ) enriched upon estrogen treatment in T-47D and ZR-75-1 cells, respectively (Figure 1-d, Supplementary 2-b). Peak annotation analysis for E2-induced GATA3 binding sites revealed 7,872 and 4,148 potential gene targets in T-47D and ZR-75-1 cells, respectively. In order to determine which of these candidate gene targets are regulated by ER-GATA3 co-occupied loci, we treated T-47D and ZR-75-1 cells for 6 hours with vehicle or E2 (10 nM) after 3 days of hormone deprivation and performed RNA-seq for differentially expressed genes (DEG). Integration of our annotated E2-induced GATA3 gene targets with E2-induced DEG genes identified 2,507 (T-47D) and 1,415 (ZR-75-1) up-regulated targets (Figure 1-e, Supplementary 2-c). Additionally, 2,733 (T-47D) and 1,303 (ZR-75-1) down-regulated targets were identified through integration of the ER-GATA3 annotated genes with RNA-seq data (Figure 1-e, Supplementary 2-c). Several common up-regulated genes found between two *in vitro* models are well-established ER target genes in breast cancer cells including *PgR* and *GREB1* (Supplementary 2-d). Visual exploration of the E2-induced GATA3 peaks co-occupied by ER through the Integrative Genomics Viewer (IGV) is shown for *PgR* and *GREB1* classic ER target genes in T-47D cells (Figure 1-g). PGR has been shown to promote the progression of cell cycle and proliferation in breast cancer as a classic ER target (Dressing et al. 2014; Kariagina et al. 2013). However, it has been shown that activated PGR has anti-proliferative effects that inhibits E2-induced growth of ER+ breast cancer cell line xenografts and ER+ primary tumors explants (Mohammed et al. 2015). GREB1 is a critical regulator of estrogen dependent breast cancer growth ('Retraction: GREB1 functions as a growth promoter and is modulated by IL6/STAT3 in breast cancer' 2014; Cheng, Michalski & Kommagani 2018; Haines et al. 2020; Rae et al. 2005). Up-regulated ER-GATA3 targets were positively associated with a set of biological processes dominated by Gene Ontology (GO) gene sets

including cell cycle and developmental growth (Figure 1-f, Supplementary 2-e). Gene set enrichment analysis (GSEA) revealed that ER-GATA3 DEG targets (up and down-regulated targets) are associated with enriched Hallmarks for estrogen response early and late, MYC-targets V1 and V2, E2F targets and G2M checkpoint in T-47D cells (Figure 1-h). Similarly, enriched hallmarks were observed in ZR-75-1 cells with ER-GATA3 up-regulated targets (Supplementary 2-f). In order to assess the effect of GATA3 knock-down on the mRNA expression level of *PgR* and *GREB1* in the condition of E2-treated cells compared with the control cells, we conducted qPCR experiments in both T-47D and ZR-75-1 breast cancer models (Figure 1-i, Supplementary 2-g). We examined the efficacy of GATA3 knock-down at the protein level in both cell lines through western blotting before conducting qPCR (Supplementary 2-h). qPCR experiments show that silencing GATA3 did not affect the basal expression level of *PgR* and *GREB1* genes, however, it significantly reduced but did not abolish the ability of E2 to increase mRNA transcript levels of classic ER target genes. Taken together, our results indicate that GATA3 enhances but is not required for ER signalling in these ER+ breast cancer models.

### **Androgen stimulation of ER+ breast cancer cells also rearranges the GATA3 cistrome**

Conducting GATA3 CHIP-seq in DHT- versus vehicle-treated ZR-75-1 and T-47D cells (Supplementary 1-a-d), we have previously shown that androgens rearrange the GATA3 binding profile in ER+ breast cancer cells (Chapter-2) (Figure 2-a, Supplementary 3-a). DHT-treatment increased the size of the GATA3 cistrome in both models (Chapter-2) (Figure 2-a, Supplementary 3-a). Principle component analysis (PCA) confirmed that DHT treatment is as the main variable in clustering the stimulated cells away from the vehicle-treated cell in both *in vitro* models (Chapter-2) (Figure 2-b, Supplementary 3-b). Therefore, for the first

time, our data discovered the effect of hormone treatment (estrogens (Chapter-3) and androgens (Chapter-2)) on shaping the binding profile of GATA3 transcription factor in ER+ breast cancer cells, indicating that GATA3 cistrome is hormone regulated in ER+ breast cancer cells. Additionally, AR has been shown to reprogram the ER binding profile in ER+ breast cancer models, which is associated with inhibition of tumor growth. Therefore, we sought to investigate the effect of AR activation on the E2-induced GATA3 cistrome. We studied the E2+DHT GATA3 cistrome compared with E2 GATA3 cistrome (Supplementary 1-a-d). Although the GATA3 cistromes under E2 and E2+DHT conditions were comparable in size (Figure 2-c), dual-receptor activation with E2+DHT treatment clearly shifted GATA3 binding events in T-47D (Figure 2-c) and ZR-75-1 cells (Supplementary 3-c). Principle component analysis (PCA) showed that DHT treatment was the major source of variation between replicates of GATA3 ChIP-seq experiments in both ER+ breast cancer models, indicating a strong impact of AR activation on the vehicle- and E2-treated GATA3 cistromes in breast cancer cells (Figure 2-d, Supplementary 3-d). Interestingly, discriminative motif analysis shows that androgen response elements (AREs) were enriched in DHT-induced GATA3 binding sites, however; estrogen response elements (EREs) were lost upon androgen stimulation in both cell line models (Figure 2-e, Supplementary 3-e), suggesting that AR activation redistributed GATA3 from ER binding sites to AR binding sites. DiffBind analysis identified 1519 GATA3 sites that were significantly altered upon treatment with E2+DHT compared with the E2-treated cells, highlighting the strong effect of androgens in modifying the E2-associated GATA3 cistrome (Figure 2-f). A total of 1331 differentially GATA3 bound loci were significantly enriched peaks, and 188 were significantly lost enrichment upon E2 and DHT co-treatment in T-47D cells (Figure 2-f). We also found 1185 differentially bound loci for GATA3 (1158 enriched and 28 lost peaks) upon E2+DHT

treatment in ZR-75-1 cells (Supplementary 3-f). In order to investigate the overlap of AR and GATA3 cistromes upon DHT treatment in ER+ breast cancer cells in the presence of E2, we integrated this data with our previously generated AR ChIP-seq data from Hickey *et al.*, 2021. We investigated the distribution of AR and GATA3 cistromes following E2+DHT stimulation in both ER+ breast cancer cells through generating two-factor MA plots, revealing five different sub-groups representing shared or unique AR and GATA3 binding events (Figure 2-g, Supplementary 3-g). Pink dots in this figure show E2+DHT-induced GATA3 binding events that are shared with the AR cistrome in each model (Figure 2-g, Supplementary 3-g). Interestingly, more than 94 % of E2+DHT-induced GATA3 loci are co-occupied with E2+DHT-induced AR sites in both breast cancer cell models (Figure 2-h, Supplementary 3-h). This suggests a potent interplay between AR and GATA3 that may contribute to AR cross-talk with ER in breast cancer cells. *RANBP3L* and *CA12* are AR target genes that showed significant enrichment of GATA3 peak upon E2+DHT treatment compared with the E2-treated cells. However, *PgR*, *GREB1*, and *MYB* are examples of ER target genes that GATA3 peaks did not altered by E2+DHT treatment in both ER+ breast cancer cells (Figure 2-g, Supplementary 3-g).

### **AR-induced ER-reprogramming is mediated through a switch in GATA3 co-regulatory function**

To investigate the effect of AR activation on GATA3 and ER cistromes in an integrated manner, we used our publicly available ER and AR cistromes generated under E2 and E2+DHT treatment conditions (Hickey *et al.* 2021). We overlapped the significant E2+DHT-induced GATA3 peaks with 17,992 and 1530 significantly (FDR  $\leq$  0.05) E2+DHT-enriched peaks for AR and ER in T-47D cells, compared with E2-treated cells (Figure 3-a). Surprisingly, all the common E2+DHT-induced ER and GATA3 binding events were co-

occupied by AR, suggesting that the ER and GATA3 binding sites gained with E2+DHT treatment are AR-dependent (Figure 3-a). According to our previous findings, AR-mediated inhibition of ER genomic signalling partially occurred due to recruitment of ER to AR binding sites associated with tumour suppressor genes (Hickey et al. 2021). We also examined the contribution between GATA3, AR and ER binding events at DHT-induced AR-ER sites that are co-occupied with GATA3 loci in T-47D (Figure 3-b), and ZR-75-1 (Supplementary 4-a,b). We observed androgen induced enrichment of the H3K27ac ChIP-seq signal at AR-GATA3-ER co-occupied loci, indicative of a more transcriptionally active chromatin state in both T-47D (Figure 3-b), and ZR-75-1 cells (Supplementary 4-c). Interestingly, assessing the accessibility of the chromatin at E2+DHT-induced AR loci including the AR-ER shared site depleted from GATA3 indicated no enrichment upon DHT-treatment, suggesting a role for GATA3 in mediating chromatin structure at AR-GATA3-ER loci in T-47D (Figure 3-c) and ZR-75-1 (Supplementary 4-d) upon E2+DHT treatment. We integrated annotated AR-GATA3-ER targets with our previously published RNA-seq data (Hickey et al. 2021) and found 537 and 211 AR-GATA3-ER gene targets that were differentially expressed in response to E2+DHT versus E2 treatment in T-47D (343 up-regulated genes and 194 down-regulated genes) and ZR-75-1 (133 up-regulated genes and 44 down-regulated genes) cells, respectively. In order to assess the biological processes associated with these DEG genes, we performed GSEA pathway analysis and found that AR-GATA3-ER up-regulated targets are enriched in epithelium development, cell-cell signalling, and regulation of cell death pathways (Figure 3-d, Supplementary 4-e). GSEA analysis showed that the down-regulated AR-GATA3-ER target genes were negatively associated with cell growth gene sets in both of the cell line models (Figure 3-e, Supplementary 4-f). Taken together, these data suggest that GATA3 may contribute with

the AR-ER cross-talk in ER+ breast cancer cells through mediating the chromatin accessibility.

In order to assess the potential involvement of GATA3 in regulating the AR good outcome signature genes described in Hickey *et al.*, 2021, we investigated whether AR-GATA3-ER targets overlap the AR signature genes (Hickey et al. 2021). The AR signature was derived from *in vivo* models and positively predicted disease survival in multiple large clinical cohorts of ER+ breast cancer, outperforming existing pan-cancer or breast cancer-specific signatures. We found that ~40 % of DEG AR signature genes, overlap with the DEG AR-GATA3-ER targets including *ZBTB16*, *SEC14L2*, *CLDN8*, and *EAF2*, suggesting a role for GATA3 in regulation of AR signature genes (Supplementary 5) in T-47D cells. We also assessed the effect of GATA3 knock-down on the expression level of AR signature genes in the condition of E2+DHT treatment compared with the E2-treated cells. Silencing GATA3 significantly reduced the E2+DHT-induced expression of both targets, however, it did not abolish the AR regulation on the expression level of these targets completely in both T-47D (Figure 3-f) and ZR-75-1 cells (Supplementary 4-g).

Together, our data suggest that AR activation relocates GATA3 and ER away from shared ER binding sites associated with proliferative genes, and induces a new subset of binding events that are common between all three transcription factors, implicating GATA3 in AR-driven tumour suppression in ER+ breast cancer cells, possibly through mediating the chromatin accessibility.

### **GATA3 knockdown reduces the growth inhibitory effects of androgens in E2-stimulated ER+ breast cancer cells**

To further investigate GATA3 involvement in AR-mediated growth inhibitory effects in ER+ breast cancer cells, we silenced GATA3 in ER+ breast cancer cell lines (mKATE red

nuclear tag labelled T-47D and ZR-75-1) to examine proliferation in real time in response to E2 and E2+DHT hormone stimulation. Transduction of T-47D and ZR-75-1 cells with mKATE red nuclear tag facilitated real-time imaging of cell growth. We simultaneously seeded and transfected the tagged cells with specific siRNAs to GATA3 (siGATA3) or to a non-targeting control (siControl). Growth assay experiments for both mKATE\_T-47D and mKATE\_ZR-75-1 cells showed that silencing GATA3 significantly reduced basal growth in both cell line models (Figure 4-a-c), suggesting a key role for GATA3 in growth of the breast cancer cells independent of ER and/or AR. Furthermore, we observed a dramatic morphology change after GATA3 knock-down irrespective to hormone stimulation in both cell lines (Figure 4-d). Knock-down of GATA3 caused genesis of gained multinuclear cells sharing one common cytoplasm. These observations suggest a new role for GATA3 in cell cycle progression and normal cytokinesis, which needs to be further studied in future. Silencing GATA3 blunted but not abolish the growth stimulatory effect of estrogens, and growth-inhibitory effects of androgens in both ER+ breast cancer cells (Figure 4-a-c). Therefore, we suggest that in the condition of estrogen stimulation, GATA3 functions as an ER co-regulator in inducing the growth of the cells; however, upon androgen stimulation GATA3 functions as an AR co-regulator, mediating the mechanism by which AR suppresses the growth of ER+ breast cancer cells. In order to investigate the effect of GATA3 knock-down on AR-mediated suppression of genes involved in E2-induced growth stimulation of ER+ breast cancer cells, we assessed the effect of GATA3 knock-down on the mRNA expression level of *PgR* and *GREB1* genes conducting qPCR experiment for E2 and E2+DHT-stimulated cells in both T-47D and ZR-75-1 breast cancer models (Figure 4-d,e). Together with what we have shown in Figure 1, our data show that silencing GATA3 significantly impairs the stimulatory effect of estrogens (Figure 1-h,i) and the inhibitory effects of

androgens (Figure 4-d,e) in regulating the mRNA transcriptional level of selected growth-associated genes. However, consistent with our growth assay results, GATA3 depletion did not abolish the effect of estrogens and androgens in increasing or reducing the mRNA expression level of selected targets, respectively (Figure 1-h,i, Figure 4-d,e). The androgen-induced switch in the GATA3 collaborative role from being an ER co-regulator to an AR co-regulator significantly affect the mRNA expression of the growth-associated ER-GATA3 target genes like *PgR* and *GREB1* in both T-47D and ZR-75-1 breast cancer cells (Figure 4-d,e), which in turn, involves in the E2-induced growth suppression of ER+ breast cancer cells. Collectively, these results suggest that GATA3 is a common ER and AR co-regulator that mediates both ER and AR signalling in regulating the growth of ER+ breast cancer cells in response to estrogen or androgen stimulation.

## Discussion

Herein, we explored the effect of estrogens and combination of estrogens and androgens in redistribution of the GATA3 binding profile in ER+ breast cancer cells. We showed that GATA3 mediates both ER and AR signalling to induce the growth or to repress the E2-induced growth in response to ER and/or AR activation in ER+ breast cancer cells. Also, we showed that GATA3 involved in AR-mediated ER reprogramming in ER+ breast cancer cells.

In this study, we generated robust, reproducible GATA3 cistromes under different hormonal conditions in two ER+ breast cancer cell lines. While GATA3 ChIP-seq has been previously performed in ER+ breast cancer cells, there are issues with the data that limit their utility. For example, Kong *et al.*, 2011 generated GATA3 cistromic data in estrogen-



treated MCF-7 cells compared with the untreated cells, but this cell line expresses both wild-type and mutant GATA3 proteins that are both pulled down with the antibodies available for GATA3 ChIP-seq. Also, the latter study provided the vehicle and E2-treated GATA3 cistromes only in one replicate (Kong et al. 2011). Additionally, Theodorou *et al.*, 2013 provide the GATA3 ChIP-seq in unstimulated MCF-7 breast cancer cells and following estrogen treatment, each in five replicates. In this study, GATA3 binding was introduced almost exclusively estrogen independent, however, regardless of using a non-appropriate cell model (MCF-7), when we analysed the public data provided from the paper we found that estrogen treatment hugely increased and rearranged the GATA3 binding profile in these cells ([Supplementary 6-e](#)). Therefore, in order to investigate the exclusive role of wild-type GATA3 in ER+ breast cancer cells we used T-47D and ZR-75-1 cells that exclusively express wild-type GATA3, and to generate a reproducible data we conducted all our ChIP-seq experiments in 3 biological replicates. We previously examined effect of androgen stimulation on GATA3 cistrome in breast cancer context in the same ER+ breast cancer models (Chapter-2). Collectively, our data indicated that estrogen- and androgen-stimulation of ER+ breast cancer cells rearrange GATA3 cistrome. We showed that E2-induced GATA3 binding sites hugely overlap with E2-induced ER cistrome, and are associated with a set of biological processes involved in cell cycle progression and growth in ER+ breast cancer cells. However, almost all E2+DHT-induced GATA3 binding sites were shared with AR in both breast cancer cell models, suggesting that GATA3 mediates AR signalling in ER+ breast cancer cells. The GATA family of transcription factors have been shown to prominently contribute to the development of many cancer types in human patients. For example, GATA2 is an AR co-regulator playing a key role in prostate cancer progression (He et al. 2014). GATA2 enhanced AR signalling in LNCaP cells and is critical for

AR expression and transcriptional activity in these cells. Silencing GATA2 markedly depleted AR mRNA and protein expression levels in different PC cell lines. Although androgens and AR suppressed GATA2 expression in PC, indicative a negative feedback loop between AR and GATA2, our data showed that AR and GATA3 does not regulate the expression of each other in breast cancer contexts (Chapter-2). Similar to GATA2 and AR in PC (He et al. 2014), we have shown that GATA3 and AR physically interact with each other in breast (Chapter-2). The colocalization of AR and GATA2 on chromatin was associated with recruitment of key transcriptional mediators including SRC-3 and p300 as well as with increased for epigenetic marks of active transcription, H3K27ac (He et al. 2014). Interestingly, AR interacts and takes the ER co-regulators SRC-3 and p300 away from the ER targets in breast cancer context (Hickey et al. 2021), while reprogramming ER cistrome in ER+ breast cancer cells. Taken together, since GATA3 is important for the function of both ER and AR transcription factors and has overlapping binding sites with both factors, our findings suggest that it is possible that the activated AR (with the background of estrogen) sequester GATA3 from the its motifs near to EREs at ER target genes to open up the nucleosome and facilitate its binding to AREs at AR target genes. Since AR interacts with all three SRC-3, p300, and GATA3 factors, a main question is still remaining to be discover that which of these three factors is most important in AR-mediated ER reprogramming in ER+ breast cancers.

## Materials and Methods

### Cell culture

ZR-75-1 (CRL-1500<sup>TM</sup>) and T-47D (HTB-133<sup>TM</sup>) breast cancer cell lines were obtained from ATCC (Manassas, VA, USA). Both cell lines were routinely tested for mycoplasma infection and authenticity confirmed by short tandem repeat profiling. Both cell lines were maintained in RPMI-1640 medium (Invitrogen) containing 10 % Foetal Bovine Serum (FBS) and 2 nM L-Glutamine (Sigma) and were incubated at 37 °C and 5 % CO<sub>2</sub>. For hormone stimulation experiments, cells were starved for 72 hours in 5 % dextran-coated charcoal stripped (DCC) FBS medium prior to hormone addition.

### Western blotting

Cells were harvested at the indicated time-points by scraping and lysed into Radioimmunoprecipitation assay (RIPA) buffer. Protein concentration was quantified using BCA protein assay (Thermo Scientific). Proteins were separated by SDS-PAGE on 10 % Criterion TGX Stain Free-gels (BioRad) and then transferred to nitrocellulose blotting membranes (GE Healthcare). Blocking was carried out in 5 % skim milk in TBST for 2 hours. Immunoblotting for GATA3 (Abcam; ab199428) and GAPDH (Merck Millipore; mab374) ([Supplementary 6-a](#)) was performed overnight and membranes were washed with PBS + 0.1% Tween-20. HRP-coupled secondary antibodies (goat anti-rabbit (Dako; 1:2000) or rabbit anti-mouse (Dako; 1:2000)) were detected with Clarity Western ECL Substrate (Bio-RAD) and visualised using a BioRad ChemiDoc imaging system.

### siRNA knockdown experiments

Pre-designed siRNAs against GATA3 and a negative control siRNA (Allstar-Neg. Control siRNA) were used for the purpose of siGATA3 studies ([Supplementary 6-b](#)). Cells

were transfected with siRNAs by reverse transfection using RNAiMAX (Invitrogen) (0.5  $\mu\text{l}/\text{cm}^2$ ) at the time of seeding for 48 hours according to the manufacturer's protocols (Thermo Fisher Scientific, Waltham, MA, USA).

### **RNA extraction and quantitative real-time PCR (qPCR)**

Cells were seeded at  $0.5 \times 10^6$  cells/well in 6-well plates and simultaneously transfected with either 10 nM of siRNAs against GATA3 or siControl (10 nM). After 48 hours of transfection, the media was changed to hormone-depleted medium supplemented with 5 % DCC FBS. After 72 hours (media refreshed at each day), both cell line models were treated for 6 hours with vehicle, 10 nM E2, or 10 nM E2+DHT before harvest. Total RNA was extracted using Trizol followed by DNase treatment using the TURBO DNase Kit (Invitrogen). Reverse transcription was performed on 500 ng of the total RNA using the iScript cDNA Synthesis Kit (Bio-Rad). The resulting cDNA was diluted 1:10 and used for qRT-PCR. Primers are listed in supplementary-6 ([Supplementary 6-c](#)). mRNA levels were normalised to *GAPDH* using the  $\Delta\Delta\text{Ct}$  method in Bio-Rad CFX-manager software, as previously described (Hickey et al. 2021).

### **ChIP-sequencing (ChIP-seq)**

ZR-75-1 and T-47D cells were seeded in 15 cm plates at  $9 \times 10^6$  cells/plate, in their normal maintenance medium. Media was changed to phenol-red-free medium supplemented with 5 % DCC -stripped FBS after 24 hours and cells were allowed to grow for a further 2 days prior to treatment with daily media changes. Cells were treated with either vehicle, 10 nM E2, 10 nM DHT, or 10 nM E2+DHT for 4 hours prior to fixation and harvest. Each experiment was done as three independent biological replicates representing consecutive passages of cells. Cells were cross-linked with 1 % formaldehyde, quenched with 2 M Glycine pH 7.5, and collected by scraping. The cells were suspended in

lysis buffer (10 mM Tris-HCl (pH 8), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % Na-Deoxycholate, and 0.5 % N-lauryl sarcosine) in the presence of protease inhibitors (Complete(R), Roche) and sonicated for 10 cycles of '30 seconds on, 30 seconds off', using a Diagenode Bioruptor. Immunoprecipitations were performed using 5 µg/IP of GATA3 (Abcam; ab199428) antibodies excluding the DNA input samples ([Supplementary 6-d](#)). Before sequencing the CHIP'd DNA samples, CHIP-PCR reactions were prepared using iQ SYBR Green Supermix (BIO-RAD) to test experimental efficacy at positive control sites. PCR was performed with the CFX384 Real Time PCR Detection System (BIO-RAD) and standard cycling conditions. CHIP-PCR data were analysed by the percentage input method and further analysed as fold enrichment over a negative control. DNA was sequenced using an Illumina NextSeq 500 (High Output) with 75 bp single-end reads. Raw data was processed in Galaxy. Trimmed FASTQ files were aligned to the hg19 genome assembly using Bowtie2 (version: 2.3.4.3, default parameters), mapped reads with a minimum MAPQ <10 and duplicate reads were removed using SAMtools. Peaks were called using MACS2 callpeak (version: 2.1.1, default settings) with a pooled input sample as the control. Only peaks found in 2 out of the 3 replicates were kept for the consensus peakset. For figures, CHIP-seq data was visualized using the Integrative Genomics Viewer (<https://igv.org/app/>). Heatmaps (Galaxy Version 3.3.2.0.1) and PCAs (Galaxy Version 3.3.2.0.0) were generated using Deeptools. Peak annotations were performed using Cisgenome (v2.0).

#### **m-KATE red nuclear tag generation**

Second generation, 3-package plasmid system was conducted to generate the m-KATE nuclear tag lentiviruses using the packaging vector of psPAX2, envelope vector of VSVG pSD11 pMD2.G, and the transfer plasmid encoding the insert of interest. HEK-293 cells were transfected with transfer plasmid and packaging plasmids using the PEI

transfection methodology. The media supplemented with transfection mix was changed to Opti-MEM media after 12 hours of incubation at 37 °C. Media containing viral particles was collected after 48 hours, purified and filtered to concentrate the m-KATE virus.

### **Stable cell line generation**

Lentiviral titration has been done for each cell line to ensure that preps are applied in sufficient quantities to obtain desired transduction efficiencies. The optimal titration for the red nuclear tag virus in both breast cancer lines was  $10^5$  IFU/ml (Infectious units (IFU)). Both ER+ (ZR-75-1 and T-47D) breast cancer cell lines were transduced with the red nuclear tag virus. Transduced lines were expanded and selected through culturing in media supplemented with G418 antibiotic. Selected cells were used for the purpose of the growth assays.

### **Cell growth assays**

Tagged T-47D ( $10 \times 10^3$ ) and ZR-75-1 ( $11 \times 10^3$ ) cells were seeded in 96-well plates in hormone stripped serum (5 % DCC FBS) in three technical replicates (in each of the two experiments) and simultaneously transfected with one of two different siRNAs to GATA3 (10 nM) or a siControl (10 nM) for 48 hours. Transfected media was removed from all wells and fresh hormone depleted media supplemented with different hormone treatments (vehicle, 10 nM E2 or 10 nM E2+DHT) was added to plates. Plates were scanned every 2 hours with the objective of 10 X for the duration of the experiment (9 days). 10 images from different time-points were selected to train the IncuCyte to count the cells in each of the models.

