Investigating the Diversity of Anti-Estrogen Resistant Breast Cancer Stem Cells

A thesis submitted to The University of Manchester
for the degree of PhD Medicine,
Division of Cancer Sciences,
School of Medical Sciences,
Faculty of Biology, Medicine and Health

2017

Aida Sarmiento Castro

Table of Contents

LIST OF FIGURES **4**

LIST OF TABLES **7**

ABSTRACT **12**

DECLARATION **13**

COPYRIGHT **13**

ACKNOWLEDGEMENTS **14**

1. CHAPTER 1: INTRODUCTION **16**

1.1 HUMAN MAMMARY GLAND DEVELOPMENT 16
1.2 MOUSE MAMMARY GLAND 17
1.3 ENDOCRINE SYSTEM 19
1.3.1 Estrogens *20*
1.4 CANCER 28
1.4.1 Breast cancer *29*
1.5 STEM CELLS 41
1.5.1 Mouse mammary stem cells *41*
1.5.2 Human mammary stem cells *44*
1.6 CANCER STEM CELLS 47
1.6.2 Breast Cancer Stem Cells as therapeutic targets *53*
1.8 ALDH ISOFORMS AND CANCER 54
1.9 CANCER CELL HETEROGENEITY 58
1.5 HYPOTHESIS 60
1.6 AIMS 60

2. CHAPTER 2: MATERIALS AND METHODS **62**

2.1 MATERIALS 62
2.1.1 Reagents *62*
2.1.2 Cell culture media *63*
2.1.3 Antibodies and FACS-staining reagents *64*
2.1.4 TaqMan Quantitative Polymerase Chain Reaction (qPCR) primers *64*
2.2 METHODS 67
2.2.1 Cell culture *67*
2.2.2 Isolation of epithelial cells from metastatic patient-derived samples *72*
2.2.3 Flow Cytometry and Flow Activated Cell sorting *75*
2.2.4 In vivo tumour-initiating capabilities from ALDH positive and ALDH negative MCF-7 cells *78*
2.2.5 Gene expression analysis *80*
2.2.6 Single cell gene expression analysis *89*
2.2.14 Statistics and bioinformatics analysis *92*

3. CHAPTER 3: STUDY OF THE ALDH POS POPULATION IN BREAST CANCER CELLS **94**

3.1 INTRODUCTION 94
3.2 RESULTS 95
3.2.1 Effects of Tamoxifen and Fulvestrant on cellular viability *95*
3.2.2 Enrichment of ALDH positive cells after anti-estrogen treatment *98*
3.2.3 Stable ALDH1A3 Knockdown by lentiviral shRNA *104*
3.2.4 ALDH1A3 isoform mediates the stem cell activity enrichment observed after anti-oestrogen treatment in MCF-7 cells. *107*
3.2.5 Characterising anti-estrogen resistant ALDH positive cells *108*
3.3 DISCUSSION 116

4. CHAPTER 4: GENE EXPRESSION ANALYSIS OF ALDH POS CELLS **120**

4.1 INTRODUCTION....120
4.2 RESULTS....121
4.2.1 Analysis of the ALDH pos and ALDH neg populations in MCF7 cells by Affymetrix Arrays....121
4.2.2 Analysis of ALDH pos and ALDH neg populations in patient-derived samples by Affymetrix Arrays....131
4.2.3 Meta-analysis of MCF7 and patient-derived samples gene expression dataset....140
4.3 DISCUSSION....143

5. CHAPTER 5: AN INVESTIGATION INTO THE CELLULAR DIVERSITY OF THE RESISTANT ALDH POS POPULATION....149

5.1 INTRODUCTION....149
5.2 RESULTS....150
5.2.1 Designing a 96 gene panel to study ALDH pos heterogeneity after anti-estrogen treatment....150
5.2.2 Single cell technology and gating strategy....153
5.2.3 Identification of cell doublets....156
5.2.4 Identification of ALDH pos subclusters with Mclust and refinement via hierarchical clustering....160
5.2.5 Investigation of the relationship between clusters from different treatments....163
5.3 DISCUSSION....174

6. CHAPTER 6: DISCUSSION AND FUTURE PERSPECTIVES....180

7. CHAPTER 7: REFERENCES....185

8. CHAPTER 8: APPENDICES....211

8.1 ALDH1A3 SHRNA KNOCK-DOWN....211
8.2 FLOW CYTOMETRY PURITY CHECK....214
8.3 CO-AUTHORED PAPERS....215

Word Count: 54,038

List of Figures

Figure 1. 1 Schematic representation of the human mammary gland and the mouse mammary fat pad (MFP). 18
Figure 1. 2 Phases and hormone distribution during the woman's menstrual cycle. 20
Figure 1. 3 Biochemical reactions for the biosynthesis of steroid hormones from cholesterol. 22
Figure 1. 4 Estrogen production in the female body. 23
Figure 1. 5 Schematic representation of Estrogen Receptor alpha (ERα) structure and mechanism of action. 25
Figure 1. 6 Tridimensional structure of the ER LBD dimer complexed to Estradiol. 27
Figure 1. 7 Types of genomic alteration leading to cancer. 29
Figure 1. 8 Structure and metabolism of Tamoxifen. 36
Figure 1. 9 Chemical structure of Fulvestrant 37
Figure 1. 10 Recurrence rate in ER+ women after receiving 5 years of Tamoxifen therapy. 40
Figure 1. 11 Possible cellular hierarchy present in tumours and its plasticity. 49
Figure 1. 12 Combination therapy strategies to target both bulk and CSCs in breast cancers. 53
Figure 1. 13 ALDH isoforms and their preferred substrate. 55

Figure 2. 1 Second generation lentiviral TRIPZ plasmid 71
Figure 2. 2 Illustration of good and bad RNA quality samples indicated by their RIN values using the 2100 Bioanalyzer. 81
Figure 2. 3 GeneChip Probe Array. 84
Figure 2. 4 Affymetrix Human Gene 1.0 st probe array design. 84
Figure 2. 5 Flowchart of the whole Affymetrix Arrays procedure. 86
Figure 2. 6 Microarray hybridisation process. 89
Figure 2. 7 Flex Six IFC vs. 96.96 IFC gene expression chips (Biomark). 91

Figure 3. 1 Anti-estrogen drugs inhibit MCF-7 cell proliferation measured using SRB assay. 97
Figure 3. 2 Six-day *in vitro* treatment time-line. 97
Figure 3. 3 MFE fold change of MCF-7 cells after 6-days Tam and Fulv treatment in monolayer. 98
Figure 3. 4 Percentage of ALDH pos cells after Tam and Fulv treatment, measured by the Aldefluor assay. 99
Figure 3. 5 Percentage of $CD44^{+}/CD24^{-}$ cells after anti-estrogen treatment. 100
Figure 3. 6 Percentage of autofluorescent cells (AutoFluo) after anti-estrogen treatment. 102
Figure 3. 7 ALDH1A3 mRNA expression after anti-estrogen treatment in MCF-7 and T47D cells. 103
Figure 3. 8 Generation of a MCF-7 stable cell line knocking down ALDH1A3. 105
Figure 3. 9 Sorting Turbo-RFP positive cells. 106
Figure 3. 10 ALDH1A3 expression in ALDH1A3KD cells after 6 days treatment. 107

Figure 3. 11 Aldefluor assay in DOX-inducible ALDH1A3KD cells after 6-days in vitro treatment. 108
Figure 3. 12 ALDEFLUOR gating strategy. 109
Figure 3. 13 MFE of ALDH pos and ALDH neg cells in MCF-7 cells. 110
Figure 3. 14 *In vivo* tumour growth from 100 ALDH neg vs. ALDH pos MCF-7 cells after control, Tam and Fulv treatment. 112
Figure 3. 15 FACS plots showing percentage of ALDH pos cells, measured by the Aldefluor assay, in metastatic patient samples. 114
Figure 3. 16 MFE between ALDH pos and ALDH neg cells from PDS. 115

Figure 4. 1 RNA Quality control of ALDH neg and ALDH pos MCF-7 samples prior to Affymetrix microarray analysis. 122
Figure 4. 2 Representative scanned image of the Human Gene 1.0 st array. 123
Figure 4. 3 Box Plots showing the distribution of intensities after microarray normalisation in MCF-7 cells using RMA. 124
Figure 4. 4 Flow diagram showing the gene expression analysis performed in MCF-7 cells. 125
Figure 4. 5 SAM plot showing differentially expressed genes between controls vs. Tam and Fulv-treated unsorted MCF-7 cells using Delta=1. 125
Figure 4. 6 Graphic display of 28 of the most representative up and downregulated genes after SAM analysis between controls and Tam and Fulv-treated unsorted MCF-7 cells. 127
Figure 4. 7 SAM plot for the two class, paired data using Delta=1 showing differentially expressed genes between ALDH pos and ALDH neg in MCF-7. 128
Figure 4. 8 ALDH1A3 mRNA expression in ALDH pos and neg populations in MCF-7 cells by qPCR. 130
Figure 4. 9 Flow diagram showing the gene expression analysis performed in patient samples. 131
Figure 4. 10 Box Plot showing distribution of intensities after microarray normalisation in patient-derived samples using RMA. 132
Figure 4. 11 Heatmap of Rank Products illustrating the top 599 differentially expressed genes between ALDH neg and ALDH pos cells. 134
Figure 4. 12 Schematic representation of the clinical information for patients BB3RC91 and BB3RC91A. 135
Figure 4. 13 Heatmap showing the largest FC between ALDH pos and ALDH neg cells in at least 5 or more PDS. 138
Figure 4. 14 Expression of 17 ALDH isoform across 9 patient samples. 139
Figure 4. 15 Gene expression of the most representative isoforms in the ALDH pos population from patient samples. 139
Figure 4. 16 ALDH1A1 and ALDH1A3 gene expression in PDS using qPCR. 140
Figure 4. 17 Venn diagram illustrates meta-analysis of the cell line data between control, Tam, Fulv groups and the patient data. 141

Figure 5. 1 Heatmap showing expression fold change of the selected 96 genes between ALDH pos and ALDH neg cells in MCF-7 cells (A) and PDS (B). 152

Figure 5. 2 Flow Cytometry gating strategy to assess ALDH pos heterogeneity at the single cell level. 153
Figure 5. 3 C1 system workflow, architecture and photographic examples of captured cells within the C1 chambers. 154
Figure 5. 4 Flow chart showing single cell data processing. 155
Figure 5. 5 PCA showing cell distribution before and after applying ComBat to the single cell data set. 156
Figure 5. 6 Photographic images of the C1 chambers to assess cell doublet ratio. 157
Figure 5. 7 Identification of cell doublets using Mclust in ALDH pos MCF-7 cells. 159
Figure 5. 8 Hierarchical clustering of single ALDH pos cells using Mclust. 161
Figure 5. 9 Hierarchical clustering of ALDH pos cells using Mclust after merging the clusters. 163
Figure 5. 10 PCA of the merged ALDH pos cell clusters. 166
Figure 5. 11 Scatter plot of the DAPC analysis for single ALDH pos MCF-7 cells after anti-estrogen treatment. 171
Figure 5. 12 PCA loading of the DAPC. 172
Figure 5. 13 Working model for ALDH pos cellular heterogeneity before and after anti-estrogen treatment in MCF-7 cells. 179

Figure 8. 1 Detailed vector map of the pTRIPZ lentiviral vector. 213
Figure 8. 2 Assessment of purity of sorted cells. 214

List of Tables

Table 2. 1 List of reagents 62
Table 2. 2 Formulation of the media used for cell culture 63
Table 2. 3 Antibodies and other staining reagents for FACS..... 64
Table 2. 4 Isotype control antibodies..... 64
Table 2. 5 Pre-designed TaqMan qPCR primer/probe from ThermoFisher. 65
Table 2. 6 Anti-estrogen drugs for *in vitro* treatment..... 68
Table 2. 7 Michigan patient sample clinical information..... 74
Table 2. 8 Manchester patient sample clinical information. 75
Table 2. 9 FACS staining combining ALDH, CD44, CD24 and DAPI. 77
Table 2. 10 Isotype control information. 77
Table 2. 11 Experiment design for *in vivo* testing the tumour capabilities of ALDH pos and ALDH neg MCF-7 cells 79
Table 2. 12 Reagents used for Reverse Transcription 81
Table 2. 13 qPCR reagents..... 83
Table 2. 14 Human Array Gene 1.0 ST specifications..... 85
Table 2. 15 Hybridisation cocktail composition..... 88

Table 3. 1 Extreme Limiting Dilution Analysis (ELDA) of ALDH pos vs. ALDH neg cells following anti-estrogen treatment. 113

Table 4. 1 Number of differentially expressed genes between ALDH pos and ALDH neg applying different Delta values. 129
Table 4. 2 Representation of the 22 differentially expressed genes identified with SAM analysis Delta=1. Values are expressed as FC between ALDH pos and ALDH neg for each group (control, Tam and Fulv) and ranked according to the control-treated group. 129
Table 4. 3 Gene ontology from paired rank products between ALDH pos and ALDH neg cells in 9 different patient-derived samples. 136
Table 4. 4 List of the 7 commonly differentially expressed genes between ALDH pos and ALDH neg cells in control, Tam and Fulv-treated MCF7 cells and PDS. 141
Table 4. 5 List of the 4 genes differentially expressed between ALDH pos and ALDH neg cells shared in samples treated with anti-estrogen drugs. 142

Table 5. 1 Percentage of ALDH pos cells in each singlet-cluster and doublet-cluster. 160
Table 5. 2 Number of ALDH pos cells that belong to each cluster and each experiment. 162
Table 5. 3 Number of ALDH pos cells per cluster after merging using Mclust. 164
Table 5. 4 List of genes differentially expressed between Clade A and Clade B identified using Mclust. 168
Table 5. 5 Common significantly differentially expressed genes between Clades C and B. 173

Table 8. 1 Identification of the Aldehyde Dehydrogenase 1 family member A3 (ALDH1A3) shRNA sequences in the *Homo sapiens* ALDH1A3 mRNA NCBI sequence (NM_000693.3).....211

List of Abbreviations

7AAD	7-actinoaminomycin
AA	Arachidonic Acid
AB	Approximately Unbiased p-value
ALDH	Aldehyde dehydrogenase
ALDH neg	ALDH Negative
ALDH pos	ALDH Positive
AML	Acute Myeloid Leukaemia
ANOVA	Analysis Of Variance
APC	Allophycocyanin
APL	Acute Promyelocytic Leukaemia
AR	Anoikis Resistant
ATCC	American Type Culture Collection
ATRA	All-trans Retinoic Acid
BAAA	Bodipy-aminoacetaldehyde
B-ALL	B cell Acute Lymphoblastic Leukaemia
BP	Bootstrap Probability
CALLA	Common Acute Lymphoblastic Leukaemia Antigen
cDNA	Complementary DNA
CNA	Copy Number Alteration
COX	Cyclooxygenase
CSC	Cancer Stem Cell
Ct	Threshold Cycles
DAPI	4',6-diamidino-2-phenylindole
DAPC	Discriminant Analysis of Principal Component
DEAB	Diethylaminobenzaldehyde
DNA	Deoxyribonucleic Acid
DOX	Doxycycline
dNTP	Deoxynucleoside Triphosphate
E_2	Estrogen
ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EMA	Epithelial Membrane Antigen
ER	Estrogen Receptor
ER+	Estrogen Receptor Positive
ERE	Estrogen Response Element
ESA	Epithelial Specific Antigen
EtOH	Ethanol
FACS	Fluorescence Activated Cell Sorting
FAM	Fluorophore 6-carboxyfluorescein
FBS	Fetal Bovine Serum
FC	Fold Change
FDR	False Discovery Rate

FITC	Fluorescein-5-isothiocyanate
FMO	Fluorescence Minus One
Fulv	Fulvestrant
G	Gauge
g	G-force centrifugal speed
GH	Growth Hormone
h	Hours
HER2	Human Epidermal Growth Factor 2
HBS	Hepes-buffered saline
HBSS	Hank's Buffered Saline Solution
HSD	Honestly Significant Difference
IFC	Integrated Fluidics Circuit
KD	Knockdown
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	Linear Discriminant
LOD	Limit of Detection
M	Molar
mAb	Monoclonal Antibody
mM	Milimolar
MFE	Mammosphere Forming Efficiency
MFP	Mammary Fat Pad
MGB	Minor Groove Binder
min	Minutes
mRNA	Messenger Ribonucleic Acid
µM	Micromolar
NFQ	Non Fluorescent Quencher
nl	Nanoliter
nM	Nanometer
NSG	Nod Scid Gamma
P value	Probability Value
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PDS	Patient-Derived Samples
PE	Phycoerythrin Cyanin-7
Pg	Progesterone
PPAR	Peroxisome Proliferator-Activated Receptor
PR	Progesterone Receptor
PRL	Prolactin
PSI	Pounds per Square Inch
qPCR	Quantitative PCR
RT-PCR	Reverse Transcription PCR

RA	Retinoic Acid
RAR	Retinoic Acid Receptor
RFP	Red Fluorescent Protein
RIN	RNA Integrity Number
RMA	Robust Multi-array Average
RNA	Ribonucleic Acid
RNases	Ribonucleases
rRNA	ribosomal RNA
RP	Rank Products
RT	Reverse Transcription
SAGE	Serial Analysis of Gene Expression
SAPE	Streptavidin-Phycoerythrin
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SEM	Standard Error of the Mean
SERM	Selective Estrogen Receptor Modulators
siRNA	Small Interfering RNA
SPIA	Single Primer Isothermal Amplification
$sPLA_2$	Secretory Phospholipase A_2
SRB	Sulforhodamine B colorimetric assay
SSEA-4	Stage-Specific Embryonal Antigen-4
STR-PCR	Short Tandem Repeat- Polymerase Chain Reaction
shRNA	Short hairpin RNA
Tam	Tamoxifen
TDLU	Terminal Ductal Lobuloalveolar Unit
TE	Tris and EDTA-based buffer
TEB	Terminal End Bud
TGFβ	Transforming Growth Factor Beta
TH	Thyroid Hormone

Abstract

The University of Manchester
Aida Sarmiento Castro
Degree of PhD Medicine
Investigating the Diversity of Anti-Estrogen Resistant Breast Cancer Stem Cells
September 2017

Introduction: Although considerable progress has been made in breast cancer research over the past few decades, more than 11,000 women died from this disease in 2014 in the UK. *De novo* or acquired resistance to standard anti-estrogen therapy is one of the main reasons for the high cancer mortality incidence in estrogen receptor positive (ER+) breast tumours. There is evidence that Cancer Stem Cells (CSCs) are the culprits for the lack of response to anti-estrogen treatments. Therefore, this study sought to determine the effects of anti-estrogens on different CSC populations. In addition, we aimed to characterise anti-estrogen resistant CSCs and to investigate their cellular diversity, in order to identify biomarkers that can be therapeutically targeted.

Methodology: The proportion of CSCs following anti-estrogen treatment in ER+ cell lines and patient-derived samples (PDS) was assessed using flow cytometry to measure cell surface antibody binding, autofluorescent cells and aldehyde dehydrogenase (ALDH) activity. The stem cell activity of ALDH positive (ALDH pos) cells was assayed using *in vivo* transplantation and *in vitro* mammosphere formation assays and their gene expression profile was studied using Affymetrix Arrays. Finally, ALDH pos cellular diversity was investigated at the single cell level using the C1 system and Biomark HD technologies (Fluidigm).

Results: ALDH pos cells are enriched in ER+ cell lines and PDS following Tam and Fulv treatment**.** Although the exact mechanism remains elusive, the ALDH1A3 isoform appears to be important at driving this enrichment in MCF-7 cells. ALDH pos cells from MCF-7 and PDS have a distinct gene expression pattern when compared to ALDH negative cells. Additionally, single cell analysis in the ALDH pos population revealed that they are not a homogeneous cellular compartment but exist in 3 clusters defined by gene expression. One of these clusters is only present after anti-estrogen treatment.

Conclusion: Anti-estrogen therapies enrich for cells with CSC properties (ALDH pos cells) which comprise at least 3 distinct populations of cells. Future studies should assess the functional relevance of these cellular subgroups and how to therapeutically target them to improve treatment of anti-estrogen resistant breast cancer.

Declaration:

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright:

Acknowledgements

First of all I would like to thank my PhD supervisor Rob Clarke who has always been extremely supportive throughout these four years, not only about scientific matters but at the personal level too. He has also given me the opportunity to travel around the world attending several international conferences and has encouraged me to present my data to the scientific community on multiple occasions. This has not only has helped me to develop as a person but has also, undoubtedly, allowed me to improve my communication skills. I am certain Rob will be one of the best bosses I will ever have!

Secondly, I want to also say a big thank you to my laboratory supervisor Bruno Simões. The constant scientific discussions I have held with him (on many occasions at unearthly hours) has added invaluable insight into my project. His patience and his scientific knowledge have been much appreciated. I am sure Bruno and I will continue to share a delightful friendship.