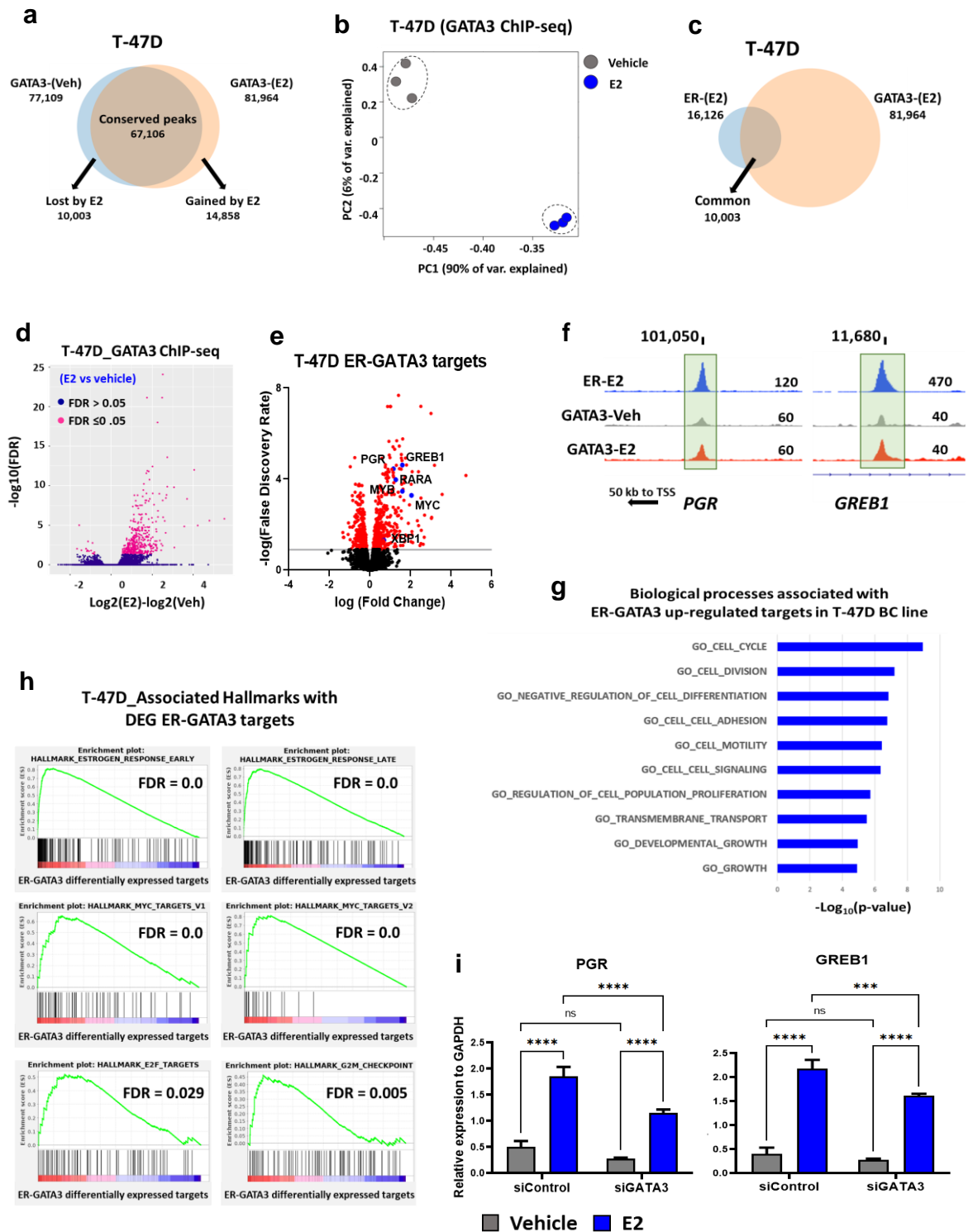
### **Statistical analyses**

Statistical calculations were performed using GraphPad Prism (GraphPad Software). Normality was assumed for all statistical tests. All tests were two-sided with a

95 % confidence interval and a P value  $<0.05$  was indicative of statistical significance. All qPCRs and growth assays in both models analysed using a two-way ANOVA, followed by Tukey's multiple comparisons test.

# Main figures

Figure 1: Estrogen stimulation of ER+ breast cancer cells rearranges the GATA3 cistrome

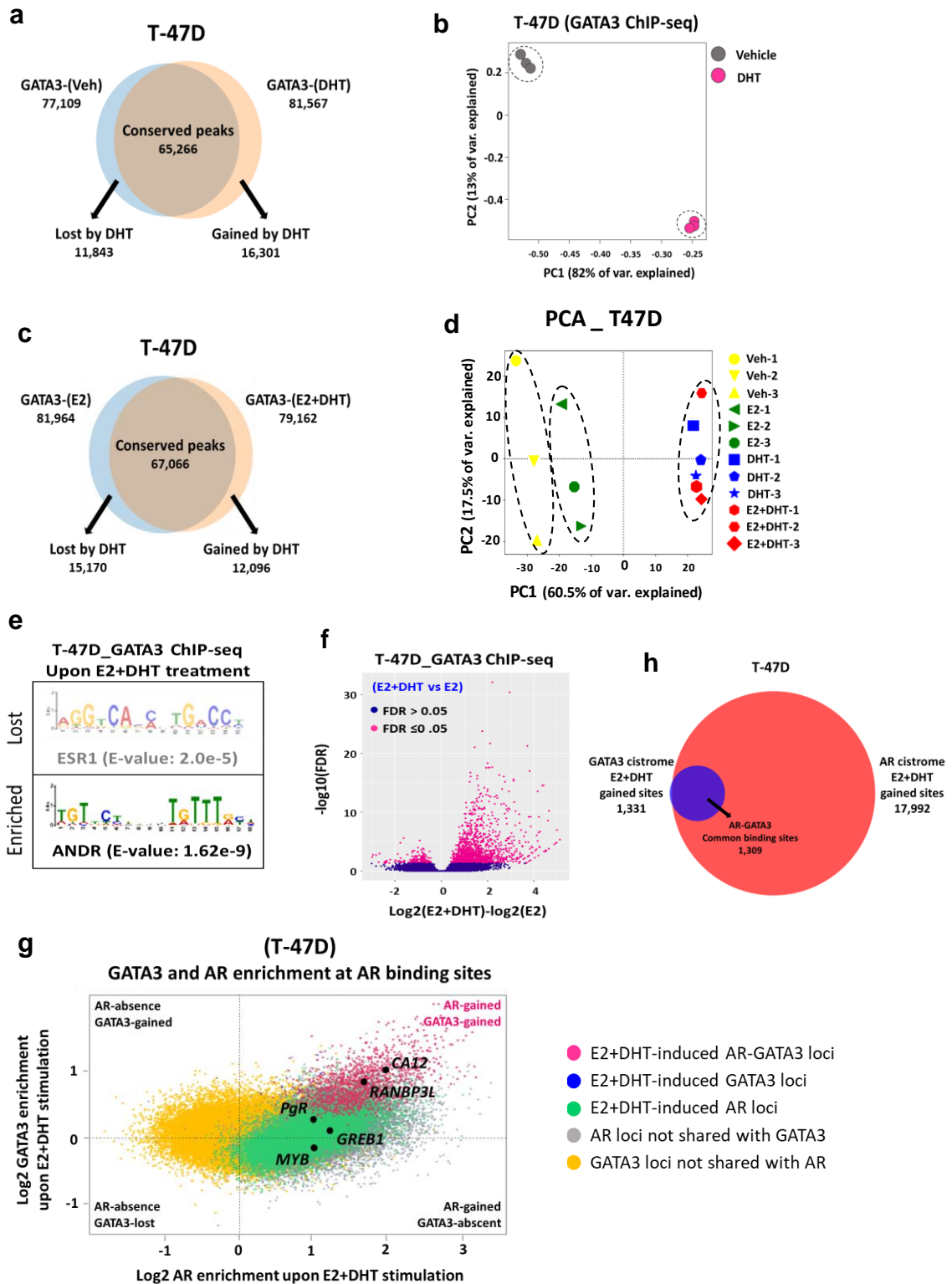




**Figure-1: Estrogen stimulation of ER+ breast cancer cells rearranges the GATA3 cistrome.**

**a)** Venn diagram shows the effect of estrogen (E2) treatment on the GATA3 binding profile. **b)** Principle component analysis (PCA) clustered the E2-treated cells away from the vehicle-treated cells in the T-47D breast cancer cell line model. Grey dots represent three independent biological replicates of vehicle-treated samples; blue dots represent 3 replicates of E2-treated samples. **c)** Venn diagram shows the overlap of ER and GATA3 cistromes under E2 treatment. **d)** Volcano plot shows the differentially enriched and depleted GATA3 binding sites ( $FDR \leq 0.05$ ) upon ER activation in T-47D breast cancer cells. The purple dots represent peaks that did not undergo a significant change ( $FDR \geq 0.05$ ) with ER activation. The pink dots represent significant ( $FDR \leq 0.05$ ) changes to GATA3 peaks in each model. **e)** Volcano plot depicts the differentially expressed E2-induced ER-GATA3 gene targets in T-47D cells. **f)** Integrative genomics viewer (IGV) snap shot shows averaged ER and GATA3 ChIP-seq signals at binding sites associated with ER target genes (*PgR*, and *GREB1*). **g)** Bar graph shows the biological pathways that are associated with E2 up-regulated ER-GATA3 gene targets in T-47D breast cancer model. **h)** Gene set enrichment analysis (GSEA) plots represent the associated Hallmarks with E2 up-regulated ER-GATA3 targets in T-47D cells. **i)** Bar graphs represent the effect of GATA3 knock-down on the mRNA expression level of *PgR* and *GREB1* genes under E2-stimulation condition compared with the vehicle cells in T-47D breast cancer model. Two-way ANOVA with Tukey's multiple comparisons test was used to determine statistically significant differences in **i**. Data shown as mean  $\pm$  SEM of three independent experiment; \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$ .

Figure 2: Androgens redistribute the GATA3 binding profile in ER+ breast cancer cells

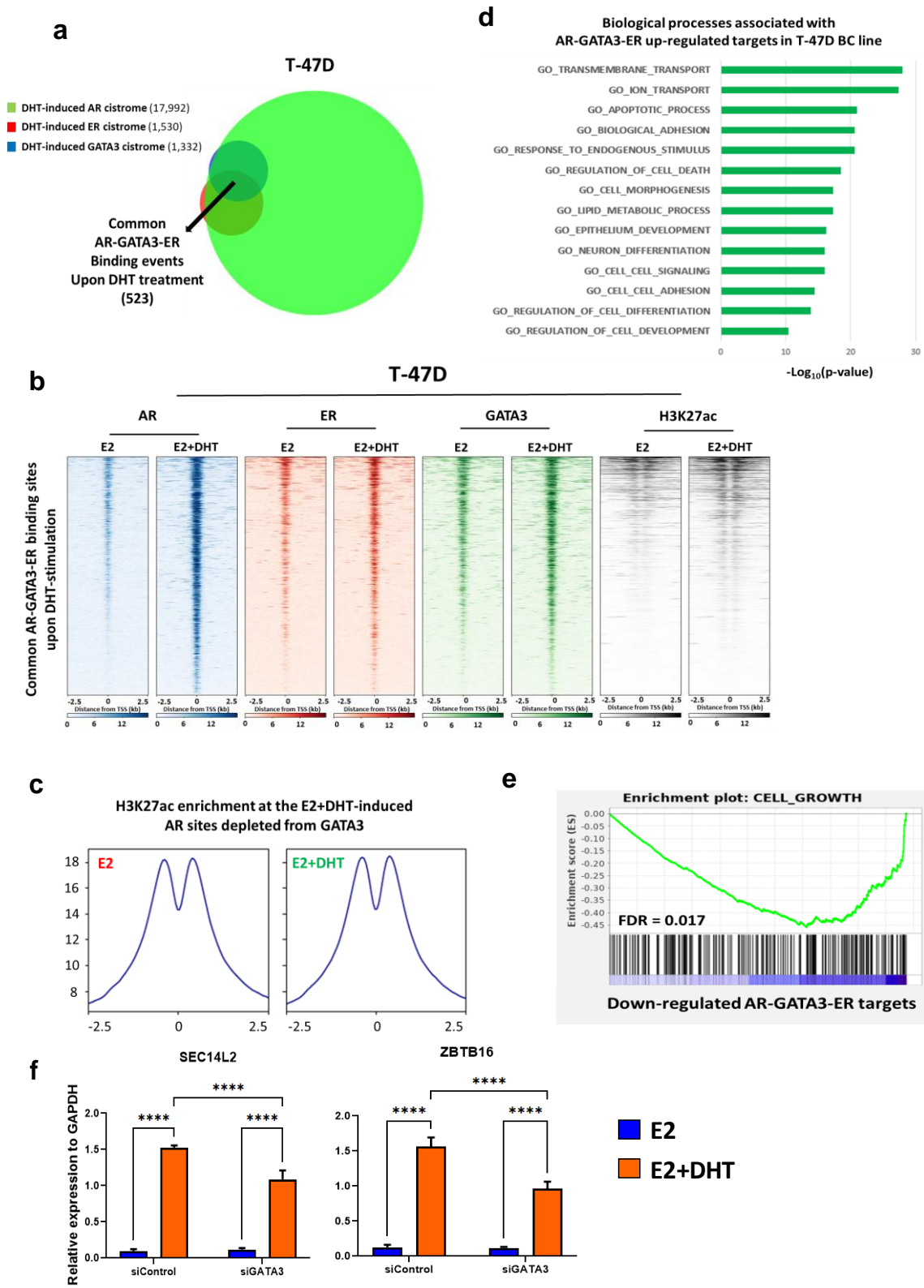


**Figure-2: Androgens redistribute the GATA3 binding profile in ER+ breast cancer cells. a)**

Venn diagram shows the effect of DHT treatment on the GATA3 cistrome. **b)** PCA plot shows the clustering of DHT-treated cells away from the vehicle-treated cells in T-47D breast cancer model. Grey dots represent three independent replicates of vehicle-treated samples; pink dots represent 3 replicates of DHT-treated samples in T-47D cells. **c)** Venn diagram shows the potent effect of DHT on re-distributing the E2-stimulated GATA3 cistrome. **d)** PCA plot shows the effect of hormones (estrogens and androgens) on clustering the treated cells in T-47D cells. Colours represent three replicates of different treatments. **e)** Motif analysis of E2+DHT-induced GATA3 binding sites revealed enrichment of the AR motif and loss of ER motif in T-47D breast cancer model. **f)** Volcano plot shows the differentially enriched and depleted GATA3 binding sites ( $FDR \leq 0.05$ ) upon AR activation in the presence of E2 in T-47D breast cancer cells. The purple dots represent peaks that did not undergo a significant change ( $FDR \geq 0.05$ ) with AR activation. The pink dots represent significant ( $FDR \leq 0.05$ ) changes to GATA3 peaks in each model. **g)** Two-factor MA plot represents distribution of the AR and GATA3 cistromes upon E2+DHT stimulation in T-47D cells. Peaks are represented by coloured dots and classified into 5 different sub-groups: grey and yellow dots represent AR and GATA3 binding sites that do not overlap with the other factor, respectively; blue dots represent E2+DHT-induced GATA3 peaks that are not shared with AR; green dots represent GATA3 binding sites that were not altered by E2+DHT stimulation but AR was recruited to these loci; and pink dots represent AR-GATA3 peaks that are significantly gained upon E2+DHT stimulation. Example loci associated with AR-GATA3-ER target genes (*RANBP3L* and *CA12* represented loci from pink group, *PgR*, *GREB1*, and *MYB* represented loci from green group) that are highlighted

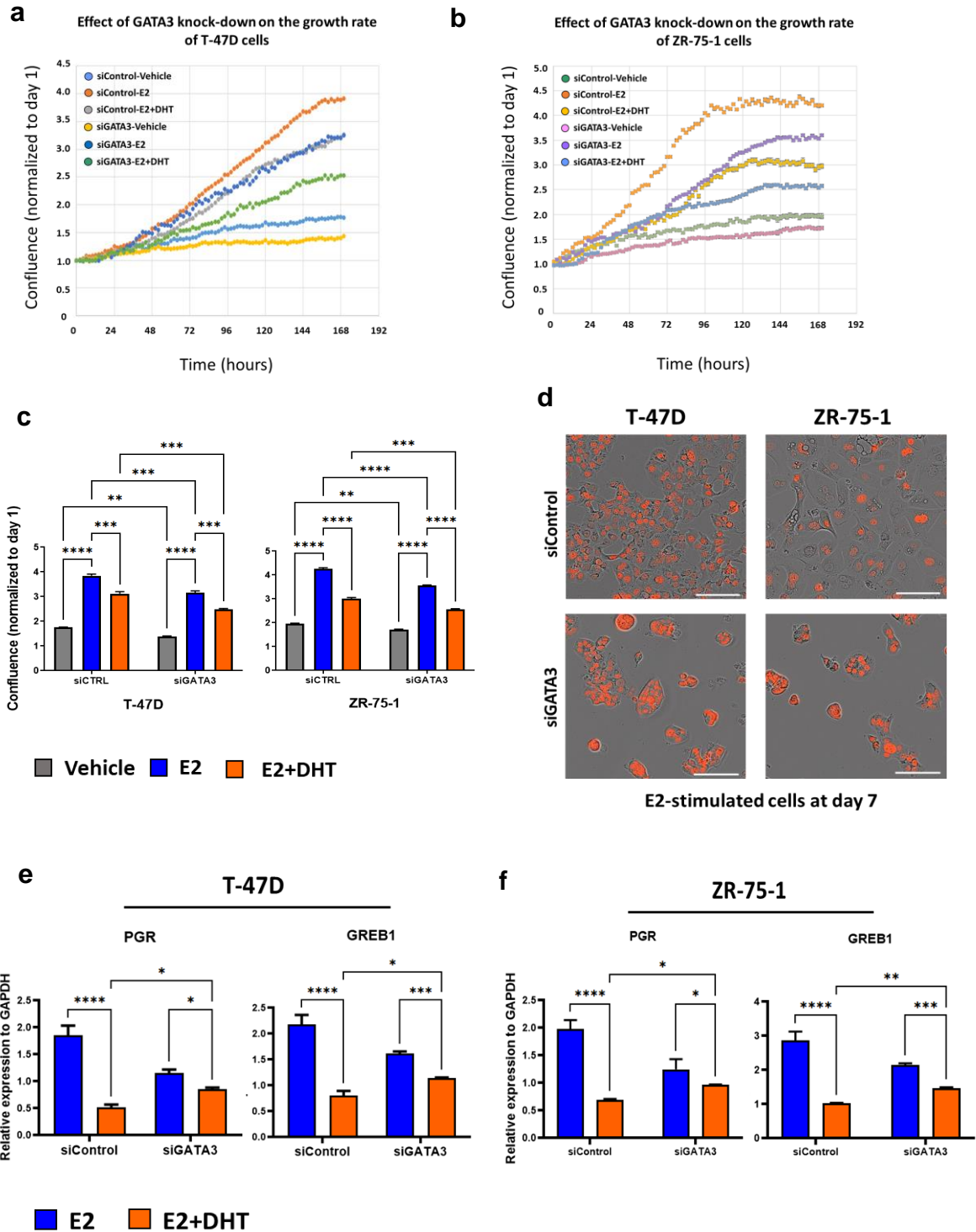
in black dots in T-47D cells. **h)** Venn diagram depicts the overlap of E2+DHT-induced GATA3 binding sites with E2+DHT-induced AR events in T-47D cells.

**Figure 3: AR-induced ER-reprogramming is mediated through a switch in GATA3 co-regulatory function**



**Figure-3: AR-induced ER-reprogramming is mediated through a switch in GATA3 co-regulatory function.** **a)** Venn diagram shows the overlap of DHT-induced AR, GATA3, and ER cistromes in T-47D cells in the presence of E2. **b)** Heatmap represents the contribution of E2+DHT-induced GATA3 peaks on E2+DHT-induced shared AR-ER binding sites in T-47D cells, and the H3K27ac enrichment at AR-GATA3-ER co-occupied loci in T-47D cells. **c)** Graph depicts the density of H3K27ac enrichment in at E2+DHT-induced AR sites depleted from GATA3 compared with the E2 stimulation condition in T-47D cells. **d)** Bar graph shows the biological pathways that are associated with up-regulated AR-GATA3-ER targets upon E2+DHT treatment in T-47D cells. **e)** GSEA analysis shows that down-regulated AR-GATA3-ER targets are negatively associated with cell growth genes in T-47D cells. **f)** Bar charts represent the effect of GATA3 knock-down on the mRNA expression level of *SEC14L2* and *ZBTB13* tumour-suppressor AR targets under E2+DHT-stimulation compared with the E2-treated cells in the T-47D breast cancer model. Two-way ANOVA with Tukey's multiple comparisons test was used to determine statistically significant differences in **f**. Data shown as mean  $\pm$  SEM of 3 independent experiments; \*\*\*\*  $p < 0.0001$ .

**Figure 4: GATA3 depletion reduces the growth inhibitory effects of androgens in E2-stimulated ER+ breast cancer cells**



**Figure-4: GATA3 depletion reduces the hormonal effects of estrogen and androgen in ER+ breast cancer cells.** IncuCyte-generated graphs show the effect of GATA3 knock-down on the growth rate of T-47D (**a, c**) and ZR-75-1 (**b, c**) cells under different hormone treatment conditions (vehicle, E2, and E2+DHT). Two-way ANOVA with Tukey's multiple comparisons test was used to determine statistically significant differences in **c**. Data shown as mean  $\pm$  SEM of 3 replicates and are representative of two independent experiments; \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$ . **d**) Images show the effect of GATA3 knock-down on morphology of the T-47D and ZR-75-1 cells. **e**) Bar graphs show the effect of GATA3 knock-down on hormone regulation of the ER target genes *PgR* and *GREB1* in T-47D (**e**) and ZR-75-1 (**f**) cells. Data shown as mean  $\pm$  SEM of 3 replicates of two independent experiments; \*  $p = 0.01$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$ .



## References

Asselin-Labat, ML, Sutherland, KD, Barker, H, Thomas, R, Shackleton, M, Forrest, NC, Hartley, L, Robb, L, Grosveld, FG, van der Wees, J, Lindeman, GJ & Visvader, JE 2007, 'Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation', *Nat Cell Biol*, vol. 9, no. 2, Feb, pp. 201-209.

Bray, F, Ferlay, J, Soerjomataram, I, Siegel, RL, Torre, LA & Jemal, A 2018, 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA Cancer J Clin*, vol. 68, no. 6, Nov, pp. 394-424.

Castrellon, AB 2017, 'Novel Strategies to Improve the Endocrine Therapy of Breast Cancer', *Oncol Rev*, vol. 11, no. 1, Mar 3, p. 323.

Cheng, M, Michalski, S & Kommagani, R 2018, 'Role for Growth Regulation by Estrogen in Breast Cancer 1 (GREB1) in Hormone-Dependent Cancers', *Int J Mol Sci*, vol. 19, no. 9, Aug 28.

Cops, EJ, Bianco-Miotto, T, Moore, NL, Clarke, CL, Birrell, SN, Butler, LM & Tilley, WD 2008, 'Antiproliferative actions of the synthetic androgen, mibolerone, in breast cancer cells are mediated by both androgen and progesterone receptors', *J Steroid Biochem Mol Biol*, vol. 110, no. 3-5, Jun, pp. 236-243.

Dixon, JM 2014, 'Endocrine Resistance in Breast Cancer', *New Journal of Science*, vol. 2014, 2014/09/17, p. 390618.

Dowsett, M, Cuzick, J, Ingle, J, Coates, A, Forbes, J, Bliss, J, Buyse, M, Baum, M, Buzdar, A, Colleoni, M, Coombes, C, Snowdon, C, Gnant, M, Jakesz, R, Kaufmann, M, Boccardo, F, Godwin, J, Davies, C & Peto, R 2010, 'Meta-analysis of breast cancer outcomes in adjuvant trials of aromatase inhibitors versus tamoxifen', *J Clin Oncol*, vol. 28, no. 3, Jan 20, pp. 509-518.

Dressing, GE, Knutson, TP, Schiewer, MJ, Daniel, AR, Hagan, CR, Diep, CH, Knudsen, KE & Lange, CA 2014, 'Progesterone receptor-cyclin D1 complexes induce cell cycle-dependent transcriptional programs in breast cancer cells', *Mol Endocrinol*, vol. 28, no. 4, Apr, pp. 442-457.

Eeckhoute, J, Keeton, EK, Lupien, M, Krum, SA, Carroll, JS & Brown, M 2007, 'Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer', *Cancer Res*, vol. 67, no. 13, Jul 1, pp. 6477-6483.

'Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials', 2005, *Lancet*, vol. 365, no. 9472, May 14-20, pp. 1687-1717.

Fisher, B, Costantino, JP, Wickerham, DL, Cecchini, RS, Cronin, WM, Robidoux, A, Bevers, TB, Kavanah, MT, Atkins, JN, Margolese, RG, Runowicz, CD, James, JM, Ford, LG & Wolmark,

N 2005, 'Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study', *J Natl Cancer Inst*, vol. 97, no. 22, Nov 16, pp. 1652-1662.

Fragomeni, SM, Sciallis, A & Jeruss, JS 2018, 'Molecular Subtypes and Local-Regional Control of Breast Cancer', *Surg Oncol Clin N Am*, vol. 27, no. 1, Jan, pp. 95-120.

Haines, CN, Klingensmith, HD, Komara, M & Burd, CJ 2020, 'GREB1 regulates PI3K/Akt signaling to control hormone-sensitive breast cancer proliferation', *Carcinogenesis*, vol. 41, no. 12, pp. 1660-1670.

He, B, Lanz, RB, Fiskus, W, Geng, C, Yi, P, Hartig, SM, Rajapakshe, K, Shou, J, Wei, L, Shah, SS, Foley, C, Chew, SA, Eedunuri, VK, Bedoya, DJ, Feng, Q, Minami, T, Mitsiades, CS, Frolov, A, Weigel, NL, Hilsenbeck, SG, Rosen, DG, Palzkill, T, Ittmann, MM, Song, Y, Coarfa, C, O'Malley, BW & Mitsiades, N 2014, 'GATA2 facilitates steroid receptor coactivator recruitment to the androgen receptor complex', *Proc Natl Acad Sci U S A*, vol. 111, no. 51, Dec 23, pp. 18261-18266.

Hickey, TE, Robinson, JL, Carroll, JS & Tilley, WD 2012, 'Minireview: The androgen receptor in breast tissues: growth inhibitor, tumor suppressor, oncogene?', *Mol Endocrinol*, vol. 26, no. 8, Aug, pp. 1252-1267.

Hickey, TE, Selth, LA, Chia, KM, Laven-Law, G, Milioli, HH, Roden, D, Jindal, S, Hui, M, Finlay-Schultz, J, Ebrahimie, E, Birrell, SN, Stelloo, S, Iggo, R, Alexandrou, S, Caldon, CE, Abdel-

Fatah, TM, Ellis, IO, Zwart, W, Palmieri, C, Sartorius, CA, Swarbrick, A, Lim, E, Carroll, JS & Tilley, WD 2021, 'The androgen receptor is a tumor suppressor in estrogen receptor-positive breast cancer', *Nat Med*, Jan 18.

Howell, A 2005, 'The future of fulvestrant ("Faslodex")', *Cancer Treat Rev*, vol. 31 Suppl 2, pp. S26-33.

Howell, A, Robertson, JF, Abram, P, Lichinitser, MR, Elledge, R, Bajetta, E, Watanabe, T, Morris, C, Webster, A, Dimery, I & Osborne, CK 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', *J Clin Oncol*, vol. 22, no. 9, May 1, pp. 1605-1613.

Kariagina, A, Xie, J, Langohr, IM, Opreanu, RC, Basson, MD & Haslam, SZ 2013, 'Progesterone stimulates proliferation and promotes cytoplasmic localization of the cell cycle inhibitor p27 in steroid receptor positive breast cancers', *Horm Cancer*, vol. 4, no. 6, Dec, pp. 381-390.

Kong, SL, Li, G, Loh, SL, Sung, WK & Liu, ET 2011, 'Cellular reprogramming by the conjoint action of ER $\alpha$ , FOXA1, and GATA3 to a ligand-inducible growth state', *Mol Syst Biol*, vol. 7, Aug 30, p. 526.

Mohammed, H, Russell, IA, Stark, R, Rueda, OM, Hickey, TE, Tarulli, GA, Serandour, AA, Birrell, SN, Bruna, A, Saadi, A, Menon, S, Hadfield, J, Pugh, M, Raj, GV, Brown, GD, D'Santos,

C, Robinson, JL, Silva, G, Launchbury, R, Perou, CM, Stingl, J, Caldas, C, Tilley, WD & Carroll, JS 2015, 'Progesterone receptor modulates ER $\alpha$  action in breast cancer', *Nature*, vol. 523, no. 7560, Jul 16, pp. 313-317.