I don't want to forget the people who have helped me along the way. Thanks to Javier Vences, Jose Gutierrez and Lili Costa for giving me the opportunity of starting my scientific career in UK and directing me on to the right path when I did not know where to start. Additionally, I am also very grateful to everyone at the University of Leicester who allowed me to become a better scientist before I started my PhD. I also want to thank every single person who has been part of the Breast Biology group throughout these four years. A special thank you to Angélica who has helped me out loads in the lab, especially in the early stages of my PhD, and to Kath whom I have chased many times to create last minute orders but has never told me off and has also helped me out several times in the lab. I would also like to thank Rachel who has always been helpful when needed and always has a smile for me, Denis who has helped me out in my *in vivo* experiments and Alice who I have enjoyed teaching and sharing time with. I would also like to thank the people who have visited us in the lab from other countries; Macarena, Ana, Francesca, Elena, Tiago and Natalia.

A special mention to Chris Clarke and Tony Banyard, members of the CRUK Manchester Institute core facilities, who have been extremely supportive and helped me beyond their duties.

I also want to thank Andy Sims for his bioinformatics support, however the real bioinformatics star is Eva Caamaño. It is thanks to her I have had my single cell data analysed bioinformatically and I will always thank her for her continuous support in all aspects of my life. She is just a great person to have near you.

I would also like to thank Professor Max Wicha who hosted me in his lab for 6 months and despite his busy agenda, he always had time for scientific discussions with me.

I must also thank my boyfriend Mark who has supported me unconditionally throughout these four years. He has been my personal English dictionary, my lab partner, the shoulder where sometimes I have cried after unsuccessfully experiments, and the person I share every day with.

Finally, I wanted to thank my parents, the best parents in the world, who have provided me with everything I have ever needed. It is because of them I have achieved everything I have in life and this thesis is entirely dedicated to them. I am so happy to make them feel proud of me.

1. Chapter 1: Introduction

1.1 Human mammary gland development

The human mammary gland begins development very early in the embryo, and it is one of the few organs whose maturity is not entirely complete at birth. Mammary development starts with the milk line growth, which is defined by the thickening of the ectodermal layer during embryonic development, which leads to the formation of the mammary anlage, the first recognizable stage of the mammary gland. When the bud is fully formed the mesenchyme cells adopt a circular disposition around the epithelial cells, which will then invade into the mammary gland to begin the rudimentary ductal branching development (Inman et al. 2015). The mammary gland undergoes isometric growth at the beginning of the neonatal period, and in spite of hormone receptors being expressed before puberty, the mammary gland grows in a hormone-independent way until puberty begins. Once ovulatory cycles have started, the highly proliferative structures known as terminal end buds (TEB) are formed at the terminal end of the ducts allowing them to invade further into the mammary gland (Figure 1.1 A) (Brisken et al. 2010). TEBs contain a layer of undifferentiated, multipotent cells (cap cells) at the invading front of the branch, responsible for the high proliferative rate of these structures. TEBs produce two major epithelial cell lineages, the secretory luminal epithelial cells and contractile myoepithelial cells, which separate the luminal cells from the basement membrane. Luminal cells surround the central lumen of TEBs and show apical-basal polarity which is maintained by the myoepithelial lineage through laminin-1 synthesis (Gudjonsson et al. 2005). Luminal cells initiate milk production and the differentiated myoepithelial cells play a key role in milk ejection during lactation. Luminal and myoepithelial cells can be distinguished depending on the level of expression of certain markers. For instance, luminal cells have been reported to express the epithelial membrane antigen (EMA), also known as MUC-1, and the epithelial specific antigen (ESA) (Gudjonsson et al. 2002), whereas

myoepithelial cells express CD10, also known as the common acute lymphoblastic leukaemia antigen (CALLA) (Gusterson et al. 1986). Furthermore, cytokeratins (KRT) such as KRT8, KRT18 and KRT19 can be used to identify luminal cells, while KRT5 and KRT14 as well as α-smooth muscle actin can be used to identify myoepithelial cells (LaBarge et al. 2007).

Estrogen (E_2), Progesterone (Pg), Prolactin (PRL), Growth Hormone (GH) and Thyroid hormone (TH) have been shown to play an important role in mammary development. These hormones are involved in ductal elongation and further branching of the ducts generating small alveoli or ductules leading to the formation of the functional unit of the human mammary gland in pre-menopausal women, the terminal duct-lobular unit (TDLU) (Figure 1.1 A). During pregnancy these alveolar buds will grow, multiply and differentiate into milk-secreting alveoli, a process which is driven by E_2, Pg PRL, GH, TH, adrenal steroids, oxytocin and insulin. After weaning, levels of lactogenic hormones drop and induce drastic tissue remodelling leading to the involution of the mammary gland. The gland is remodelled by different proteases, of which metalloproteinase-3 is the most prominent one. Therefore, the mammary gland and the surrounding tissue architecture undergo significant remodelling during pregnancy (Anderson 2002) (Watson 2006) (Inman et al. 2015). Finally, the mammary gland will go through the postmenopausal involution where some of the lobules and ducts are substituted by fat and intralobular stroma is replaced by collagen (Gusterson et al. 2000).

1.2 Mouse mammary gland

The mouse mammary gland, also referred to as the mouse mammary fat pad (MFP) appears as a rudimentary ductal structure during the first three weeks after birth. During the pubertal stage following hormone stimulus, TEB structures emerge and undergo elongation for about 7 to 9 more weeks (Figure 1.1 B). In contrast to humans, the mouse mammary gland does not possess TDLU, instead

alveolar buds are formed throughout puberty to eventually differentiate during pregnancy. Furthermore, the MFP has significantly more adipocytes, whereas human mammary gland contains more connective tissue. Following weaning, the MFP undergoes involution which normally takes around 2 weeks.

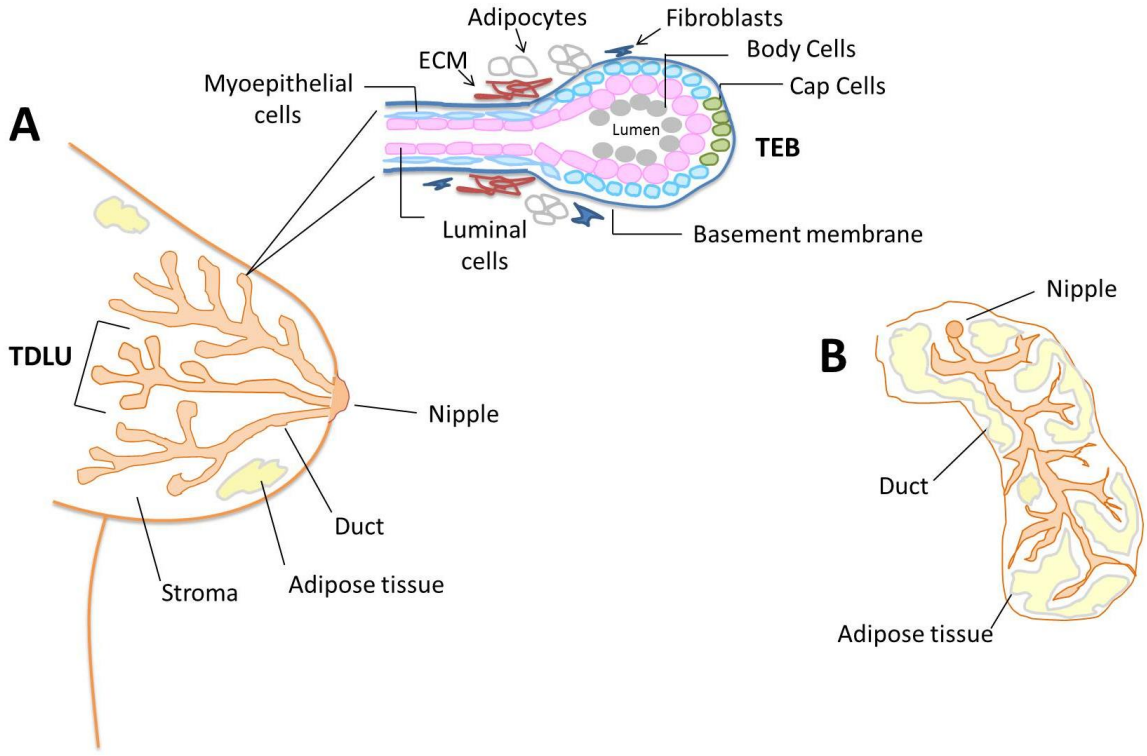


Figure 1. 1 Schematic representation of the human mammary gland and the mouse mammary fat pad (MFP).
A) The human mammary gland contains a ductal branching network that radiates from the mammary papilla or nipple. At the terminal end of the ducts there are highly proliferative structures called Terminal End Buds (TEBs), a bilayered structure that consists of luminal epithelial cells (pink cells) that produce milk during pregnancy, separated from the basement membrane (blue line) by the myoepithelial cells (blue cells). At the end of the TEB there are undifferentiated, multipotent cap cells responsible for cell proliferation. TEBs are surrounded by extracellular matrix (ECM) and connective tissue. Within the surrounding stroma there are various cell types including adipocytes, and fibroblasts. The functional portion of the human mammary gland is the Terminal Ductal Lobular Unit (TDLU), which forms from the TEB and consists of lactiferous duct, intralobular collecting duct, terminal ductules or acini and intralobular and interlobular stroma. **B)** The Mouse mammary fat pad consists of TEBs, the functional unit of the mouse mammary gland, however it lacks TDLU, and also contains significantly more adipocytes compared to humans.

The MFP has been used as a model to mimic and study human disease, and although differences in structure are evident, the basic biology of human and mouse mammary gland is similar. In mice, 5 pairs of mammary fat pads are distributed along the milk line, and usually number 4, located on the abdominal wall, is the one used for transplantation experiments as it is easy to find and access. The first mammary epithelial cell transplantation experiment was performed in 1959 by DeOme and colleagues. This experimental technique is explained in section 1.5.1 (Deome et al. 1959).

1.3 Endocrine system

The hypothalamic-pituitary-gonadal axis plays a significant role in mammary development. The hypothalamus, located in the brain, produce gonadotropin-releasing hormone (GnRH), which in turn stimulates the pituitary gland to secrete the gonadotropin hormones known as the luteinising hormone (LH) and the follicle-stimulating hormone (FSH). This stimulates the ovaries to begin E_2 and Pg synthesis (Figure 1.4). The entire hormone network is differentially secreted during the two phases of the ovarian cycle contributing to breast epithelium proliferation; the follicular phase characterised by an E_2 peak, and the luteal phase, which is characterised by a decrease in E_2 levels but high Pg levels (Figure 1.2). Cell proliferation occurs during luteal phase, specifically in the highly proliferative structures TDLU of the mammary gland. E_2 plays a major role in cellular proliferation and Pg receptor expression (Clarke et al. 1997), whereas Pg can both inhibit and promote proliferation of breast cancer cells. Pg was shown to induce cell proliferation in the mouse mammary gland and in human cells through a mechanism which involves Cyclin D1 and also by a paracrine mechanism mediated by receptor activator of NF-KB-ligand-dependent (RANKL) (Beleut et al. 2010, Nolan et al. 2016). The finding that the vast majority of breast cancers originate in the TDLU suggests a role for menstrual cycle changes in hormone levels in the development of breast cancer. Furthermore, different studies have reported an association between early menarche

and late menopause, and therefore higher exposure to menstrual hormones throughout the life-time, and a higher breast cancer incidence (Collaborative Group on Hormonal Factors in Breast Cancer 2012).

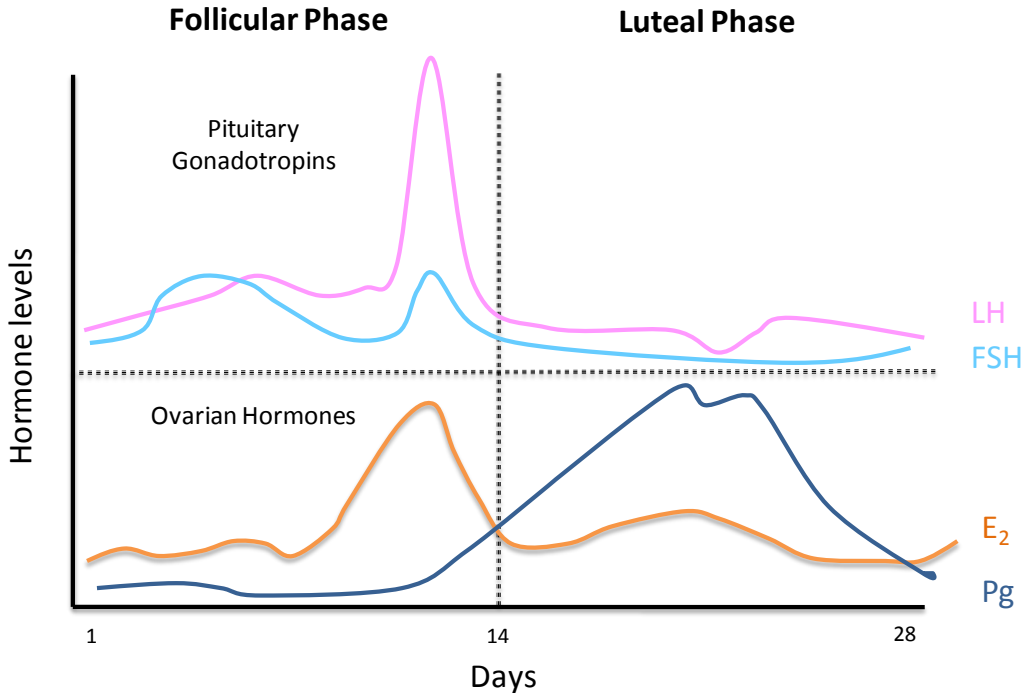


Figure 1. 2 Phases and hormone distribution during the woman's menstrual cycle.
Changes in levels of luteinising hormone (LH) and follicle-stimulating hormone (FSH) (pituitary gonadotropin hormones) are shown in the top panel, whereas levels of estrogen (E_2) and progesterone (Pg) (ovarian hormones) are in the bottom panel. Hormone fluctuations are shown over a 28-day period to represent one women menstrual cycle. The first 14 days constitute the follicular phase, whereas the last 14 represent the luteal phase.

1.3.1 Estrogens

E_2, known as C18 steroids (Figure 1.3) were described in 1931 by A. Butenandt, as important female hormones (Butenandt 1931). E_2 is derived from cholesterol, like all steroid hormones, and its biosynthesis is represented in Figure 1.3. E_2 is required for the normal reproductive processes, as well as more general metabolic roles, such as maintenance of bone mass and cognitive function. E_2 also promotes growth and differentiation in a wide range of tissues such as the mammary gland

(Nelson et al. 2001). E_2 are constituted of three major hormones; estrone, estradiol and estriol. Estradiol has the highest estrogenic activity and therefore is the one this piece of work will be referring to from now on as E_2. As mentioned previously, E_2 is produced in the ovaries, however it has also been shown that E_2 production is not exclusively associated with ovaries or even with females. Extragonadal sites such as adipose tissue as well as the brain appear to be another important source of E_2 production in the body. In post-menopausal women, E_2 biosynthesis is driven in these peripheral sites by the aromatase enzyme which converts androstenedione into estradiol (Figure 1.4) (Simpson et al. 2000).

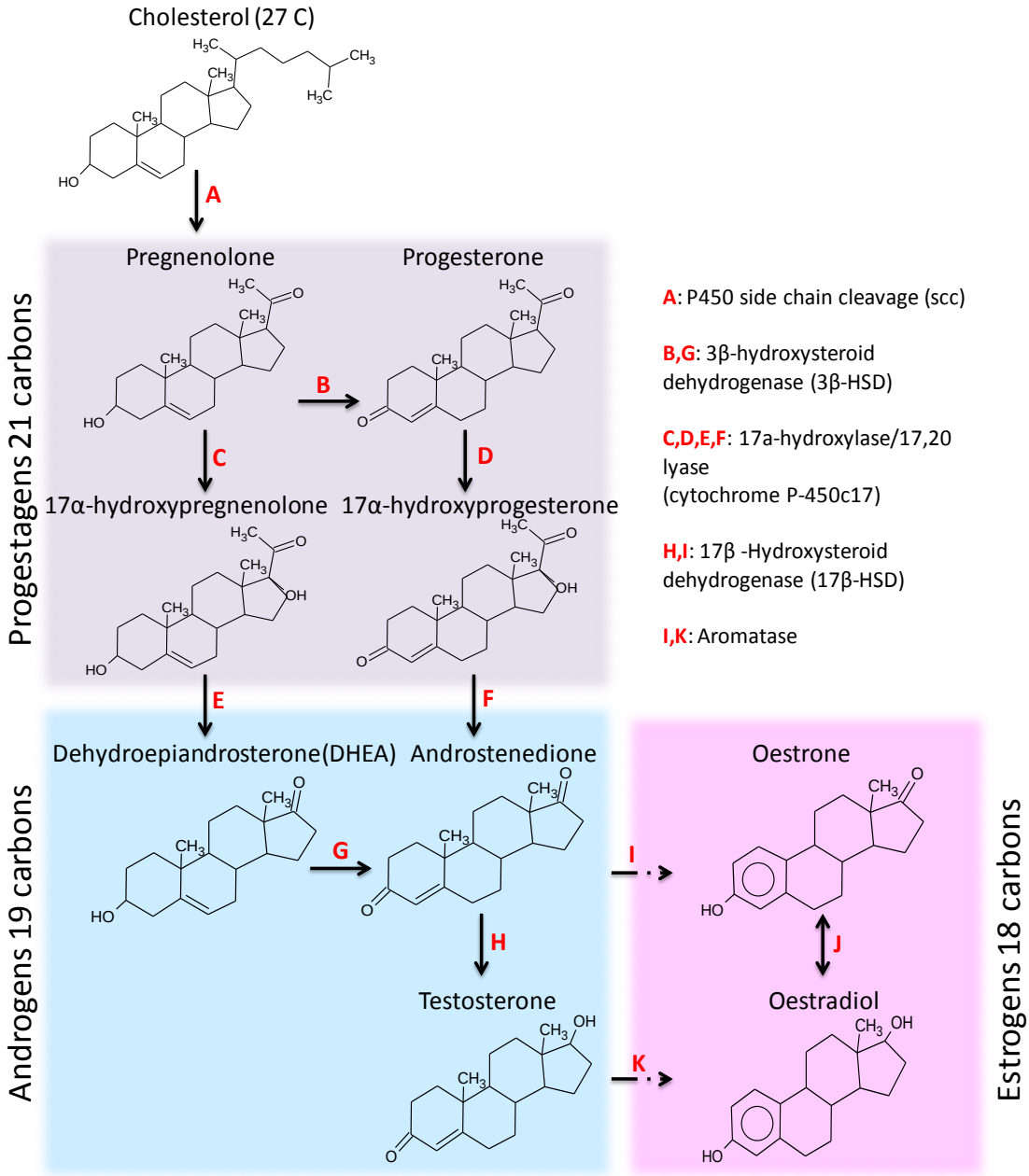


Figure 1. 3 Biochemical reactions for the biosynthesis of steroid hormones from cholesterol.
Steroid hormones derive from cholesterol. Cytochrome P450 scc cleaves the side chain of cholesterol converting it into pregnenolone. 17α-hydroxylase/17,20 lyase (cytochrome P-450c17) is the enzyme responsible for the conversion of C21 progestagens into C19 androgens. Aromatase enzyme converts androgens into E_2 by creating an aromatic ring structure. Solid arrows represent direct conversion, broken arrows indicate steps where intermediate metabolites have been omitted to simplify the diagram.

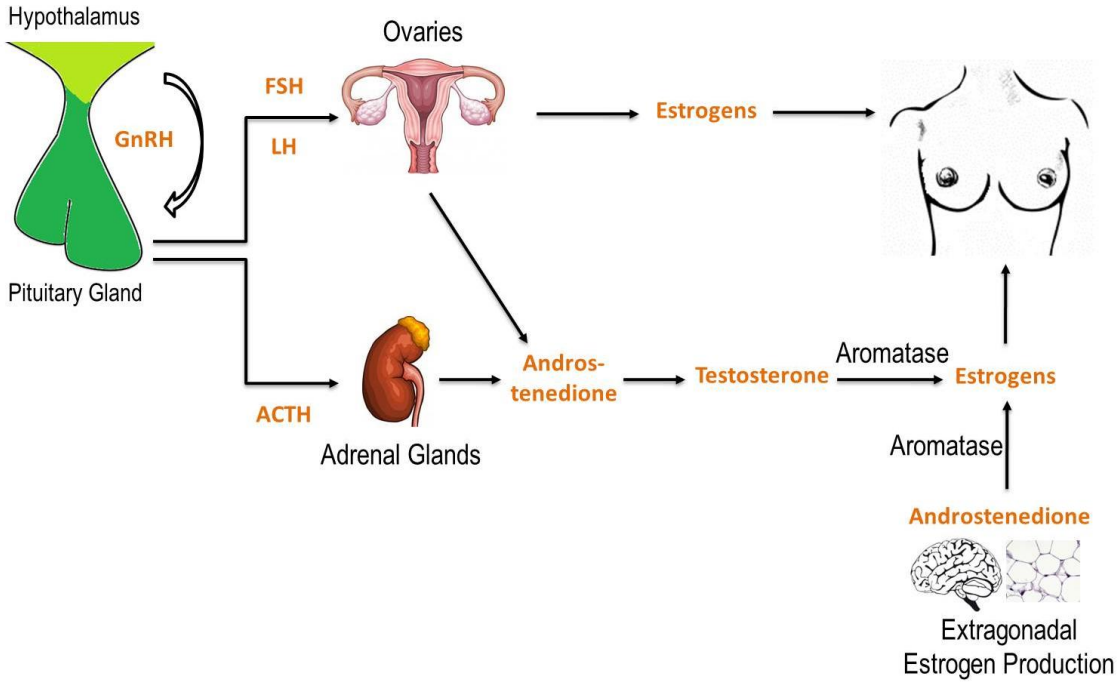


Figure 1. 4 Estrogen production in the female body.
The hypothalamus produces gonadotropin-releasing hormone (GnRH), which in turn stimulates the secretion of the luteinizing hormone (LH), the follicle-stimulating hormone (FSH) and the adenocorticotropic hormone (ACTH) in the pituitary gland. FSH and LH stimulate the ovaries to begin estrogen and progesterone synthesis. Another estrogen source comes from extra-gonadal sites such as the brain, adipose tissue or the bone. The production of adrenal androstenedione (the precursor of testosterone, refer to Figure 1.3) is driven by ACTH, whereas the production of gonadal androstenedione is mediated by the gonadotropins. Androstenedione is converted into testosterone which in turn produces E_2. At the peripheral sites, androstenedione is also converted into E_2 by the aromatase enzyme.

1.3.1.1 Estrogen receptor

E_2 action is mediated by the estrogen receptor (ER), which belongs to the nuclear hormone receptor family. The function of the members of this family relies on binding to steroid and thyroid hormones among others, to eventually enhance the transcriptional expression of target genes. ER has two isoforms, the estrogen receptor alpha (ERα) encoded by the ESR1 gene with a molecular weight of 66 KDa and the estrogen receptor beta (ERβ) encoded by ESR2 gene and a molecular weight of 54.2 KDa. Both isoforms contain multiple functional domains, first elucidated by Kumar et al., in 1986. They consist of an N-terminal domain (NTD), a DNA binding domain (DBD) that comprises two zinc

fingers, a ligand binding domain (LBD) and two activation sites; activation function 1 and activation function 2 (AF-1 and AF-2, respectively) (Figure 1.5). ERα and ERβ present a high homology degree in the DBD (95% of amino acid identity); however the two isoforms show less conservation in the LBD and in the amino-terminal domain (55% and 15% respectively) (Bulun et al. 2005, Kumar et al. 2011). ERα is the isoform used in the clinic to indicate the hormone dependence of breast tumours and therefore the term ER will refer to ERα exclusively from here onwards.

Due to its small size and lipophilic nature, estradiol freely crosses the cell membrane. It circulates through the cytoplasm and enters the nucleus where it binds the nuclear transcription factor ER in the LBD (Welshons et al. 1984). Upon binding to the receptor a conformational change occurs and homodimerization takes place. This leads to the receptor binding to different estrogen response elements (ERE), specific sequences of DNA located in the promoter or enhancer region of certain genes. Subsequently, several coactivators such as steroid receptor coactivator (SRC-1), glucocorticoid receptor-interacting protein 1 (GRIP) and activator protein (AP1) as well as transcription factors such as GATA3 and FOXA1 (Carroll et al. 2005) are recruited and interact with AF-2 to stimulate the association of the transcription initiation complex. This structural complex promotes the incorporation of the RNA polymerase II and transcription begins (Shiau et al. 1998). There are other molecules required for the appropriate function of the ER. The heat-shock protein HSP90, a chaperone molecule mainly located in the cytoplasm but also found in the nucleus, has been shown to be essential for keeping ER inactive during absence of E_2, as well as playing a key role providing an efficient hormonal response (Figure 1.5 B) (Devin-Leclerc et al. 1998).

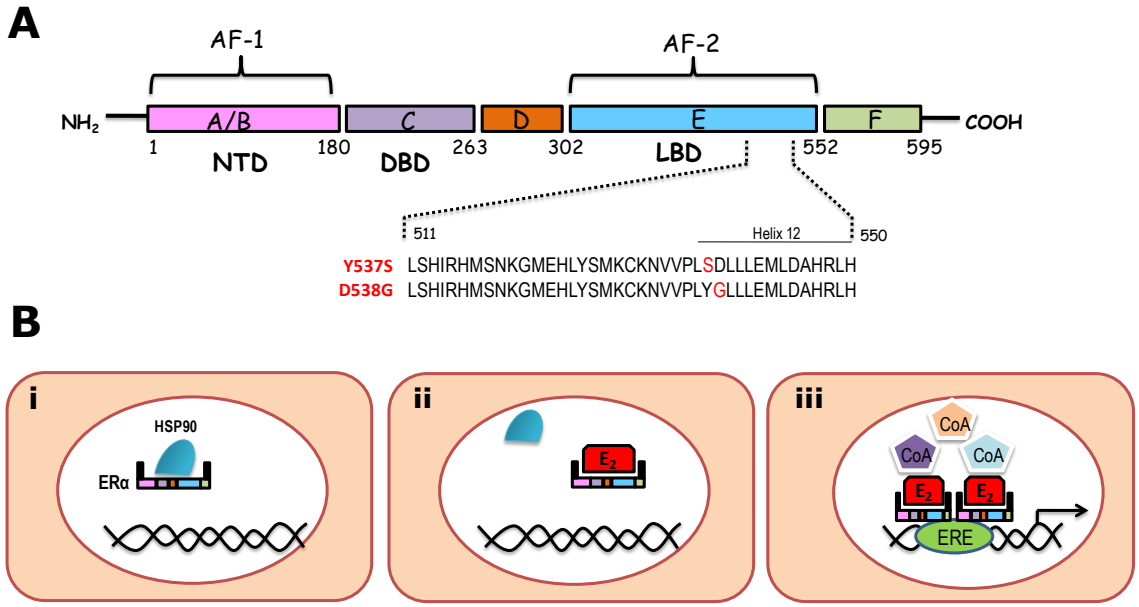


Figure 1. 5 Schematic representation of Estrogen Receptor alpha (ERα) structure and mechanism of action.
A) ER consists of 595 amino acids containing 5 structural domains; N-terminal domain (NTD), DNA binding domain (DBD), hinge region (D), ligand binding domain (LBD) and F region (F). The hinge region is important for providing a flexible region between the DBD and LBD, whereas the F region has shown to play an important role in receptor dymerisation. It also has two activation functions AF-1 and AF-2, located within the NTD and LBD respectively, which regulate the transcriptional activity of ER. A short amino acid sequence is represented in the LBD to show the two most relevant breast cancer mutations in the ER, Y537S and D538G (Robinson, Wu et al. 2013, Toy, Shen et al. 2013) **B)** ER is kept inactive by binding to HSP90 in the absence of E_2 (i). Upon binding to E_2 (ii), homodimerization occurs and the receptor binds to the estrogen response elements (iii). Subsequently, coactivators are recruited and transcription begins (iii).

Around 10-15% of luminal epithelial cells express ER in the normal breast. Clarke et al. performed double immunofluorescent labelling to identify ER-expressing cells and proliferative cells. Since E_2 induces proliferation in the mammary gland one might think that a proportion of ER positive cells would also be proliferating cells. However, these experimental approaches discovered that ER-positive cells and proliferative cells constituted different populations in the normal human breast epithelium. These results indicate that the ER positive cell population control the proliferation of the non-expressing ER cells by

paracrine signalling, probably involving growth factors such as amphiregulin (Clarke et al. 1997) (Ciarloni et al. 2007).

1.3.1.1.1 3D structure and mutations of the LBD in the ER

The LBD of ER consists of 12 helices (H1-H12), arranged in a three-layered antiparallel structure creating a hydrophobic core where E_2 positions. Helix 12 functions as a lid of the cavity known as the ligand-binding pocket (Figure 1.6 A). The A-phenol ring of E_2 binds to a specific rigid region in the ligand-binding pocket, which only accepts planar structures. Glu353, Arg394 and His524 are important amino acids of the LBD, which bind the hydroxyl groups that E_2 contains (Figure 1.6 B). Helix 12 plays a major role in the functioning of ER and its position depends on the chemical structure of the ligand that is binding to the receptor (i.e. whether it is agonist or antagonist) (Gronemeyer et al. 2004). Helix 12 is not only important at sealing the ligand-binding cavity for a perfect LBD-E_2 interaction, but also plays a key role at recruiting transcriptional coactivators (Brzozowski et al. 1997).

Toy and colleagues reported two mutations Tyr537Ser and Asp538Gly in the helix 12 of ER by performing next generation sequencing (Figure 1.5). These mutants, located in the LBD, were found to be present in approximately 20% of metastatic ER+ breast cancer patients who had previously been treated with endocrine therapy, mainly aromatase inhibitors. By looking at the tridimensional structure, an enhancement of the agonist conformation of ER was observed due to the hydrogen bonds formed between the amino acids of the Tyr537Ser and Asp538Gly (helix 12) mutants and the amino acids of Asp 351 (helix 3). Consequently, ER signalling was up-regulated even in the absence of ligand and although anti-estrogen therapy could revert this effect to a certain extent the cellular residual activity was still evident (Toy et al. 2013). Robinson et al., also studied acquired resistance in metastatic breast cancer and identified the same mutants (Robinson et al. 2013).

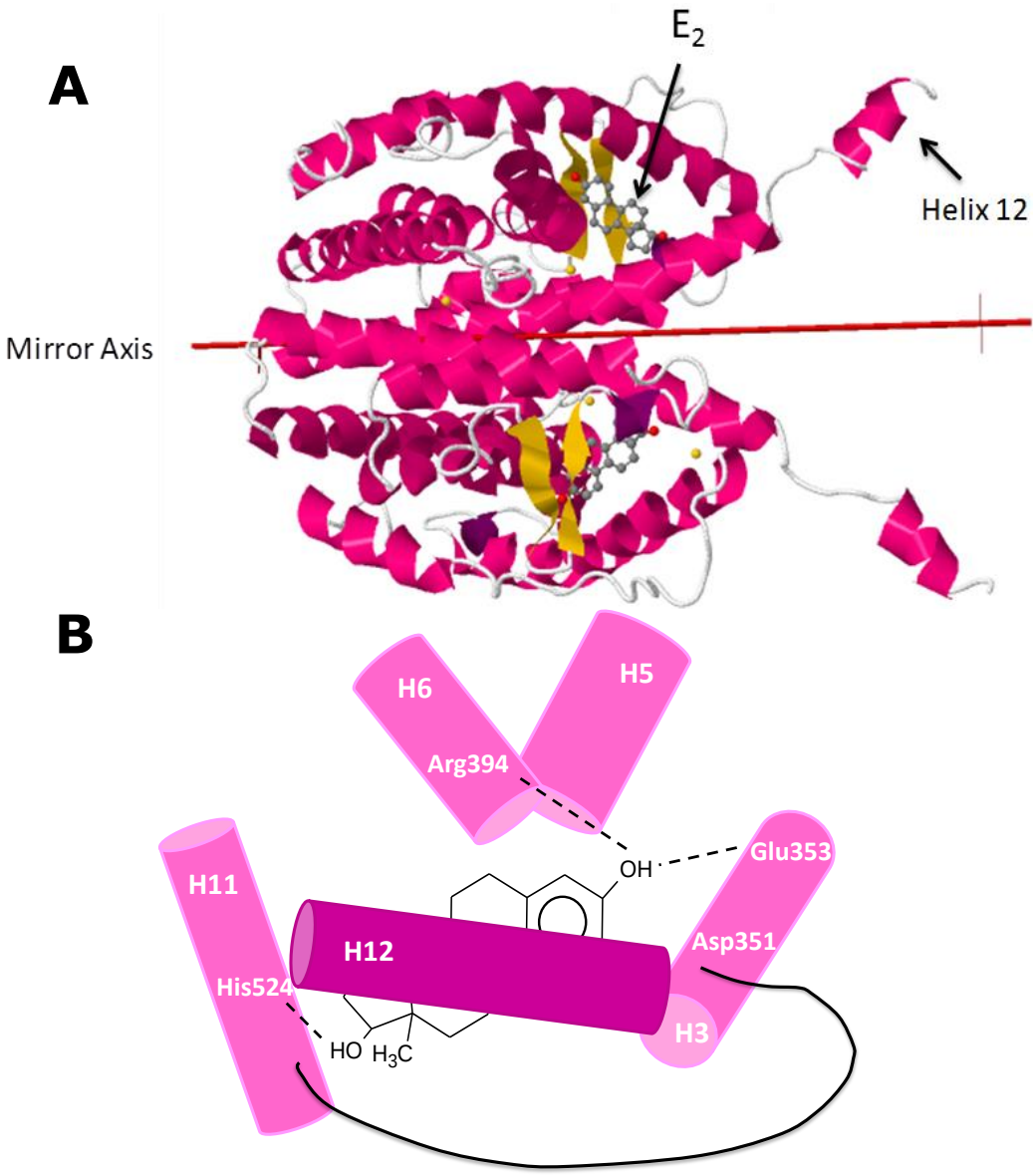


Figure 1. 6 Tridimensional structure of the ER LBD dimer complexed to Estradiol.
A) Alpha helixes are illustrated in pink and beta sheets in yellow. Mirror axis is shown to separate the two monomers. Two E_2 molecules are shown in the diagram, each one bound to a different subunit. Upon binding to E_2, helix 12 is orientated so that coregulator proteins can also bind to the ER. Figure generated from the Protein Data Bank (Protein Data Bank ID 1A52)(Berman et al. 2000). **B)** The diagram is based on the ER- E_2 crystallographic structure of the ligand-binding pocket. His524, Arg394, Glu353 and Asp351 are crucial amino acids of the LBD, which make hydrogen-bonding interactions (dashed lines) with E_2.

1.4 Cancer

The word "cancer" comes from the Greek word "carcinoma" which means crab, referring to the pattern exhibited by tumours when spread into the body, which looks similar to crab's legs. Cancer is an umbrella term, which comprises a group of diseases classified based upon their site of origin and is characterised by uncontrolled growth. Various hallmarks of cancer were proposed by Hanahan and Weinberg in 2000 (Hanahan et al. 2000), which were updated in 2011 (Hanahan et al. 2011). These were set to help provide a logical understanding of tumour development, and are considered essential qualities required for tumours to develop. Genetic instability and inflammation are thought to provide the driving force, promoting the hallmarks of cancer, such as the ability to overcome growth repression, resist cell death, induce replicative immortality and angiogenesis, and invade local tissue and form distant metastasis.

Peter Nowell first suggested the clonal evolution, also known as the stochastic model, in 1976 to help explain the origin of cancers. This cancer model was described as a process driven by any single cell that acquires multiple mutations or "hits" conferring selective growth advantages to a specific clone of cells. Subsequent rounds of clonal selections would lead to the enrichment of the most aggressive clones through developing metastatic properties and this would be facilitated by the obliteration of the more sensitive population by cancer therapies (Nowell 1976). In 1929, the German cytologist Boveri was the first scientist who established that chromosomal defects might lead to the uncontrolled proliferation of cells (Boveri 1929) and gave rise to the concept that cancer may arise from the disrupted function of normal cellular genes (proto-oncogenes). Somatic nucleotide polymorphism (SNPs) may allow tumour cells to gain malignant properties, however SNPs may also result in a loss of function, which normally happens in tumour supressor genes, such as TP53. Copy number alterations (CNAs) (gain or loss) are a form of structural variation in the genome, compared to the general population. The increase in gene copy number of ERBB2, which occurs in 18% of breast cancer patients, correlates with worse

patient outcome (Slamon et al. 1987). Chromosomal derangements are also frequent in cancer. For instance, a translocation between chromosomes 9 and 22, known as the Philadelphia chromosome, is present in patients with chronic myelogenous leukaemia (Figure 1.7). The gene *ABL1* from chromosome 9 is juxtaposed onto the *BCR* gene on chromosome 22. The resulting hybrid protein plays a role in the tyrosine kinase signalling and causes cells to divide uncontrollably, which leads to the development of cancer (Macconaill et al. 2010).

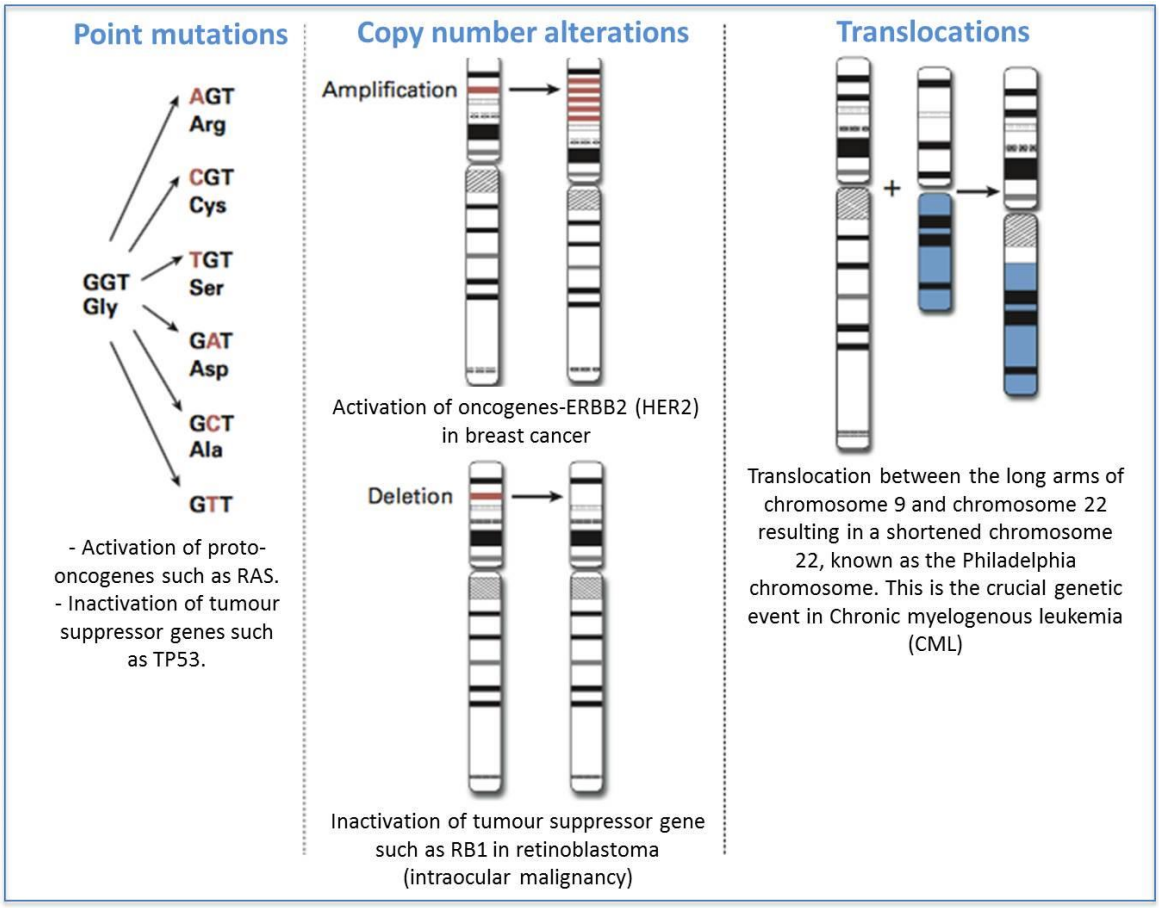


Figure 1. 7 Types of genomic alteration leading to cancer.
Adapted from MacConaill et al., 2010.

1.4.1 Breast cancer

Breast cancer is the most common female cancer in UK. Breast cancer accounts for more than 30% of all new cancer cases in women and the life-time risk of developing breast cancer is one in nine for

women. Furthermore, in 2014 there were 63,100 new cases of both invasive and in situ breast cancers and 11,433 deaths in the UK, which shows the need of further investigating the disease to improve prognosis and treatment (Cancer Research UK 2017).

Breast cancer represents a very heterogeneous disease not only between different tumours but also within a single tumour. This cellular and molecular heterogeneity provided the basis for the classification of these tumours within different subtypes resulting in more individualised therapies and better patient prognosis. Tumour diversity has been shown to emerge from different factors. Histopathology examination of the tumour was employed to study breast cancer heterogeneity throughout the vast majority of the 20th century, and even though it is not a very informative method it can predict the preferential organ site for metastasis. By looking at the gene expression profile of patient derived samples Perou et al., managed to identify four different groups of samples known as luminal, basal, HER2 and normal breast. And it was a year later, in 2001, when Sorlie et al., discovered the existence of substantial differences in gene expression patterns within the luminal subtype, which led them to create two new subgroups; luminal subtype A and luminal subtype B (Perou et al. 2000, Sorlie et al. 2001). Luminal tumours generally express ER (with higher expression in Luminal A compared to Luminal B) irrespective of PR expression, whereas basal-like tumours do not express these hormone receptors and conversely exhibit expression of cytokeratins 5, 14 and 17. On the other hand, HER2+ tumours are characterized by the over-expression of the member of the epidermal growth factor receptor (EGFR) due to amplification of chromosome 8. Sims et al., argue that differences within these subtypes might arise from differences in the cell of origin. The most differentiated subtype, the luminal subtype, would originate from the ER-positive progenitor cells, whereas the poorly differentiated basal subtype would arise from the primitive ER negative cancer stem cell (CSC) (Sims et al. 2007). In 1980s, standard Immunohistochemistry (IHC) and Fluorescence In Situ Hybridisation (FISH) of ER PR and HER2 markers began to be used for molecular subtyping profile of breast cancers and

they are considered an essential tool in order to evaluate endocrine therapy response (Dai et al. 2015).