Park, S, Koo, JS, Kim, MS, Park, HS, Lee, JS, Lee, JS, Kim, SI, Park, BW & Lee, KS 2011, 'Androgen receptor expression is significantly associated with better outcomes in estrogen receptor-positive breast cancers', *Ann Oncol*, vol. 22, no. 8, Aug, pp. 1755-1762.

Perou, CM, Sørlie, T, Eisen, MB, van de Rijn, M, Jeffrey, SS, Rees, CA, Pollack, JR, Ross, DT, Johnsen, H, Akslen, LA, Fluge, O, Pergamenschikov, A, Williams, C, Zhu, SX, Lønning, PE, Børresen-Dale, AL, Brown, PO & Botstein, D 2000, 'Molecular portraits of human breast tumours', *Nature*, vol. 406, no. 6797, Aug 17, pp. 747-752.

Peters, AA, Buchanan, G, Ricciardelli, C, Bianco-Miotto, T, Centenera, MM, Harris, JM, Jindal, S, Segara, D, Jia, L, Moore, NL, Henshall, SM, Birrell, SN, Coetzee, GA, Sutherland, RL, Butler, LM & Tilley, WD 2009, 'Androgen receptor inhibits estrogen receptor-alpha activity and is prognostic in breast cancer', *Cancer Res*, vol. 69, no. 15, Aug 1, pp. 6131-6140.

Rae, JM, Johnson, MD, Scheys, JO, Cordero, KE, Larios, JM & Lippman, ME 2005, 'GREB 1 is a critical regulator of hormone dependent breast cancer growth', *Breast Cancer Res Treat*, vol. 92, no. 2, Jul, pp. 141-149.

Reis-Filho, JS & Pusztai, L 2011, 'Gene expression profiling in breast cancer: classification, prognostication, and prediction', *Lancet*, vol. 378, no. 9805, Nov 19, pp. 1812-1823.

'Retraction: GREB1 functions as a growth promoter and is modulated by IL6/STAT3 in breast cancer', 2014, *PLoS One*, vol. 9, no. 7, p. e102287.

Ricciardelli, C, Bianco-Miotto, T, Jindal, S, Butler, LM, Leung, S, McNeil, CM, O'Toole, SA, Ebrahimie, E, Millar, EKA, Sakko, AJ, Ruiz, AI, Vowler, SL, Huntsman, DG, Birrell, SN, Sutherland, RL, Palmieri, C, Hickey, TE & Tilley, WD 2018, 'The Magnitude of Androgen Receptor Positivity in Breast Cancer Is Critical for Reliable Prediction of Disease Outcome', *Clin Cancer Res*, vol. 24, no. 10, May 15, pp. 2328-2341.

Takaku, M, Grimm, SA & Wade, PA 2015, 'GATA3 in Breast Cancer: Tumor Suppressor or Oncogene?', *Gene Expr*, vol. 16, no. 4, pp. 163-168.

Theodorou, V, Stark, R, Menon, S & Carroll, JS 2013, 'GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility', *Genome Res*, vol. 23, no. 1, Jan, pp. 12-22.

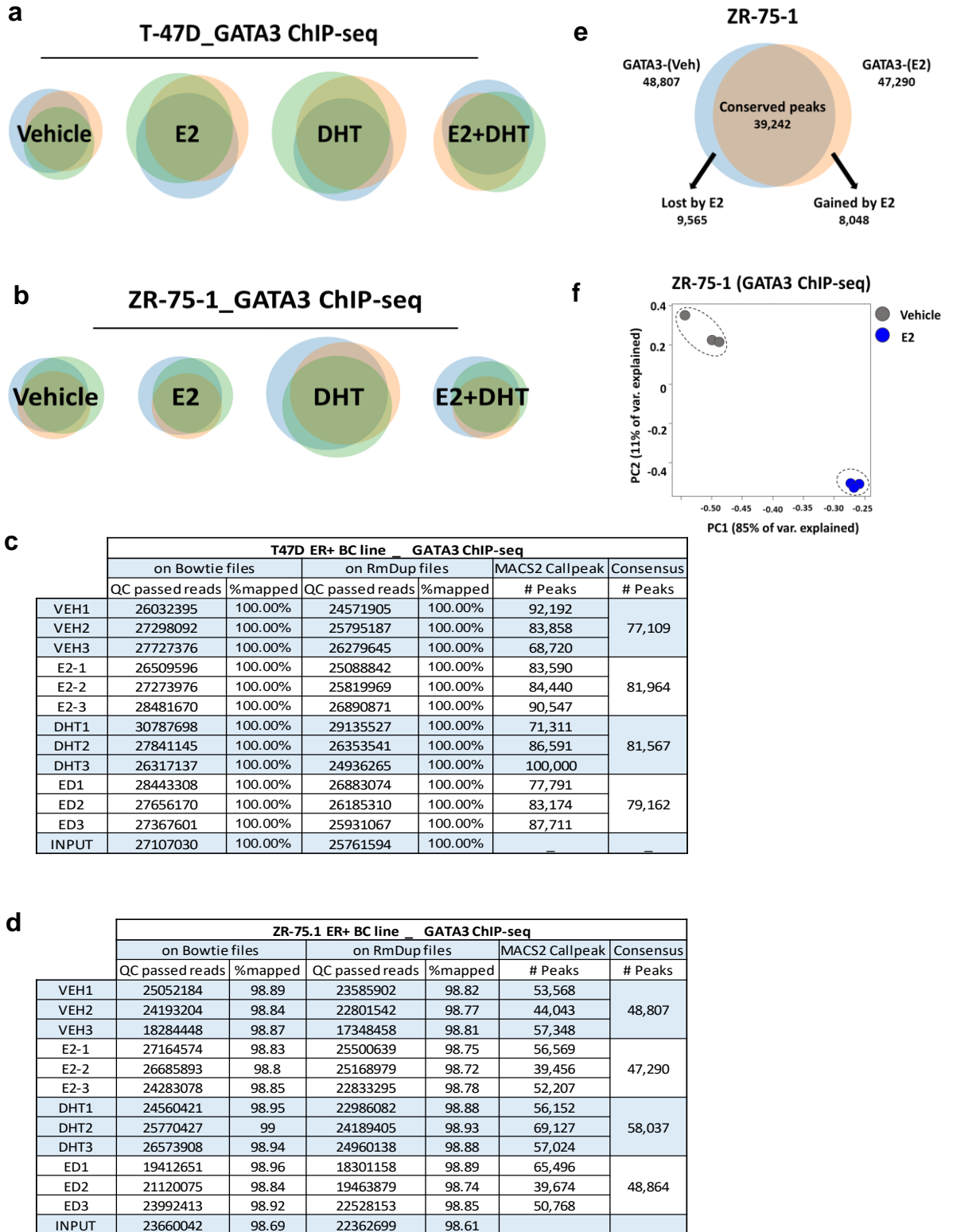
Thürlimann, B, Keshaviah, A, Coates, AS, Mouridsen, H, Mauriac, L, Forbes, JF, Paridaens, R, Castiglione-Gertsch, M, Gelber, RD, Rabaglio, M, Smith, I, Wardley, A, Price, KN & Goldhirsch, A 2005, 'A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer', *N Engl J Med*, vol. 353, no. 26, Dec 29, pp. 2747-2757.

Vergote, I & Abram, P 2006, 'Fulvestrant, a new treatment option for advanced breast cancer: tolerability versus existing agents', *Ann Oncol*, vol. 17, no. 2, Feb, pp. 200-204.

Yu, Q, Niu, Y, Liu, N, Zhang, JZ, Liu, TJ, Zhang, RJ, Wang, SL, Ding, XM & Xiao, XQ 2011, 'Expression of androgen receptor in breast cancer and its significance as a prognostic factor', *Ann Oncol*, vol. 22, no. 6, Jun, pp. 1288-1294.

## Supplementary figures

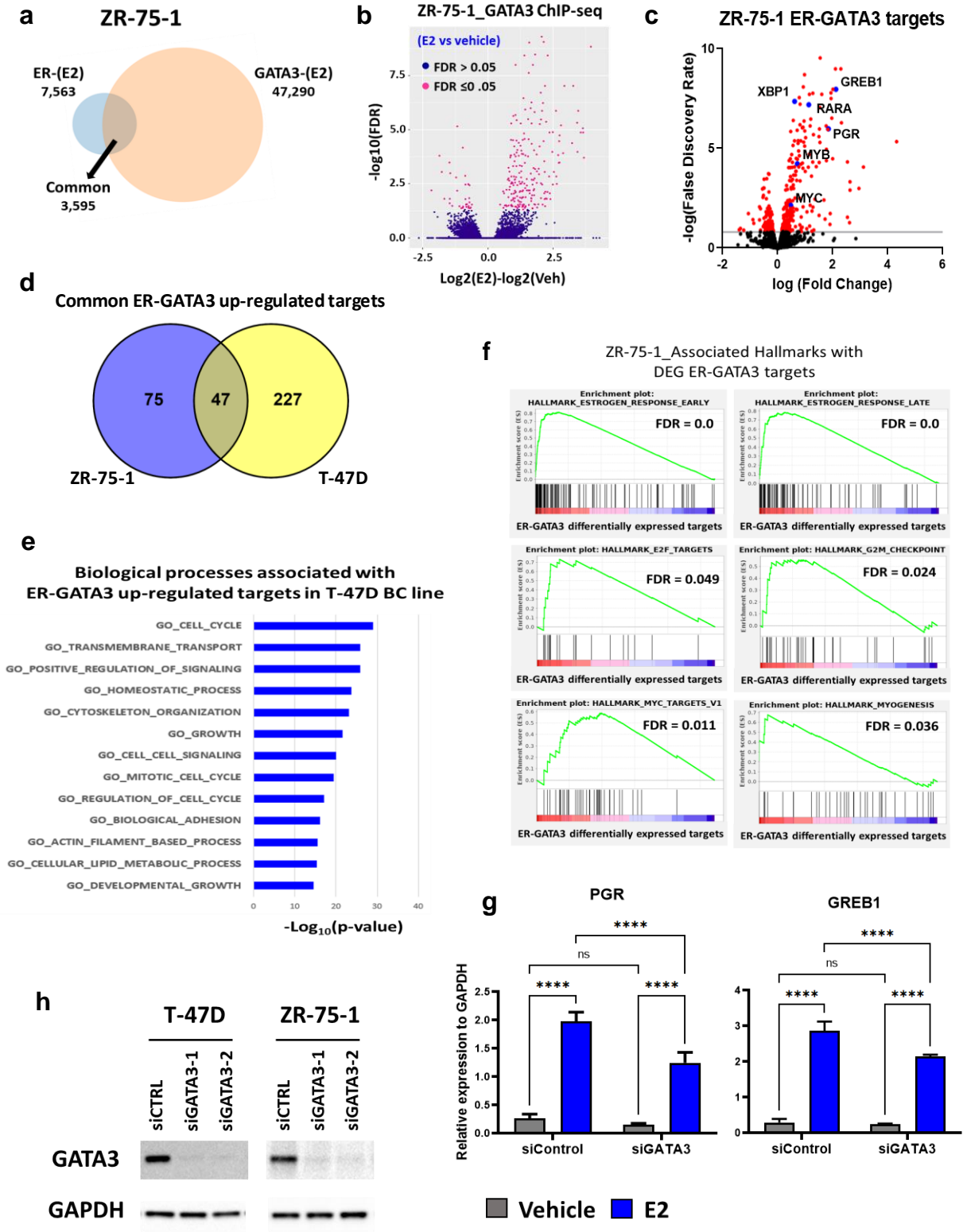
### Supplementary 1: GATA3 cistrome in different hormone treated conditions in ER+ breast cancer cells





**Supplementary-1: GATA3 cistrome in different hormone treated conditions in ER+ breast cancer cells.** Venn diagrams represent the concordance of the 3 biological replicates of each vehicle, E2, DHT, and E2+DHT treated GATA3 CHIP-seq experiment in two ER+ breast cancer *in vitro* cells, T-47D **(a)** and ZR-75-1 **(b)**. Tables provide information regards the GATA3 CHIP-seq data files that have been processed through GALAXY Australia in T-47D **(c)** and ZR-75-1 **(d)** cells. Size of the consensus files for each treatment condition in both ER+ breast cancer cells have been shown in tables **c** and **d**. **e)** Venn diagram shows the effect of estrogen treatment on the arrangement of GATA3 binding profile in ZR-75-1 cells. **f)** PCA clustering the estrogen-treated cells away from the vehicle-treated cells in ZR-75-1 breast cancer model, characterizing the trends exhibited by the E2 compared with the vehicle treated cells in this model. Grey representing three replicates of vehicle-treated samples, however, blue representing 3 replicates of E2-treated samples in ZR-75-1 cells.

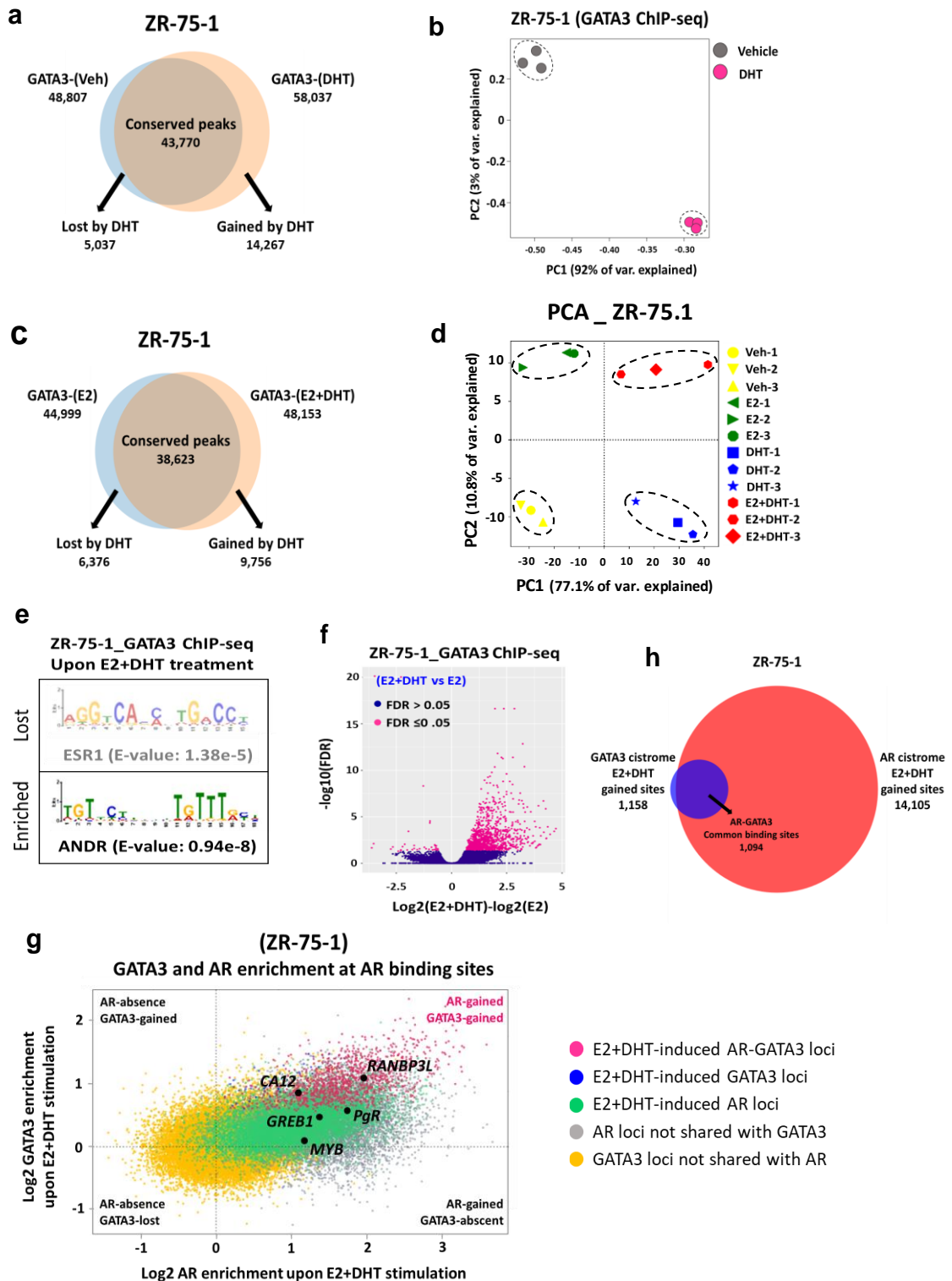
Supplementary 2: Estrogen stimulation of ZR-75-1 cells rearranges the GATA3 cistrome



**Supplementary-2: Estrogen stimulation of ZR-75-1 cells rearranges the GATA3 cistrome.**

**a)** Venn diagram represents the overlap ER and GATA3 cistromes under E2 treatment in ZR-75-1 cells. **b)** Volcano plot shows the differentially enriched and depleted GATA3 binding sites ( $FDR \leq 0.05$ ) upon ER activation in ZR-75-1 breast cancer cells. The purple dots represent peaks that did not undergo a significant change ( $FDR \geq 0.05$ ) with ER activation. The pink dots represent significant ( $FDR \leq 0.05$ ) changes to GATA3 peaks in this model. **c)** Volcano plot depicts the differentially expressed ER-GATA3 targets in ZR-75-1 cells. **d)** Venn diagram show the overlap of up-regulated ER-GATA3 targets in two ER+ breast cancer cells, T-47D and ZR-75-1. **e)** Bar graph shows the biological pathways that are associated with ER-GATA3 targets in ZR-75-1 breast cancer model. **f)** Gene set enrichment analysis (GSEA) plots represent the associated Hallmarks with up-regulated ER-GATA3 targets in ZR-75-1 cells. **g)** Bar charts represent the effect of GATA3 knock-down on the mRNA expression level of *PgR* and *GREB1* genes under E2-stimulation condition compared with the vehicle cells in ZR-75-1 breast cancer model. Two-way ANOVA with Tukey's multiple comparisons test was used to determine statistically significant differences in **g**. Data shown as mean  $\pm$  SEM of 3 independent experiments; \*\*\*\*  $p < 0.0001$ . **h)** Western blotting images show the efficacy of GATA3 silencing in both T-47D and ZR-75-1 cells, also images represent the effect of GATA3 knock-down on ER and AR protein expression levels in T-47D and ZR-75-1 cells.

**Supplementary 3: Androgens redistribute the GATA3 binding profile in ZR-75-1 breast cancer cells**



**Supplementary-3: Androgens redistribute the GATA3 binding profile in ZR-75-1 breast**

**cancer cells. a)** Venn diagram shows the effect of DHT treatment on the GATA3 cistrome.

**b)** PCA plot shows the clustering of DHT-treated cells away from the vehicle-treated cells in ZR-75-1 breast cancer model. Grey dots represent three independent replicates of vehicle-treated samples; pink dots represent 3 replicates of DHT-treated samples in ZR-75-

1 cells. **c)** Venn diagram shows the potent effect of DHT on re-distributing the E2-stimulated GATA3 cistrome. **d)** PCA plot shows the effect of hormones (estrogens and androgens) on clustering the treated cells in ZR-75-1 cells. Colours represent three

replicates of different treatments. **e)** Motif analysis of E2+DHT-induced GATA3 binding sites revealed enrichment of the AR motif and loss of ER motif in ZR-75-1 breast cancer model. **f)** Volcano plot shows the differentially enriched and depleted GATA3 binding sites

( $FDR \leq 0.05$ ) upon AR activation in the presence of E2 in ZR-75-1 breast cancer cells. The purple dots represent peaks that did not undergo a significant change ( $FDR \geq 0.05$ ) with AR activation. The pink dots represent significant ( $FDR \leq 0.05$ ) changes to GATA3 peaks in each

model. **g)** Two-factor MA plot represents distribution of the AR and GATA3 cistromes upon E2+DHT stimulation in ZR-75-1 cells. Peaks are represented by coloured dots and classified

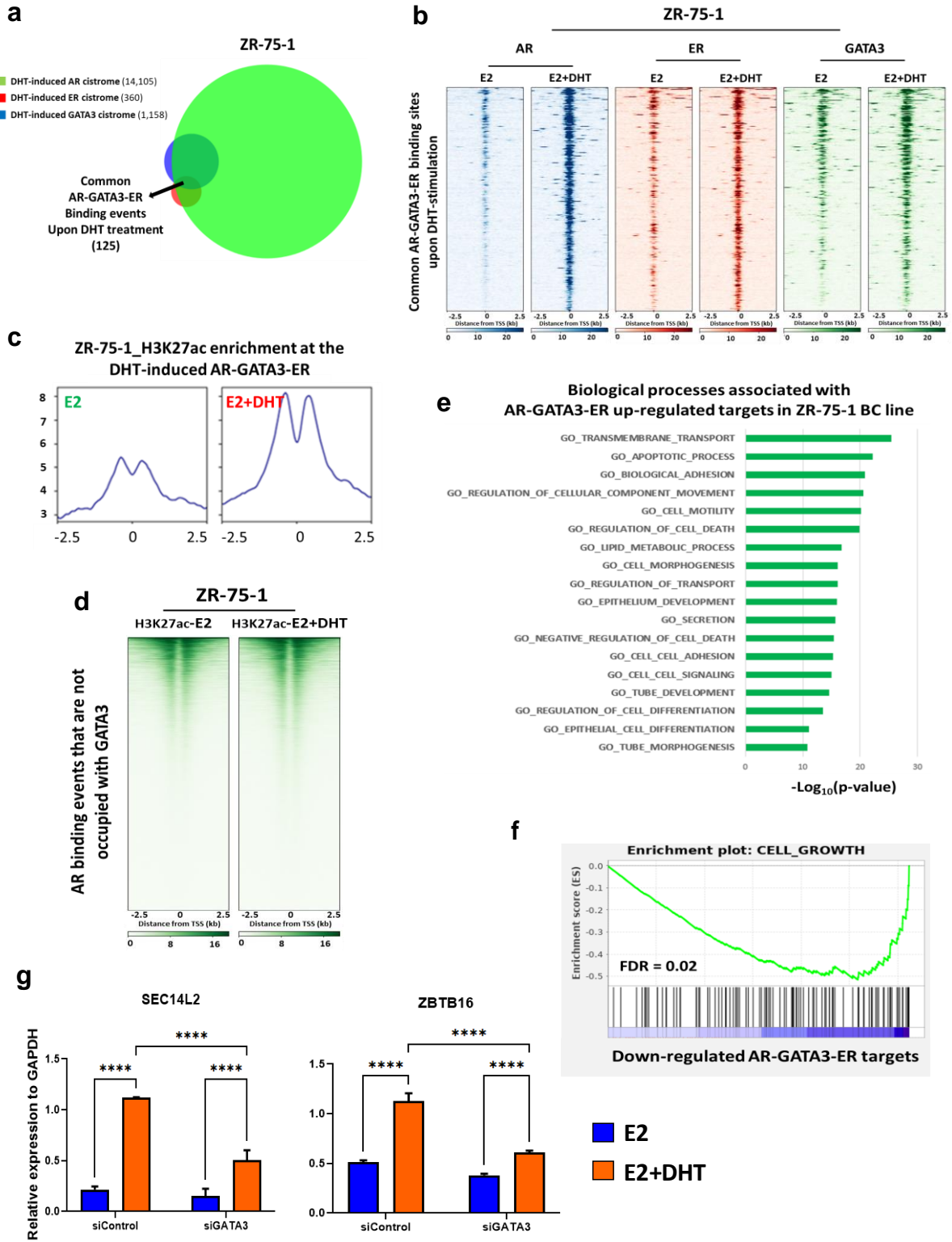
into 5 different sub-groups: grey and yellow dots represent AR and GATA3 binding sites that do not overlap with the other factor, respectively; blue dots represent E2+DHT-induced GATA3 peaks that are not shared with AR; green dots represent GATA3 binding

sites that were not altered by E2+DHT stimulation but AR was recruited to these loci; and pink dots represent AR-GATA3 peaks that are significantly gained upon E2+DHT stimulation. Example loci associated with AR-GATA3-ER target genes (*RANBP3L* and *CA12*

represented loci from pink group, *PgR*, *GREB1*, and *MYB* represented loci from green group) that are highlighted in black dots in ZR-75-1 cells. **h)** Venn diagram depicts the

overlap of E2+DHT-induced GATA3 binding sites with E2+DHT-induced AR events in ZR-75-1 cells.

# Supplementary 4: AR-induced ER-reprogramming is mediated through a switch in GATA3 co-regulatory function



**Supplementary-4: AR-induced ER-reprogramming is mediated through a switch in GATA3 co-regulatory function.** **a)** Venn diagram shows the overlap DHT-induced AR, GATA3, and ER cistromes in ZR-75-1 cells. **b)** Heatmap represents the contribution of DHT-induced GATA3 peaks on DHT-induced shared AR-ER binding sites in ZR-75-1. **c)** Read density plots represent the H3K27ac enrichment in E2- and E2+DHT-treated ZR-75-1 cells. **d)** Heatmap show the density of H3K27ac enrichment in at E2+DHT-induced AR sites depleted from GATA3 compared with the E2 stimulation condition in ZR-75-1 cells. **e)** Bar graph shows the biological pathways that are associated with up-regulated AR-GATA3-ER targets upon E2+DHT treatment in ZR-75-1 cells. **f)** GSEA analysis represents that down-regulated AR-GATA3-ER targets are associated with down-regulation of cell growth in ZR-75-1 cells. **g)** Bar graphs represent the effect of GATA3 knock-down on the mRNA expression level of *SE14L2* and *ZBTB16* tumour-suppressor AR targets under E2+DHT-stimulation compared with the E2-treated cells in ZR-75-1 breast cancer model. Two-way ANOVA with Tukey's multiple comparisons test was used to determine statistically significant differences in **f**. Data shown as mean  $\pm$  SEM of 3 independent experiments; \*\*\*\*  $p < 0.0001$ .



**Supplementary 5: Shared DEG AR signatures with AR-GATA3 targets in T-47D cells**

Up-regulated AR signature genes shared with AR-GATA3-ER targets	Down-regulated AR signature genes shared with AR-GATA3-ER targets
<ol style="list-style-type: none"> <li>1. AKR1C3</li> <li>2. AKR1D1</li> <li>3. ALDH1A3</li> <li>4. C1orf21</li> <li>5. EAF2</li> <li>6. ECHDC2</li> <li>7. ELOVL5</li> <li>8. ENPP3</li> <li>9. FAR2</li> <li>10. GDPD3</li> <li>11. HPGD</li> <li>12. MTSS1</li> <li>13. MUC15</li> <li>14. PIP</li> <li>15. S100P</li> <li>16. SCP2</li> <li>17. SH3RF2</li> <li>18. SLC9A2</li> <li>19. SMPDL3A</li> <li>20. THRSP</li> <li>21. ZNF385B</li> <li>22. ZNF689</li> <li>23. CLDN8</li> <li>24. EMP1</li> <li>25. FMO5</li> <li>26. MCCC2</li> <li>27. MYBPC1</li> <li>28. SEC14L2</li> <li>29. STEAP2</li> <li>30. TPD52</li> <li>31. ZBTB16</li> </ol>	<ol style="list-style-type: none"> <li>1. ABI3BP</li> <li>2. ASPM</li> <li>3. CCNA2</li> <li>4. CKAP2L</li> <li>5. E2F2</li> <li>6. E2F8</li> <li>7. EXO1</li> <li>8. FAM196A</li> <li>9. FAM83D</li> <li>10. IQCA1</li> <li>11. ITPR1</li> <li>12. LMNB1</li> <li>13. LYPD6B</li> <li>14. ME3</li> <li>15. MGP</li> <li>16. RAPGEF4</li> <li>17. RTN1</li> <li>18. TPRG1</li> <li>19. ASB13</li> <li>20. CTNND2</li> <li>21. RAB37</li> <li>22. RERG</li> <li>23. SLC30A8</li> </ol>

**Supplementary-5: Shared DEG AR signatures with AR-GATA3 targets in T-47D cells.** Table depicts the up- and down-regulated AR signature genes that are common with AR-GATA3 targets in T-47D breast cancer cell line model.