A significant amount of mutated genes seem to be present in breast cancers. Moreover, certain somatic mutations are characteristic of specific breast cancer subtypes. For example, *PIK3CA* has been found to be mutated in approximately 45% of luminal A tumours, whereas only 9% of basal-like tumours seem to exhibit this mutation. On the other hand, the tumour suppressor gene *TP53* is predominantly mutated in basal-like cancers (80%), while the same mutation is only present in 12% of the luminal A cases (Koboldt 2012). Inherited mutations, known as germline mutations, play a key role in the development of some specific types of breast cancer such as BRCA1 mutations, which have been associated with a basal epithelial phenotype (Foulkes et al. 2003). Curtis and colleagues studied over 2,000 primary breast cancer specimens and observed that CNAs affect gene expression in breast cancers. Several known drivers such as *PTEN*, *MYC*, *CCND1* or *HER2* and putative driver copy number aberrations such as *MDM1*, *CDK3*, *CDK4* or *NCOR1* were found in this study. Likewise, *PPP2R2A* and *MTAP* showed heterozygous and homozygous deletions, respectively, in some of the breast cancer samples investigated. By combining the CNA and gene expression data from all patient samples, scientists identified 10 clusters, and each of them could be correlated with distinct clinical outcomes (Curtis et al. 2012). In addition, Nik-Zainal et al., identified a total of 1,682 likely driver mutations in 93 cancer genes. The 10 most commonly mutated genes, which comprise 63% of all drivers, are *TP53*, *PIK3CA*, *MYC*, *CCND1*, *PTEN*, *ERBB2*, *FGFR1*, *GATA3*, *RB1* and *MAP3K1* (Nik-Zainal et al. 2016).

As previously mentioned, there is a high percentage of women who die from breast cancer in UK, however primary tumours are not normally responsible for the high morbidity of breast cancers but the metastasis at distant sites. Between 10-15% of breast cancer patients develop a metastatic process within three years from the diagnosis of the primary tumour (Weigelt et al. 2005). The most common metastatic sites are lung (71%), bone (71%), lymph nodes (67%) and liver (62%) (Lee 1983).

1.4.1.1 Breast cancer therapies for estrogen receptor positive tumours

Although breast cancer has been known since the Egyptians more than 3000 years B.C, the first therapeutic approach ever undertaken to fight this ancient disease was surgical removal of the tumour. It was in 1757 when Henri Le Dran first recommended that excising the tumour with surgery could help treat breast cancer and later in 1867 Charles More introduced the technique of whole breast removal to eradicate the "centrifugal spread" of the tumour. In 1896, George Beatson was the first to show that removal of ovaries (oophorectomy) would lead to remission of metastatic breast cancer (Beatson 1896) and Stanley Boyd argued that ovaries' secretion was the culprit in promoting cancer development (Boyd 1899). Shortly after the discovery of the radioactive element radium at the beginning of the twentieth century by Marie Curie (1898), researchers found that radiotherapy was associated with successful cancer treatment and Ralston Paterson made a huge contribution to the field by delivering the most effective dose of radiation in cancers (Paterson et al. 1959). Edith Paterson, who carried out her research at the Christie hospital in Manchester, led the world´s first clinical trial of stilbestrol, a chemical analogue of E_2, to treat breast cancer. Ovarian radiation and stilbestrol were then used as therapeutic approaches between 1930 and 1940, followed by the use of triphenylamines, the precursor of Tamoxifen (Tam) (Nathanson 1946). In 1950, researchers began to observe that only one third of the patients could be cured by using surgery and radiotherapy treatments alone or in combination. Subsequently, researcher's efforts focused on finding chemicals that could improve patient's outcome, a treatment approach known as chemotherapy. Together, these treatments turned out to be more effective targeting breast cancer than when applied alone in most cases, which led to the establishment of the adjuvant therapy as a successful therapeutic approach against breast cancer. These findings, in association with improved diagnostic tools, such as modern mammography methods, and the introduction of clinical trials, managed to decrease the number of deaths for the first time in history in 1991 (DeVita et al. 2012).

Different researchers observed that administration of E_2 could induce tumours in some rodents which led for the first time to the association of the hormone steroid and cancer development (Lacassagne 1932, Lupulescu 1981) and by 1967 Elwood Jensen first identified the ER in breast cancer cells (Jensen et al. 1968). Contrary to normal breast tissue, ER-expressing cancer cells are also positive for the proliferative marker Ki67, which suggests that cancer cells enter the cell cycle in a cell autonomous manner in ER+ breast tumours (Shoker et al. 1999). A major breakthrough in breast cancer research occurred 4 decades ago when the hormone receptor measurements were established in the clinic. It was accepted that cancer growth of ER+ breast tumours is driven by endogenous E_2 and can often respond to endocrine therapy. These findings led to the introduction of antiestrogen drugs in the clinic. Therefore, ERα is a well-established predictive marker in breast cancer and its expression is important to decide whether anti-estrogen therapy should be given to breast cancer patients. Contrary to ERα, ERβ appears to be downregulated in lesions such as ductal hyperplasia and DCIS (Roger et al. 2001). A recent study has shown that ERβ is upregulated in breast CSCs (Ma et al. 2017), however the usefulness of ERβ in breast cancer prognosis and treatment remains to be fully defined.

Selective Estrogen Receptor Modulators (SERMs) are antiestrogen drugs that compete with E_2 for binding to the ER. These compounds exert estrogenic or antiestrogenic activities depending on the tissue. The most extensively used of these drugs is Tam. Tam, first known as ICI46474, is a non-steroidal compound which exhibits a dual activity, as it has been shown that it can act as agonist in the bone, liver and endometrium and as an antagonist in breast tissue. Interestingly, Tam can induce the expression of some estrogen-responsive genes such as transforming growth factor β (TGFβ), progesterone receptor (PR) and FOS even when it functions as an antagonist in the breast. The mechanism responsible for the dual behaviour of Tam appears to be cell-specific and mainly due to differences in the activation function AF-1 of ER (Hodges et al. 2003). Tam captured the attention of Drs Harper, Richardson and Walpole due to its potent antiestrogenic activity in animal models. This drug was first used as a post coital contraceptive method, however following clinical

testing studies, mainly based at the Christie Hospital in Manchester, led to the first clinical use of Tam by Moya Cole and Ian Todd and the subsequent approval of Tam for treatment of advanced breast cancer for the first time in UK in 1978 (Cole et al. 1971). In consequence, a clinical trial recruiting 1285 women with ER+ diseases was launched. Women were treated with Tam in the adjuvant setting for 2 years and results showed that Tam significantly delayed recurrence in breast cancer at 21 months follow up (Baum et al. 1983) . Later on in 1988, a meta-analysis of 28 clinical trials demonstrated that Tam effectively reduced 5-year mortality (Group 1988). NSABP P-1 clinical trial demonstrated Tam's efficacy as a preventive agent causing an average reduction of 69% occurrence in ER+ tumours in women at risk (Fisher et al. 1998). The IBIS-1 study found that Tam could prevent the development of breast cancer in high-risk women, although several side effects were shown (Cuzick et al. 2002). Findings from the ATLAS trial showed that 5-year Tam treatment significantly decreased breast cancer mortality throughout the first 15 years after diagnosis. Moreover, continuing Tam intake for an extra 5 years lead to a further reduction in recurrence and mortality (Davies et al. 2013). Tumour removal followed by Tam treatment could prevent any residual micrometastasis from developing leading to tumour relapse. In addition to Tam's ability for suppressing breast cancer recurrence, Tam has also shown to decrease contralateral primary breast cancers (Rutqvist et al. 1991).

Despite the benefits exerted by Tam as mentioned above, this anti-estrogen agent has also been linked to venous thrombolic events and endometrial cancer. Tam was also found to develop hepatic tumours in rats through a genotoxic mechanism involving Tam-induced DNA damage. Nonetheless, there is not substantial evidence for a role of Tam at promoting liver cancer in humans. Likewise, several studies have reported an increase in Type II endometrial carcinomas, which are non-E2 related but instead rely on chromosomal instability and TP53 mutations, after Tam exposure (Brown 2009).

Tam is a triphenylethylene molecule extensively metabolised by members of the superfamily cytochrome P450 (CYP), especially CYP2D6 and CYP3A4, into 4-Hydroxytamoxifen and N-desmethyltamoxifen.

However, the main metabolite that exerts antiestrogenic properties has been demonstrated to be Endoxifen (Figure 1.8) (Wu et al. 2009). Therefore, mutations in the CYP2D6 enzyme, which occurs in approximately 8% of Caucasian women, have been suggested to play a role in Tam-response failure (Hoskins et al. 2009). It has been reported that below the range $2x10^{-6}$ and 10^{-5} M of Tam concentration the drug exhibits cytostatic effects, whereas a higher concentration of Tam would induce cytotoxic effects independent of ER (Etienne et al. 1989). Since its approval, Tam has earned a place on a global stage due to its efficacy and affordability and it has become part of the World Health Organisation´s list of essential drugs for breast cancer treatment (Robertson et al. 2016).

Raloxifene is another SERM associated with a decrease in the incidence of breast cancer in high-risk women. The STAR study (study of tamoxifen and raloxifene) carried out in 2006 by the US National Centre Institute published that raloxifene (60 mg/day) has the same benefits as Tam (20 mg/day), and furthermore the former had a lower incidence of uterus cancer, however this drug is not used very often in the clinic (Vogel 2009).

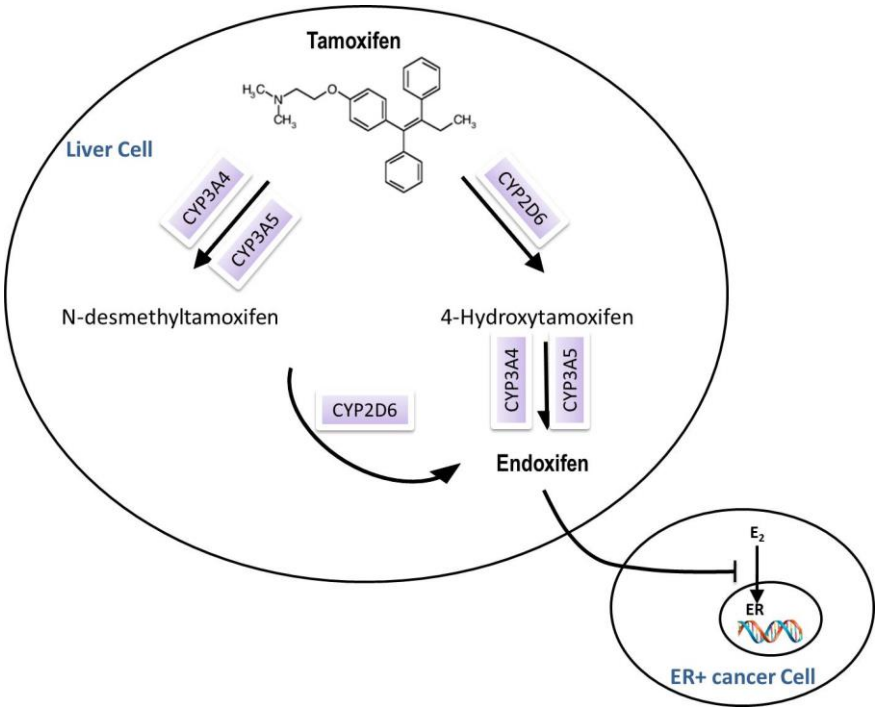


Figure 1. 8 Structure and metabolism of Tamoxifen.
Tam presents relatively little affinity for ER and needs to be metabolised in the liver. CYP3A4 and CYP3A5 are the more efficient enzymes responsible for Tam's transformation, however there are also other isoforms like CYP2C19, CYP2C9, CYP2B6 and CYP1A2 involved in this pathway that have been omitted in the diagram. The key metabolites of Tam are N-desmethyltamoxifen, 4-Hydroxytamoxifen and Endoxifen. The latter has 100 times more affinity for the ER than Tam and it will prevent E_2 from binding to the ER in ER+ cancer cells.

Efforts were made towards finding a substitute agent of Tam which would erase its agonist effects but at the same time maintaining its potency. The search for such a drug led to the discovery of several compounds, which were then named as pure anti-estrogens. The best known is Fulvestrant (Fulv), which was first named as ICI 182,780, and was later marketed by AstraZeneca as Faslodex. Fulv (7α-[9-(4,4,5,5,5-pentafluoro-pentylsulfinyl)nonyl]estra-1,3,5(10)-triene-3,17β-diol) is a steroidal compound analog of E_2 but with a different side-chain at the 7α position (Figure 1.9), which binds to the ER with much higher affinity than Tam (Dowsett et al. 2005). Fulv is a full antagonist whose mechanism of action consists of ER degradation with the subsequent impairment of the estrogenic transcriptional activity. Fulv is normally administrated by a monthly 500 mg 5 ml intramuscular injection.

Since the promising results of the 0020 and 0021 trials (Osborne et al. 2002) and the fact that there is not cross-resistance associated to Tam and Fulv, the later was adopted in 2002 as an anti-estrogen treatment for post-menopausal women with ER+ advanced breast cancer after unsuccessful Tam therapy (Howell 2006). It is important to highlight that Fulv is never given to breast cancer patients in the adjuvant setting (following surgery of the primary tumour).

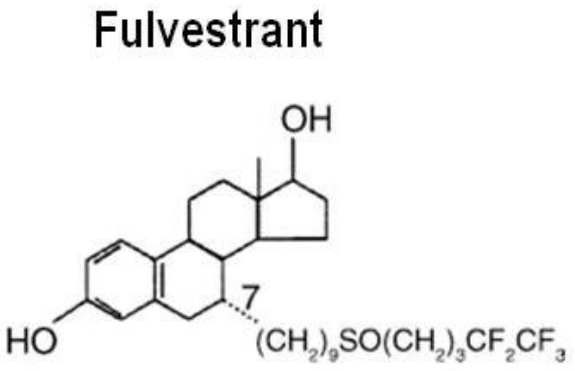


Figure 1. 9 Chemical structure of Fulvestrant
Figure adapted from Anthony Howell, 2005.

E_2 levels in postmenopausal women (10-20 pg/mL) are much lower than those in reproductive-age women (ranging from 50-500 pg/mL depending on the stage of the menstrual cycle) (Santanam et al. 1998). However, breast cancer cells contain 10-20 times more E_2 than plasma in post-menopausal women (Geisler 2003). The main E_2 present in the plasma of postmenopausal women is sulfate estrone, a weak inactive E_2 compound synthesised from androgens at the peripheral sites by an enzyme called aromatase. Breast cancer cells will then turn the inactive form of estrone into its active form (Pasqualini et al. 1988), which will be further metabolised to 17β-estradiol through the 17β-hydroxisteroid dehydrogenase type 1 (Brodie et al. 1997).

Aromatase inhibitors (AIs) function by inhibiting aromatase activity which leads to a 98% decrease in circulating E_2 levels. The use of AIs has been restricted to post-menopausal women and it has replaced Tam as the therapy of choice in post-menopausal women with ER+ breast cancer (Goldhirsch et al. 2009). Although AIs can stop peripheral E_2 biosynthesis

from androgens in pre-menopausal women, low levels of E_2 triggers a compensatory feedback increasing gonadotropin secretion from the hypothalamus and eventually restoring E_2 levels through ovarian function (Jordan 2006). Furthermore, levels of the enzyme are much higher in ovarian cells in reproductive-age women, whereas aromatase appears to be more present in adipose tissue in post-menopausal women (Bulun et al. 2005).

Three different generations of AI have been developed; first generation (Aminoglutethimide), second generation (Fadrozole and Vorozole) and third generation (Letrozole and Anastrozole, which are non-steroidal and Exemestane, which is steroidal). Several health issues such as bone loss as well as a lack of enzyme specificity have been associated with the two first generation drugs. It is important to emphasise that around 80% of breast cancers occur in women older than 50, which normally correspond with post-menopausal women, hence highlighting the need for an effective therapy in this group of patients with AI providing the best results (Musgrove et al. 2009).

Inhibitors of the cyclin-dependent kinase 4 and 6 (CDK4/6) such as Palbociclib, block the progression of ER+ breast cancer cells from G1 to S phase of the cell cycle. It has been shown that the combination of Palbociclib with anti-estrogen therapy has synergistic effects in ER+ breast cancer cell lines. Moreover, the addition of Palbociclib to letrozole (Hu et al. 2017), and also to Fulv (Cristofanilli et al. 2016) has been recently approved as a novel therapy for women with advance breast cancer diseases.

Several unfinished and completed clinical trials have tested inhibitors of the most frequently active pathways in breast cancer. For instance, the mTOR pathway inhibitor Everolimus was given to postmenopausal women in combination with Tam and results showed a superior clinical benefit rate (proportion of patients with a complete or partial response or with stable disease at week 24) in the Everolimus plus Tam group, compared to Tam alone (Bachelot et al. 2012). In a separate clinical trial, patients received Everolimus with Exemestane or Exemestane alone and researchers reported benefits in progression free survival and clinical benefit rate (complete response + partial response +

stable disease for ≥ 24 weeks) in favour of Exemestane with Everolimus (Piccart et al. 2014). Inhibitors of other cellular processes such as AKT and PI3K pathways are also being tested in patients in combination with anti-estrogen drugs.

Overall, patients with early ER+ breast tumours are given drug therapy before (neoadjuvant therapy) or after surgical removal of the primary tumour (adjuvant therapy). Tam is given to premenopausal women following surgery and chemotherapy, , whereas aromatase inhibitors are given to postmenopausal women. Patients with advanced breast tumours will receive anti-estrogen therapy as first-line treatment, and chemotherapeutic agents such as Capecitabine as second or third-line treatment.

1.4.1.1.1 Anti-estrogen resistance

ER+ breast cancer patients tend to have a better prognosis than patients lacking ER expression, including longer disease-free interval and overall survival (Knight et al. 1980). Despite the fact that patients with ER+ breast cancers have the greatest chances of response to anti-estrogen therapy, a high percentage of patients will have *de novo* resistance or develop acquired resistance after an initial response to the treatment. In 2011, Davies and colleagues reported the results of a meta-analysis from over 10,000 women with ER+ diseases, where Tam was given as adjuvant therapy. The results, illustrated in Figure 1.10, showed that despite the evident benefit shown by Tam at reducing recurrence rate by 28,5%, there are still a large proportion of women who will develop resistance to the therapy within 15 years (71,5%) (Davies et al. 2011). In some of the cases, patients can respond to first-line endocrine therapies, however the mechanisms triggering this event are still unknown.

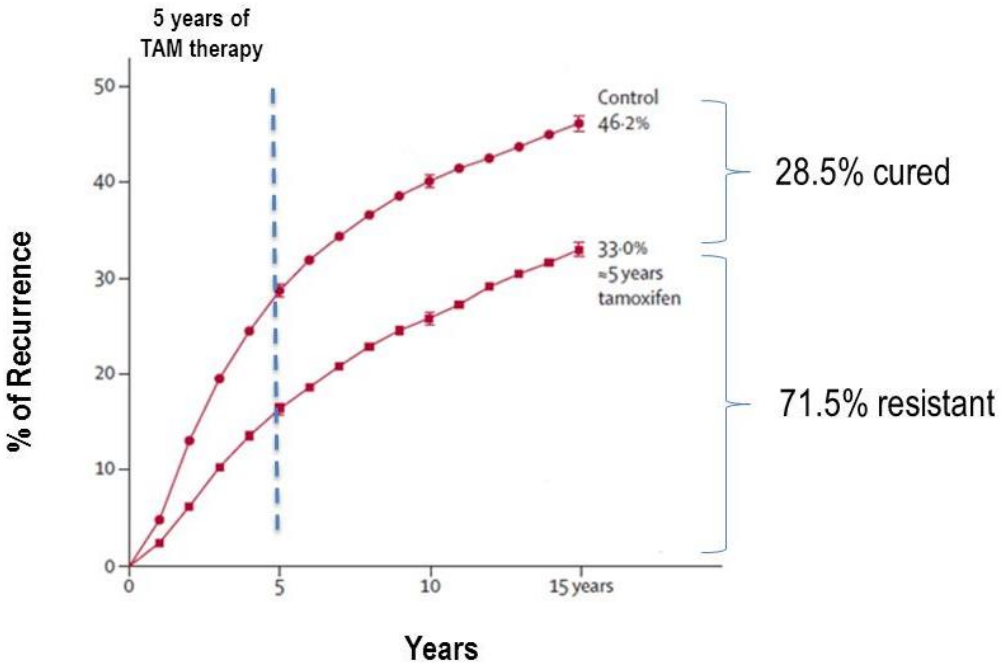


Figure 1. 10 Recurrence rate in ER+ women after receiving 5 years of Tamoxifen therapy.
Women were considered ER+ when measurements of ER were ≥ 10fmol/mg cytosol protein. Figure adapted from Early Breast Cancer Trialists' Collaborative Group, 2011.