**Supplementary 6: List of antibodies, siRNAs and primers used in this study**

**a**

Western blotting				
	Company	Cat numbers	Spices	Dilluton factor
<b>GATA3</b>	Abcam	ab199428	Rabbit	1 in 500
<b>GAPDH</b>	Merck Millipore	mab374	Mouse	1 in 2000

**b**

	Scene	Anti-scene	Company
<b>siGATA3-JC-1</b>	AAACUAGGUCUGAUUUCUU	UGAAUAUCAGACCUAGUUUU	Sigma
<b>siGATA3-JC-2</b>	CUUUUUUGCAUCUGGGUAG	CUACCCAGAUGCAAUAAAG	Sigma
<b>siControl</b>	Cat No: 1027281	Lot No: 284526154	Allstar-Neg. Control siRNA

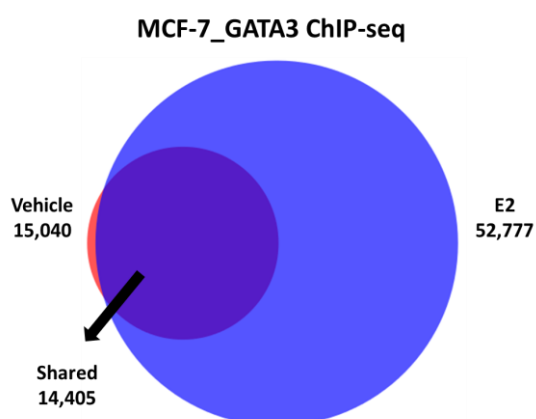
**c**

	Forward	Reverse	Metling TM	Company
<b>SEC14L2</b>	GCCGAATCCAGATGACTATTTTCT	GATGTTGTCAATGTCCTTTTGCTT	55	Sigma
<b>ZBTB16</b>	GAGATCCTCTTCCACCGCAAT	CCGCATACAGCAGGTCATC	55	Sigma
<b>PgR</b>	ACAACCACCAGTCACCTTCC	CCATCCACACGCTCCAGAAT	55	Sigma
<b>GREB1</b>	CATCTCTGCCCTTTGAAA CAAAA	GGGCATCACCCGAAACAAG	55	Sigma

**d**

ChIP-seq				
	Company	Cat numbers	Spices	Volume
<b>GATA3</b>	Abcam	ab199428	Rabbit	5ug / IP

**e**



Theodorou *et al.*, *Genome Res*, 2013

## **Chapter-4: General discussion**

## 4. Chapter-4: General discussion

### 4.1. Introduction

The majority of breast cancers are dependent on estrogen for their growth and approximately 80 % of tumours express the ER, which is the main nuclear transcription factor that mediates the oncogenic effects of estrogens (Lim et al. 2016). ER elicits its transcriptional activity within a transcriptional complex comprised of multiple factors (Mohammed et al. 2013; Papachristou et al. 2018). In the recent past, the GATA3 transcription factor has been characterized as an important regulator of ER signalling in ER+ breast cancer that functions in collaboration with the ER transcriptional complex at multiple loci of the genome to regulate ER activity (Theodorou et al. 2013). GATA3 has an important role in mediating enhancer accessibility at regulatory regions involved in ER-mediated transcription (Theodorou et al. 2013). GATA3 is expressed exclusively in the luminal epithelial cells of the mammary gland (Shackleton et al. 2006) and plays an essential role in mammary development and specification (Kouros-Mehr, Bechis, et al. 2008; Kouros-Mehr et al. 2006). In the adult mammary glands of mice, GATA3 expression is necessary to maintain the differentiated state of luminal epithelial cells after completion of mammary development (Kouros-Mehr et al. 2006). The majority of breast cancers also express AR (Hickey, TE et al. 2021). AR has been shown to inhibit estrogen-induced growth of ER+ breast cancers (Hickey, TE et al. 2021) and its expression is associated with improved disease-free survival and more benign clinical and pathologic factors (e.g. lower tumor grade and smaller tumor) in ER- BC (Hu, XQ et al. 2017; Luo et al. 2010). ER and AR share co-regulatory proteins including FOXA1 (Robinson, JL & Carroll 2012), SRC-3 and P300 (Hickey, TE et al. 2021) supporting my hypothesis that since GATA3 is an important factor in ER signalling, it might also play key role in AR signalling in breast cancers. To date, the

role of GATA3 in AR signalling in the context of BC, in the presence or absence of ER, has not been investigated. Therefore, in my PhD thesis I explored the role of GATA3 in AR signalling using a variety of models representing different breast cancer subtypes to determine whether GATA3 is important for AR-mediated growth inhibition of ER+ breast cancer subtypes and to explore its role in the context of ER- breast cancer. In summary, my findings show that GATA3 is a novel AR interacting protein in normal and malignant breast epithelial cells and co-operates with AR in driving luminal-lineage identity in ER+ and ER- breast cancer cell lines. My findings provide a mechanistic explanation for the association between AR and luminal gene profiles in clinical breast cancers. Also, my data provides novel insight into the cross-talk between ER, GATA3, and AR in ER+ BC, highlighting the signalling complexity of these TFs in this disease. Findings of this thesis show the involvement of GATA3 in AR-mediated tumour-suppression in ER+ breast cancer contexts, which in turn highlights the role of common co-regulators like GATA3 in mediating AR and ER cross-talk in breast cancer progression.

## **4.2. Major findings of this thesis**

### **4.2.1. GATA3 is a novel AR interacting protein independent of ER expression in normal mammary tissues and different BC subtypes.**

AR functions within a transcriptional complex along with several co-regulatory proteins that are implicated in a diverse number of cellular functions, however, no study, to date has investigated the AR interacting proteins in different breast cancer subtypes. Therefore, for the purpose of this thesis, I used an unbiased proteomic technique called RIME (Mohammed et al. 2016) to identify proteins that interact with AR while bound to chromatin and discovered GATA3 as a novel AR interacting factor across different breast

cancer subtypes (Chapter-2). The interaction of AR and GATA3 transcription factors was confirmed through PLA and Co-IP techniques in both ER+ and ER- cell lines and clinical samples of ER+ and ER- breast cancer (Chapter-2). Additionally, my findings show that AR activation with DHT induces nuclear translocation both AR and GATA3 compared with the vehicle-treated cells in ER+ and ER- breast cancer models, which suggests that AR and GATA3 function is a complex in breast cancer (Chapter-2). I also observed AR and GATA3 interactions in normal breast clinical samples, which suggests a key role for the AR-GATA3 complex in normal mammary gland function (Chapter-2). Taken together, studying a comprehensive catalogue of AR interacting proteins in various breast cancer cell models provided a basis for the in-depth and quantitative characterisation of protein interactome dynamics of AR complex networks deciphering the mechanisms involved in this disease. Identifying GATA3 as an AR interacting protein, possibly within a functional complex with AR, suggests a new insight into the complexity of the AR transcription factor networks in breast cancer. Cooperative action of transcription factors creates complex gene regulatory networks to maintain cell characteristics. Disruption of these regulatory networks is often associated with human diseases such as cancer. For instance, alteration to the expression and activity of AR co-regulators including FOXA1 is an important mechanism in disease progression in both prostate (Paltoglou et al. 2017) and breast cancer (Robinson, J et al. 2011). AR plays oncogenic functions in male oesophageal squamous cell carcinoma (ESCC). In a very recent research, GATA3 has been shown to form a complex with ligand-activated AR on chromatin in ESCC cell line models, increased recruitment of corepressors SMRT and HDAC3 leading to pronounced repression of tumor suppressor genes (e.g., DUSP4 and FOSB) and enhanced ESCC growth (Huang, et al. 2021). Also, GATA3 positivity is highly correlated with AR expression regardless of ER presence in breast cancer tumours (Boto &

Harigopal 2018; Kim et al. 2016). Therefore, further investigations around the importance of AR and GATA3 interactions in breast cancer may reveal how any alteration in GATA3 expression, as an AR co-regulator, would affect the AR signalling in various breast cancer contexts.

#### **4.2.2. Stimulation of breast cancer cells with estrogen or androgen hormones reprograms the GATA3 cistrome in breast cancer cell line models.**

It has never been reported previously whether hormone stimulation (estrogens and/or androgens) may affect the GATA3 binding profile and function in breast cancer. Therefore, in order to explore the possible effect of hormone stimulations on the GATA3 cistrome, I mapped the genome-wide binding profile of GATA3 through ChIP-seq experiments under different hormone treatment conditions following a period of hormone depletion, including vehicle, E2, DHT and E2+DHT treatments, in two ER+ and two ER- breast cancer cell line models (Chapter-2 and 3). I also performed GATA3 ChIP-seq in two ER+ PDX models of breast cancer that had been treated with vehicle or an AR agonist (Chapter-2). My findings clearly show that estrogen shifts the GATA3 cistrome towards a new subset of binding sites compared with the vehicle-treated cells in ER+ breast cancer cell line models, T-47D and ZR-75-1 (Chapter-3). Also, AR-agonist induction of ER+ and ER- breast cancer cell lines and ER+ PDX breast tumours show a potent effect of AR agonists (DHT or Enobosarm) in redistributing the GATA3 cistrome in these models (Chapter-2 and 3). Interestingly, GATA3 ChIP-seq data associated with E2+DHT treatment in ER+ breast cancer cells, T-47D and ZR-75-1, compared with the E2-treated cells show that AR activation rearranges the E2-induced GATA3 binding events in these models (Chapter-3),



demonstrating an antagonistic effect. Generally, my findings, for the first time, demonstrate that GATA3 binding profile is hormone regulated (by estrogens (Chapter-3) and androgens (Chapter-2)) in breast cancer context. Therefore, this would shed a light into the fact that there may be a same hormonal effect on chromatin binding preferences of other AR and / or ER co-regulators in breast cancer, which mediate the signalling and function of the master regulators driving breast cancers. Also, this highlights the importance of selecting appropriate models and having precise experimental design for investigating transcription factors function in a complex in hormone-regulated cancers like breast disease.

#### **4.2.3. GATA3 is bound and plays a regulatory role at genomic loci associated with AR-mediated growth inhibition in ER+ BC to inhibit E2-stimulated growth.**

My data show that the E2-induced GATA3 cistrome has a high degree of overlap with the E2-induced ER cistrome, an association that is important for regulation of genes associated with cell cycle progression and growth of ER+ breast cancer cell (Chapter-3). However, dual-receptor activation (ER and AR) with E2+DHT treatment clearly shifted E2-induced GATA3-ER binding events in ER+ breast cancer cells away from growth regulatory loci. These data suggest that the potent interplay between ligand-activated AR and GATA3 facilitates the ability of AR signalling to antagonise ER signalling in ER+ breast cancer cells (Chapter-3). Interestingly, AR-GATA3-ER co-occupied loci are associated with up-regulated targets upon E2+DHT treatment, and were positively enriched in gene sets associated with epithelium development, cell-cell signalling, and regulation of cell death pathways. In contrast, down-regulated AR-GATA3-ER gene targets were negatively associated with cell

growth gene sets in both of the ER+ cell line models. Also, my data show that AR-ER loci co-occupied with GATA3 are located in transcriptional active chromatin upon DHT treatment, however, loci depleted from GATA3 does not show enrichment of H3K27ac, an indicative of active sites. Together, my findings suggest that AR activation relocates GATA3 away from shared ER binding sites associated with proliferative genes, and induces a new subset of binding events that are common between all three transcription factors, implicating GATA3 in AR-driven tumour suppression in ER+ breast cancer cells (Chapter-3), possibly through facilitating AR DNA binding.

#### **4.2.4. A collaborative role for GATA3 and AR in promotion of luminal epithelial differentiation in breast cells regardless of ER expression.**

My findings show that the majority of GATA3 binding events induced by AR activation were co-occupied by AR and associated with an increase in H3K27ac signal, indicative of an active transcriptional state of chromatin at these loci (Chapter-2). This finding is consistent with the effect of AR activation on interaction and nuclear translocation of GATA3 in breast cancer cells across all ER+/ER- cell line models (Chapter-2). Common AR-GATA3 active loci were associated with genes within biological gene sets involved in development and differentiation of mammary epithelial tissue including mammary gland development, mammary gland duct morphogenesis and epithelial cell differentiation (Chapter-2). Performing GATA3 ChIP-PCR following AR knock-down showed that silencing AR abolished DHT-induced enrichment of GATA3 at representative loci, indicating that GATA3 binding at these AR target genes is dependent on AR. The reciprocal experiment with GATA3 knock-down, showed that AR binding was decreased but not abolished at these loci, indicating that GATA3 facilitates, but is not critical for AR binding

(Chapter-2). Accordingly, GATA3 knock-down significantly reduced expression levels of the representative AR target genes, indicating that the reduction of AR binding at associated loci had a functional impact on AR transactivation capacity. Investigation of AR-GATA3 regulated targets revealed an association with up-regulation of luminal epithelial markers and down-regulation basal epithelial markers, indicative of luminal phenotype promotion (Chapter-2). Mechanistically, my data revealed that DHT stimulation induces AR-mediated translocation of GATA3 into the nucleus, resulting in a gain of new GATA3 binding sites, where GATA3 co-regulates AR function in promoting the expression of luminal epithelial driver genes and repressing the expression of the basal markers in breast cancer cells (Chapter-2).

## **4.3. Future directions**

### **4.3.1. Investigating the structural basis of AR and GATA3 protein-protein interaction**

Protein-protein interactions (PPIs) are physical contacts of high specificity established between two or more protein molecules that play a key role in predicting the function of a particular target protein and the ability to drug it (Ivanov, Khuri & Fu 2013). Thus, pathway perturbation, through the disruption of PPIs critical for cancer, offers a novel and effective strategy to curtail the transmission of oncogenic signals (Ivanov, Khuri & Fu 2013). As our understanding of cancer biology has significantly increased in recent years, interest in targeting PPIs as anti-cancer strategies has increased as well. This thesis provided the first data demonstrating the AR-GATA3 interaction in breast cancer, however, it did not investigate whether this is a direct interaction without having any bridging factor in between. Also, I did not investigate the details of AR-GATA3 protein interactions to assess which domains of these factors are required for their protein interaction properties. In addition, I did not explore the possible effect of GATA3 mutations in disrupting AR-GATA3 interaction and its consequences in breast cancer progression. Therefore, it would be critical to further investigate the basis of AR-GATA3 interactions and possible effect of GATA3 mutations on this interaction in order to better understand the role of AR-GATA3 interactions breast cancer biology.

### **4.3.2. AR-GATA3 function in normal mammary glands**

AR has been suggested to increase in the population of luminal MECs in mouse mammary glands treated with Enobosarm (Tarulli et al. 2019). Also, a significant increase

in the proportion of p63+ basal MECs and decrease in ER+ luminal MECs was observed in AR-null mammary epithelium of mice, suggesting an alteration in the balance of MEC differentiation (Tarulli et al. 2019). AR transcription factor is shown to be enriched in breast cancer cells that were engrafted in the mammary gland through intraductal injection that maintains luminal profiles in cancer cells (Sflomos et al. 2016). Also, in human breast cancer, AR signalling has been shown to promote the maintenance of a luminal state. Additionally, AR lineage tracing studies in mouse prostate identified AR expressing basal cells that are required for re-population of the luminal niche after regression (Xie et al. 2017). In addition, GATA3 express in the differentiated luminal epithelial cells lining the breast ductal structures and associated with luminal cell identity and function in mammary glands (Kouros-Mehr et al. 2006; Kouros-Mehr & Werb 2006). Interestingly, my data confirm that the AR-GATA3 interaction occurs in normal clinical breast tissues as well as in malignant breast cancer cells, which suggests a key role for the AR-GATA3 complex in normal breast biology. According to the important role of AR-GATA3 complex in promoting the luminal phenotype in breast cancer context, I propose that their interaction in normal mammary would be critical in inducing the luminal cell fate determination of the breast cells. I suggest to investigate the details of AR-GATA3 interactions during mammary gland development using *MMTV-Cre* mice models that may lack one or two copy(ies) of their *GATA3* and / or *AR* genes and study that how would deletion of each factor influence the population of both alveolar and ductal progenitor cells compared with the basal progenitors and ultimately affect the mammary cell fate.

### 4.3.3. Investigating the effect of GATA3 depletion on mRNA expression level of basal lineage targets

My findings identified numerous up-regulated luminal epithelial markers as well as down-regulated basal marker genes, indicative of luminal phenotype promotion within transcriptionally active regions of chromatin at common AR-GATA3 regulated loci. My data showed that silencing AR significantly abolished DHT-induced enrichment of GATA3 at representative loci associated with luminal genes in both ER+ and ER- breast cancer models, and that silencing GATA3 negatively impacted the DHT-induced enrichment of AR binding at the selected luminal representative loci. Also, knock-down of AR or GATA3 significantly reduced the DHT-induced mRNA expression level of luminal genes *in vitro*. However, due to the lack of reagents, and COVID-19 related lab restrictions, I did not have a chance to expand my ChIP-PCR and qPCR experiments to examine the effect of AR and GATA3 knock-down on binding enrichment of the other factor and on the mRNA expression levels of their basal target genes. However, I believe this would be critical to be able to make a definite conclusion with regard to the role of the AR-GATA3 complex in promotion of luminal phenotype in breast cancer context. Although I could not process the experiments related to the basal genes, I expect to see that AR knock-down abolishes DHT-induced enrichment of GATA3 at basal representative loci and also that GATA3 knock-down negatively impacted the DHT-induced enrichment of AR binding at the selected basal representative loci in both ER+ and ER- breast cancer models. Also, I expect that silencing AR or GATA3 would increase mRNA expression level of basal makers in both ER+ and ER- breast cancer models.

#### **4.3.4. Exploring the role of GATA3 in cytokinesis and cell cycle progression of breast cancer cells**

While assessing the effect of GATA3 knock-down on the growth-stimulatory and / or growth-inhibitory effect of estrogens and androgens in ER+ breast cancer cell proliferation, respectively, I observed dramatic morphology changes in both T-47D and ZR-75-1 ER+ breast cancer cells upon GATA3 silencing, causing multinuclear cells in a common cytoplasm. Failure to complete cytokinesis has been proposed to promote tumorigenesis (Fujiwara et al. 2005; Ganem, Storchova & Pellman 2007; Steigemann et al. 2009) by leading to tetra-ploidy and ensuing chromosomal instability (Caldwell, Green & Kaplan 2007; Fujiwara et al. 2005; Ganem, Storchova & Pellman 2007). For instance, GATA6 is another member of the GATA family of transcription factors that has been shown to function in early embryonic stem cell differentiation (Cai et al. 2009). GATA6 is usually lost in ovarian cancer and leads in cytokinesis failure, deformation of the nuclear envelope and formation of polyploid and aneuploid cells (Cai et al. 2009). Thus, GATA6 is also linked to tumorigenesis. Also, the cytokinesis regulator FYVE-CENT is usually mutated in breast cancers (Sjöblom et al. 2006). Depletion of FYVE-CENT results in an increased number of binuclear and multinuclear profiles, as well as cells arrested in cytokinesis, indicating its important role in cytokinesis (Sagona et al. 2010). Therefore, I suggest that depletion of GATA3 in these ER+ breast cell line models caused a failure in completion of normal cytokinesis, which in turn led to polynuclear giant cells irrespective to the hormone treatment condition of the cells. I believe that it would be crucial to further investigate the role of GATA3 in normal cytokinesis and cell cycle progression in breast cancer.

#### 4.3.5. Assessing the effect of GATA3 ectopic expression in ER+ and ER- breast cancer cells

I have generated nine doxycycline-inducible ER+ (T-47D and ZR-75-1) and ER- (MDA-MB-453) breast cancer lines, that are able to over-express wild-type or T308 mutant GATA3 protein as well as cells expressing the negative control insertion (Appendix-5.1.1). T308 GATA3 mutation is a truncation mutation in ZnFn2 domain of the gene, that can recapitulate the clinical consequences of other mutations with a similar truncated protein product like what we see in MCF-7 cells. According to the recent large-scale genomic profiling of breast tumors, frequent mutations has been identified in GATA3 (Pereira et al. 2016). More than 10 % of breast tumors carry GATA3 mutations (Ciriello et al. 2015). However, it is having been difficult to specifically investigate the role of mutant GATA3 due to lack of specific antibodies that exclusively pick the mutant form away from the wild-type GATA3 protein (e.g. in MCF-7 cells and primary tumours). Currently, available GATA3 ChIP specific antibodies detect both mutant and wild-type forms. R330fs GATA3 (a heterozygous frameshift mutation at arginine 330 of ZnFn2 domain of GATA3 and naturally exists in MCF-7 cells) mutant disrupts the cooperative action with ER, FOXA1, and rearrange ER and FOXA1 chromatin localization, which is associated with altered chromatin architecture and differential gene expression in GATA3 mutant cells (Takaku et al. 2020). Although, I did not get a chance to investigate the role of ectopic expression of wild-type and T308 mutant GATA3 in association with breast cancer progression, I strongly believe that my generated transduced breast cancer cell lines are among the best models for future mutation-related studies including:



- a) Investigating the effect of mutant GATA3 on its interactions with other transcription factors (e.g. AR, ER), AR signalling, AR binding profile, and AR function in different breast cancer contexts,
- b) Assessing the effect of mutant GATA3 on shaping the GATA3 cisome and its functional consequences in breast cancer cells in response to ER and or AR activation,
- c) Examining the effect of over-expression of wild-type and / or mutant GATA3 on growth rate of ER+ and ER- breast cancer cells.

#### **4.4. Concluding remarks**

Taken together, this study aimed to reveal the role of GATA3 in AR signalling in the context of breast cancer, and to characterise the cross-talk between AR and GATA3 in the presence or absence of activated ER. My data show a cooperative role for AR and GATA3 in suppressing estrogen-induced tumour growth of ER+ breast cancer and in driving the luminal-lineage identity in normal mammary glands and all breast cancer contexts. Also, my findings provide novel insight into the complexity of ER, GATA3, and AR cross-talk in breast cancer, preparing a basis for further investigations into AR and GATA3 co-operative genomic activity and the direct consequences of this on breast cancer progression and differentiation.

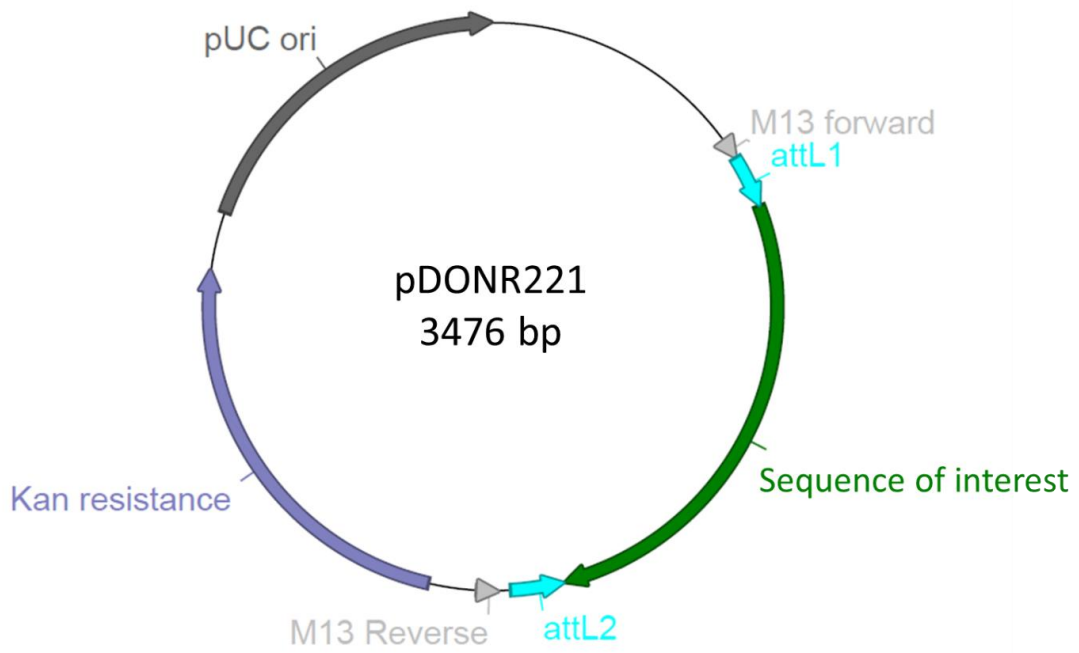
## Chapter-5: Appendix

## **5. Chapter-5: Appendix**

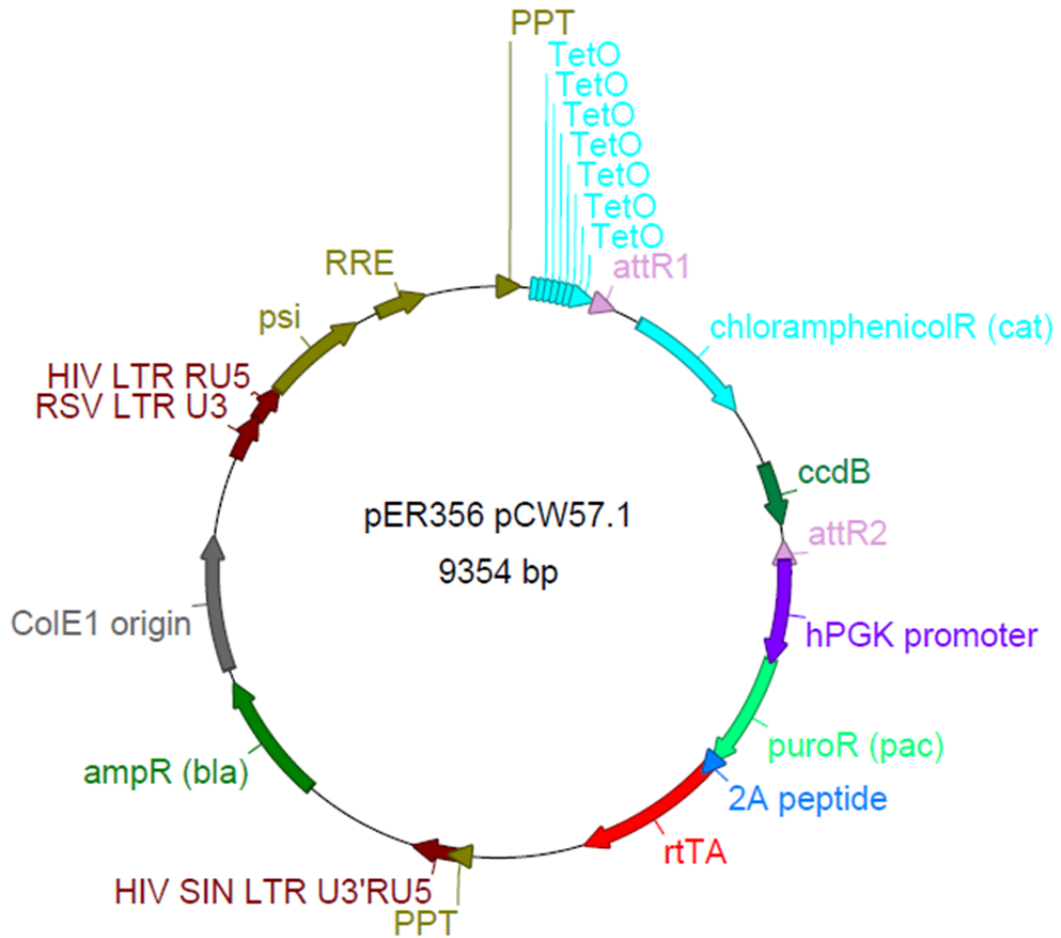
### **5.1. Appendix – Generating stable knock-down of factors in breast cancer cell lines**