There is a wide range of mechanisms suggested to be involved in endocrine resistance. Some of these mechanisms are loss of ER expression (<20% of the cases) (Osborne et al. 2011), mutations in the ER, over expression of ER coactivators leading to constitutive ER-mediated transcription, (Musgrove et al. 2009), antiestrogen binding sites (AEBS) (Pavlik et al. 1992) and multidrug resistance protein 1 (mdr1) (Clarke et al. 2001). A role of the cancer associated fibroblasts (CAFs) present in the stroma (Busch et al. 2012) and also of the epidermal growth factor receptor (EGFR) pathway has been implicated in endocrine resistance (Fan et al. 2007). Work from Larsen and colleagues found that the tyrosine kinase SRC was upregulated in patients who had received Tam as first line-therapy. These results suggest a role for SRC as a predictive biomarker for Tam resistance (Larsen et al. 2015). Finally, breast CSCs have also been shown to be unresponsive to anti-estrogen therapy and be responsible for driving tumour repopulation and relapse (refer to section 1.6.1 and 1.6.1.1 for further information regarding CSCs and anti-estrogen resistance).

1.5 Stem Cells

In 1877, E. Haeckle was the first researcher who used the term stem cells (*Stammzelle*) to describe the fertilized egg as the cell from which arise the rest of the cells that form a developing organism (Maehle 2011). Nowadays, the concept of stem cells can refer to normal or pathological, natural or artificial stem cells. Stemness refers to two cellular properties: the ability to self-renew, which means the ability to generate at least one new stem cell during cell division to maintain the stem cell pool, and the ability to differentiate in order to give rise to more specialised cells (Laplane 2016).

1.5.1 Mouse mammary stem cells

The mammary gland undergoes tissue remodelling during pregnancy to generate the lobuloalveoli capable of secreting milk. Hormones such as prolactin and progesterone are essential for this developmental process. After pregnancy, the mammary gland suffers a postlactional involution to return to the pre-pregnant state. This huge expansion of the epithelial content of the mammary gland and the subsequent cell death during involution highlights the presence of a stem cell population, known as the mammary stem cell.

An enormous progress has been made in the mammary stem cell field, with large contributions from *in vivo* studies in mice. The MFP transplantation experimental approach, pioneered by DeOme and co-workers, has shown to be the gold standard to test for stem/progenitor cells, by assessing their ability to regenerate the structure and cell lineages of the mammary gland. In this technique the mouse MFP is de-epithelialised and subsequently donor explants or disaggregated cells are transplanted to eventually produce mammary outgrowth. DeOme's experimental approach in 1959, showed that small epithelial duct fragments from any portion of the mammary gland were capable of repopulating the entire glandular structure which led to our current understanding of the mammary stem cell field. In 1998, Kordon and Smith provided further evidence for the clonality of mammary outgrowths

by retrovirally tagging epithelial fragments using the mouse mammary tumour virus (MMTV) (Kordon et al. 1998). MMTV would randomly insert its provirus into the somatic cell DNA of CzechII mice and provided evidence that a functional mammary gland could be generated from a single stem cell. A few years later, Shackleton and colleagues also reported the existence of mouse mammary stem cells, as they were able to generate a functional mammary gland in mice from a single stem cell marked as Lin^- (negative for the endothelial marker CD31 and the hematopoietic antigens CD45 and TER119) $CD29^{high}CD24^+$ (Shackleton et al. 2006). They also identified a luminal mammary epithelial population marked as $CD29^{lo}CD24^+$. Shackleton et al., were able to enrich the mammary repopulating units (MRU) from 1/4900 to 1/64 by using those two cell surface markers. Moreover, in a later publication it was found that the $CD29^{lo}CD24^+$ population could be further subdivided using CD61 (β3-integrin), which represents a marker for progenitors (Asselin-Labat et al. 2007).

Stingl and co-workers observed that mammary stem cells could be enzymatically isolated from adult mouse mammary tissue and regenerated mammary outgrowth when implanted into a secondary mouse. Further characterisation and selection of these MRU led to the identification of a population of cells $CD45^-Ter119^-CD31^-Sca\text{-}1^{low}CD24^{med}CD49f^{high}$ with the ability of generating mammary outgrowth from a single cell. Moreover, these cells were found to be in G1 or S/G2/M fraction showing that these mammary stem cells are a cycling population (Stingl et al. 2006). These results are in contradiction with earlier findings from Welm et al., who applied the Sca-1 antigen to enrich for mammary stem cells ($Sca\text{-}1^+$) with regenerative potential when transplanted into cleared mouse mammary fat pads (Welm et al. 2002).

Staining with the cell surface marker CD24 in addition to lineage-specific markers allowed the identification of three different populations in the mouse mammary gland; non-epithelial ($CD24^-$), myoepithelial/basal ($CD24^{low}$) and luminal epithelial cells ($CD24^{high}$). Mouse mammary stem cells were found to belong to the basal compartment as shown by the exclusive ability of $CD24^{low}$ cells to repopulate the cleared mammary fat pad (Sleeman et al. 2006).

Nevertheless, all studies mentioned previously are based on transplantation experimental approaches, which might not mirror totally the stem cell behaviour and fate *in situ*. Van Keymeulen et al., performed genetic lineage-tracing *in vivo* using mice, and demonstrated that the mammary gland contains luminal and basal unipotent stem cells (Van Keymeulen et al. 2011). Furthermore, Van Amerongen et al., performed lineage-tracing analysis using the Wnt-target gene Axin2 to label and trace mammary stem cells and also found that both luminal and basal stem cells only give rise to luminal and basal differentiated cells respectively during physiological conditions. However, when these unipotent basal stem cells were transplanted into cleared fat pads they behaved as multipotent stem which suggests that transplantation assays triggers a regenerative potential that is not present during developmental conditions (Van Amerongen et al. 2012).

A recent publication identified PROCR as a novel target of the Wnt signaling pathway. $PROCR^{+}$ cells are a basal population (although they show lower expression levels of KRT5/KRT14 compared with other types of basal cells) with an epithelial-to-mesenchymal transition (EMT) phenotype that has been shown to allow multipotent capabilities, demonstrated by *in vivo* lineage-tracing (Wang et al. 2015).

In 2014, evidence for the existence of multipotent mammary stem cells were provided by Rios and colleagues. In this occasion, luminal cells ($CD29^{lo}/CD24^{+}$) were selected based on the expression of the luminal transcription factor *ELF5* and through applying confocal imaging, researchers showed that ELF5/GFP cells were restricted to the inner luminal layer. On the other hand, the expression of the basal cells ($CD29^{high}/CD24^{+}$) was directed through the KRT5 promoter and KRT5/GFP expression was confined to the outer myoepithelial layer. In order to track the *in vivo* fate of both luminal ELF5/GFP and basal KRT5/GFP cells, researchers used a multicolour Confetti cre-recombinase system in combination with novel 3D imaging. It was found that clonal epithelial patches contained both myoepithelial and luminal cells which derived from a common basal precursor (Rios et al. 2014).

The controversy regarding the mouse mammary lineage continued when Wuidart et al., used two novel approaches to study the mouse

mammary gland. Researchers employed multicolour lineage tracing to assess multipotency and another method called lineage tracing at saturation to study stem cell fate and it was shown that unipotent stem cells drive mammary gland development and adult tissue remodelling (Wuidart et al. 2016). These findings refute previous published results where researchers provided evidence for the existence of multipotent mammary stem cells (Rios et al. 2014), and support previous research by Van Keymeulen (Van Keymeulen et al. 2011).

Giraddi and colleagues administered labelled synthetic nucleosides in mouse and observed the rate of uptake of these nucleosides by the different cellular compartments in the mouse mammary gland. It was found that the luminal cell compartment in the non-pregnant adult mouse consisted of three distinct populations, maintained by their own-lineage restricted committed progenitors that contributed differently to the mammary outgrowth. Surprisingly, the $Sca1^{+}/CD49b^{-}$ non-clonogenic luminal cells were responsible for the largest increase in cell number during the mouse oestrus cycle. Around 80% of the $Sca1^{+}/CD49b^{-}$ cells are ER+, however less than 1% of those cells are also positive for the proliferative marker Ki67, therefore cellular proliferation must be driven by the non-clonogenic ER- luminal cells (Giraddi et al. 2015).

A very recent publication investigated how ER+ and ER- luminal populations were formed in the mouse mammary gland. It was found that *SOX9*+ cells gave rise to ER- luminal cells, whereas *PROM-1*+ cells, also known as CD133, originated ER+ luminal cells instead. Both *SOX9*+ and *PROM-1*+ cell compartments were only able to sustain their respective lineages.

In summary, there is substantial controversy surrounding the study of the mammary epithelium and its hierarchy, which highlights the need for further research on this field.

1.5.2 Human mammary stem cells

The clonal origin of normal human breast epithelium was elucidated by the observation that lobular and ductal regions contained identical chromosomal alterations (Tsai et al. 1996). In 1998, Stingl and

colleagues identified three distinct progenitor cell populations based on their levels of expression of MUC-1 (also known as EMA), CALLA and ESA. The MUC-1^{+}/CALLA^{-}/ESA^{+} population gave rise to cells with luminal features when seeded at low clonal density in two-dimensional and three-dimensional cultures, whereas MUC-$1^{-to\pm}$/CALL$A^{\pm to+}$/ESA^{+} (± = intermediate expression) generated mixed colonies containing cells with luminal and myoepithelial characteristics (Stingl et al. 1998). The third type of progenitor cells, which expressed MUC-1^{-}/CALLA^{+}/ESA^{-} were able to generate cells with myoepithelial features only. A later publication from the same first author further confirmed the existence of bipotent human mammary epithelial progenitor cells. Cells positive for epithelial cell adhesion molecule (EPCAM+), isolated from women undergoing reduction mammoplasties, where single-cell FACS sorted into 96 well plates and the colonies generated were stained for luminal and myoepithelial cell markers. The appearance of mixed colonies (containing both luminal and myoepithelial markers) clearly established the existence of bipotent human mammary epithelial progenitor cells. Interestingly, these mixed colonies consisted of a core of cells expressing the luminal markers KRT19, EPCAM and variable levels of MUC-1, which were surrounded by cells expressing the myoepithelial marker KRT14. In addition, researchers also found luminal- and myoepithelial-restricted colonies (Stingl et al. 2001).

Different publications have reported that breast stem cells do not express ER in human breast tissue (Lim et al. 2009) (Clayton et al. 2004). Nevertheless, Clarke and colleagues argued that cells with Hoechst dye-effluxing "side population" (SP) properties, characteristic of mammary stem cells, showed higher ERα expression than non-SP cells (Clarke et al. 2005).

Eirew et al., developed a method for transplanting human mammary stem cells along with irradiated fibroblasts within collagen under the kidney capsule of immunocompromised mice. A population characterised for the expression of CD49f^{high}/ESA^{low} was found to regenerate duct-like structures which also expressed established luminal and myoepithelial markers (Eirew et al. 2008). Moreover, Lim et al., also

identified a population of human breast cells defined by the expression of $CD49f^{high}/ESA^{-}$ with clonogenic activity *in vitro* (Lim et al. 2009).

Villadsen and colleagues demonstrated the existence of a stem cell zone identified in the mammary ducts of human breasts, characterised by the accumulation of $KRT19^{+}/KRT14^{+}$ cells. This $KRT19^{+}/KRT14^{+}$ population coincided with the $CD49f^{hi}/ESA^{+}$ phenotype and also with the expression of the embryonic stem cell marker stage-specific embryonal antigen-4 (SSEA-4) (Villadsen et al. 2007).

More recently, single-cell studies enabled researchers to further characterise normal mammary epithelial cells. Three different populations were identified using reduction mammoplasty samples: a basal/stem population defined as $Lin^{-}/CD49f^{hi}/EPCAM^{lo}/cKit^{-}$, a luminal population marked as $Lin^{-}/CD49f^{lo}/EPCAM^{hi}/cKit^{-}$ and a luminal progenitor cell compartment defined as $Lin^{-}/CD49f^{med}/EPCAM^{med}/cKit^{+}$ (Lawson et al. 2015).

Another marker that identifies cells with stem cell properties is the activity of the aldehyde dehydrogenase 1 (ALDH1). Ginestier et al., humanised the mouse MFP by pre-injecting immortalised human fibroblasts and observed that ALDH positive (ALDH pos) human mammary epithelial cells regenerated the MFP and formed mammospheres with greater efficiency than ALDH negative cells. ALDH1+ cells were able to give rise to both myoepithelial ($CD10^{+}/ESA^{-}$) and luminal epithelial cells ($CD10^{-}/ESA^{+}$) (Ginestier et al. 2007).

The mammosphere assay is an *in vitro* method used to grow undifferentiated stem cells in suspension, and was developed by Dontu et al., (Dontu et al. 2003) based on the neurosphere assay (Reynolds et al. 1996). Mammosphere-derived cells were tested for their differentiation capabilities by growing them at clonogenic densities in differentiating conditions. Three different types of colonies were generated: colonies containing ductal epithelial cells only, colonies containing myoepithelial cells only and mixed colonies containing both lineages. However, only a small number of the mammosphere-derived cells retained the capacity to recreate the entire ductal-acinar architecture of the mammary tree. Mammosphere formation efficiency in serially passaged cells (secondary mammospheres) assesses self-renewal properties of CSCs and

researchers observed that secondary mammospheres consisted almost entirely of multipotent cells. In addition, transcriptional profile studies of the mammosphere-derived cells revealed overlapping genetic programs with other stem and progenitor cells. These results supported the use of the mammosphere assay for stem cell enrichment in normal breast tissue (Dontu et al. 2003).

1.6 Cancer Stem Cells

CSC are cancer cells with stem cell properties of self-renewal, however this designation does not imply that CSCs possess the multipotent capabilities demonstrated by normal adult stem cells. The CSC concept has become very popular throughout the twenty-first century as reflected in a steady increase of the number of publications per year (there were only a few publications in the early 1990s, however in 2013 the number of publications raise to more than 6,000 per year). The high number of publications on CSCs reveals that the concept is now extensively supported and generally accepted by the scientific community (Laplane 2016).

For a cell to be defined as a CSC there is a need to demonstrate its ability to form a heterogeneous tumour containing both CSCs and differentiated bulk cells. Serial transplantation of defined cell populations in immune-compromised mice has provided a benchmark for this definition. Such cells that have this capacity have also been referred to in the literature as tumour initiating cells. CSC can divide symmetrically, producing two identical daughter cells leading to a CSC expansion, or conversely produce two differentiated cells and be lost from the stem cell pool. They can also undergo asymmetrical division giving rise to one CSC and one progenitor cell. The resemblance between normal stem cells and CSCs implies that CSCs might arise from their normal counterpart by acquiring transforming mutations, however this might vary depending on the type of cancer. It has also been reported that leukaemia stem cells might come from more committed progenitor cells that reacquire the

stem cell properties by activating a self-renewal program (Krivtsov et al. 2006).

By definition, CSCs are a subpopulation of cells that can successfully develop tumour formation and initiate cellular growth at secondary sites (Reya et al. 2001, Al-Hajj et al. 2003). *In vivo* transplantation assays are the gold standard technique to identify CSCs. CSCs are generally xenografted in the orthotopic site and assessed for their ability to produce tumours. Often, a histologically stained section is used to confirm the xenograft morphology phenotypically resembles the patient tumour the cells were derived from. After the defined population of cells has been implanted and a tumour is formed, the same population can be isolated and transplanted into a second mouse to further demonstrate long-term self-renewal properties. However there are obvious limitations in these assays. The integrity of the cells and their intrinsic pathways can be altered after the cells have being disrupted from the original tumour. Differences between species can also affect the interaction between growth factors and their ligands. Furthermore, some studies have suggested that this assay selects for the most aggressive clones, as the majority of the cells would not survive in a foreign environment (Clarke et al. 2006). Several researchers have argued that CSCs might not be such a rare population; for instance, modifications of the transplantation assay conditions, such as the use of more immunocompromised mouse models, can considerably increase the percentage of tumour-initiating cells in melanomas to the extent that even unsorted single cells can form tumours in 27% of the cases (Quintana et al. 2008). This was then disputed by Ishizawak et al., where they demonstrated CSCs are a rare population in other solid tumours (Ishizawa et al. 2010). It could also be possible that the number of CSCs within a tumour is being underestimated and that the capability of a cell to be able to proliferate and develop tumours is context-dependent; forming tumours only under certain environmental conditions. It is also very important to underline that not all tumourigenic cells contribute to tumour growth. Some CSCs might belong to a more quiescent clone with competitive disadvantages that conversely, upon transplantation might begin to play an important role in tumour promotion. The CSC model

states that tumours are hierarchically distributed with the CSCs at the top of the apex, however, this tumour hierarchy might significantly vary from one tumour to another as shown in the figure below (Meacham et al. 2013).

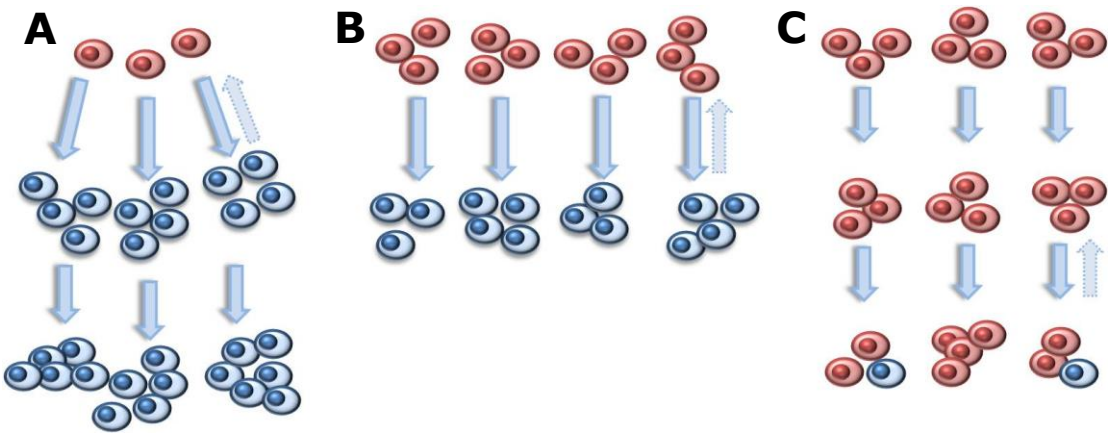


Figure 1. 11 Possible cellular hierarchy present in tumours and its plasticity.
A) Some tumours might consist of a few CSCs (red cells) which give rise to many bulk cells (blue cells). **B)** Other tumours might exhibit a less sharp hierarchical distribution where many CSCs form a small number of bulk cells. **C)** Also it is possible that certain tumours have almost no hierarchy and give rise to very few bulk cells. Broken arrows show that the differentiation of CSCs cells into non-tumourigenic cells can be reversible.

The existence of CSCs was confirmed in 1997 when a small but variable subpopulation of cells that expressed $CD34^{+}CD38^{-}$ cell surface phenotype was found to induce human acute myeloid leukaemia in immunodeficient mice (Lapidot et al. 1994, Bonnet et al. 1997). Remarkably, these cells expressed the same cell surface marker as the normal hematopoietic progenitor cell population (Bonnet et al. 1997). Accumulation of the fluorescent vitamin, Riboflavin, in cytoplasmic vesicles was reported as a novel tool for identification and isolation of cells with CSC properties in pancreatic, colorectal, hepatocellular and non-small cell lung cancer carcinomas. Autofluorescent cells were highly metastatic *in vivo* and expressed pluripotency-associated genes such as *KLF4*, *NANOG*, *OCT3/4* and SOX2 (Miranda-Lorenzo et al. 2014). Another

feature used to identify the presence of CSCs is by studying the side population (SP) using flow cytometry. CSCs have been identified by their capacity to efflux the Hoechst dye 33342 as a consequence of expressing ATP-binding cassette transporters. The presence of these membrane proteins have been argued to play a key role in body protection as they are able to rapidly efflux any damaging cellular metabolism product and also to pump out any anti-cancer drugs (Huls et al. 2009).

An *in vitro* assay regularly used to enrich for CSCs is their ability to grow spheres under non-adherent conditions in the presence of serum-free media (Ponti et al. 2005). Ponti and colleagues first developed this technique in breast cancer cells based on the mammosphere assay developed by Dontu et al., in normal breast tissue (Dontu et al. 2003). Under these circumstances the majority of cells will die from apoptosis induced by anchorage-dependent cells after loss of attachment of the extracellular matrix (ECM), an event known as anoikis, whereas the remaining population will be anoikis resistant (AR), which is considered as the stem cell-like population. This assay supports the growth and expansion of cancer stem and progenitor cells that response to the cytokines EGF, bFGF or both. These spheres can then be disaggregated and reseeded again at clonogenic density to propagate secondary spheres and demonstrate self-renewal (Shaw et al. 2012).

Hence, a growing body of literature has examined the use of numerous cell surface markers and assays to isolate CSCs in different types of cancer; however, formal proof of their self-renewal and tumour-initiating properties is required.

1.6.1 Breast cancer stem cells

Identification of CSCs in solid tumours has been more challenging than in blood malignancies, due in part to difficulties in tumour dissociation. However, in 2003 Clarke and Wicha showed that human breast cancers contain a population of cells expressing the cell surface marker $CD44^+/CD24^-$ and as few as 100 cells containing this phenotype successfully formed tumours in mice compared to 20000 $CD44^+/CD24^+$ (Al-Hajj et al. 2003). The ability of CSCs to retain the lipophilic dye

PKH26 was first described by Pece and colleagues in 2010. It was reported that poorly differentiated breast cancers showed a greater content of PKH26 positive cells, when compared to more differentiated tumours (Pece et al. 2010).

Aldehyde dehydrogenase 1 (ALDH1) is an enzyme responsible for the oxidation of intracellular aldehydes and its activity, measured employing a fluorescence-based enzymatic assay combined with Flow cytometry (Aldefluor assay), is another functional marker used in the study of CSCs. In their seminal work, Ginestier and colleagues showed that breast cancer cells with increased ALDH activity have stem/progenitor characteristics. ALDH1 positive cells are able to self-renew and to form tumours with higher frequency than ALDH1 negative cells. Furthermore, ALDH1 expression from human breast carcinomas correlated with poor prognosis (Ginestier et al. 2007). Interestingly, Liu et al., observed two distinct populations of cells, an EMT-like population characterised as ALDH1+ and a more MET-like population expressing $CD44^+/CD24^-$ markers in breast cancers. Results from Affymetrix array Hu133 plus 2.0 revealed downregulation of *Vimentin* and *ZEB1* and upregulation of *Claudin 3* in ALDH+ cells compared to $CD44^+/CD24^-$ cells. These two cell populations appeared to be highly plastic as they could transition between each other endowing these cells with the capabilities for invasion and growth at distant sites. It was also reported the existence of a population of cells, comprising 1.6% of the total human breast cancer cells, which co-expressed both markers, ALDH1 and $CD44^+/CD24^-$ with even greater tumour-initiating capabilities (Liu et al. 2014).

1.6.1.1 Breast Cancer Stem Cells and Anti-estrogen resistance

Breast CSCs are thought to be unresponsive to standard therapy including chemo-, radio- and endocrine therapies. Harrison and colleagues have shown that E_2 treatment expands the CSC population and its activity in breast cancer. Taking into consideration these results and the fact that several researchers have reported that CSCs do not express ER, scientists wondered what the mechanism that mediates the signals

is. It has been suggested that CSC respond to E_2 stimulation through paracrine signalling, mainly mediated by EGF and Notch signalling pathways (O'Brien et al. 2009, Harrison et al. 2013). It seems clear to hypothesise that therefore CSCs do not respond directly to endocrine therapies, as drugs will not be able to bind to ER in these cells. Following endocrine treatments, CSCs might enter in a dormant state and later on they become the drivers to repopulate tumour growth after therapy. Of interest, it has recently been reported that even though ALDH pos cells do not express the classical 66 KD ERα (ERα66), they might express ERα36, an alternative ER that is localised in the plasma membrane and is capable of directly activating mitogenic signaling (Zhang et al. 2012). Interestingly, Zhang and colleagues observed later on that Tam treatment increases ERα36 expression in breast cancer cells. High ERα36 protein levels induced hypersensitivity to E_2 leading to ERK phosphorylation and activation, which could explain why breast cancer cells escape Tam therapy (Zhang et al. 2013).Xiaoxian Li et al., observed the proportion of breast CSCs in patients before and after receiving chemotherapy and reported an increase in the percentage of $CD44^+/CD24^-$ population, as well as an increase in the mammosphere formation efficiency (MFE) of breast cancer cells (Li et al. 2008). Furthermore, several studies have suggested that the proportion of CSC, indicated by ALDH1 activity, might predict resistance to chemotherapy (Gong et al. 2010).

Our group has reported that ALDH pos cells are enriched following anti-estrogen treatment in breast cancer cells. These resistant ALDH pos cells show high JAG1 ligand and NOTCH4 receptor protein levels. It was also found that high ALDH1 predicts anti-estrogen resistance in women treated with Tam and that the gene signature *NOTCH4*/*HES*/*HEY* is correlated with poor prognosis (Simões et al. 2015). Previously, Piva and colleagues studied Tam resistant cells derived from MCF-7 and found that these cells formed more primary and secondary mammospheres than the control cells. In addition, it was found that Tam resistant cells showed high levels of the stem cell marker *SOX2* by qPCR and western blotting and that SOX2 downregulation using small interfering RNA (siRNA) resulted in a significant inhibition of mammosphere formation and a

reduction of the $CD44^{+}/CD24^{-}$ cell content. Using SOX2-overexpressing cells and SOX2 shRNAs researchers reported how Wnt signalling pathway is important for breast CSCs protection against the anti-proliferative effects of Tam (Piva et al. 2014).

1.6.2 Breast Cancer Stem Cells as therapeutic targets

Normally, clinicians assess the size of the tumour mass to evaluate the efficacy of a particular drug. However, this strategy does not take into account the small population of CSCs present in the tumour. Ideally, combination therapy targeting both bulk tumour cells and the CSCs would reduce the probabilities of tumour relapse. CSCs sustain themselves on different signaling pathways such as Wnt, Notch or Hedgehog to survive and proliferate. Therefore, pharmacological inhibitors that target these pathways would be a good therapeutic approach (Figure 1.12) (Brooks et al. 2015).

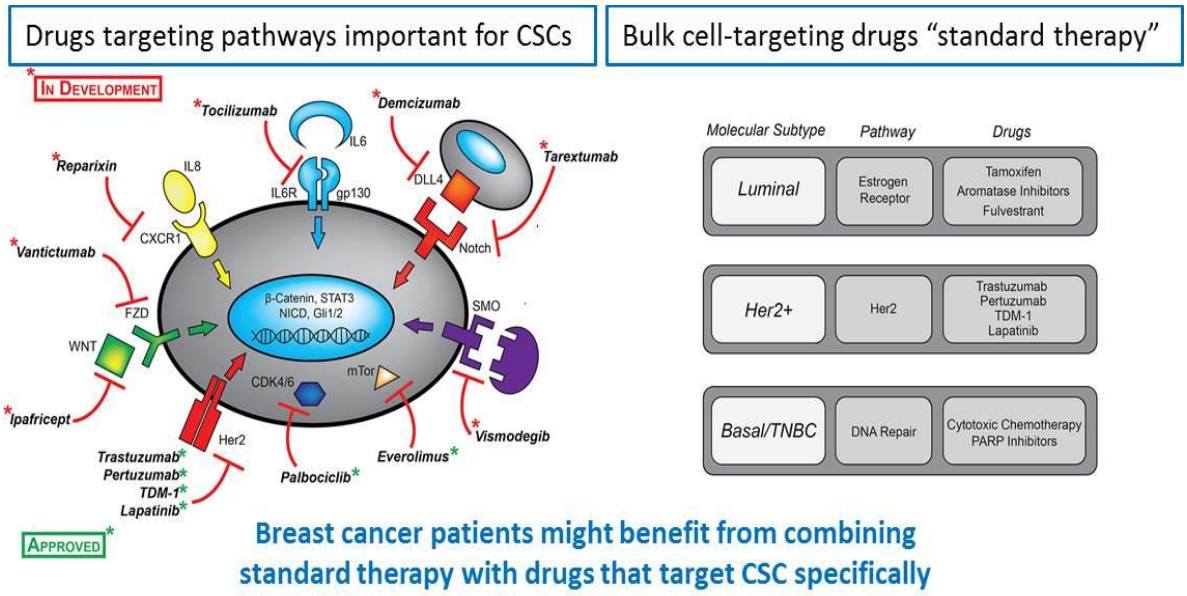


Figure 1. 12 Combination therapy strategies to target both bulk and CSCs in breast cancers.
This illustration describes CSC-targeting drugs either approved or in development and, on the right, FDA approved drugs to treat bulk cells in each of the molecular subtypes existent in breast cancer. Figure adapted from Brooks et al., 2015.

As mentioned in section 1.4.1.1, the CDK4/6 inhibitor Palbociclib in combination with anti-estrogen therapy has successfully been used in the clinic to treat breast cancers due to its ability to inhibit Cyclin D, the downstream target of E_2 mitogenic signaling. However, it has recently been reported that the Cyclin D-CDK4/CDK6 complex is also important in the regulation of CSCs in breast tumours (Figure 1.12) (Lamb et al. 2013). This highlights the importance of combining multiple drugs to target bulk and CSCs in order to improve breast cancer therapeutics.

1.8 ALDH isoforms and cancer

ALDH activity is an important marker that identifies breast CSCs as mentioned previously. There are 19 ALDH isoforms, with different preferred substrate, which are found in all subcellular regions such as cytosol, mitochondria, endoplasmic reticulum or the nucleus (Figure 1.13) (Braun et al. 1987) (Koppaka et al. 2012). Due to their important role in aldehyde metabolism, mutations and alteration of expression of ALDH isoforms are linked to a number of diseases. Moreover, the expression of different ALDH isoforms has been associated with a worse patient outcome. Both ALDH1A1 and ALDH1A3 have been shown to detoxify anti-cancer drugs and diminish drug effectiveness. ALDH1A1 enzyme commonly detoxifies the chemotherapeutic agent cyclophosphamide, which leads to drug resistance due to the upregulation of ALDH1A1 and ALDH1A3 isoforms in the tumour cells (Sreerama et al. 1997). Moreover, when ALDH activity was inhibited using the DEAB inhibitor or retinoic acid (ATRA) supplementation, triple negative breast cancer cells became sensitised to chemotherapy and radiotherapy (Croker et al. 2012).

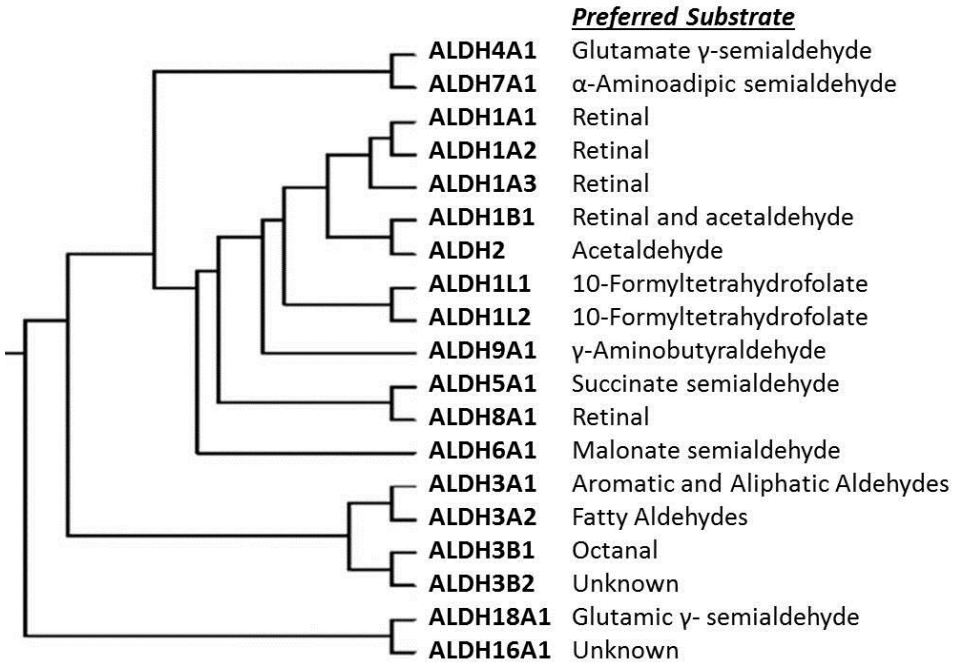


Figure 1. 13 ALDH isoforms and their preferred substrate.
Figure adapted from Koppaka et al., 2012.

As mentioned above, ALDHs have been implicated as contributors to CSC-associated Aldefluor fluorescence. This assay was first developed to isolate hematopoietic stem cells from the human umbilical cord, a tissue with high ALDH1A1 expression, however, while specific ALDH isoforms have some preferred substrate (refer to Figure 1.13), they also show cross-reactivity. Therefore, in all likelihood the Aldefluor assay detects ALDH activity from one or more isoforms (Storms et al. 1999). However, only increased expression of ALDH1A1 and ALDH1A3 isoforms has been reported in ALDH pos breast cancer cells. In 2011, Marcato and colleagues observed that ALDH activity in breast cancer cell lines is mainly due to the ALDH1A3 isoform (Marcato et al. 2011). ALDH1A3 expression levels have been shown to correlate to tumour grade, metastasis and cancer stage. Similarly, a recent meta-analyses from 15 different publications found ALDH1A1 to be a biomarker that predicts tumour progression and poor survival of breast cancer patients (Liu et al. 2014). Qiu et al., studied the members of the ALDH family and found that only ALDH1A3 expression was a predictive marker of poor clinical outcome in breast cancer (Qiu et al. 2014).

In spite of the role of ALDH isoforms in malignant disease and their high activity in CSCs, there have only been developed pharmacological inhibitors for 3 of those isoforms to treat alcohol abuse and cancer (Koppaka et al. 2012). ALDH isoforms show a high degree of similarity amongst them, therefore the majority of the activators or inhibitors designed to target ALDHs are not specific for a particular isoform. For instance, Disulfiram, which has been used in the clinic for decades to treat alcohol abuse, inhibits the breakdown of acetaldehyde by ALDH1A1 and ALDH2. Disulfiram has strong anti-cancer activity *in vitro* and in particular, it has been shown to reduce breast cancer cell growth in MDA-MB-231 cells (Chen et al. 2006). Citral was recently discovered to be an effective inhibitor of ALDH1A3-mediated colony formation and nanoparticle-encapsulated Citral successfully impaired *in vitro* growth of MDA-MB-231 cells overexpressing ALDH1A3 (Thomas et al. 2016). Thanks partially to the discovery of some of the ALDH crystal structures (Moore et al. 1998) (Moretti et al. 2016) different residues present in their respective substrate-binding tunnel have been identified, which will potentially allow the design of more isoform-specific compounds.

Overall, there is a growing body of literature regarding the involvement of ALDH1A1 and ALDH1A3 in cancer stemness, poor prognosis and tumour progression. Therefore, ALDH1A1 and ALDH1A3 inhibitors may be an effective adjuvant therapy in the treatment of breast cancers.

The ALDH superfamily is responsible for the $NAD(P)^+$-dependent oxidation of endogenous and exogenous aldehydes into carboxylic acids. Endogenous aldehydes are formed during the cellular metabolism of aminoacids, lipids, vitamins or steroids. For instance, oxidative stress often leads to lipid peroxidation generating over 200 aldehydes. Exogenous aldehydes on the other hand, are mainly generated as consequence of human activities (i.e. cigarette smoke or motor vehicle exhaust) (Morgan et al. 2015). Aldehydes interact with thiol and amino groups that might lead to cytotoxic or carcinogenic processes, therefore ALDH can be seen as a detoxifying cellular agent. However, there are situations when ALDHs catalyse reactions, which generate chemically reactive products harmful for the organism.

ALDHs also mediate the irreversible oxidation of retinaldehyde, also known as retinal, to retinoic acid (RA), which will bind the nuclear receptors (Vasiliou et al. 2000). RA transactivates various response pathways via the nuclear Retinoic Acid Receptors (RARs) and Peroxisome Proliferator-Activated Receptor (PPARs) receptors by forming heterodimers with the Retinoid X Receptors (RXRs). RA generally signals through RAR/RXR heterodimers and binds to DNA-response elements known as RA response elements (RAREs). When RA binds RAR leads to inhibition of cell proliferation, hence this explains why RA is currently being used as a therapeutic agent against myeloid acute leukaemia, since RA stimulates differentiation of leukemic cells into normal granulocytes (Tallman 1996). In addition, the RA analog Fenretinide has shown to decrease risk of second breast cancers in premenopausal women (Veronesi et al. 2006). However, RA can also bind PPARs to promote survival in cells that show high levels of fatty acid-binding protein 5 (FABP5) (Schug et al. 2007). RA has also been shown to promote cell growth of the basal keratinocytes of the skin and enhance skin tumour formation (Verma et al. 1982). Therefore, RA can exert opposing cellular effects via binding to different nuclear receptors.

Hua et al observed an interaction between E_2 and RA signalling using Chip-seq in the breast cancer cell line, MCF-7. Researchers found that activation of the RA signalling pathway antagonises E_2 signalling in breast cancer and also that RAR binding to chromatin was dependent on ER (Hua et al. 2009). RAR-α protein expression has been associated to Tam resistance in human breast cancers and its expression could be used as a potential predictive marker for adjuvant Tam therapy, however there was no correlation between RAR-α mRNA and protein levels. Researchers also found that Tam-resistant cells were more sensitive to RAR ligands than parental MCF-7 cells and argued that this could be due to a change in the protein interactome around ER and RAR-α during the acquisition of Tam resistance (Johansson et al. 2013). PPAR receptors play a key role in lipid and energy metabolism and recently, Wang and colleagues found that inhibition of PPAR-γ with synthetic antagonists lead to a reduction of the ALDH pos population in HER2 positive cancer cells (Wang et al. 2013).

Altogether, ALDHs drive RA production, which either contributes or prevents tumour formation via its cognate nuclear receptors. More studies are needed in order to elucidate the exact mechanism of action and how this pathway can be therapeutically targeted to tackle breast cancer therapeutic resistance.

1.9 Cancer cell heterogeneity

Tumour heterogeneity has been reported to exist not only between different cancer patients (inter-tumour heterogeneity) but also within a single tumour (intra-tumour heterogeneity). Tumours contain cellular niches enriched for distinct phenotypic properties such as quiescence, self-renewal or adaptation to hypoxia, which increases the complexity of cancer cellular diversity (Patel et al. 2014). Clonal heterogeneity has been found to correlate with increased risk of progression in esophageal cancer patients (Maley et al. 2006). According to the plasticity model, differentiated cells can reacquire stem cell characteristics via cell reprogramming (Takahashi, 2006). This reversible transition between the CSC and non-CSC phenotype can introduce another level of complexity regarding cellular heterogeneity.

Genomic profiling of CSCs is revealing a landscape of altered cellular signals and transduction cascades. Different signalling pathways such as Wnt, Notch and Hedgehog, have been shown to be important in CSC function and maintenance (Takebe et al. 2015). Nevertheless, the vast majority of the transcriptomics and proteomics research carried out in CSCs has been done using population-averaged assays, which reflect the dominant biological mechanism of the population, however this approach fails at capturing the behaviour of individual CSCs. Single-cell RNA sequencing is a powerful tool to interrogate mRNA heterogeneity. By stochastically attaching labels to complementary DNA (cDNA) molecules, it is possible to determine the quantity of cDNA molecules present in single cells (Fu et al. 2011).

In recent years several new platforms have been developed for measuring different features at the single cell level. For instance, the

CyTOF (Fluidigm) is a novel mass cytometry technology for single cell proteomics, which allows the analysis of over 100 different proteins per single cell. In their research, Giesen and colleagues employed immunohistochemistry, high-resolution laser ablation and CyTOF mass cytometry to show cell-cell interactions and heterogeneity within the different human breast cancer samples studied (Giesen et al. 2014). The C1 system (Fluidigm), used in this work, is an automated device for single-cell genomics, which enables researchers to capture single cells and perform targeted pre-amplification of the cDNA. Subsequently, these samples can be further analysed by doing targeted expression using the BioMark HD system or RNA sequencing (RNAseq) employing the Illumina sequencing system. Xin et al., reported the identification of cell-type-specific transcription factors and pathways in single mouse pancreatic islet cells through using the C1 system platform (Xin et al. 2016). Likewise, by combining single cell-FACS sorting and the BioMark HD technology, scientists revealed a subset of cells, present in low-burden metastatic tissues, with distinct stem-cell gene expression signature and enhanced tumour-initiating capabilities, compared to metastatic cells from high-burden tissues. These results supported the idea that metastasis is initiated by CSCs (Lawson et al. 2015).

Over the past few years, there has been considerable progress regarding single cancer cell research. Although many technical challenges are still encountered, it is anticipated that different layers of "omics" single cell data, in combination with information regarding cell-to-cell interaction and communication will help us understand and treat cancer patients better.

1.5 Hypothesis

My hypothesis is that anti-estrogen resistant breast CSCs are heterogeneous in their cellular phenotype and the way they respond to signalling pathways.

1.6 Aims

Although considerable progress has been made in breast cancer research over the past few decades, more than 11,000 women died from this disease in 2014 in the UK. Tamoxifen (Tam) and Fulvestrant (Fulv) are the standard therapy for ER+ breast tumours, however these agents do not target CSCs. The following aims were set to address my hypothesis:

1) The first aim of this PhD project is to investigate which particular CSC population is enriched following anti-estrogen treatment with Tam and Fulv in ER+ breast cancer cells.

2) The second aim is to characterise the enriched CSC population in ER+ breast cancer cells testing their stem cell capabilities *in vivo* and *in vitro* and also interrogating their gene expression profile as bulk population aiming at identified targetable biomarkers.

3) The third and final aim of the present work is to investigate the enriched CSC population at the single cell level to identify any existing cellular diversity that otherwise would be hidden away analysing bulk cells. For this purpose, changes in mRNA transcripts upon anti-estrogen treatment were measured in single breast CSCs.