#### **5.1.1. Appendix-1: Plasmid constructs**

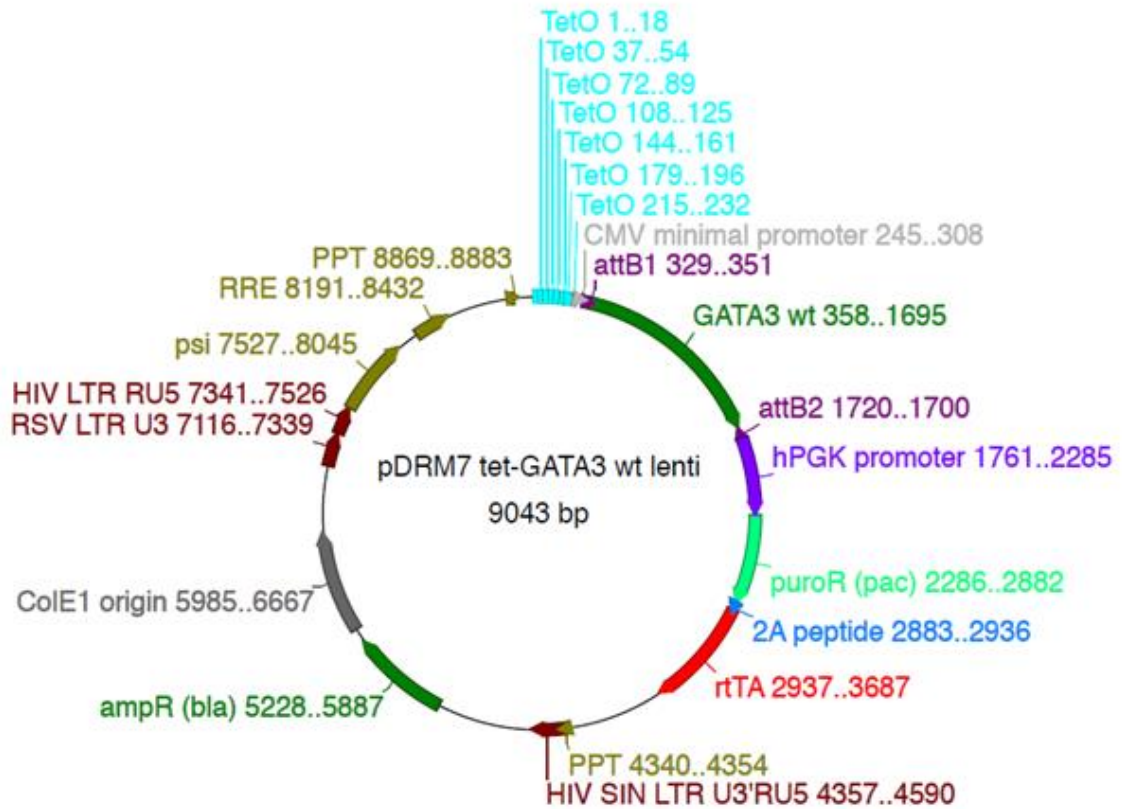
The composition and construction of plasmid vectors generated for future experimental purposes related to GATA3 study are described and illustrated on the following pages.



**pDONR221** was the entry clone used in Gateway cloning containing Kan resistance gene (for selection in BUG culture) that accepted our sequence of interest (e.g. wtGATA3, mutGATA3, or the GUS-negative control).



**pER365** was the destination vector used in Gateway cloning containing Amp resistance gene (for selection in BUG culture) and puroR resistance gene (for selection in cell culture) that received our sequences of interest (e.g. wtGATA3, mutGATA3, or the GUS-negative control) from the entry clone.



**pDRM7-tet-GATA3-wt-lenti** was the final construct containing puroR resistance gene (for selection in cell culture) and the sequence of wtGATA3 to over-express the wild-type *GATA3* gene.

➤ Wild-type GATA3 sequence inserted in pDRM7-tet-GATA3-wt-lenti:

```

360      370      380      390      400      410      420      430      440      450
* * * * *
358 atggaggtdgacggcggaccagccgctgggtgagccaccaccaccccgccgtgctcaacgggcagcaccgggacacgaccaccgggctcagccact 457
M E V T A D Q P R W V S H H H P A V L N G Q H P D T H H P G L S H S
1
GATA3 wt

460      470      480      490      500      510      520      530      540      550
* * * * *
458 cctacatggcggcggcagcagcagccgctgcccggagggtggatgagctctttttaaaccatcagcggcgaaggcaaccacgctccggccctactacgaaactc 557
Y M D A A Q Y P L P E E V D V L F N I D G Q G N H V P P Y Y G N S
35
GATA3 wt

560      570      580      590      600      610      620      630      640      650
* * * * *
558 ggtcaggccacgggtgcagagggtaccctccgaccaccacgggagccagggtgtccgcccgcctctgttccatggatccctaccctggctggagggggc 657
V R A T V Q R Y P P T H H G S Q V C R P P L L H G S L P W L D G G
68
GATA3 wt

660      670      680      690      700      710      720      730      740      750
* * * * *
658 aaagccctgggcagccaccacccgctccccctggaaatctcagccccttctccaagacgtccatccaccacggcctccccggggccctctccgctctacc 757
K A L G S H H T A S P W N L S P F S K T S I H H G S P G P L S V Y P
101
GATA3 wt

760      770      780      790      800      810      820      830      840      850
* * * * *
758 cccgggctctgtctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 857
P A S S S S L S G G H A S P H L F T F P P T P P K D V S P D P S L
135
GATA3 wt

860      870      880      890      900      910      920      930      940      950
* * * * *
858 gtccaccccagcctcggcggcctcggcccgccaggaagagagtgccctcaagtaccaggtgcccctgcccagacagcatgaagctggagctgctccca 957
S T P G S A G S A R Q D E K E C L K Y Q V P L P D S M K L E S S H
168
GATA3 wt

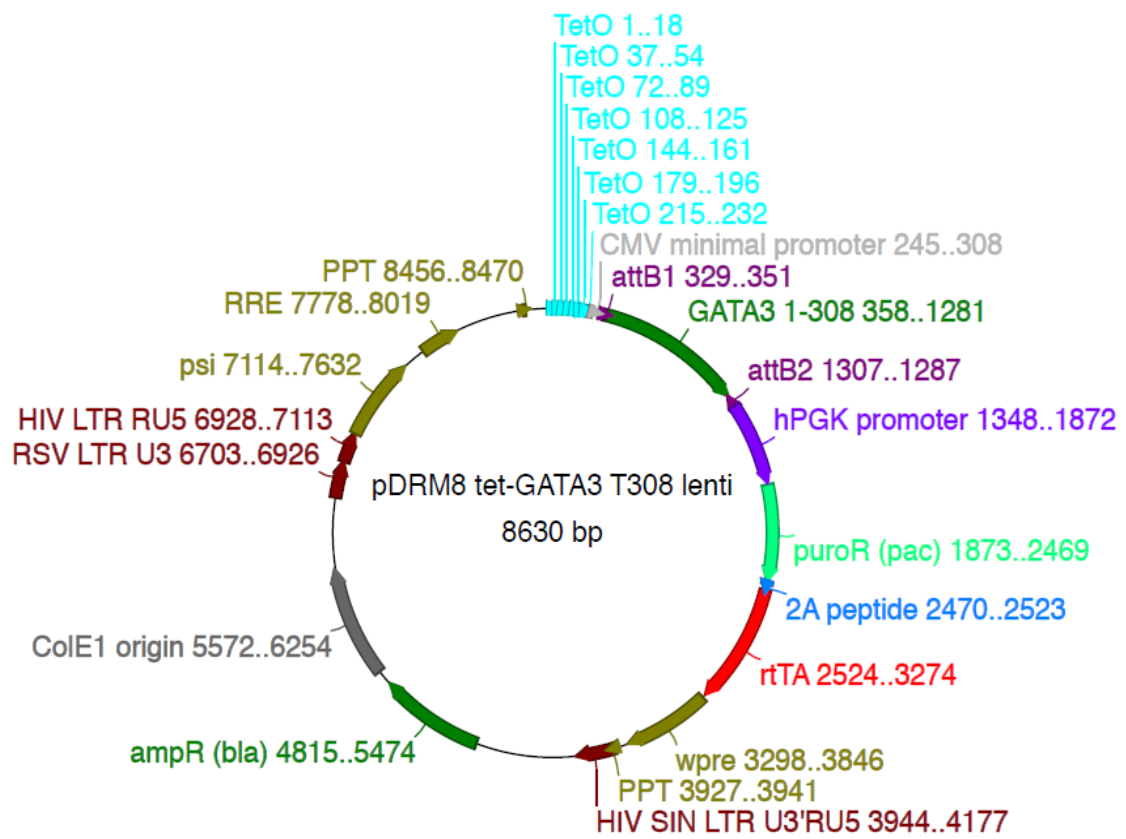
960      970      980      990      1000      1010      1020      1030      1040      1050
* * * * *
958 cccctgdcacatgacccggccgggtggaggccctcctcgtggagccaccaccaccatcaccaccctaccgctccctaccggtgcccggatcagagctccggactc 1057
S R G S M T A L G G A S S S T H H P I T T Y P P Y V P E Y S S G L F
201
GATA3 wt

1060      1070      1080      1090      1100      1110      1120      1130      1140      1150
* * * * *
1058 ccccccagcagcctgctgggggggtccccaccggctccggatgcaagtcaggcccaaggcccggtccagcacagaagccagggagctgtgaaactc 1157
P P S S L L G G S P T G F G C K S R P K A R S S T E G R E C V N C
235
GATA3 wt

```







**pDRM8-tet-GATA3-T308-lenti** was the final construct containing puroR resistance gene (for selection in cell culture) and the sequence of mutGATA3 to over-express the mutant GATA3 gene.

➤ T308 mutant GATA3 sequence inserted in pDRM8-tet-GATA3-T308-lenti:

```

360      370      380      390      400      410      420      430      440      450
* * * * *
358 atggaggtagcggcggcagccgctgggtgagccaccaccaccgcccgtgctcaacgggcagcaccgggacagcaccaccggggcctcagccact 457
M E V T A D Q P R W V S H H H P A V L N G Q H P D T H H P G L S H S
1
GATA3 1-308

460      470      480      490      500      510      520      530      540      550
* * * * *
458 ctacatggagcggcggcagcaccgctggcgggaggaggatggtccttlltcaacatcagcgggtcaaggccaaccagctccggccctactaaggaaact 557
Y M D A A Q Y P L P E E V D V L F N I D G Q G N H V P P Y Y G N S
35
GATA3 1-308

560      570      580      590      600      610      620      630      640      650
* * * * *
558 ggtcagggcccaagggtcagaggttaacctccgagcccaaccaaggagaccaggtgtgcccggccgctctgcttccatggatccctaccctggtggaagggg 657
V R A T V Q R Y P P T H H G S Q V C R P P L L H G S L P W L D G G
68
GATA3 1-308

660      670      680      690      700      710      720      730      740      750
* * * * *
658 aaagccctgggcagccaccacaccgctcccccctgggaatctcagccccttctccaagacgtccatccaccacggctccccggggcccctctccggtctacc 757
K A L G S H H T A S P W N L S P F S K T S I H H G S P G P L S V Y P
101
GATA3 1-308

760      770      780      790      800      810      820      830      840      850
* * * * *
758 cccggcctctgctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 857
P A S S S S L S G G H A S P H L F T F P P T P P K D V S P D P S L
135
GATA3 1-308

860      870      880      890      900      910      920      930      940      950
* * * * *
858 gtcaccccaggctcggcgggctcggccggcaggaagaagaagtgctcgaagtaccaggtgcccctgcccagacagatgaagctggaagctggtcccaac 957
S T P G S A G S A R Q D E K E C L K Y Q V P L P D S M K L E S S H
168
GATA3 1-308

960      970      980      990      1000     1010     1020     1030     1040     1050
* * * * *
958 cccgctggcagcatgaccgcccctgggtggagcctcctcgtcagaccaccaccaccatcaccacctaccggccctacgtgcccaggtacagctccggactct 1057
S R G S M T A L G G A S S S T H H P I T T Y P P Y V P E Y S S G L F
201
GATA3 1-308

1060     1070     1080     1090     1100     1110     1120     1130     1140     1150
* * * * *
1058 cccccccagcagcctgctggcggctccccccaccggctctggatgcaagtccagggccaaggccgggtccagccacagaaggcagggagtggtgaaact 1157
P P S S L L G G S P T G F G C K S R P K A R S S T E G R E C V N C
235
GATA3 1-308

```

```

1160      1170      1180      1190      1200      1210      1220      1230      1240      1250
  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
1158 ggggcaacctcgacccactgtggcggcgagatggcacgggacactacctgtgcaacgcctgctctatcacaaaatgaacggacagaacccggcc 1257
  G A T S T P L W R R D G T G H Y L C N A C G L Y H K M N G Q N R P
268

```

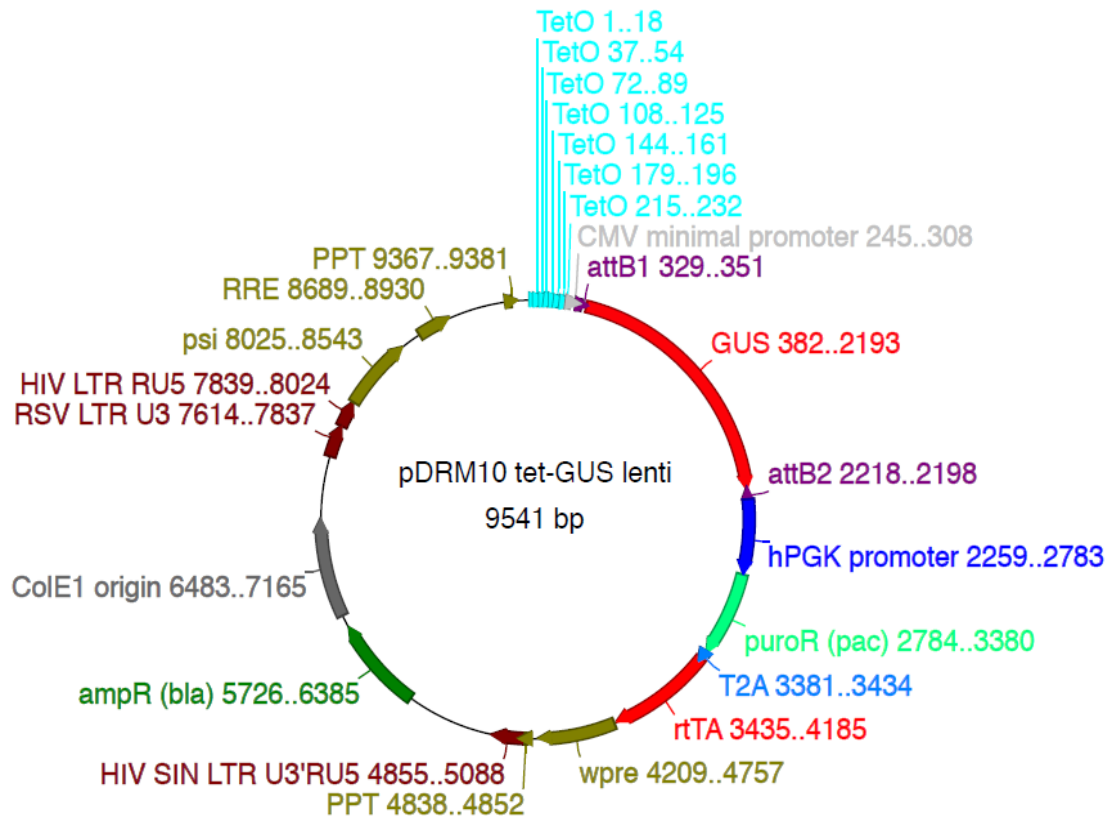
GATA3 1-308

```

1260      1270      1280
  *  *  *  *  *
1258 tcattaaagcccaagcgaaggct 1281
  L I K P K R R L
301

```

GATA3 1-308



**pDRM10-tet-GUS-lenti** was the final construct containing puroR resistance gene (for selection in cell culture) and the GUS sequence as a negative control.





### **5.1.2. Appendix-2: Generating lentiviral transfer plasmid encoding the insert of interest, viral transfection, and transduction in desired cell lines**

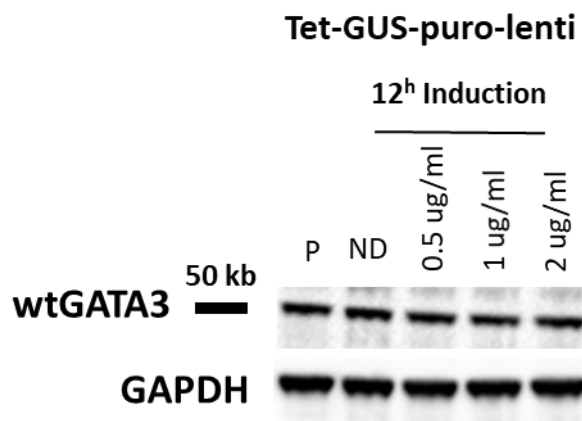
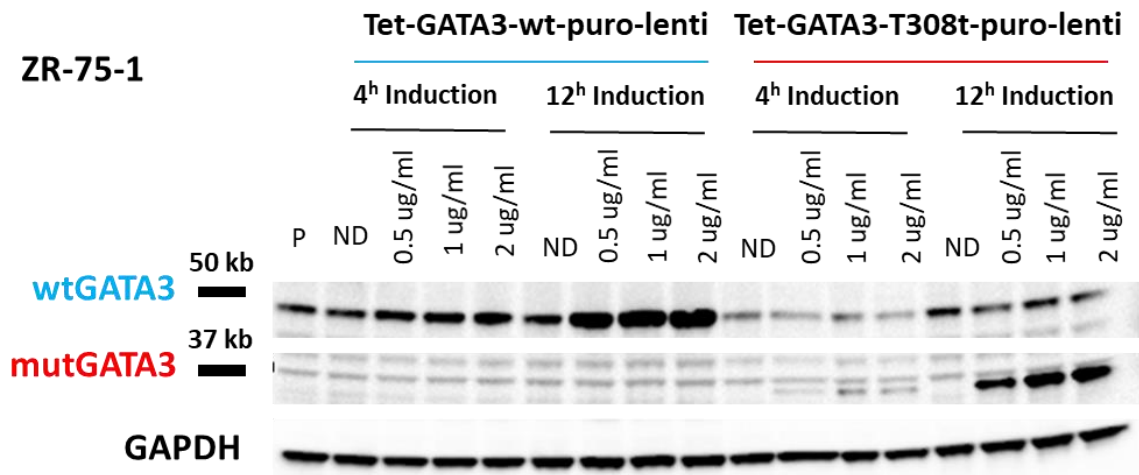
Second generation, 3-package plasmid system was conducted to generate the desired recombinant lentiviruses using the packaging vector of psPAX2, envelope vector of VSVG pSD11 pMD2.G, and the transfer plasmid encoding the insert of interest. HEK-293 cells were transfected with transfer plasmid and packaging plasmids that provides all the viral proteins required for virus particle production within a packaging cell, using the PEI transfection methodology. 10 uM chloroquine was added to the normal media of the HEK-293 cells in order to increase the efficacy of transfection. The media supplemented with transfection mix was changed to Opti-MEM media after 12 hours of incubation at 37 °C. Media containing viral particles was collected after 48 hours, purified and filtered to concentrate the virus.

Lentiviral titration has been done for each cell line to ensure that preps are applied in sufficient quantities to obtain desired transduction efficiencies. Two ER+ (ZR-75-1 and T-47D) and one ER- (MDA-MB-453) breast cancer cell lines were transduced with the appropriate Doxycycline inducible viruses (e.g. wtGATA3, mutGATA3, or the GUS negative control vector). The optimal titration for all the generated viruses in all the breast cancer lines were  $10^5$  IFU/ml (Infectious units (IFU)).

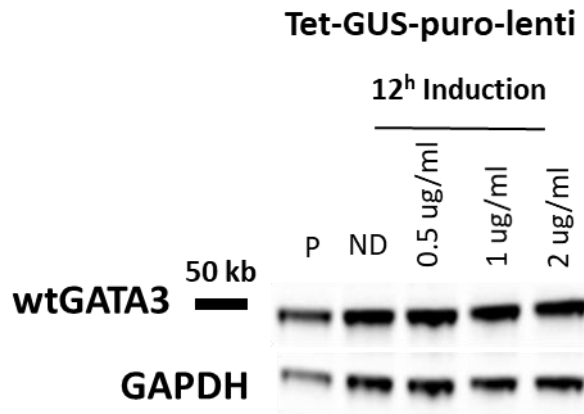
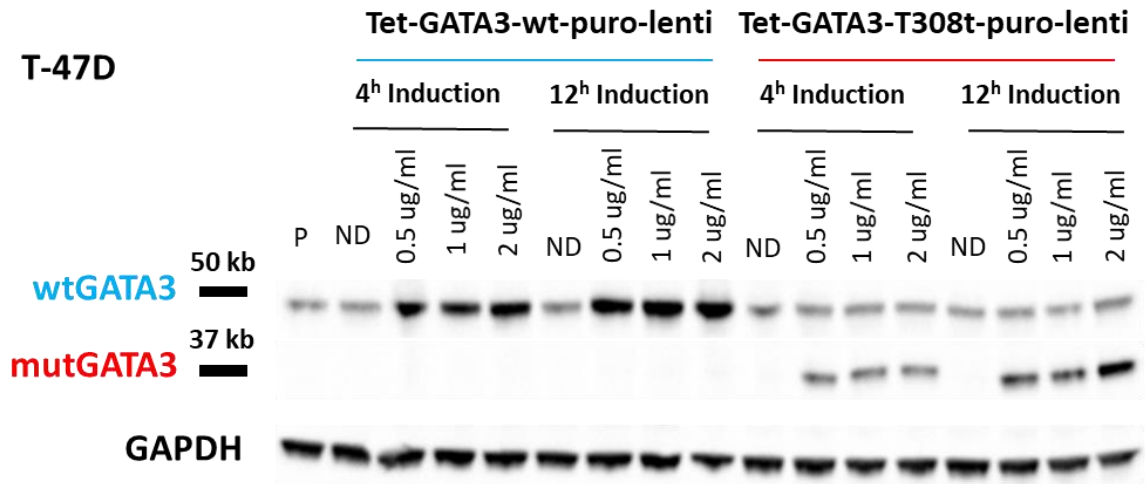
Transduced lines were expanded and selected through culturing in media supplemented with puromycin. Selected cells finally induced with different doses of Doxycycline in various induction timepoints to select the best dose/timepoint for further experiments. The efficacy of induction has been checked through western blotting. ZR-75-1 and MDA-MB-453 breast cancer cells had the most wtGATA3 or mutGATA3 protein

expression level after 12 hours of induction with 1 ug/ml of Doxycycline. T-47D breast cancer cells had the most wtGATA3 or mutGATA3 protein expression level after 12 hours of induction with 1 ug/ml of Doxycycline for the wild-type and 2 ug/ml of Doxycycline for the mutant.

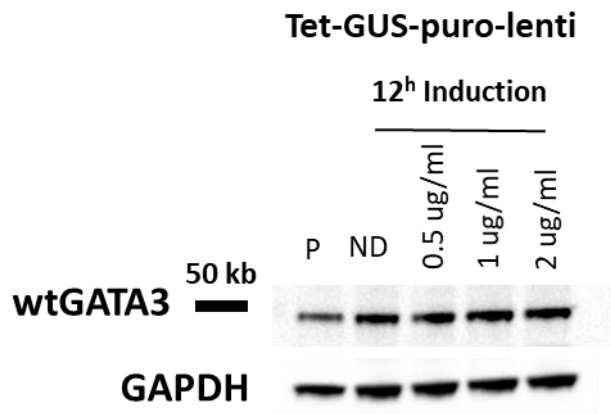
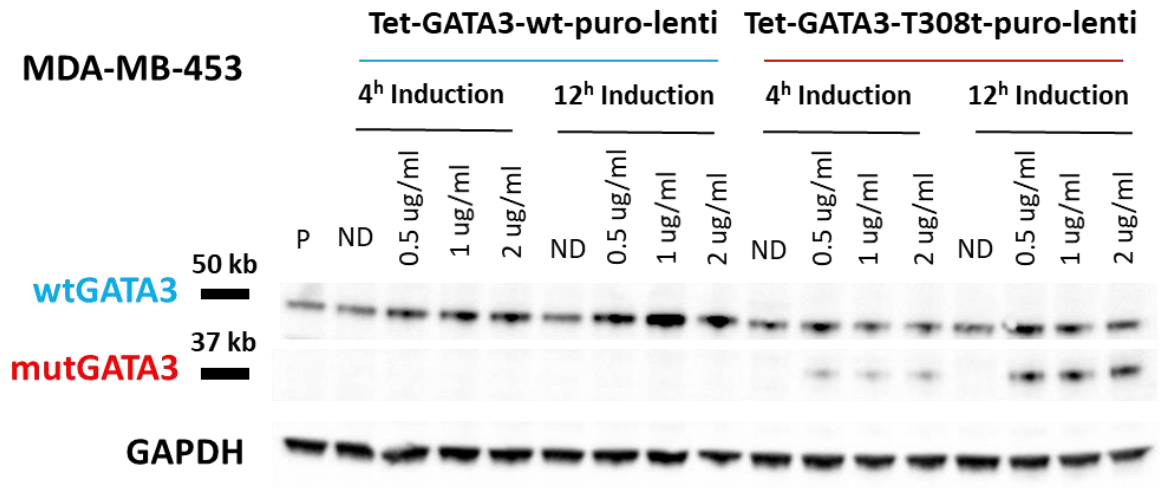




- ND: no DOX induction



- ND: no DOX induction



- ND: no DOX induction

## Bibliography

(WHO), WHO 2020, *Breast cancer now most common form of cancer: WHO taking action*, <<https://www.who.int/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action>>.