The present work should help understand the mechanisms underlying anti-estrogen resistance in ER+ breast cancer and also provide a better insight into the gene expression profile of breast CSCs

and how to target them. The ultimate goal of this project is to provide the scientific basis to help improve ER+ breast cancer patient's outcome.

2. Chapter 2: Materials and methods

2.1 Materials

2.1.1 Reagents

Table 2. 1 List of reagents

Reagent	Suppliers	Catalogue Number
7-Aminoactinomycin D (7-AAD)	BD bioscience	559925
17-ß estradiol pellets	Innovative Research of America	NE-121
17-β estradiol for drinking water	Sigma	E8875
2 Mercaptopethanol (99%)	Sigma	M7522
Aldefluor kit reagent	STEMCELL	01700
B27	Gibco	7504044
BioMark HD Chips 10 - 96.96 Gene Expression	Fluidigm	BMKM10-96.96
BCA protein assay kit	Thermo Scientific	23225
C1 Single-Cell Auto Prep Array for Pre-Amp (10-17um)	Fluidigm	100-5480
C1 Single-Cell Auto Prep Reagent Preamplification Module 1 and Module 2	Fluidigm	100-5319
cDNA preparation with Reverse Transcription MasterMix	Fluidigm	100-6472
DAPI	Sigma	D9542
Dimethyl Sulfoxide (DMSO)	Sigma	D5879
DRAQ5	Biostatus	DR05500
Dulbecco's Modified Eagle's Medium (DMEM)	Sigma	D5546
Elution Buffer	Applied Biosystems	4305893
Endofree Plasmid Maxi Kit	Qiagen	12362
Ethanol absolute	VWR chemicals	UN1170
Foetal Bovine Serum	Gibco	10270-106
ICI 182,780 / Fulvestrant	TOCRIS bioscience	1047
Hanks Buffered Saline Solution (HBSS)	Sigma	RNBC3010
Hoechst 33342	Sigma	14533
Human Epidermal Growth Factor (hEGF)	MACS Miltenyi Biotec	130-097-751
L-Glutamine	Gibco	25030-024
Library efficiency DH5α Competent Cells	Thermo Scientific	18263-012

Matrigel	Corning	356230
Lymphoprep	Axis Shield	1114544
MicroAmp Optical Adhesive Film	Applied Biosystems	201310 207
Needle (25 G)	Terumo	NN-2525R
pMD.2 plasmid	Addgene	12259
pCMVR8.74 plasmid	Addgene	22036
Poly-Hema	Sigma	P3932
Puromycin	Sigma	P8833
Reverse Transcription Kit	Fluidigm	100-6297
RNaseZap	Sigma	R2020
RNeasy Plus micro kit	Qiagen	74034
RNeasy Plus mini kit	Qiagen	74134
Single Cell-to-Ct kit	Ambion	4458236
Syringe (5 ml)	Terumo	SS05S1
Sulforhodamine B	Sigma	S1402
TaqMan® Fast Advanced Master Mix For BioMark	Thermo Scientific	4444556
TaqMan Transcription Reagents	Applied Biosystems	N808-0234
TaqMan Universal PCR master mix	Roche	4304437
TRIPZ inducible lentiviral shRNA	Dharmacon	RHS4740
Tris-base (Trizma)	Sigma	T1503
Trypan Blue	Sigma	T8154
Trypsin	Sigma	SLBH4853
(Z)-4-Hydroxytamoxifen	Sigma	H7904
6-well culture plate	Corning	3516
75 cm^2 flask	Corning	430641
225 cm^2 flask	Corning	3001
150x25mm tissue culture dish	Falcon	353025
30 ml Universal Tube	SLS Select	SLS7500
384-well reaction plate with no barcode	Applied Biosystems	4343370

2.1.2 Cell culture media

Table 2. 2 Formulation of the media used for cell culture

	Medium
Monolayer MCF-7, T47D and HEK-293T cell culture	Complete Media: Dulbecco's Modified Eagle´s Medium (DMEM) with Phenol red, 10% Fetal Bovine Serum (FBS), 2 mM Glutamine
Mammosphere culture	Phenol-red free DMEM F12, B27 supplement, rEGF

2.1.3 Antibodies and FACS-staining reagents

Table 2. 3 Antibodies and other staining reagents for FACS

Antibody/ Reagent	Fluoro-chrome	Clone	Species	Company	Catalogue Number	Concentration
CD24	PE-Cy7	ML5	Mouse	Biolegend	311120	200 ng/µl
CD44	APC	G44-26	Mouse	BD	559942	3 ng/µl
DAPI	DAPI	n/a	n/a	Sigma Aldrich	D9542	0.1 mg/ml
Aldefluor	BAA	n/a	n/a	STEMCELL	01700	5 µl BAAA per 1 ml buffer

Table 2. 4 Isotype control antibodies

		Species	Company	Catalogue Number	Concentration
PE-Cy7	Isotype IgG1	Mouse IgG1	BD	348788	200 ng/µl
APC	Isotype IgG1	Mouse IgG1	BD	340442	3 ng/µl

2.1.4 TaqMan Quantitative Polymerase Chain Reaction (qPCR) primers

All primers used for qPCR were best coverage TaqMan assays purchased from Thermofisher and are listed in Table 2.5. Best coverage primers were designed based on the following rules: primers do not detect gene products with similar sequence neither off-target sequences, primers are designed across exon-exon junction, they generate a short amplicon for more efficient PCR reactions, primers do not map to multiple genes and they do not target the 5´untranslated region. TaqMan probes contained the fluorophore 6-carboxyfluorescein (FAM), a minor groove binder (MGB) and a 3´nonfluorescent quencher (NFQ). All probes span exons whenever possible.

Table 2. 5 Pre-designed TaqMan qPCR primer/probe from ThermoFisher.

Gene Name	Primer ID	Amplicon Length
18S	Hs03928990_g1	69
ABCG2	Hs01053790_m1	83
AKT1	Hs00178289_m1	66
ALDH1A1	Hs00946916_m1	61
ALDH1A3	Hs00167476_m1	60
ANXA3	Hs00974395_m1	209
AR	Hs00171172_m1	72
BRCA1	Hs01556191_m1	96
Ca12	Hs01080909_m1	70
CCL5	Hs00982282_m1	70
CCND1	Hs00765553_m1	57
CD24	Hs03044178_g1	146
CD44	Hs01075856_m1	78
CD66c/CEACAM6	Hs03645554_m1	67
CDH1	Hs01023894_m1	61
CDH2	Hs00983056_m1	66
CDH3	Hs00354998_m1	132
CTNNB1	Hs00355049_m1	67
CXCR1	Hs01921207_s1	169
CXCR4	Hs00607978_s1	153
CYR61	Hs00155479_m1	88
DLL1	Hs00194509_m1	74
DLL4	Hs00184092_m1	78
EGFR	Hs01076078_m1	60
ENAH	Hs00403109_m1	89
EPCAM	Hs00901885_m1	95
EPGN	Hs04334113_m1	82
ESR1	Hs00174860_m1	62
EZH2	Hs00544833_m1	86
FBXW7	Hs00217794_m1	76
GAPDH	Hs02758991_g1	93
GATA3	Hs00231122_m1	80
GJA1	Hs00748445_s1	142
GPRC5A	Hs01551896_m1	66
GSK3B	Hs01047719_m1	65
HER2	Hs01001580_m1	60
HES1	Hs00172878_m1	78
HEY1	Hs01114113_m1	82
HEY2	Hs00232622_m1	111
HPRT1	Hs02800695_m1	82
ID1	Hs03676575_s1	100
IGFBP5	Hs00181213_m1	85
IL1R1	Hs00991002_m1	152
IL6	Hs00985639_m1	66
IL6R	Hs01075666_m1	69
IL6ST	Hs00174360_m1	72
IL8	Hs99999034_m1	81
ITGA6	Hs01041011_m1	64

ITGB3	Hs01001469_m1	59
JAG1	Hs01070032_m1	72
JAG2	Hs00171432_m1	110
KRT18	Hs02827483_g1	64
KRT19	Hs00761767_s1	116
KRT5	Hs00361185_m1	133
KRT8	Hs01595539_g1	164
LIN28A	Hs00702808_s1	143
LOX	Hs00942480_m1	81
MET	Hs01565584_m1	73
MKI67	Hs01032443_m1	66
MKP1/DUSP1	Hs00610256_g1	63
MTOR	Hs00234508_m1	103
MUC1	Hs00159357_m1	84
NANOG	Hs04260366_g1	99
NES	Hs04187831_g1	58
NFKB1	Hs00765730_m1	66
NOTCH1	Hs01062014_m1	80
NOTCH2	Hs01050702_m1	60
NOTCH3	Hs01128541_m1	81
NOTCH4	Hs00965892_m1	96
NUMB	Hs01105433_m1	71
PCNA	Hs00427214_g1	138
PGR	Hs01556702_m1	77
PIK3CA	Hs00907957_m1	83
PLA2G10	Hs00358567_m1	67
PLA2G2A	Hs00179898_m1	100
POU5F1	Hs00999632_g1	77
PTGIS	Hs00919949_m1	56
RAB7A	Hs01115139_m1	93
SERPINE1	Hs01126606_m1	79
SERPING1	Hs00163781_m1	70
SLPI	Hs00268204_m1	69
SNAI1	Hs00195591_m1	66
SNAI2	Hs00950344_m1	86
SOCS3	Hs02330328_s1	89
SOX2	Hs01053049_s1	91
SPARC	Hs00234160_m1	76
TAZ	Hs00794094_m1	87
TGFB1	Hs00998133_m1	57
TGFBR1	Hs00610320_m1	73
TMPRSS4	Hs00212669_m1	102
TP53	Hs01034249_m1	108
TP63	Hs00978343_m1	85
TSPAN6	Hs01073456_m1	138
TWIST1	Hs01675818_s1	85
UXT	Hs00188238_m1	89
YAP1	Hs00902712_g1	62
ZEB1	Hs00232783_m1	63

2.2 Methods

2.2.1 Cell culture

2.2.1.1 Adherent culture

MCF-7 and T47D cells were purchased from the American Type Culture Collection (ATCC). MCF-7 and T47D cells are human ER+/PR+/HER2- breast cancer cell lines derived from metastatic pleural effusions. These two cell lines are molecularly classified as luminal A. All cell lines were maintained at 37 ºC in 5% CO_2. Cells were passaged periodically when 80% confluency was reached under sterile conditions, using a class II safety cabinet (Labotal). Cell lines were authenticated using a Short Tandem Repeat- Polymerase Chain Reaction (STR-PCR) with fragment analysis on the 3130xl genetic analyser and tested regularly for mycoplasma contamination by PCR through the CRUK Manchester Institute Molecular Biology Core Facility.

2.2.1.2 Mammosphere culture

Unsorted cells: MCF-7 cells were detached using trypsin, which was neutralised using complete media, followed by cell centrifugation for 4 minutes (min) at 400 g. After supernatant was discarded, cells were syringed with a 25 G needle to ensure a single cell suspension was formed then counted and resuspended with the adequate amount of mammosphere media. Cells were then plated at a cell density of 3,000 cells per well in polyHema-coated 6 well plates and incubated at 37°C and 5% CO_2 for 7 days. Mammosphere forming efficiency (MFE) was assessed after 7 days of culture by counting the number of mammospheres greater than 50 µM of diameter, at 40x magnification using the Olympus CK2-TR microscope.

Sorted cells: Cells were sorted into 1.5 ml eppendorfs containing 2% FBS in Hank's Buffered Saline Solution (HBSS). Sorted cells were spun at 400 g for 4 min and counted using trypan blue. 5,000 sorted patient cells or 3,000 sorted MCF-7 cells were seeded per well in poly-hema coated plates containing mammosphere media. Plates were

incubated for 7 days and mammospheres bigger than 50 μM diameter were counted at 40x magnification.
Mammosphere forming efficiency percentage was calculated as follows:
(Number of mammospheres per well/ number of cells seeded per well) x 100

Poly-hema coated plates were made by dissolving 12 g of poly-Hema in 1 litre of 95% ethanol and stirring on a heated plate for 24 hours (h). One ml of the poly-Hema solution was added into each well and 6 well plates were then incubated in the oven at 40°C for 48 h until plates were dried. Remaining poly-hema solution was stored at room temperature.

2.2.1.3 In vitro anti-estrogen treatments

4-OH-Tamoxifen (Tam) was dissolved in 100% ethanol at a concentration of 10^{-2} M and stored in amber eppendorfs at -20 ºC as a stock solution. Fulvestrant (Fulv) was also dissolved in ethanol at the same concentration and kept at -20ºC until its use (Table 2.6).

6 days *in vitro* treatment: MCF-7 cells were seeded in 225 cm^2 tissue culture dishes and cells were exposed to either Tam at a concentration of $1x10^{-6}$ μM or Fulv at $1x10^{-7}$ μM for 6 days. Note that cells were split after 3 days and reseeded at the same cell density for another 3 days (refer to Figure 3.2 for schematic representation of 6 days *in vitro* treatment). For each set of experiments, an ethanol control was set up.

Table 2. 6 Anti-estrogen drugs for *in vitro* treatment

Drug	Solvent	Final Concentration (μM)	Effect
Tam	ETOH	1	ER antagonist
Fulv	ETOH	0.1	ER full antagonist

2.2.1.4 Sulforhodamine B colorimetric assay

The Sulforhodamine B colorimetric assay (SRB) was used to assess the effects of the ER-inhibitors, Tam and Fulv, on cell growth by measuring the cellular protein content. 5,000 cells were seeded per well with the adequate drug concentration in a 96 well plate for 24, 48 and 72 h. For fixation, cells were incubated with 50 µl/well of 50% trichloracetic acid for 1 h at 4 ºC. Subsequently, cells were washed with water five times and plates were left to dry out overnight. Fixed cells were then stained with 100 µl/well of 0.4% SRB dissolved in 1% acetic acid for 30 min at room temperature. Plates were then washed 3 times with 1% acetic acid to remove residual SRB and left to air dry. SRB was then solubilised with the addition of 100 µl/well of 10 mM Tris-base (pH 10.5), for 20 min at room temperature. Absorbance was measured at 550 nanometers (nm) using a UVmax plate reader and data was represented in a growth kinetics curve.

2.2.1.5 Stable Gene Knockdown using short hairpin RNA

2.2.1.5.1 Transformation of plasmid into competent E-coli cells

Library efficiency DH5α competent cells were thawed and 100 µl of the competent cells were mixed with 8 ng of the plasmid of interest gently tapping the tube to mix, followed by 30 min incubation on ice. Cells were then heat-shocked at 42 °C for 45 seconds followed by 2 min incubation on ice. 900 µl of Lysogeny Broth (LB) medium was added onto the cells and cells were shook at 225 rpm for 37 °C for 1 h. 100 µl of the solution was spread onto LB plates with 100 µg/ml of Ampicillin and plates were incubated at 37 °C overnight. One of the colonies was picked using a P2 tip, added onto 5 ml of LB media containing Ampicillin (100 µg/ml) and incubated for 8 h at 37 °C and 225 rpm. Media containing bacteria was then transferred into a conical flask with 400 ml of LB media with the appropriate antibiotic and incubated overnight at 37 °C shaking at 225 rpm. Plasmid DNA was isolated using the Endofree Plasmid Maxi Kit (Qiagen).

2.2.1.5.2 Lentiviral production and infection

In order to stably knock down ALDH1A3, second generation TRIPZ inducible lentiviral short hairpin RNA (shRNA) was used for the transduction procedure. Three different shRNAs were used named as virus 1, virus 4 and virus 5 (RHS4740-EG220, Dharmacon; V3THS_378581: TGTTGATAAATATCTTGGT, V3THS_378584: TGGTTGAAGAACACTCCCT, V3THS_378585: ATCTTGGTGAACTTGACCT). Refer to appendix 8.2 to visualise the three shRNA sequences mapped to the ALDH1A3 mRNA. This vector is engineered to be Tet-Off so that the shRNA expression is induced by doxycycline (DOX) and can be visualised through a turbo red fluorescent protein (RFP) fluorescent reporter. Moreover, the vector carries genes, which provide antibiotics resistance (Puromycin for mammalian selection and Ampicillin for bacterial selection) (Figure 2.1 A and Appendix Figure 8.3). Lentiviruses are a subclass of Retrovirus, which are able to reverse transcribe RNA into DNA and integrate it into the genome of dividing and also non-dividing cells. For safety reasons lentiviral vectors don't carry the required genes for their replication. TRIPZ are second generation lentiviruses, which means that 3 different plasmids are required for virus production; a packaging plasmid, a transfer plasmid and an envelope plasmid (Figure 2.1 B). HEK-293T, a variant of the human embryonic kidney 293 cell line, was the cell line of choice for the co-transfection of the plasmids and the generation of the lentiviral particles.

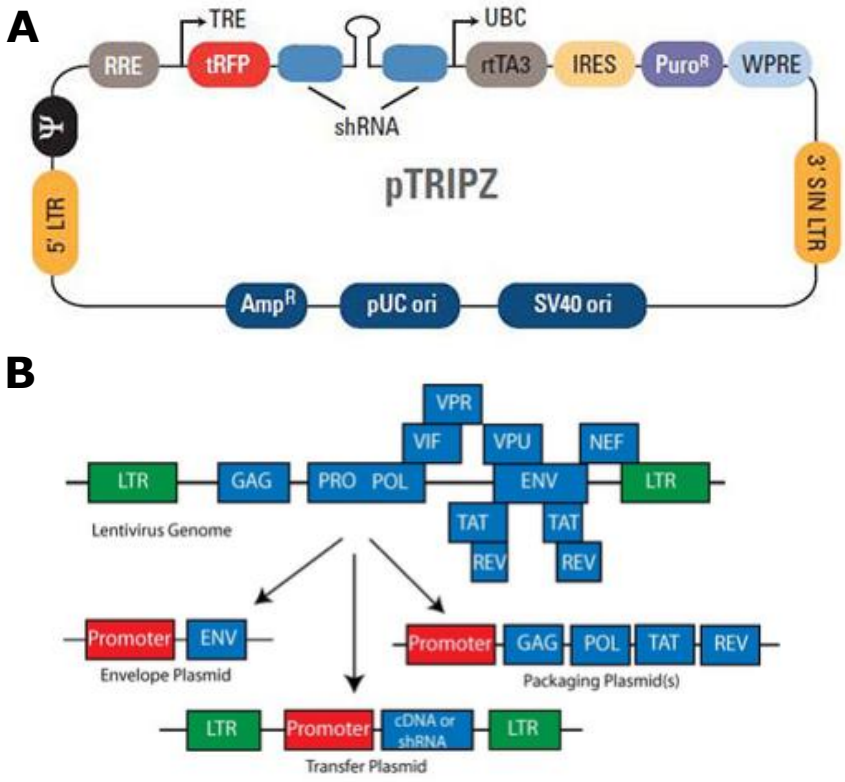


Figure 2. 1 Second generation lentiviral TRIPZ plasmid
A) Representation of the TRIPZ inducible lentiviral vector. *5´LTR:* 5´long terminal repeat; *Ψ*: psi packaging sequence; *RRE*: Rev response element; *TRE*: Tetracycline-inducible promoter; *tRFP*: Turbo-RFP reporter; *shRNA*: microRNA-adapted shRNA (based on miR-30) for gene knockdown; *UBC*: Human ubiquitin C promoter for constitutive expression of rtTA3 and Puromycin resistance genes; *rtTA3*: reverse tetracycline-transactivator 3 for tetracycline-dependent induction of the TRE promoter. *IRES*: Internal ribosomal entry site; *Puro^R*: Puromycin resistance; *WPRE*: woodchuck hepatitis posttranscriptional regulatory element. *3´SIN LTR:* 3´self-inactivating long terminal repeat for increased lentivirus safety. **B)** Schematic representation of the three different second generation plasmids (envelope, packaging and transfer plasmid), required for the generation of lentiviral particles. Figures adapted from dharmacon.gelifesciences.com and addgene.org, respectively.

Using the calcium phosphate precipitation method, DNA from all three plasmids was transfected into 293T cells. 10 µg of viral construct, 3.5 µg of pDM2.G, which encodes the envelope vesicular stomatitis virus glycoprotein (VSV-G), and 6.5 µg of pCMVΔ8.74, which expresses HIV-1 Gag, Pol and accessory proteins except Vpu, were combined with 87 µl of 2M $CaCl_2$ and water to a final volume of 700 µl (solution A). 700 µl of

HBS (Hepes-buffered saline) at pH 7 (solution B) were added drop wisely onto solution A while vortexing. Solution A+B was incubated for 12 min at room temperature. Subsequently, solution was gently vortexed and 1,400 µl were added dropwise to each culture plate medium followed by overnight incubation at 37 ºC. Fresh media was added after 12 h and virus was collected after 24, 48 and 72 h post-transfection. Media containing lentiviral particles was spun down for 5 min at 400 g to get rid of cell debris, filtered through a 0.45 µm filter and stored at 4 ºC. Lentiviral particles were concentrated by ultracentrifugation for 2 h at 20,000 rpm at 4 ºC while applying vacuum (Sorvall Ultra Pro80, AH-629 rotor). Pellet was then resuspended in 1/1000 of the original volume in PBS and stored at -80 °C. In order to calculate the right amount of viral particles necessary to infect MCF-7 cells, titration of the virus was performed by performing serial dilutions using polybrene at a concentration of 8 µg/ml. This process allowed us to achieve an optimal Multiplicity of Infection (MOI); the number of virions per cell in a transduction. DOX was added to the culture at a concentration of 1 µg/µl for at least 48 h to induce expression of Turbo-RFP. Percentage of red cells was assessed by Flow cytometry using the PE channel. Puromycin was added to the cell culture at a concentration of 1 µg/ml to select the transduced cells.