Adomas, AB, Grimm, SA, Malone, C, Takaku, M, Sims, JK & Wade, PA 2014, 'Breast tumor specific mutation in GATA3 affects physiological mechanisms regulating transcription factor turnover', *BMC Cancer*, vol. 14, Apr 22, p. 278.

Agoff, SN, Swanson, PE, Linden, H, Hawes, SE & Lawton, TJ 2003, 'Androgen receptor expression in estrogen receptor-negative breast cancer. Immunohistochemical, clinical, and prognostic associations', *Am J Clin Pathol*, vol. 120, no. 5, Nov, pp. 725-731.

Ali, S, Rasool, M, Chaoudhry, H, P, NP, Jha, P, Hafiz, A, Mahfooz, M, Abdus Sami, G, Azhar Kamal, M, Bashir, S, Ali, A & Sarwar Jamal, M 2016, 'Molecular mechanisms and mode of tamoxifen resistance in breast cancer', *Bioinformation*, vol. 12, no. 3, pp. 135-139.

Allred, DC, Brown, P & Medina, D 2004, 'The origins of estrogen receptor alpha-positive and estrogen receptor alpha-negative human breast cancer', *Breast Cancer Res*, vol. 6, no. 6, pp. 240-245.

Ando, S, De Amicis, F, Rago, V, Carpino, A, Maggiolini, M, Panno, M & Lanzino, M 2002, 'Breast cancer: from estrogen to androgen receptor', *Molecular and Cellular Endocrinology*, vol. 193, pp. 121-128.

'Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials', 2015, *Lancet*, vol. 386, no. 10001, Oct 3, pp. 1341-1352.

Aspinall, SR, Stamp, S, Davison, A, Shenton, BK & Lennard, TW 2004, 'The proliferative effects of 5-androstene-3 beta,17 beta-diol and 5 alpha-dihydrotestosterone on cell cycle analysis and cell proliferation in MCF7, T47D and MDAMB231 breast cancer cell lines', *J Steroid Biochem Mol Biol*, vol. 88, no. 1, Jan, pp. 37-51.

Asselin-Labat, M-L, Sutherland, KD, Barker, H, Thomas, R, Shackleton, M, Forrest, NC, Hartley, L, Robb, L, Grosveld, FG & van der Wees, J 2007, 'Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation', *Nature cell biology*, vol. 9, no. 2, pp. 201-209.

Australia, C 2020, <<https://www.canceraustralia.gov.au/affected-cancer/cancer-types/breast-cancer/statistics>>.

Bain, DL, Heneghan, AF, Connaghan-Jones, KD & Miura, MT 2007, 'Nuclear receptor structure: implications for function', *Annu Rev Physiol*, vol. 69, pp. 201-220.

Ballaré, C, Uhrig, M, Bechtold, T, Sancho, E, Di Domenico, M, Migliaccio, A, Auricchio, F & Beato, M 2003, 'Two domains of the progesterone receptor interact with the estrogen receptor and are required for progesterone activation of the c-Src/Erk pathway in mammalian cells', *Mol Cell Biol*, vol. 23, no. 6, Mar, pp. 1994-2008.

Banerji, S, Cibulskis, K, Rangel-Escareno, C, Brown, KK, Carter, SL, Frederick, AM, Lawrence, MS, Sivachenko, AY, Sougnez, C, Zou, L, Cortes, ML, Fernandez-Lopez, JC, Peng, S, Ardlie, KG, Auclair, D, Bautista-Pina, V, Duke, F, Francis, J, Jung, J, Maffuz-Aziz, A, Onofrio, RC, Parkin, M, Pho, NH, Quintanar-Jurado, V, Ramos, AH, Rebollar-Vega, R, Rodriguez-Cuevas, S, Romero-Cordoba, SL, Schumacher, SE, Stransky, N, Thompson, KM, Uribe-Figueroa, L, Baselga, J, Beroukhim, R, Polyak, K, Sgroi, DC, Richardson, AL, Jimenez-Sanchez, G, Lander, ES, Gabriel, SB, Garraway, LA, Golub, TR, Melendez-Zajgla, J, Toker, A, Getz, G, Hidalgo-Miranda, A & Meyerson, M 2012, 'Sequence analysis of mutations and translocations across breast cancer subtypes', *Nature*, vol. 486, no. 7403, Jun 20, pp. 405-409.

Barcellos-Hoff, M 2013, 'Does microenvironment contribute to the etiology of estrogen receptor-negative breast cancer?', *Clinical Cancer Research*, vol. 19, pp. 541-548.

Bates, DL, Chen, Y, Kim, G, Guo, L & Chen, L 2008, 'Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA', *J Mol Biol*, vol. 381, no. 5, Sep 19, pp. 1292-1306.

Bernardo, GM, Lozada, KL, Miedler, JD, Harburg, G, Hewitt, SC, Mosley, JD, Godwin, AK, Korach, KS, Visvader, JE, Kaestner, KH, Abdul-Karim, FW, Montano, MM & Keri, RA 2010, 'FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis', *Development*, vol. 137, no. 12, Jun, pp. 2045-2054.

Birrell, SN, Bentel, JM, Hickey, TE, Ricciardelli, C, Weger, MA, Horsfall, DJ & Tilley, WD 1995, 'Androgens induce divergent proliferative responses in human breast cancer cell lines', *J Steroid Biochem Mol Biol*, vol. 52, no. 5, May, pp. 459-467.

Blokzijl, A, ten Dijke, P & Ibanez, CF 2002, 'Physical and functional interaction between GATA-3 and Smad3 allows TGF-beta regulation of GATA target genes', *Curr Biol*, vol. 12, no. 1, Jan 8, pp. 35-45.

Boto, A & Harigopal, M 2017, 'GATA3 Expression in Breast Cancers is Strongly Associated with AR Expression in a Large TMA Study', *IJCP*, vol. 103.

Boto, A & Harigopal, M 2018, 'Strong androgen receptor expression can aid in distinguishing GATA3+ metastases', *Hum Pathol*, vol. 75, May, pp. 63-70.

Bray, F, Ferlay, J, Soerjomataram, I, Siegel, RL, Torre, LA & Jemal, A 2018, 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA Cancer J Clin*, vol. 68, no. 6, Nov, pp. 394-424.

Brisken, C & O'Malley, B 2010, 'Hormone action in the mammary gland', *Cold Spring Harb Perspect Biol*, vol. 2, no. 12, Dec, p. a003178.

Broudy, VC 1997, 'Stem cell factor and hematopoiesis', *Blood*, vol. 90, no. 4, Aug 15, pp. 1345-1364.

Brown, CJ, Goss, SJ, Lubahn, DB, Joseph, DR, Wilson, EM, French, FS & Willard, HF 1989, 'Androgen receptor locus on the human X chromosome: regional localization to Xq11-12 and description of a DNA polymorphism', *Am J Hum Genet*, vol. 44, no. 2, Feb, pp. 264-269.

Burger, H 2002, 'Androgen production in women', *Fertility and Sterility*, vol. 77, no. Suppl 4, pp. S3-S5.

Cai, KQ, Caslini, C, Capo-chichi, CD, Slater, C, Smith, ER, Wu, H, Klein-Szanto, AJ, Godwin, AK & Xu, XX 2009, 'Loss of GATA4 and GATA6 expression specifies ovarian cancer histological subtypes and precedes neoplastic transformation of ovarian surface epithelia', *PLoS ONE*, vol. 4, no. 7, Jul 31, p. e6454.

Caldwell, CM, Green, RA & Kaplan, KB 2007, 'APC mutations lead to cytokinetic failures in vitro and tetraploid genotypes in Min mice', *J Cell Biol*, vol. 178, no. 7, Sep 24, pp. 1109-1120.

Cardiff, RD & Wellings, SR 1999, 'The comparative pathology of human and mouse mammary glands', *J Mammary Gland Biol Neoplasia*, vol. 4, no. 1, Jan, pp. 105-122.

Cardoso, F, Costa, A, Norton, L, Senkus, E, Aapro, M, André, F, Barrios, CH, Bergh, J, Biganzoli, L, Blackwell, KL, Cardoso, MJ, Cufer, T, El Saghir, N, Fallowfield, L, Fenech, D, Francis, P, Gelmon, K, Giordano, SH, Gligorov, J, Goldhirsch, A, Harbeck, N, Houssami, N, Hudis, C, Kaufman, B, Krop, I, Kyriakides, S, Lin, UN, Mayer, M, Merjaver, SD, Nordström, EB, Pagani, O, Partridge, A, Penault-Llorca, F, Piccart, MJ, Rugo, H, Sledge, G, Thomssen, C,



Van't Veer, L, Vorobiof, D, Vrieling, C, West, N, Xu, B & Winer, E 2014, 'ESO-ESMO 2nd international consensus guidelines for advanced breast cancer (ABC2)', *Breast*, vol. 23, no. 5, Oct, pp. 489-502.

Carr, JR, Kiefer, MM, Park, HJ, Li, J, Wang, Z, Fontanarosa, J, DeWaal, D, Kopanja, D, Benevolenskaya, EV, Guzman, G & Raychaudhuri, P 2012, 'FoxM1 regulates mammary luminal cell fate', *Cell Rep*, vol. 1, no. 6, Jun 28, pp. 715-729.

Carroll, JS, Liu, XS, Brodsky, AS, Li, W, Meyer, CA, Szary, AJ, Eeckhoute, J, Shao, W, Hestermann, EV & Geistlinger, TR 2005, 'Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1', *Cell*, vol. 122, no. 1, pp. 33-43.

Carroll, JS, Meyer, CA, Song, J, Li, W, Geistlinger, TR, Eeckhoute, J, Brodsky, AS, Keeton, EK, Fertuck, KC & Hall, GF 2006, 'Genome-wide analysis of estrogen receptor binding sites', *Nature genetics*, vol. 38, no. 11, pp. 1289-1297.

Castellano, I, Allia, E, Accortanzo, V, Vandone, AM, Chiusa, L, Arisio, R, Durando, A, Donadio, M, Bussolati, G, Coates, AS, Viale, G & Sapino, A 2010, 'Androgen receptor expression is a significant prognostic factor in estrogen receptor positive breast cancers', *Breast Cancer Res Treat*, vol. 124, no. 3, Dec, pp. 607-617.

Castrellon, AB 2017, 'Novel Strategies to Improve the Endocrine Therapy of Breast Cancer', *Oncol Rev*, vol. 11, no. 1, Mar 3, p. 323.

Chanock, SJ, Burdett, L, Yeager, M, Llaca, V, Langerod, A, Presswalla, S, Kaaresen, R, Strausberg, RL, Gerhard, DS, Kristensen, V, Perou, CM & Borresen-Dale, AL 2007, 'Somatic sequence alterations in twenty-one genes selected by expression profile analysis of breast carcinomas', *Breast Cancer Res*, vol. 9, no. 1, p. R5.

Chawla, A, Repa, JJ, Evans, RM & Mangelsdorf, DJ 2001, 'Nuclear receptors and lipid physiology: opening the X-files', *Science*, vol. 294, no. 5548, Nov 30, pp. 1866-1870.

Chen, Y, Bates, DL, Dey, R, Chen, PH, Machado, AC, Laird-Offringa, IA, Rohs, R & Chen, L 2012, 'DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation', *Cell Rep*, vol. 2, no. 5, Nov 29, pp. 1197-1206.

Cheng, G, Weihua, Z, Warner, M & Gustafsson, JA 2004, 'Estrogen receptors ER alpha and ER beta in proliferation in the rodent mammary gland', *Proc Natl Acad Sci U S A*, vol. 101, no. 11, Mar 16, pp. 3739-3746.

Chook, YM & Blobel, G 2001, 'Karyopherins and nuclear import', *Curr Opin Struct Biol*, vol. 11, no. 6, Dec, pp. 703-715.

Chou, J, Lin, JH, Brenot, A, Kim, JW, Provot, S & Werb, Z 2013, 'GATA3 suppresses metastasis and modulates the tumour microenvironment by regulating microRNA-29b expression', *Nat Cell Biol*, vol. 15, no. 2, Feb, pp. 201-213.

Chou, J, Provot, S & Werb, Z 2010, 'GATA3 in development and cancer differentiation: cells GATA have it!', *J Cell Physiol*, vol. 222, no. 1, Jan, pp. 42-49.

Ciriello, G, Gatza, ML, Beck, AH, Wilkerson, MD, Rhie, SK, Pastore, A, Zhang, H, McLellan, M, Yau, C, Kandoth, C, Bowlby, R, Shen, H, Hayat, S, Fieldhouse, R, Lester, SC, Tse, GM, Factor, RE, Collins, LC, Allison, KH, Chen, YY, Jensen, K, Johnson, NB, Oesterreich, S, Mills, GB, Cherniack, AD, Robertson, G, Benz, C, Sander, C, Laird, PW, Hoadley, KA, King, TA & Perou, CM 2015, 'Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer', *Cell*, vol. 163, no. 2, Oct 8, pp. 506-519.

Clark, KL, Halay, ED, Lai, E & Burley, SK 1993, 'Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5', *Nature*, vol. 364, no. 6436, pp. 412-420.

Cochrane, D, Bernales, S, Jacobsen, B, Cittelly, D, Howe, E, D Amato, N, Spoelstra, N, Edgerton, S, Jean, A & Guerrero, J 2014, 'Role of the androgen receptor in breast cancer and preclinical analysis of enzalutamide', *Breast Cancer Research*, vol. 16, p. R7.

Cohen, H, Ben-Hamo, R, Gidoni, M, Yitzhaki, I, Kozol, R, Zilberberg, A & Efroni, S 2014, 'Shift in GATA3 functions, and GATA3 mutations, control progression and clinical presentation in breast cancer', *Breast Cancer Res*, vol. 16, no. 6, Nov 20, p. 464.

'Comprehensive molecular portraits of human breast tumours', 2012, *Nature*, vol. 490, no. 7418, Oct 4, pp. 61-70.

Conti, E, Uy, M, Leighton, L, Blobel, G & Kuriyan, J 1998, 'Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha', *Cell*, vol. 94, no. 2, Jul 24, pp. 193-204.

Cops, EJ, Bianco-Miotto, T, Moore, NL, Clarke, CL, Birrell, SN, Butler, LM & Tilley, WD 2008, 'Antiproliferative actions of the synthetic androgen, mibolerone, in breast cancer cells are mediated by both androgen and progesterone receptors', *J Steroid Biochem Mol Biol*, vol. 110, no. 3-5, Jun, pp. 236-243.

Crispino, JD, Lodish, MB, Thurberg, BL, Litovsky, SH, Collins, T, Molkenin, JD & Orkin, SH 2001, 'Proper coronary vascular development and heart morphogenesis depend on interaction of GATA-4 with FOG cofactors', *Genes Dev*, vol. 15, no. 7, Apr 1, pp. 839-844.

Curtis Hewitt, S, Couse, JF & Korach, KS 2000, 'Estrogen receptor transcription and transactivation: Estrogen receptor knockout mice: what their phenotypes reveal about mechanisms of estrogen action', *Breast Cancer Res*, vol. 2, no. 5, pp. 345-352.

Davies, C, Godwin, J, Gray, R, Clarke, M, Cutter, D, Darby, S, McGale, P, Pan, HC, Taylor, C, Wang, YC, Dowsett, M, Ingle, J & Peto, R 2011, 'Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials', *Lancet*, vol. 378, no. 9793, Aug 27, pp. 771-784.

De Bosscher, K, Desmet, SJ, Clarisse, D, Estébanez-Perpiña, E & Brunsveld, L 2020, 'Nuclear receptor crosstalk - defining the mechanisms for therapeutic innovation', *Nat Rev Endocrinol*, vol. 16, no. 7, Jul, pp. 363-377.

De Las Rivas, J & Fontanillo, C 2010, 'Protein-protein interactions essentials: key concepts to building and analyzing interactome networks', *PLoS Comput Biol*, vol. 6, no. 6, Jun 24, p. e1000807.

De Vos, P, Claessens, F, Peeters, B, Rombauts, W, Heyns, W & Verhoeven, G 1993, 'Interaction of androgen and glucocorticoid receptor DNA-binding domains with their response elements', *Mol Cell Endocrinol*, vol. 90, no. 2, Jan, pp. R11-16.

Denayer, S, Helsen, C, Thorrez, L, Haelens, A & Claessens, F 2010, 'The rules of DNA recognition by the androgen receptor', *Mol Endocrinol*, vol. 24, no. 5, May, pp. 898-913.

Dimitrakakis, C & Bondy, C 2009, 'Androgens and the breast', *Breast Cancer Res*, vol. 11, no. 5, p. 212.

Dimitrakakis, C, Zhou, J, Wang, J, Belanger, A, LaBrie, F, Cheng, C, Powell, D & Bondy, C 2003, 'A physiologic role for testosterone in limiting estrogenic stimulation of the breast', *Menopause*, vol. 10, no. 4, Jul-Aug, pp. 292-298.

Doane, A, Danso, M, Lal, P, Donaton, M, Zhang, L, Hudis, C & Gerald, W 2006, 'An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen', *Oncogene*, vol. 25, pp. 3994-4008.

Dydensborg, A, Rose, A, Wilson, B, Grote, D, Paquet, M, Giguere, V, Siegel, P & Bouchard, M 2009, 'GATA3 inhibits breast cancer growth and pulmonary breast cancer metastasis', *Oncogene*, vol. 28, no. 29, pp. 2634-2642.

Echeverria, PC & Picard, D 2010, 'Molecular chaperones, essential partners of steroid hormone receptors for activity and mobility', *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, vol. 1803, no. 6, pp. 641-649.

Edwards, DP 2000, 'The role of coactivators and corepressors in the biology and mechanism of action of steroid hormone receptors', *J Mammary Gland Biol Neoplasia*, vol. 5, no. 3, Jul, pp. 307-324.

Eckhoute, J, Carroll, JS, Geistlinger, TR, Torres-Arzayus, MI & Brown, M 2006, 'A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer', *Genes Dev*, vol. 20, no. 18, Sep 15, pp. 2513-2526.

Eckhoute, J, Keeton, EK, Lupien, M, Krum, SA, Carroll, JS & Brown, M 2007, 'Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer', *Cancer Res*, vol. 67, no. 13, Jul 1, pp. 6477-6483.

'Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials', 2005, *Lancet*, vol. 365, no. 9472, May 14-20, pp. 1687-1717.

Elhaji, YA, Gottlieb, B, Lumbroso, R, Beitel, LK, Foulkes, WD, Pinsky, L & Trifiro, MA 2001, 'The polymorphic CAG repeat of the androgen receptor gene: a potential role in breast cancer in women over 40', *Breast Cancer Res Treat*, vol. 70, no. 2, Nov, pp. 109-116.

Ellis, MJ, Ding, L, Shen, D, Luo, J, Suman, VJ, Wallis, JW, Van Tine, BA, Hoog, J, Goiffon, RJ, Goldstein, TC, Ng, S, Lin, L, Crowder, R, Snider, J, Ballman, K, Weber, J, Chen, K, Koboldt, DC, Kandoth, C, Schierding, WS, McMichael, JF, Miller, CA, Lu, C, Harris, CC, McLellan, MD, Wendl, MC, DeSchryver, K, Allred, DC, Esserman, L, Unzeitig, G, Margenthaler, J, Babiera, GV, Marcom, PK, Guenther, JM, Leitch, M, Hunt, K, Olson, J, Tao, Y, Maher, CA, Fulton, LL, Fulton, RS, Harrison, M, Oberkfell, B, Du, F, Demeter, R, Vickery, TL, Elhammali, A, Piwnica-Worms, H, McDonald, S, Watson, M, Dooling, DJ, Ota, D, Chang, LW, Bose, R, Ley, TJ, Piwnica-Worms, D, Stuart, JM, Wilson, RK & Mardis, ER 2012, 'Whole-genome analysis informs breast cancer response to aromatase inhibition', *Nature*, vol. 486, no. 7403, Jun 10, pp. 353-360.

Ellsworth, RE, Blackburn, HL, Shriver, CD, Soon-Shiong, P & Ellsworth, DL 2017, 'Molecular heterogeneity in breast cancer: State of the science and implications for patient care', *Semin Cell Dev Biol*, vol. 64, Apr, pp. 65-72.

Espinosa Fernandez, JR, Eckhardt, BL, Lee, J, Lim, B, Pearson, T, Seitz, RS, Hout, DR, Schweitzer, BL, Nielsen, TJ, Lawrence, OR, Wang, Y, Rao, A & Ueno, NT 2020, 'Identification of triple-negative breast cancer cell lines classified under the same molecular subtype using different molecular characterization techniques: Implications for translational research', *PLoS ONE*, vol. 15, no. 4, p. e0231953.

Fan, W, Chang, J & Fu, P 2015, 'Endocrine therapy resistance in breast cancer: current status, possible mechanisms and overcoming strategies', *Future Med Chem*, vol. 7, no. 12, Aug, pp. 1511-1519.

Fang, SH, Chen, Y & Weigel, RJ 2009, 'GATA-3 as a marker of hormone response in breast cancer', *J Surg Res*, vol. 157, no. 2, Dec, pp. 290-295.

Farmer, P, Bonnefoi, H, Becette, V, Tubiana-Hulin, M, Fumoleau, P, Larsimont, D, Macgrogan, G, Bergh, J, Cameron, D & Goldstein, D 2005, 'Identification of molecular apocrine breast tumours by microarray analysis', *Oncogene*, vol. 24, pp. 4660-4671.

Feng, Y, Manka, D, Wagner, KU & Khan, SA 2007, 'Estrogen receptor-alpha expression in the mammary epithelium is required for ductal and alveolar morphogenesis in mice', *Proc Natl Acad Sci U S A*, vol. 104, no. 37, Sep 11, pp. 14718-14723.

Feng, Y, Spezia, M, Huang, S, Yuan, C, Zeng, Z, Zhang, L, Ji, X, Liu, W, Huang, B, Luo, W, Liu, B, Lei, Y, Du, S, Vuppapapati, A, Luu, HH, Haydon, RC, He, TC & Ren, G 2018, 'Breast cancer



development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis', *Genes Dis*, vol. 5, no. 2, Jun, pp. 77-106.

Fisher, B, Costantino, JP, Wickerham, DL, Cecchini, RS, Cronin, WM, Robidoux, A, Bevers, TB, Kavanah, MT, Atkins, JN, Margolese, RG, Runowicz, CD, James, JM, Ford, LG & Wolmark, N 2005, 'Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study', *J Natl Cancer Inst*, vol. 97, no. 22, Nov 16, pp. 1652-1662.

Fragomeni, SM, Sciallis, A & Jeruss, JS 2018, 'Molecular Subtypes and Local-Regional Control of Breast Cancer', *Surg Oncol Clin N Am*, vol. 27, no. 1, Jan, pp. 95-120.

Frelin, C, Herrington, R, Janmohamed, S, Barbara, M, Tran, G, Paige, CJ, Benveniste, P, Zuniga-Pflucker, JC, Souabni, A, Busslinger, M & Iscove, NN 2013, 'GATA-3 regulates the self-renewal of long-term hematopoietic stem cells', *Nat Immunol*, vol. 14, no. 10, Oct, pp. 1037-1044.

Fujiwara, T, Bandi, M, Nitta, M, Ivanova, EV, Bronson, RT & Pellman, D 2005, 'Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells', *Nature*, vol. 437, no. 7061, Oct 13, pp. 1043-1047.

Ganem, NJ, Storchova, Z & Pellman, D 2007, 'Tetraploidy, aneuploidy and cancer', *Curr Opin Genet Dev*, vol. 17, no. 2, Apr, pp. 157-162.

Gaynor, KU, Grigorieva, IV, Allen, MD, Esapa, CT, Head, RA, Gopinath, P, Christie, PT, Nesbit, MA, Jones, JL & Thakker, RV 2013, 'GATA3 mutations found in breast cancers may be associated with aberrant nuclear localization, reduced transactivation and cell invasiveness', *Hormones and Cancer*, vol. 4, no. 3, pp. 123-139.

Giguère, Y, Dewailly, E, Brisson, J, Ayotte, P, Laflamme, N, Demers, A, Forest, VI, Dodin, S, Robert, J & Rousseau, F 2001, 'Short polyglutamine tracts in the androgen receptor are protective against breast cancer in the general population', *Cancer Res*, vol. 61, no. 15, Aug 1, pp. 5869-5874.

Glass, CK 1994, 'Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers', *Endocr Rev*, vol. 15, no. 3, Jun, pp. 391-407.

Goldfarb, DS, Corbett, AH, Mason, DA, Harreman, MT & Adam, SA 2004, 'Importin alpha: a multipurpose nuclear-transport receptor', *Trends Cell Biol*, vol. 14, no. 9, Sep, pp. 505-514.

Gonzalez-Angulo, AM, Stemke-Hale, K, Palla, SL, Carey, M, Agarwal, R, Meric-Berstam, F, Traina, TA, Hudis, C, Hortobagyi, GN, Gerald, WL, Mills, GB & Hennessy, BT 2009, 'Androgen receptor levels and association with PIK3CA mutations and prognosis in breast cancer', *Clin Cancer Res*, vol. 15, no. 7, Apr 1, pp. 2472-2478.

Green, S, Walter, P, Kumar, V, Krust, A, Bornert, JM, Argos, P & Chambon, P 1986, 'Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A', *Nature*, vol. 320, no. 6058, Mar 13-19, pp. 134-139.

Greene, GL, Gilna, P, Waterfield, M, Baker, A, Hort, Y & Shine, J 1986, 'Sequence and expression of human estrogen receptor complementary DNA', *Science*, vol. 231, no. 4742, Mar 7, pp. 1150-1154.

Greeve, MA, Allan, RK, Harvey, JM & Bentel, JM 2004, 'Inhibition of MCF-7 breast cancer cell proliferation by 5alpha-dihydrotestosterone; a role for p21(Cip1/Waf1)', *J Mol Endocrinol*, vol. 32, no. 3, Jun, pp. 793-810.

Grimm, SL & Nordeen, SK 1998, 'Mouse mammary tumor virus sequences responsible for activating cellular oncogenes', *J Virol*, vol. 72, no. 12, Dec, pp. 9428-9435.

Grote, D, Boualia, SK, Souabni, A, Merkel, C, Chi, X, Costantini, F, Carroll, T & Bouchard, M 2008, 'Gata3 acts downstream of beta-catenin signaling to prevent ectopic metanephric kidney induction', *PLoS Genet*, vol. 4, no. 12, Dec, p. e1000316.

Gucalp, A, Tolaney, S, Isakoff, S, Ingle, J, Liu, M, Carey, L, Blackwell, K, Rugo, H, Nabell, L & Forero, A 2013, 'Phase II trial of bicalutamide in patients with androgen receptor-positive, estrogen receptor-negative metastatic breast cancer', *Clinical Cancer Research*, vol. 19, pp. 5505-5512.

Gustin, JP, Miller, J, Farag, M, Rosen, DM, Thomas, M, Scharpf, RB & Lauring, J 2017, 'GATA3 frameshift mutation promotes tumor growth in human luminal breast cancer cells and induces transcriptional changes seen in primary GATA3 mutant breast cancers', *Oncotarget*, vol. 8, no. 61, Nov 28, pp. 103415-103427.

Hackenberg, R, Luttcens, S, Hofmann, J, Kunzmann, R, Holzel, F & Schulz, K 1991, 'Androgen sensitivity of the new human breast cancer cell line MFM-223', *Cancer research*, vol. 51, pp. 5722-5727.

Hammond, M, Hayes, D, Dowsett, M, Allred, D, Hagerty, K, Badve, S, Fitzgibbons, P, Francis, G, Goldstein, N & Hayes, M 2010, 'American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version)', *Archives of Pathology & Laboratory Medicine*, vol. 134, pp. e48-e72.

Hanahan, D & Weinberg, RA 2011, 'Hallmarks of cancer: the next generation', *Cell*, vol. 144, no. 5, Mar 4, pp. 646-674.

He, B, Lanz, RB, Fiskus, W, Geng, C, Yi, P, Hartig, SM, Rajapakshe, K, Shou, J, Wei, L, Shah, SS, Foley, C, Chew, SA, Eedunuri, VK, Bedoya, DJ, Feng, Q, Minami, T, Mitsiades, CS, Frolov, A, Weigel, NL, Hilsenbeck, SG, Rosen, DG, Palzkill, T, Ittmann, MM, Song, Y, Coarfa, C, O'Malley, BW & Mitsiades, N 2014, 'GATA2 facilitates steroid receptor coactivator recruitment to the androgen receptor complex', *Proc Natl Acad Sci U S A*, vol. 111, no. 51, Dec 23, pp. 18261-18266.

Helguero, LA, Faulds, MH, Gustafsson, JA & Haldosén, LA 2005, 'Estrogen receptors alfa (ERalpha) and beta (ERbeta) differentially regulate proliferation and apoptosis of the normal murine mammary epithelial cell line HC11', *Oncogene*, vol. 24, no. 44, Oct 6, pp. 6605-6616.

Henderson, BE, Ponder, B & Ross, RK 2003, *Hormones, Genes, and Cancer*.

Hickey, T, Robinson, J, Carroll, J & Tilley, W 2012, 'Minireview: the androgen receptor in breast tissues: growth inhibitor, tumor suppressor, oncogene?', *Molecular Endocrinology*, vol. 26, pp. 1252-1267.

Hickey, TE, Selth, LA, Chia, KM, Laven-Law, G, Milioli, HH, Roden, D, Jindal, S, Hui, M, Finlay-Schultz, J, Ebrahimie, E, Birrell, SN, Stelloo, S, Iggo, R, Alexandrou, S, Caldon, CE, Abdel-Fatah, TM, Ellis, IO, Zwart, W, Palmieri, C, Sartorius, CA, Swarbrick, A, Lim, E, Carroll, JS & Tilley, WD 2021, 'The androgen receptor is a tumor suppressor in estrogen receptor-positive breast cancer', *Nat Med*, Jan 18.

Higgins, MJ & Stearns, V 2009, 'Understanding resistance to tamoxifen in hormone receptor-positive breast cancer', *Clin Chem*, vol. 55, no. 8, Aug, pp. 1453-1455.

Hoch, RV, Thompson, DA, Baker, RJ & Weigel, RJ 1999, 'GATA-3 is expressed in association with estrogen receptor in breast cancer', *Int J Cancer*, vol. 84, no. 2, Apr 20, pp. 122-128.

Howell, A 2005, 'The future of fulvestrant ("Faslodex")', *Cancer Treat Rev*, vol. 31 Suppl 2, pp. S26-33.

Howell, A, Robertson, JF, Abram, P, Lichinitser, MR, Elledge, R, Bajetta, E, Watanabe, T, Morris, C, Webster, A, Dimery, I & Osborne, CK 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', *J Clin Oncol*, vol. 22, no. 9, May 1, pp. 1605-1613.

Hruschka, N, Kalisz, M, Subijana, M, Graña-Castro, O, Del Cano-Ochoa, F, Brunet, LP, Chernukhin, I, Sagrera, A, De Reynies, A, Kloesch, B, Chin, SF, Burgués, O, Andreu, D, Bermejo, B, Cejalvo, JM, Sutton, J, Caldas, C, Ramón-Maiques, S, Carroll, JS, Prat, A, Real, FX & Martinelli, P 2020, 'The GATA3 X308\_Splice breast cancer mutation is a hormone context-dependent oncogenic driver', *Oncogene*, vol. 39, no. 32, Aug, pp. 5455-5467.

Hu, R, Dawood, S, Holmes, M, Collins, L, Schnitt, S, Cole, K, Marotti, J, Hankinson, S, Colditz, G & Tamimi, R 2011, 'Androgen receptor expression and breast cancer survival in postmenopausal women', *Clinical Cancer Research*, vol. 17, pp. 1867-1874.

Hu, XQ, Chen, WL, Ma, HG & Jiang, K 2017, 'Androgen receptor expression identifies patient with favorable outcome in operable triple negative breast cancer', *Oncotarget*, vol. 8, no. 34, Aug 22, pp. 56364-56374.

Huang, F, Chen, H, Zhu, X, Gong, T, Li, X, Hankey, W, Wang, H, Chen, Z, Wang, Q, & Liu, Z, 2021, 'The oncogenomic function of androgen receptor in esophageal squamous cell carcinoma is directed by GATA3', *Cell Research*, vol. 31, p. 362-365.

Iggo, R 2011, 'New insights into the role of androgen and oestrogen receptors in molecular apocrine breast tumours', *Breast Cancer Research*, vol. 13, p. 318.

Iggo, R 2018, 'Classification of breast tumours into molecular apocrine, luminal and basal groups based on an explicit biological model', *BioRxiv*, p. 270975.

Ingle, JN, Twito, DI, Schaid, DJ, Cullinan, SA, Krook, JE, Mailliard, JA, Tschetter, LK, Long, HJ, Gerstner, JG, Windschitl, HE & et al. 1991, 'Combination hormonal therapy with tamoxifen plus fluoxymesterone versus tamoxifen alone in postmenopausal women with metastatic breast cancer. An updated analysis', *Cancer*, vol. 67, no. 4, Feb 15, pp. 886-891.

Inman, JL, Robertson, C, Mott, JD & Bissell, MJ 2015, 'Mammary gland development: cell fate specification, stem cells and the microenvironment', *Development*, vol. 142, no. 6, Mar 15, pp. 1028-1042.

Ivanov, AA, Khuri, FR & Fu, H 2013, 'Targeting protein-protein interactions as an anticancer strategy', *Trends Pharmacol Sci*, vol. 34, no. 7, Jul, pp. 393-400.

Jiang, YZ, Yu, KD, Zuo, WJ, Peng, WT & Shao, ZM 2014, 'GATA3 mutations define a unique subtype of luminal-like breast cancer with improved survival', *Cancer*, vol. 120, no. 9, pp. 1329-1337.

Jing, H, Vakoc, CR, Ying, L, Mandat, S, Wang, H, Zheng, X & Blobel, GA 2008, 'Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus', *Mol Cell*, vol. 29, no. 2, Feb 1, pp. 232-242.

Jozwik, KM & Carroll, JS 2012, 'Pioneer factors in hormone-dependent cancers', *Nat Rev Cancer*, vol. 12, no. 6, May 4, pp. 381-385.

Kaufman, CK, Zhou, P, Pasolli, HA, Rendl, M, Bolotin, D, Lim, KC, Dai, X, Alegre, ML & Fuchs, E 2003, 'GATA-3: an unexpected regulator of cell lineage determination in skin', *Genes Dev*, vol. 17, no. 17, Sep 1, pp. 2108-2122.

Khan, SA, Rogers, MA, Khurana, KK, Meguid, MM & Numann, PJ 1998, 'Estrogen receptor expression in benign breast epithelium and breast cancer risk', *J Natl Cancer Inst*, vol. 90, no. 1, Jan 7, pp. 37-42.

Kilker, RL, Hartl, MW, Rutherford, TM & Planas-Silva, MD 2004, 'Cyclin D1 expression is dependent on estrogen receptor function in tamoxifen-resistant breast cancer cells', *J Steroid Biochem Mol Biol*, vol. 92, no. 1-2, Sep, pp. 63-71.

Kim, S, Moon, BI, Lim, W, Park, S, Cho, MS & Sung, SH 2016, 'Expression patterns of GATA3 and the androgen receptor are strongly correlated in patients with triple-negative breast cancer', *Hum Pathol*, vol. 55, Sep, pp. 190-195.

Kinyamu, HK & Archer, TK 2003, 'Estrogen receptor-dependent proteasomal degradation of the glucocorticoid receptor is coupled to an increase in mdm2 protein expression', *Mol Cell Biol*, vol. 23, no. 16, Aug, pp. 5867-5881.



Kong, SL, Li, G, Loh, SL, Sung, WK & Liu, ET 2011, 'Cellular reprogramming by the conjoint action of ERalpha, FOXA1, and GATA3 to a ligand-inducible growth state', *Mol Syst Biol*, vol. 7, Aug 30, p. 526.

Kouros-Mehr, H, Bechis, SK, Slorach, EM, Littlepage, LE, Egeblad, M, Ewald, AJ, Pai, SY, Ho, IC & Werb, Z 2008, 'GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model', *Cancer Cell*, vol. 13, no. 2, Feb, pp. 141-152.

Kouros-Mehr, H, Kim, JW, Bechis, SK & Werb, Z 2008, 'GATA-3 and the regulation of the mammary luminal cell fate', *Curr Opin Cell Biol*, vol. 20, no. 2, Apr, pp. 164-170.

Kouros-Mehr, H, Slorach, EM, Sternlicht, MD & Werb, Z 2006, 'GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland', *Cell*, vol. 127, no. 5, Dec 1, pp. 1041-1055.

Kouros-Mehr, H & Werb, Z 2006, 'Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis', *Dev Dyn*, vol. 235, no. 12, Dec, pp. 3404-3412.

Kuiper, GG, Enmark, E, Peltö-Huikko, M, Nilsson, S & Gustafsson, JA 1996, 'Cloning of a novel receptor expressed in rat prostate and ovary', *Proc Natl Acad Sci U S A*, vol. 93, no. 12, Jun 11, pp. 5925-5930.

Lacroix, M & Leclercq, G 2004, 'About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer', *Mol Cell Endocrinol*, vol. 219, no. 1-2, Apr 30, pp. 1-7.

Laganière, J, Deblois, G, Lefebvre, C, Bataille, AR, Robert, F & Giguère, V 2005, 'From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response', *Proc Natl Acad Sci U S A*, vol. 102, no. 33, Aug 16, pp. 11651-11656.

Lanzino, M, De Amicis, F, McPhaul, M, Marsico, S, Panno, M & Ando, S 2005, 'Endogenous coactivator ARA70 interacts with estrogen receptor  $\alpha$  (ER $\alpha$ ) and modulates the functional ER $\alpha$ /androgen receptor interplay in MCF-7 cells', *Journal of Biological Chemistry*, vol. 280, pp. 20421-20430.

Lanzino, M, Sisci, D, Morelli, C, Garofalo, C, Catalano, S, Casaburi, I, Capparelli, C, Giordano, C, Giordano, F, Maggiolini, M & Andò, S 2010, 'Inhibition of cyclin D1 expression by androgen receptor in breast cancer cells--identification of a novel androgen response element', *Nucleic Acids Res*, vol. 38, no. 16, Sep, pp. 5351-5365.

Lapointe, J, Fournier, A, Richard, V & Labrie, C 1999, 'Androgens down-regulate bcl-2 protooncogene expression in ZR-75-1 human breast cancer cells', *Endocrinology*, vol. 140, pp. 416-421.

Lehmann, BD, Bauer, JA, Chen, X, Sanders, ME, Chakravarthy, AB, Shyr, Y & Pietenpol, JA 2011, 'Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies', *J Clin Invest*, vol. 121, no. 7, Jul, pp. 2750-2767.

Lentjes, MH, Niessen, HE, Akiyama, Y, de Bruïne, AP, Melotte, V & van Engeland, M 2016, 'The emerging role of GATA transcription factors in development and disease', *Expert Rev Mol Med*, vol. 18, Mar 8, p. e3.

Lim, E, Tarulli, G, Portman, N, Hickey, TE, Tilley, WD & Palmieri, C 2016, 'Pushing estrogen receptor around in breast cancer', *Endocrine-related cancer*, vol. 23, no. 12, pp. T227-T241.

Lin, EY, Jones, JG, Li, P, Zhu, L, Whitney, KD, Muller, WJ & Pollard, JW 2003, 'Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases', *Am J Pathol*, vol. 163, no. 5, Nov, pp. 2113-2126.

Lloyd-Lewis, B, Harris, OB, Watson, CJ & Davis, FM 2017, 'Mammary Stem Cells: Premise, Properties, and Perspectives', *Trends Cell Biol*, vol. 27, no. 8, Aug, pp. 556-567.

Loibl, S, Muller, B, von Minckwitz, G, Schwabe, M, Roller, M, Darb-Esfahani, S, Ataseven, B, du Bois, A, Fissler-Eckhoff, A & Gerber, B 2011, 'Androgen receptor expression in primary breast cancer and its predictive and prognostic value in patients treated with neoadjuvant chemotherapy', *Breast Cancer Research and Treatment*, vol. 130, pp. 477-487.

Lonard, DM & O'Malley, BW 2006, 'The expanding cosmos of nuclear receptor coactivators', *Cell*, vol. 125, no. 3, May 5, pp. 411-414.

Lonard, DM & O'Malley, BW 2007, 'Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation', *Molecular cell*, vol. 27, no. 5, pp. 691-700.

Luo, X, Shi, Y, Li, Z & Jiang, W 2010, 'Expression and clinical significance of androgen receptor in triple negative breast cancer', *Chinese Journal of Cancer*, vol. 29, pp. 585-590.

Lupien, M & Brown, M 2009, 'Cistromics of hormone-dependent cancer', *Endocr Relat Cancer*, vol. 16, no. 2, Jun, pp. 381-389.

Lupien, M, Eeckhoute, J, Meyer, CA, Wang, Q, Zhang, Y, Li, W, Carroll, JS, Liu, XS & Brown, M 2008, 'FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription', *Cell*, vol. 132, no. 6, Mar 21, pp. 958-970.

Lydon, JP, DeMayo, FJ, Funk, CR, Mani, SK, Hughes, AR, Montgomery, CA, Jr., Shyamala, G, Conneely, OM & O'Malley, BW 1995, 'Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities', *Genes Dev*, vol. 9, no. 18, Sep 15, pp. 2266-2278.

Macedo, LF, Guo, Z, Tilghman, SL, Sabnis, GJ, Qiu, Y & Brodie, A 2006, 'Role of Androgens on MCF-7 Breast Cancer Cell Growth and on the Inhibitory Effect of Letrozole', *Cancer research*, vol. 66, no. 15, pp. 7775-7782.

Manavathi, B, Samanthapudi, VS & Gajulapalli, VN 2014, 'Estrogen receptor coregulators and pioneer factors: the orchestrators of mammary gland cell fate and development', *Front Cell Dev Biol*, vol. 2, p. 34.

Maneechotesuwan, K, Xin, Y, Ito, K, Jazrawi, E, Lee, KY, Usmani, OS, Barnes, PJ & Adcock, IM 2007, 'Regulation of Th2 cytokine genes by p38 MAPK-mediated phosphorylation of GATA-3', *J Immunol*, vol. 178, no. 4, Feb 15, pp. 2491-2498.

Masoud, V & Pagès, G 2017, 'Targeted therapies in breast cancer: New challenges to fight against resistance', *World J Clin Oncol*, vol. 8, no. 2, Apr 10, pp. 120-134.

Massie, CE, Adryan, B, Barbosa-Morais, NL, Lynch, AG, Tran, MG, Neal, DE & Mills, IG 2007, 'New androgen receptor genomic targets show an interaction with the ETS1 transcription factor', *EMBO Rep*, vol. 8, no. 9, Sep, pp. 871-878.

Mazaira, GI, Zgajnar, NR, Lotufo, CM, Daneri-Becerra, C, Sivils, JC, Soto, OB, Cox, MB & Galigniana, MD 2018, 'The Nuclear Receptor Field: A Historical Overview and Future Challenges', *Nucl Receptor Res*, vol. 5.

McNamara, K, Harwood, D, Simanainen, U, Walters, K, Jimenez, M & Handelsman, D 2010, 'Measurement of sex steroids in murine blood and reproductive tissues by liquid chromatography tandem mass spectrometry', *Journal of Steroid Biochemistry and Molecular Biology*, vol. 121, pp. 611-618.

McNamara, KM, Moore, NL, Hickey, TE, Sasano, H & Tilley, WD 2014, 'Complexities of androgen receptor signalling in breast cancer', *Endocrine-related cancer*, vol. 21, no. 4, pp. T161-T181.

Mehra, R, Varambally, S, Ding, L, Shen, R, Sabel, MS, Ghosh, D, Chinnaiyan, AM & Kleer, CG 2005, 'Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis', *Cancer Res*, vol. 65, no. 24, Dec 15, pp. 11259-11264.

Melmed, S, Polonsky, KS, Larsen, PR & Kronenberg, HM 2015, *Williams textbook of endocrinology*, Elsevier Health Sciences.

Miller, WR 2003, 'Aromatase inhibitors: mechanism of action and role in the treatment of breast cancer', *Semin Oncol*, vol. 30, no. 4 Suppl 14, Aug, pp. 3-11.

Miranda, TB, Voss, TC, Sung, MH, Baek, S, John, S, Hawkins, M, Grøntved, L, Schiltz, RL & Hager, GL 2013, 'Reprogramming the chromatin landscape: interplay of the estrogen and glucocorticoid receptors at the genomic level', *Cancer Res*, vol. 73, no. 16, Aug 15, pp. 5130-5139.

Mohammed, H, D'Santos, C, Serandour, AA, Ali, HR, Brown, GD, Atkins, A, Rueda, OM, Holmes, KA, Theodorou, V, Robinson, JL, Zwart, W, Saadi, A, Ross-Innes, CS, Chin, SF, Menon, S, Stingl, J, Palmieri, C, Caldas, C & Carroll, JS 2013, 'Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor', *Cell Rep*, vol. 3, no. 2, Feb 21, pp. 342-349.

Mohammed, H, Russell, IA, Stark, R, Rueda, OM, Hickey, TE, Tarulli, GA, Serandour, AA, Birrell, SN, Bruna, A, Saadi, A, Menon, S, Hadfield, J, Pugh, M, Raj, GV, Brown, GD, D'Santos, C, Robinson, JL, Silva, G, Launchbury, R, Perou, CM, Stingl, J, Caldas, C, Tilley, WD & Carroll, JS 2015, 'Progesterone receptor modulates ER $\alpha$  action in breast cancer', *Nature*, vol. 523, no. 7560, Jul 16, pp. 313-317.

Mohammed, H, Taylor, C, Brown, GD, Papachristou, EK, Carroll, JS & D'Santos, CS 2016, 'Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes', *Nat Protoc*, vol. 11, no. 2, Feb, pp. 316-326.

Moinfar, F, Okcu, M, Tsybrovskyy, O, Regitnig, P, Lax, SF, Weybora, W, Ratschek, M, Tavassoli, FA & Denk, H 2003, 'Androgen receptors frequently are expressed in breast carcinomas: potential relevance to new therapeutic strategies', *Cancer*, vol. 98, no. 4, Aug 15, pp. 703-711.

Morrisey, EE, Ip, HS, Tang, Z & Parmacek, MS 1997, 'GATA-4 activates transcription via two novel domains that are conserved within the GATA-4/5/6 subfamily', *J Biol Chem*, vol. 272, no. 13, Mar 28, pp. 8515-8524.

Munugalavadla, V, Dore, LC, Tan, BL, Hong, L, Vishnu, M, Weiss, MJ & Kapur, R 2005, 'Repression of c-kit and its downstream substrates by GATA-1 inhibits cell proliferation during erythroid maturation', *Mol Cell Biol*, vol. 25, no. 15, Aug, pp. 6747-6759.

Murashima, A, Kishigami, S, Thomson, A & Yamada, G 2015, 'Androgens and mammalian male reproductive tract development', *Biochim Biophys Acta*, vol. 1849, no. 2, Feb, pp. 163-170.

Murphy, CG & Dickler, MN 2016, 'Endocrine resistance in hormone-responsive breast cancer: mechanisms and therapeutic strategies', *Endocr Relat Cancer*, vol. 23, no. 8, Aug, pp. R337-352.

Naderi, A, Chia, K & Liu, J 2011, 'Synergy between inhibitors of androgen receptor and MEK has therapeutic implications in estrogen receptor-negative breast cancer', *Breast Cancer Research*, vol. 13, p. R36.

Naderi, A & Hughes-Davies, L 2008, 'A functionally significant cross-talk between androgen receptor and ErbB2 pathways in estrogen receptor negative breast cancer', *Neoplasia*, vol. 10, pp. 542-548.

Naylor, MJ & Ormandy, CJ 2007, 'Gata-3 and mammary cell fate', *Breast Cancer Research*, vol. 9, no. 2, p. 302.

Need, E, Selth, L, Harris, T, Birrell, S, Tilley, W & Buchanan, G 2012, 'Research resource: interplay between the genomic and transcriptional networks of androgen receptor and estrogen receptor  $\alpha$  in luminal breast cancer cells', *Molecular Endocrinology*, vol. 26, pp. 1941-1952.



Ni, M, Chen, Y, Fei, T, Li, D, Lim, E, Liu, X & Brown, M 2013, 'Amplitude modulation of androgen signaling by c-MYC', *Genes and Development*, vol. 27, pp. 734-748.

Ni, M, Chen, Y, Lim, E, Wimberly, H, Bailey, S, Imai, Y, Rimm, D, Liu, X & Brown, M 2011, 'Targeting androgen receptor in estrogen receptor-negative breast cancer', *Cancer Cell*, vol. 20, pp. 119-131.

Niemeier, L, Dabbs, D, Beriwal, S, Striebel, J & Bhargava, R 2010, 'Androgen receptor in breast cancer: expression in estrogen receptor-positive tumors and in estrogen receptor-negative tumors with apocrine differentiation', *Modern Pathology*, vol. 23, pp. 205-212.

Nilsson, S, Makela, S, Treuter, E, Tujague, M, Thomsen, J, Andersson, G, Enmark, E, Pettersson, K, Warner, M & Gustafsson, JA 2001, 'Mechanisms of estrogen action', *Physiol Rev*, vol. 81, no. 4, Oct, pp. 1535-1565.

O'Hare, MJ, Ormerod, MG, Monaghan, P, Lane, EB & Gusterson, BA 1991, 'Characterization in vitro of luminal and myoepithelial cells isolated from the human mammary gland by cell sorting', *Differentiation*, vol. 46, no. 3, Apr, pp. 209-221.

Oakes, SR, Gallego-Ortega, D & Ormandy, CJ 2014, 'The mammary cellular hierarchy and breast cancer', *Cell Mol Life Sci*, vol. 71, no. 22, Nov, pp. 4301-4324.

Ogawa, S, Inoue, S, Watanabe, T, Hiroi, H, Orimo, A, Hosoi, T, Ouchi, Y & Muramatsu, M 1998, 'The Complete Primary Structure of Human Estrogen Receptor  $\beta$  (hER $\beta$ ) and Its

Heterodimerization with ER  $\alpha$  in Vivo and in Vitro', *Biochemical and biophysical research communications*, vol. 243, no. 1, pp. 122-126.

O'Nate, SA, Tsai, SY, Tsai, MJ & O'Malley, BW 1995, 'Sequence and characterization of a coactivator for the steroid hormone receptor superfamily', *Science*, vol. 270, no. 5240, Nov 24, pp. 1354-1357.

Ortmann, J, Prifti, S, Bohlmann, M, Rehberger-Schneider, S, Strowitzki, T & Rabe, T 2002, 'Testosterone and 5 $\alpha$ -dihydrotestosterone inhibit in vitro growth of human breast cancer cell lines', *Gynecological Endocrinology*, vol. 16, no. 2, pp. 113-120.

Osborne, CK, Pippen, J, Jones, SE, Parker, LM, Ellis, M, Come, S, Gertler, SZ, May, JT, Burton, G, Dimery, I, Webster, A, Morris, C, Elledge, R & Buzdar, A 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', *J Clin Oncol*, vol. 20, no. 16, Aug 15, pp. 3386-3395.

Osborne, CK & Schiff, R 2011, 'Mechanisms of endocrine resistance in breast cancer', *Annu Rev Med*, vol. 62, pp. 233-247.

Overmoyer, B, Sanz-Altamira, P, Taylor, RP, Hancock, ML, Dalton, JT, Johnston, MA & Steiner, MS 2014, 'Enobosarm: A targeted therapy for metastatic, androgen receptor

positive, breast cancer', *Journal of Clinical Oncology*, vol. 32, no. 15\_suppl, 2014/05/20, pp. 568-568.

Pace, P, Taylor, J, Suntharalingam, S, Coombes, RC & Ali, S 1997, 'Human estrogen receptor beta binds DNA in a manner similar to and dimerizes with estrogen receptor alpha', *J Biol Chem*, vol. 272, no. 41, Oct 10, pp. 25832-25838.

Paltoglou, S, Das, R, Townley, SL, Hickey, TE, Tarulli, GA, Coutinho, I, Fernandes, R, Hanson, AR, Denis, I, Carroll, JS, Dehm, SM, Raj, GV, Plymate, SR, Tilley, WD & Selth, LA 2017, 'Novel Androgen Receptor Coregulator GRHL2 Exerts Both Oncogenic and Antimetastatic Functions in Prostate Cancer', *Cancer Res*, vol. 77, no. 13, Jul 1, pp. 3417-3430.

Pan, YF, Wansa, KD, Liu, MH, Zhao, B, Hong, SZ, Tan, PY, Lim, KS, Bourque, G, Liu, ET & Cheung, E 2008, 'Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements', *J Biol Chem*, vol. 283, no. 47, Nov 21, pp. 32977-32988.

Panet-Raymond, V, Gottlieb, B, Beitel, L, Pinsky, L & Trifiro, M 2000, 'Interactions between androgen and estrogen receptors and the effects on their transactivational properties', *Molecular and Cellular Endocrinology*, vol. 167, pp. 139-150.

Papachristou, EK, Kishore, K, Holding, AN, Harvey, K, Roumeliotis, TI, Chilamakuri, CSR, Omarjee, S, Chia, KM, Swarbrick, A, Lim, E, Markowitz, F, Eldridge, M, Siersbaek, R, D'Santos, CS & Carroll, JS 2018, 'A quantitative mass spectrometry-based approach to

monitor the dynamics of endogenous chromatin-associated protein complexes', *Nat Commun*, vol. 9, no. 1, Jun 13, p. 2311.

Park, S, Koo, J, Kim, M, Park, H, Lee, J, Lee, J, Kim, S, Park, B & Lee, K 2011, 'Androgen receptor expression is significantly associated with better outcomes in estrogen receptor-positive breast cancers', *Annals of Oncology*, vol. 22, pp. 1755-1762.

Park, S, Koo, J, Park, H, Kim, J, Choi, S, Lee, J, Park, B & Lee, K 2010, 'Expression of androgen receptors in primary breast cancer', *Annals of Oncology*, vol. 21, pp. 488-492.

Parker, JS, Mullins, M, Cheang, MC, Leung, S, Voduc, D, Vickery, T, Davies, S, Fauron, C, He, X, Hu, Z, Quackenbush, JF, Stijleman, IJ, Palazzo, J, Marron, JS, Nobel, AB, Mardis, E, Nielsen, TO, Ellis, MJ, Perou, CM & Bernard, PS 2009, 'Supervised risk predictor of breast cancer based on intrinsic subtypes', *J Clin Oncol*, vol. 27, no. 8, Mar 10, pp. 1160-1167.

Patient, RK & McGhee, JD 2002, 'The GATA family (vertebrates and invertebrates)', *Current opinion in genetics & development*, vol. 12, no. 4, pp. 416-422.

Pawlak, M, Lefebvre, P & Staels, B 2012, 'General molecular biology and architecture of nuclear receptors', *Curr Top Med Chem*, vol. 12, no. 6, pp. 486-504.

Pereira, B, Chin, SF, Rueda, OM, Vollan, HK, Provenzano, E, Bardwell, HA, Pugh, M, Jones, L, Russell, R, Sammut, SJ, Tsui, DW, Liu, B, Dawson, SJ, Abraham, J, Northen, H, Peden, JF, Mukherjee, A, Turashvili, G, Green, AR, McKinney, S, Oloumi, A, Shah, S, Rosenfeld, N,

Murphy, L, Bentley, DR, Ellis, IO, Purushotham, A, Pinder, SE, Børresen-Dale, AL, Earl, HM, Pharoah, PD, Ross, MT, Aparicio, S & Caldas, C 2016, 'The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes', *Nat Commun*, vol. 7, May 10, p. 11479.

Perou, CM, Sorlie, T, Eisen, MB, van de Rijn, M, Jeffrey, SS, Rees, CA, Pollack, JR, Ross, DT, Johnsen, H, Akslen, LA, Fluge, O, Pergamenschikov, A, Williams, C, Zhu, SX, Lonning, PE, Borresen-Dale, AL, Brown, PO & Botstein, D 2000, 'Molecular portraits of human breast tumours', *Nature*, vol. 406, no. 6797, Aug 17, pp. 747-752.

Peters, AA, Buchanan, G, Ricciardelli, C, Bianco-Miotto, T, Centenera, MM, Harris, JM, Jindal, S, Segara, D, Jia, L, Moore, NL, Henshall, SM, Birrell, SN, Coetzee, GA, Sutherland, RL, Butler, LM & Tilley, WD 2009, 'Androgen receptor inhibits estrogen receptor-alpha activity and is prognostic in breast cancer', *Cancer Res*, vol. 69, no. 15, Aug 1, pp. 6131-6140.

Peters, AA, Ingman, WV, Tilley, WD & Butler, LM 2011, 'Differential effects of exogenous androgen and an androgen receptor antagonist in the peri- and postpubertal murine mammary gland', *Endocrinology*, vol. 152, no. 10, Oct, pp. 3728-3737.

Poulin, R, Baker, D & Labrie, F 1988, 'Androgens inhibit basal and estrogen-induced cell proliferation in the ZR-75-1 human breast cancer cell line', *Breast Cancer Research and Treatment*, vol. 12, pp. 213-225.

Poulin, R, Simard, J, Labrie, C, Petitclerc, L, Dumont, M, Lagace, L & Labrie, F 1989, 'Down-regulation of estrogen receptors by androgens in the ZR-75-1 human breast cancer cell line', *Endocrinology*, vol. 125, no. 1, Jul, pp. 392-399.

Purdie, CA, Quinlan, P, Jordan, LB, Ashfield, A, Ogston, S, Dewar, JA & Thompson, AM 2014, 'Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study', *Br J Cancer*, vol. 110, no. 3, Feb 4, pp. 565-572.

Qu, Q, Mao, Y, Fei, X & Shen, K 2013, 'The impact of androgen receptor expression on breast cancer survival: a retrospective study and meta-analysis', *PLoS ONE*, vol. 8, p. e82650.

Rastinejad, F 2001, 'Retinoid X receptor and its partners in the nuclear receptor family', *Curr Opin Struct Biol*, vol. 11, no. 1, Feb, pp. 33-38.

Reis-Filho, JS & Puzstai, L 2011, 'Gene expression profiling in breast cancer: classification, prognostication, and prediction', *Lancet*, vol. 378, no. 9805, Nov 19, pp. 1812-1823.

Rhee, HS & Pugh, BF 2011, 'Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution', *Cell*, vol. 147, no. 6, Dec 9, pp. 1408-1419.

Rhee, HS & Pugh, BF 2012, 'Genome-wide structure and organization of eukaryotic pre-initiation complexes', *Nature*, vol. 483, no. 7389, Jan 18, pp. 295-301.

Ricciardelli, C, Bianco-Miotto, T, Jindal, S, Butler, LM, Leung, S, McNeil, CM, O'Toole, SA, Ebrahimie, E, Millar, EKA, Sakko, AJ, Ruiz, AI, Vowler, SL, Huntsman, DG, Birrell, SN, Sutherland, RL, Palmieri, C, Hickey, TE & Tilley, WD 2018, 'The Magnitude of Androgen Receptor Positivity in Breast Cancer Is Critical for Reliable Prediction of Disease Outcome', *Clin Cancer Res*, vol. 24, no. 10, May 15, pp. 2328-2341.

Robertson, JF, Osborne, CK, Howell, A, Jones, SE, Mauriac, L, Ellis, M, Kleeberg, UR, Come, SE, Vergote, I, Gertler, S, Buzdar, A, Webster, A & Morris, C 2003, 'Fulvestrant versus anastrozole for the treatment of advanced breast carcinoma in postmenopausal women: a prospective combined analysis of two multicenter trials', *Cancer*, vol. 98, no. 2, Jul 15, pp. 229-238.

Robinson, J, MacArthur, S, Ross-Innes, C, Tilley, W, Neal, D, Mills, I & Carroll, J 2011, 'Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1', *EMBO Journal*, vol. 30, pp. 3019-3027.

Robinson, JL & Carroll, JS 2012, 'FoxA1 is a key mediator of hormonal response in breast and prostate cancer', *Front Endocrinol (Lausanne)*, vol. 3, p. 68.

Robinson, JL, Hickey, TE, Warren, AY, Vowler, SL, Carroll, T, Lamb, AD, Papoutsoglou, N, Neal, DE, Tilley, WD & Carroll, JS 2014, 'Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype', *Oncogene*, vol. 33, no. 50, Dec 11, pp. 5666-5674.

Russo, J & Russo, IH 2004, 'Development of the human breast', *Maturitas*, vol. 49, no. 1, Sep 24, pp. 2-15.

Sagona, AP, Nezis, IP, Pedersen, NM, Liestøl, K, Poulton, J, Rusten, TE, Skotheim, RI, Raiborg, C & Stenmark, H 2010, 'PtdIns(3)P controls cytokinesis through KIF13A-mediated recruitment of FYVE-CENT to the midbody', *Nat Cell Biol*, vol. 12, no. 4, Apr, pp. 362-371.

Sahu, B, Laakso, M, Ovaska, K, Mirtti, T, Lundin, J, Rannikko, A, Sankila, A, Turunen, JP, Lundin, M & Konsti, J 2011, 'Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer', *The EMBO journal*, vol. 30, no. 19, pp. 3962-3976.

Sanchez, R, Nguyen, D, Rocha, W, White, JH & Mader, S 2002, 'Diversity in the mechanisms of gene regulation by estrogen receptors', *Bioessays*, vol. 24, no. 3, Mar, pp. 244-254.

Santagata, S, Thakkar, A, Ergonul, A, Wang, B, Woo, T, Hu, R, Harrell, JC, McNamara, G, Schwede, M, Culhane, AC, Kindelberger, D, Rodig, S, Richardson, A, Schnitt, SJ, Tamimi, RM & Ince, TA 2014, 'Taxonomy of breast cancer based on normal cell phenotype predicts outcome', *J Clin Invest*, vol. 124, no. 2, Feb, pp. 859-870.

Sanyal, A, Lajoie, BR, Jain, G & Dekker, J 2012, 'The long-range interaction landscape of gene promoters', *Nature*, vol. 489, no. 7414, Sep 6, pp. 109-113.



Sekiya, T & Zaret, KS 2007, 'Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo', *Molecular cell*, vol. 28, no. 2, pp. 291-303.

Serandour, AA, Brown, GD, Cohen, JD & Carroll, JS 2013, 'Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties', *Genome Biol*, vol. 14, no. 12, Dec 27, p. R147.

Sflomos, G, Dormoy, V, Metsalu, T, Jeitziner, R, Battista, L, Scabia, V, Raffoul, W, Delaloye, JF, Treboux, A, Fiche, M, Vilo, J, Ayyanan, A & Brisken, C 2016, 'A Preclinical Model for ER $\alpha$ -Positive Breast Cancer Points to the Epithelial Microenvironment as Determinant of Luminal Phenotype and Hormone Response', *Cancer Cell*, vol. 29, no. 3, Mar 14, pp. 407-422.

Shackleton, M, Vaillant, F, Simpson, KJ, Stingl, J, Smyth, GK, Asselin-Labat, ML, Wu, L, Lindeman, GJ & Visvader, JE 2006, 'Generation of a functional mammary gland from a single stem cell', *Nature*, vol. 439, no. 7072, Jan 5, pp. 84-88.

Shaoxian, T, Baohua, Y, Xiaoli, X, Yufan, C, Xiaoyu, T, Hongfen, L, Rui, B, Xiangjie, S, Ruohong, S & Wentao, Y 2017, 'Characterisation of GATA3 expression in invasive breast cancer: differences in histological subtypes and immunohistochemically defined molecular subtypes', *J Clin Pathol*, vol. 70, no. 11, Nov, pp. 926-934.

Shiina, H, Matsumoto, T, Sato, T, Igarashi, K, Miyamoto, J, Takemasa, S, Sakari, M, Takada, I, Nakamura, T, Metzger, D, Chambon, P, Kanno, J, Yoshikawa, H & Kato, S 2006, 'Premature ovarian failure in androgen receptor-deficient mice', *Proc Natl Acad Sci U S A*, vol. 103, no. 1, Jan 3, pp. 224-229.

Shoker, BS, Jarvis, C, Sibson, DR, Walker, C & Sloane, JP 1999, 'Oestrogen receptor expression in the normal and pre-cancerous breast', *J Pathol*, vol. 188, no. 3, Jul, pp. 237-244.

Sjöblom, T, Jones, S, Wood, LD, Parsons, DW, Lin, J, Barber, TD, Mandelker, D, Leary, RJ, Ptak, J, Silliman, N, Szabo, S, Buckhaults, P, Farrell, C, Meeh, P, Markowitz, SD, Willis, J, Dawson, D, Willson, JK, Gazdar, AF, Hartigan, J, Wu, L, Liu, C, Parmigiani, G, Park, BH, Bachman, KE, Papadopoulos, N, Vogelstein, B, Kinzler, KW & Velculescu, VE 2006, 'The consensus coding sequences of human breast and colorectal cancers', *Science*, vol. 314, no. 5797, Oct 13, pp. 268-274.

Smalley, M & Ashworth, A 2003, 'Stem cells and breast cancer: A field in transit', *Nat Rev Cancer*, vol. 3, no. 11, Nov, pp. 832-844.

Smith, CL & O'Malley, BW 2004, 'Coregulator function: a key to understanding tissue specificity of selective receptor modulators', *Endocr Rev*, vol. 25, no. 1, Feb, pp. 45-71.

Sørli, T, Perou, CM, Tibshirani, R, Aas, T, Geisler, S, Johnsen, H, Hastie, T, Eisen, MB, van de Rijn, M, Jeffrey, SS, Thorsen, T, Quist, H, Matese, JC, Brown, PO, Botstein, D, Lønning,

PE & Børresen-Dale, AL 2001, 'Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications', *Proc Natl Acad Sci U S A*, vol. 98, no. 19, Sep 11, pp. 10869-10874.

Sorlie, T, Tibshirani, R, Parker, J, Hastie, T, Marron, JS, Nobel, A, Deng, S, Johnsen, H, Pesich, R, Geisler, S, Demeter, J, Perou, CM, Lonning, PE, Brown, PO, Borresen-Dale, AL & Botstein, D 2003, 'Repeated observation of breast tumor subtypes in independent gene expression data sets', *Proc Natl Acad Sci U S A*, vol. 100, no. 14, Jul 8, pp. 8418-8423.

Souza, PCT, Textor, LC, Melo, DC, Nascimento, AS, Skaf, MS & Polikarpov, I 2017, 'An alternative conformation of ER $\beta$  bound to estradiol reveals H12 in a stable antagonist position', *Sci Rep*, vol. 7, no. 1, Jun 14, p. 3509.

Spilianakis, CG, Lalioti, MD, Town, T, Lee, GR & Flavell, RA 2005, 'Interchromosomal associations between alternatively expressed loci', *Nature*, vol. 435, no. 7042, Jun 2, pp. 637-645.

Stallcup, MR, Kim, JH, Teyssier, C, Lee, YH, Ma, H & Chen, D 2003, 'The roles of protein-protein interactions and protein methylation in transcriptional activation by nuclear receptors and their coactivators', *J Steroid Biochem Mol Biol*, vol. 85, no. 2-5, Jun, pp. 139-145.

Steigemann, P, Wurzenberger, C, Schmitz, MH, Held, M, Guizetti, J, Maar, S & Gerlich, DW 2009, 'Aurora B-mediated abscission checkpoint protects against tetraploidization', *Cell*, vol. 136, no. 3, Feb 6, pp. 473-484.

Stephens, PJ, Tarpey, PS, Davies, H, Van Loo, P, Greenman, C, Wedge, DC, Nik-Zainal, S, Martin, S, Varela, I, Bignell, GR, Yates, LR, Papaemmanuil, E, Beare, D, Butler, A, Cheverton, A, Gamble, J, Hinton, J, Jia, M, Jayakumar, A, Jones, D, Latimer, C, Lau, KW, McLaren, S, McBride, DJ, Menzies, A, Mudie, L, Raine, K, Rad, R, Chapman, MS, Teague, J, Easton, D, Langerod, A, Lee, MT, Shen, CY, Tee, BT, Huimin, BW, Broeks, A, Vargas, AC, Turashvili, G, Martens, J, Fatima, A, Miron, P, Chin, SF, Thomas, G, Boyault, S, Mariani, O, Lakhani, SR, van de Vijver, M, van 't Veer, L, Foekens, J, Desmedt, C, Sotiriou, C, Tutt, A, Caldas, C, Reis-Filho, JS, Aparicio, SA, Salomon, AV, Borresen-Dale, AL, Richardson, AL, Campbell, PJ, Futreal, PA & Stratton, MR 2012, 'The landscape of cancer genes and mutational processes in breast cancer', *Nature*, vol. 486, no. 7403, May 16, pp. 400-404.

Stingl, J, Eirew, P, Ricketson, I, Shackleton, M, Vaillant, F, Choi, D, Li, HI & Eaves, CJ 2006, 'Purification and unique properties of mammary epithelial stem cells', *Nature*, vol. 439, no. 7079, Feb 23, pp. 993-997.

Stochaj, U & Silver, P 1992, 'Nucleocytoplasmic traffic of proteins', *Eur J Cell Biol*, vol. 59, no. 1, Oct, pp. 1-11.

Swinstead, EE, Paakinaho, V & Hager, GL 2018, 'Chromatin reprogramming in breast cancer', *Endocr Relat Cancer*, vol. 25, no. 7, Jul, pp. R385-R404.

Takaku, M, Grimm, SA, De Kumar, B, Bennett, BD & Wade, PA 2020, 'Cancer-specific mutation of GATA3 disrupts the transcriptional regulatory network governed by Estrogen Receptor alpha, FOXA1 and GATA3', *Nucleic Acids Res*, vol. 48, no. 9, May 21, pp. 4756-4768.

Takaku, M, Grimm, SA, Roberts, JD, Chrysovergis, K, Bennett, BD, Myers, P, Perera, L, Tucker, CJ, Perou, CM & Wade, PA 2018, 'GATA3 zinc finger 2 mutations reprogram the breast cancer transcriptional network', *Nature communications*, vol. 9, no. 1, pp. 1-14.

Takaku, M, Grimm, SA, Shimbo, T, Perera, L, Menafra, R, Stunnenberg, HG, Archer, TK, Machida, S, Kurumizaka, H & Wade, PA 2016, 'GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler', *Genome Biology*, vol. 17, no. 1, 2016/02/27, p. 36.

Takaku, M, Grimm, SA & Wade, PA 2015, 'GATA3 in Breast Cancer: Tumor Suppressor or Oncogene?', *Gene Expr*, vol. 16, no. 4, pp. 163-168.

Takimoto, GS, Tung, L, Abdel-Hafiz, H, Abel, MG, Sartorius, CA, Richer, JK, Jacobsen, BM, Bain, DL & Horwitz, KB 2003, 'Functional properties of the N-terminal region of progesterone receptors and their mechanistic relationship to structure', *J Steroid Biochem Mol Biol*, vol. 85, no. 2-5, Jun, pp. 209-219.

Tanaka, H, Takizawa, Y, Takaku, M, Kato, D, Kumagawa, Y, Grimm, SA, Wade, PA & Kurumizaka, H 2020, 'Interaction of the pioneer transcription factor GATA3 with nucleosomes', *Nature communications*, vol. 11, no. 1, 2020/08/18, p. 4136.

Tarulli, GA, Laven-Law, G, Shehata, M, Walters, KA, Denis, IM, Rahman, MM, Handelsman, DJ, Dean, NR, Tilley, WD & Hickey, TE 2019, 'Androgen Receptor Signalling Promotes a Luminal Phenotype in Mammary Epithelial Cells', *J Mammary Gland Biol Neoplasia*, vol. 24, no. 1, Mar, pp. 99-108.

Theodorou, V, Stark, R, Menon, S & Carroll, JS 2013, 'GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility', *Genome Res*, vol. 23, no. 1, Jan, pp. 12-22.

Thürlimann, B, Keshaviah, A, Coates, AS, Mouridsen, H, Mauriac, L, Forbes, JF, Paridaens, R, Castiglione-Gertsch, M, Gelber, RD, Rabaglio, M, Smith, I, Wardley, A, Price, KN & Goldhirsch, A 2005, 'A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer', *N Engl J Med*, vol. 353, no. 26, Dec 29, pp. 2747-2757.

Tilley, WD, Marcelli, M, Wilson, JD & McPhaul, MJ 1989, 'Characterization and expression of a cDNA encoding the human androgen receptor', *Proc Natl Acad Sci U S A*, vol. 86, no. 1, Jan, pp. 327-331.

Tindemans, I, Serafini, N, Di Santo, JP & Hendriks, RW 2014, 'GATA-3 function in innate and adaptive immunity', *Immunity*, vol. 41, no. 2, Aug 21, pp. 191-206.

Tormey, DC, Lippman, ME, Edwards, BK & Cassidy, JG 1983, 'Evaluation of tamoxifen doses with and without fluoxymesterone in advanced breast cancer', *Ann Intern Med*, vol. 98, no. 2, Feb, pp. 139-144.

Toy, W, Shen, Y, Won, H, Green, B, Sakr, RA, Will, M, Li, Z, Gala, K, Fanning, S, King, TA, Hudis, C, Chen, D, Taran, T, Hortobagyi, G, Greene, G, Berger, M, Baselga, J & Chandarlapaty, S 2013, 'ESR1 ligand-binding domain mutations in hormone-resistant breast cancer', *Nat Genet*, vol. 45, no. 12, Dec, pp. 1439-1445.

Trainor, CD, Ghirlando, R & Simpson, MA 2000, 'GATA zinc finger interactions modulate DNA binding and transactivation', *J Biol Chem*, vol. 275, no. 36, Sep 8, pp. 28157-28166.

Tremblay, M, Sanchez-Ferras, O & Bouchard, M 2018, 'GATA transcription factors in development and disease', *Development*, vol. 145, no. 20, Oct 22.

Tseng, LM, Chiu, JH, Liu, CY, Tsai, YF, Wang, YL, Yang, CW & Shyr, YM 2017, 'A comparison of the molecular subtypes of triple-negative breast cancer among non-Asian and Taiwanese women', *Breast Cancer Res Treat*, vol. 163, no. 2, Jun, pp. 241-254.

Usary, J, Llaca, V, Karaca, G, Presswala, S, Karaca, M, He, X, Langerød, A, Kåresen, R, Oh, DS & Dressler, LG 2004, 'Mutation of GATA3 in human breast tumors', *Oncogene*, vol. 23, no. 46, pp. 7669-7678.

Usary, J, Llaca, V, Karaca, G, Presswala, S, Karaca, M, He, X, Langerod, A, Karesen, R, Oh, DS, Dressler, LG, Lonning, PE, Strausberg, RL, Chanock, S, Borresen-Dale, AL & Perou, CM 2004, 'Mutation of GATA3 in human breast tumors', *Oncogene*, vol. 23, no. 46, Oct 7, pp. 7669-7678.

Vakoc, CR, Letting, DL, Gheldof, N, Sawado, T, Bender, MA, Groudine, M, Weiss, MJ, Dekker, J & Blobel, GA 2005, 'Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1', *Mol Cell*, vol. 17, no. 3, Feb 4, pp. 453-462.

Valastyan, S & Weinberg, RA 2011, 'Tumor metastasis: molecular insights and evolving paradigms', *Cell*, vol. 147, no. 2, Oct 14, pp. 275-292.

van Doorninck, JH, van Der Wees, J, Karis, A, Goedknecht, E, Engel, JD, Coesmans, M, Rutteman, M, Grosveld, F & De Zeeuw, CI 1999, 'GATA-3 is involved in the development of serotonergic neurons in the caudal raphe nuclei', *J Neurosci*, vol. 19, no. 12, Jun 15, p. RC12.

Van Esch, H, Groenen, P, Nesbit, MA, Schuffenhauer, S, Lichtner, P, Vanderlinden, G, Harding, B, Beetz, R, Bilous, RW, Holdaway, I, Shaw, NJ, Fryns, JP, Van de Ven, W, Thakker, RV & Devriendt, K 2000, 'GATA3 haplo-insufficiency causes human HDR syndrome', *Nature*, vol. 406, no. 6794, Jul 27, pp. 419-422.

Vera-Badillo, F, Templeton, A, de Gouveia, P, Diaz-Padilla, I, Bedard, P, Al-Mubarak, M, Seruga, B, Tannock, I, Ocana, A & Amir, E 2014, 'Androgen receptor expression and



outcomes in early breast cancer: a systematic review and meta-analysis', *Journal of the National Cancer Institute*, vol. 106, p. djt319.

Visvader, JE 2009, 'Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis', *Genes Dev*, vol. 23, no. 22, Nov 15, pp. 2563-2577.

Walters, KA, Simanainen, U & Handelsman, DJ 2010, 'Molecular insights into androgen actions in male and female reproductive function from androgen receptor knockout models', *Hum Reprod Update*, vol. 16, no. 5, Sep-Oct, pp. 543-558.

Wang, Q, Li, W, Liu, XS, Carroll, JS, Janne, OA, Keeton, EK, Chinnaiyan, AM, Pienta, KJ & Brown, M 2007, 'A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth', *Mol Cell*, vol. 27, no. 3, Aug 3, pp. 380-392.

Wang, Q, Li, W, Zhang, Y, Yuan, X, Xu, K, Yu, J, Chen, Z, Beroukhim, R, Wang, H, Lupien, M, Wu, T, Regan, MM, Meyer, CA, Carroll, JS, Manrai, AK, Jänne, OA, Balk, SP, Mehra, R, Han, B, Chinnaiyan, AM, Rubin, MA, True, L, Fiorentino, M, Fiore, C, Loda, M, Kantoff, PW, Liu, XS & Brown, M 2009, 'Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer', *Cell*, vol. 138, no. 2, Jul 23, pp. 245-256.

Wang, Z, Gerstein, M & Snyder, M 2009, 'RNA-Seq: a revolutionary tool for transcriptomics', *Nat Rev Genet*, vol. 10, no. 1, Jan, pp. 57-63.

Wilkinson-White, L, Lester, KL, Ripin, N, Jacques, DA, Mitchell Guss, J & Matthews, JM 2015, 'GATA1 directly mediates interactions with closely spaced pseudopalindromic but not distantly spaced double GATA sites on DNA', *Protein Sci*, vol. 24, no. 10, Oct, pp. 1649-1659.

Winer, EP 2005, 'Optimizing endocrine therapy for breast cancer', *J Clin Oncol*, vol. 23, no. 8, Mar 10, pp. 1609-1610.

Xie, Q, Liu, Y, Cai, T, Horton, C, Stefanson, J & Wang, ZA 2017, 'Dissecting cell-type-specific roles of androgen receptor in prostate homeostasis and regeneration through lineage tracing', *Nat Commun*, vol. 8, Jan 23, p. 14284.

Yan, W, Cao, QJ, Arenas, RB, Bentley, B & Shao, R 2010, 'GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition', *Journal of Biological Chemistry*, vol. 285, no. 18, pp. 14042-14051.

Yang, HY & Evans, T 1992, 'Distinct roles for the two cGATA-1 finger domains', *Mol Cell Biol*, vol. 12, no. 10, Oct, pp. 4562-4570.

Yang, Z, Gu, L, Romeo, PH, Bories, D, Motohashi, H, Yamamoto, M & Engel, JD 1994, 'Human GATA-3 trans-activation, DNA-binding, and nuclear localization activities are organized into distinct structural domains', *Mol Cell Biol*, vol. 14, no. 3, Mar, pp. 2201-2212.

Yeh, S, Hu, YC, Wang, PH, Xie, C, Xu, Q, Tsai, MY, Dong, Z, Wang, RS, Lee, TH & Chang, C 2003, 'Abnormal mammary gland development and growth retardation in female mice and MCF7 breast cancer cells lacking androgen receptor', *J Exp Med*, vol. 198, no. 12, Dec 15, pp. 1899-1908.

Yoon, NK, Maresh, EL, Shen, D, Elshimali, Y, Apple, S, Horvath, S, Mah, V, Bose, S, Chia, D, Chang, HR & Goodglick, L 2010, 'Higher levels of GATA3 predict better survival in women with breast cancer', *Hum Pathol*, vol. 41, no. 12, Dec, pp. 1794-1801.

Yu, Q, Niu, Y, Liu, N, Zhang, J, Liu, T, Zhang, R, Wang, S, Ding, X & Xiao, X 2011, 'Expression of androgen receptor in breast cancer and its significance as a prognostic factor', *Annals of Oncology*, vol. 22, pp. 1288-1294.

Yuan, Y, Lee, JS, Yost, SE, Frankel, PH, Ruel, C, Egelston, CA, Guo, W, Gillece, JD, Folkerts, M, Reining, L, Highlander, SK, Robinson, K, Padam, S, Martinez, N, Tang, A, Schmolze, D, Waisman, J, Sedrak, M, Lee, PP & Mortimer, J 2020, 'A Phase II Clinical Trial of Pembrolizumab and Enobosarm in Patients with Androgen Receptor-Positive Metastatic Triple-Negative Breast Cancer', *Oncologist*, Nov 3.

Zaidan, N & Ottersbach, K 2018, 'The multi-faceted role of Gata3 in developmental haematopoiesis', *Open Biol*, vol. 8, no. 11, Nov 21.

Zhou, J, Ng, S, Adesanya-Famuiya, O, Anderson, K & Bondy, CA 2000, 'Testosterone inhibits estrogen-induced mammary epithelial proliferation and suppresses estrogen receptor expression', *FASEB J*, vol. 14, no. 12, Sep, pp. 1725-1730.

|