2.2.2 Isolation of epithelial cells from metastatic patient-derived samples

Metastatic fluid samples were collected at the Christie NHS Foundation Trust (UK). All patients provided written informed consent before samples were taken. Fluid samples (local reference number: 11-DOG02-226) from the Christie NHS Foundation Trust were collected by the Biobank (local reference number: 12-ROCL-01) following the local research ethics committee guidelines (05/Q1402/25).

Ascites or pleural effusions were spun for 10 min at 1000 g and pellets were resuspended in Phosphate Buffered Saline (PBS). White

blood cells, also known as leukocytes, were excluded from the epithelial tumour cells using the density gradient Lymphoprep (Stemcell Technologies) following manufacturer's protocol. Subsequently, epithelial cells were frozen down in liquid nitrogen until needed. Of note, due to the lack of metastatic samples provided by the biobank at certain stages of this project, patient samples were also collected from the University of Michigan Comprehensive Cancer Center (USA).

The patient samples from Manchester: 68, 69, 90, 90A, 91, 91A, 94 are ascitic fluids, whereas samples 71 and 89 are pleural effusions. The patient information from samples collected in Michigan does not specify the nature of the fluid.

Clinical information about the patients is shown in Table 2.7 and Table 2.8. Quick Score (QS), also known as Allred score, is a clinical instrument based on the intensity and percentage of cells that express ER through immunohistochemistry (Leake et al. 2000). Summing up the scores for proportion and intensity gives us a QS value between 0 and 8. The cut off for positivity using Allred score is ≥3.82. QS scoring enables clinicians to stratify patients into cancers that are likely to respond to anti-estrogen treatments. QS values for some of the patient samples used in this work are shown in Table 2.8.

Score for proportion is as follows:
0 = no staining
1 = < 1% nuclei staining
2 = 1–10% nuclei staining
3 = 11–33% nuclei staining
4 = 34–66% nuclei staining
5 = 67–100% nuclei staining

Score for intensity is as follows:
0 = no staining
1 = weak staining
2 = moderate staining
3 = strong staining

For some patient samples there is only information regarding the percentage of ER+/PR+ cancer cells or their positivity (+) for the hormonal markers. All patient samples are ER+/PR+/HER2- except patient sample 71 which is ER+/PR+/HER2+. Hormone receptor assessment using immunohistochemistry was performed in primary tumours not in the metastatic lesions. Previous research has demonstrated that the receptor profiles between the primary and metastatic tumours change in around 20% of the cases (Raica et al. 2014), therefore the data shown in Table 2.7 and 2.8 might not accurately reflect hormone receptors expression in the metastatic cells isolated from the pleural or ascytic fluids.

Table 2. 7 Michigan patient sample clinical information.
ND: No data available. PDS: Patient-derived sample. *=discontinued

PDS	ER	PR	Chemotherapy	Endocrine Therapy	Bone Therapy	Other Therapy	Metastasis
49	+	+	Doxorubicin Cyclophosphamide Taxol Gemcitabine Capecitabine Etoposide Ixabepilone Eribulin	Goserelin Anastrozole Fulv	Zometa Zolendronic acid	Anti-DLL4 antibody OMP-21M18*	Bone Brain Pleura
53	+	+	Doxorubicin Cyclophosphamide Taxol	Tam Anastozole	Zometa	ND	Liver Bone
56	+	+	Capecitabine Vinorelbine Carboplatinum Paclitaxel	Letrazole Tam Exemestane	Denosumab	ND	Chest wall Lymph Bone
61	+	+	Doxorubicin Cyclophosphamide Docetaxel	Anastrozole Exemestane	Denosumab	ND	Bone Pleura
66	+	+	Capecitabine Gemcitabine Carboplatin	Tam Arimidex	Denosumab	ND	

Table 2. 8 Manchester patient sample clinical information.
Second and third columns show percentage of ER+ and PR+ cancer cells, respectively. ND: No data available. Samples 90 and 90A belong to the same patient but samples were taken at different time points. The same thing applies to samples 91 and 91A. QS: Quick score.

PDS	ER	PR	Chemo-therapy	Endocrine Therapy	Bone Therapy	Other Therapy	Metastasis
68	QS 8		Capecitabine FEC	Tam Fulv Anastrazole	ND	ND	Lymph nodes Bladder Peritoneum Liver Lung
69	96	77	ND	Tam Fulv Anastrazole Letrozole	Pamidronate Zoledronic Acid	ND	Peritoneum Bone Lymph node
71	54	72	FEC Taxol Capecitabine Vinorelbine Taxotere Eribulin	Tam Fulv Anastrazole Letrozole Exemestan Goserelin	Pamidronate Zoledronic Acid Ibandronic Acid	Herceptin Lapatinib	ND
89	QS 8	QS 8	Capecitabine FEC Taxol	Tam Letrozole	ND	ND	Liver Lung Bone
90 - 90 A	QS 8	QS 8	Capecitabine	ND	ND	ND	Liver Bone Pleura
91 - 91 A	96	98	FEC Docetaxel	Tam Anastrazole Letrozole Exemestan	ND	Everolimus	Bone Liver Peritoneum Omentum
94	+	+	Treatment naïve				ND

2.2.3 Flow Cytometry and Flow Activated Cell sorting

Cells were syringed using a 25-Gauge (G) needle and filtered with a 40 μM filter before passing them through the Flow cytometry equipment to avoid blockage. Samples were then resuspended in FACS buffer (2 % FBS, 1mM EDTA, 0.1 % sodium azide in PBS) at $5x10^6$ cells per ml (300 μl was the minimum volume cells were resuspended with) before fluorescence was measured. If the aldefluor assay was performed (see below) the cell pellet was resuspended in Aldefluor buffer instead.

2.2.3.1 Aldefluor assay

The Aldefluor kit is a fluorescent reagent system (Stemcell Technologies) used for the detection of the ALDH activity in cells. This kit was employed to isolate the CSC population on the basis of their ALDH activity. The protocol was carried out following the manufacturer's guidelines. Cells were trypsinised, neutralised with media and passed through a 25 G syringe to obtain a single cell suspension. Cells were counted using the vital stain trypan blue and the required cell number was pelleted at 400 g for 4 min.

A maximum of $1x10^6$ cells were stained per tube. Cells were resuspended in 0.5 ml of Aldefluor buffer and 10 µl of the ALDH inhibitor Diethylaminobenzaldehyde (DEAB) was added to one of the eppendorfs for each treatment, which was used to correct for background staining and to define the ALDH pos region in the flow cytometer. The Aldefluor reagent Bodipy-aminoacetaldehyde (BAAA) was added to all eppendorfs (2.5 µl) and samples were incubated in the dark at 37 ºC for 45 min. BAAA entered the living cells through passive diffusion and was then transformed by intracellular ALDH into the negatively charged Bodipy-aminoacetate (BAA) which is kept inside the cells. This reaction leads to the formation of fluorescence which is proportional to the activity of ALDH inside the cell. After incubation, samples were washed twice, centrifuged at 400 g for 3 min and after supernatant was removed pellets were resuspended in 300 µl of Aldefluor buffer. Five µl of the cell viability stain 7-actinoaminomycin (7-AAD) was added per $1x10^6$ cells for cell death exclusion. ALDH-bright cells were detected in the green fluorescence channel (520-540 nm) with the LSRII (BD Bioscience) or the InFlux (BD Bioscience) cytometers. Unstained cells were always used to ensure accurate gating. Cell debris and doublets were gated out based on cell size and complexity. Data was analysed using the FlowJo software package.

2.2.3.2 Aldefluor, CD44-APC, CD24-PE-CY7 and DAPI staining

For single cell and *in vivo* experiments, MCF-7 cells were stained following a protocol established in the laboratory of Professor Max S.

Wicha (University of Michigan, Comprehensive Cancer Centre). A cell suspension was prepared at a concentration of $1x10^7$ cells/ml in Aldefluor buffer and the conjugated antibodies, Aldefluor reagent and the cell viability stain 4',6-diamidino-2-phenylindole (DAPI) were added according to Table 2.9. Cells were then incubated at 37 °C for 40 min, washed twice and resuspended with Aldefluor reagent at $5x10^6$ cells per ml. Isotype controls were used to gate the $CD44^+/CD24^-$ population at the concentration specified below (Table 2.10). Single staining and fluorescence minus one (FMO) controls, which contain all the antibodies tested minus one of them, were used for compensation and gating strategies respectively.

Table 2. 9 FACS staining combining ALDH, CD44, CD24 and DAPI.

Channel	Marker	Company	Catalogue	Final Concentration (µM)
FITC	Aldefluor	STEMCELL	01701	5 µl
PE-Cy7	CD24	BD	348788	200 ng/µl
APC	CD44	BD	340442	3 ng/µl
DAPI	DAPI	Sigma	D9542	0.1 mg/ml
Total MasterMix volume				120 µl
Cells				880 µl
Total				1000 µl

Table 2. 10 Isotype control information.

Channel	Marker	MAb	Company	Catalogue	Concentration
PE-Cy7	Isotype	Mouse IgG1	BD	348788	200 ng/µl
APC	Isotype	Mouse IgG1	BD	340442	3 ng/µl

2.2.3.3 Flow Activated Cell Sorting (FACS)

Cells were sorted using the InFlux (BD Bioscience) and HBSS as sheath fluid. Cells were sorted at 16 pounds per square inch (PSI) using a 100 µm nozzle and collected into 1.5 ml eppendorfs containing 2 % FBS in HBSS. When cells were used for RNA extraction, cells were sorted into RNA lysis buffer instead.

Following cell sorting, purity check of the sorted population was performed to assess percentage of sorting efficiency. The post-sort purity check was performed by sorting a small aliquot of cells into an eppendorf with Aldefluor buffer, which contains efflux inhibitors that inhibit the activity of ABC transporters and therefore retains the fluorescence substrate inside the cells. Sorting efficiency was always greater than 75%. A representative example of a purity check FACS plot is represented in Appendix Figure 8.4.

2.2.4 *In vivo* tumour-initiating capabilities from ALDH positive and ALDH negative MCF-7 cells

MCF-7 cells were treated in monolayer for 6 days with ethanol control, Tam or Fulv (according to the experimental design described in Figure 3.2), followed by FACS sorting of the ALDH pos and ALDH neg populations into 1.5 ml eppendorfs. Subsequently, cells were spun down, resuspended in mammosphere media and counted using trypan blue. Serial cell dilutions were made according to Table 2.11. Matrigel was added to the cell suspension to produce 0.2 ml of matrigel and mammosphere media at a 1:1 ratio. Cells were inoculated subcutaneously into the left and right flanks of 4-5 week old female Nod Scid Gamma (NSG) mice using a 25 G needle, following the experimental design represented below.

Table 2. 11 Experiment design for *in vivo* testing the tumour capabilities of ALDH pos and ALDH neg MCF-7 cells

Mouse number	Left Flank	Right Flank	Cell number per flank	*In vitro* pre-treatment with
1-4	ALDH neg	ALDH pos	1000	Ethanol Control
5-8	ALDH neg	ALDH pos	100	
9-12	ALDH neg	ALDH pos	10	
13-16	ALDH neg	ALDH pos	1000	Tam
17-20	ALDH neg	ALDH pos	100	
21-24	ALDH neg	ALDH pos	10	
25-28	ALDH neg	ALDH pos	1000	Fulv
29-32	ALDH neg	ALDH pos	100	
33-36	ALDH neg	ALDH pos	10	
37-38	Unsorted cells		1000/100	Untreated

Unsorted untreated cells were injected into two control mice (37-38 mice) as a quality control step, to ensure that the sorting procedure was not a cause for lack of engraftment. 17β-estradiol 90 day slow-release pellets (0.18 mg) were implanted subcutaneously in the neck scruff region of the mice seven days before injection of the tumour cells. After 90 days, when the pellets were exhausted, 17β-estradiol was administered in the drinking water (8 μg/ml) at all times. Tumour measurements were taken three times a week after tumour cells were implanted using a calliper and tumour size was calculated using the formula 0.5 x Length x $Width^2$.

2.2.5 Gene expression analysis

RNA extraction: Total RNA extraction from cells cultured in 6-well plates (unsorted ≥5x10^5 cells) was carried out using the RNasePlus Mini Kit (Qiagen) and total RNA extraction from sorted cells (≤5x10^5 cells) was performed using the RNasePlus Micro Kit (Qiagen), according to manufacturer's instructions. All surfaces were cleaned using the RNaseZap (Sigma Life Science), a cleaning agent for removing ribonucleases (RNases), before starting. RNA concentration and purity of the unsorted cells was assessed using the Nanodrop (NanoDrop ND-1000 spectrophotometer, Thermo Scientific). The ratio of the readings at 260 nm and 280 nm (A_{260}/A_{280}) provides an estimate of the purity of RNA and a ratio of ~2.0 identifies pure RNA. The ratio A_{260}/A_{230} is used as a secondary measure of nucleic acid purity and should be greater than 2.0. RNA samples were stored at -80 ºC until needed.

Quality Control: The 2100 Bioanalyzer (Agilent 2100 Bioanalyzer system, Agilent Technologies) and the Qubit (Thermofisher Scientific) were used for quantitation and quality control of the RNA from sorted cells due to the reduce yield and volume of the sample. The 2100 Bioanalyzer is a chip-based acid nucleic system used to assess RNA quality by looking at the ribosomal RNA (rRNA) peaks (18S and 28S for eukaryotic rRNA) in a miniaturised column electrophoresis. To calculate the RNA concentration, the area under the RNA electropherogram is assessed and compared to that from the ladder. This system also provides with an RNA Integrity Number (RIN) value where RIN=10 indicates highest RNA integrity and RIN=0 shows highly degraded RNA. An example of an optimal and a poor RIN value is shown in Figure 2.2. When assessing RNA integrity from samples with very low RNA concentrations the RNA 6000 Nano chip was used as only requires around 25 ng of RNA. RIN values ≥ 7 are considered optimal for downstream applications such as microarrays.

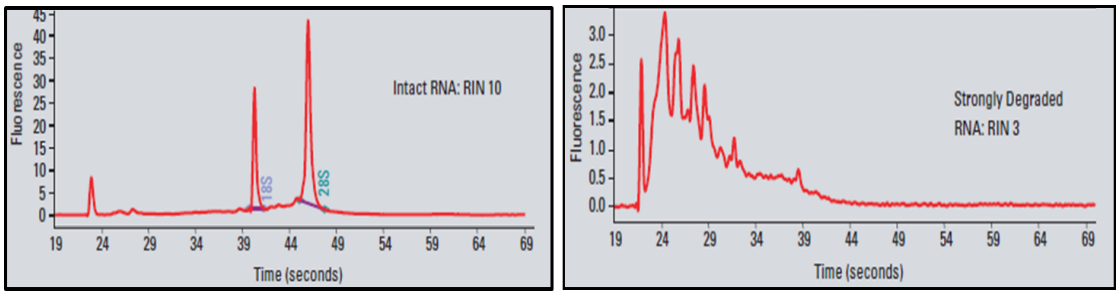


Figure 2. 2 Illustration of good and bad RNA quality samples indicated by their RIN values using the 2100 Bioanalyzer.
The graph on the left shows perfect distinguishable peaks for both 18s and 28s rRNA and therefore a RIN value of 10 is assigned to the sample, whereas the graph shown on the right does not present distinct peaks, therefore it is assigned a low RIN value of 3. Figure taken from Agilent Technologies.

Reverse Transcription: The synthesis of the first strand cDNA was performed using the TaqMan Transcription Reagents (Applied Biosystems) (Table 2.12). One µg of RNA was reversed in a 50 µl reaction by adding the reagents and the volumes specified in the table below. RNase free water was added to the reaction mix for a final volume of 50 µl.

Table 2. 12 Reagents used for Reverse Transcription

Reagents	**Volume**	**Final Concentration**
10x TaqMan Buffer	5 µl	1x
$MgCl_2$(25mM)	11 µl	5.5 mM
dNTPs (10mM)	10 µl	2 mM
OligodT (50 µM)	2.5 µl	2.5 µM
RNase inhibitor (20 U/µl)*	1 µl	0.4 U/µl
Reverse Transcriptase	1.25 µl	

***** 1 enzyme unit (U) = 1 µmol min^{-1}

Samples were loaded onto the PTC-200 PCR thermal cycler (MJ Research) and run for 10 min at 25 ºC, 30 min at 48 ºC and 5 min at 95 ºC. For each set of samples a negative control, containing the reaction mix but no RNA, was set up. Samples were then kept at -80 ºC until further use. For low amounts of RNA, generally from sorted samples, the

cDNA preparation Reverse Transcription Kit (Fluidigm) was used instead, which uses a mixture of random primers and oligo dT.

cDNA pre-amplification: Targeted pre-amplification was performed using the Pre-Amp mastermix (Fluidigm) and the TaqMan assays, following manufacturer's guidelines. Diluted reaction products were stored at -20 ºC until needed. When interrogating ALDH1A3 gene expression from sorted ALDH pos and ALDH neg MCF-7 cells, cDNA pre-amplification was necessary to increase the number of cDNA molecules and detect gene expression using the FlexSix technology (Fluidigm).

2.2.5.1 Quantitative Polymerase Chain Reaction

Expression of target genes was assessed by quantitative PCR (qPCR) on the ABI 7900HT Real time PCR system (Applied Biosystems) or on the QuantStudio real time PCR systems (Applied Biosystems) using the TaqMan Probe-Based Gene Expression Analysis (Applied Biosystems)**.** Pairs of unlabeled PCR primers specific for each transcript and specific TaqMan probes carrying FAM dye label on the 5' end and MGB and NFQ on the 3' end were used. 50 ng of cDNA was amplified in each well using 384-well plates. The table shown below (Table 2.13) contains the volumes needed for each reagent per well. Samples were loaded in triplicate and two housekeeping genes were used for each PCR reaction. Conditions used for amplification of cDNA fragments were as follows: 50 ºC for 2 min, 95 ºC for 10 min, 40 cycles of amplification, 95 ºC for 15 seconds, and 60 º C for 1 min. For details of TaqMan primers refer to Table 2.5.

Table 2. 13 qPCR reagents

Reagents	Volume	Final Concentration
cDNA (0.02µg/µl)*	5 µl	5 ng/µl
TaqMan Master Mix (2x)	4.4 µl	1x
Forward & Reverse Primer pair (20 pmol/µl)	0.5 µl	500 nM
Universal Probe (10µM)	0.1 µl	100 nM

*50 µl of cDNA (0.02µg/µl) resulting from the Reverse Transcription reaction were diluted in 50 µl of ddH_20 prior using it in the qPCR.

2.2.5.2 Affymetrix Microarrays

Whole transcript array was used to assess the mRNA expression profile in bulk cells and between ALDH pos and ALDH neg MCF-7 cells after ethanol control, Tam and Fulv treatment. The gene expression differences between ALDH pos and ALDH neg samples from PDS were also investigated using Affymetrix arrays.. RNA extraction was performed as explained at the beginning of section 2.2.5. The AffymetrixGeneChip (Figure 2.3) system enabled us to interrogate the whole transcriptome using a technology that includes 25-mer oligonucleotide probes arrays placed on quartz chips of normally 1.28 cm^2 area. GeneChip Human Array Gene 1.0 ST (Affymetrix) consists of single square-shaped feature known as probe cells of typically 11 µm. These arrays consist of perfect match-only designed probes (it does not contain mismatch probes) that hybridise to sense targets (Figure 2.4). The Human Array Gene 1.0 ST provides an accurate measurement of protein coding and long non-coding RNA transcripts. The specifications for this gene chip array are described in Table 1.

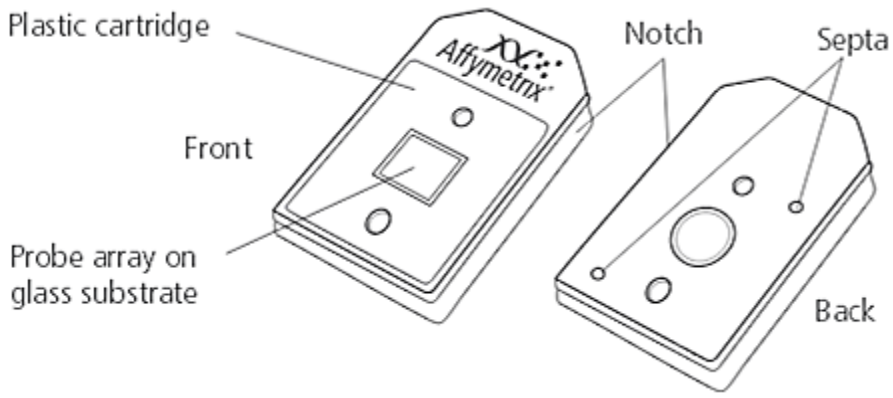


Figure 2. 3 GeneChip Probe Array.
Figure taken from Affymetrix.

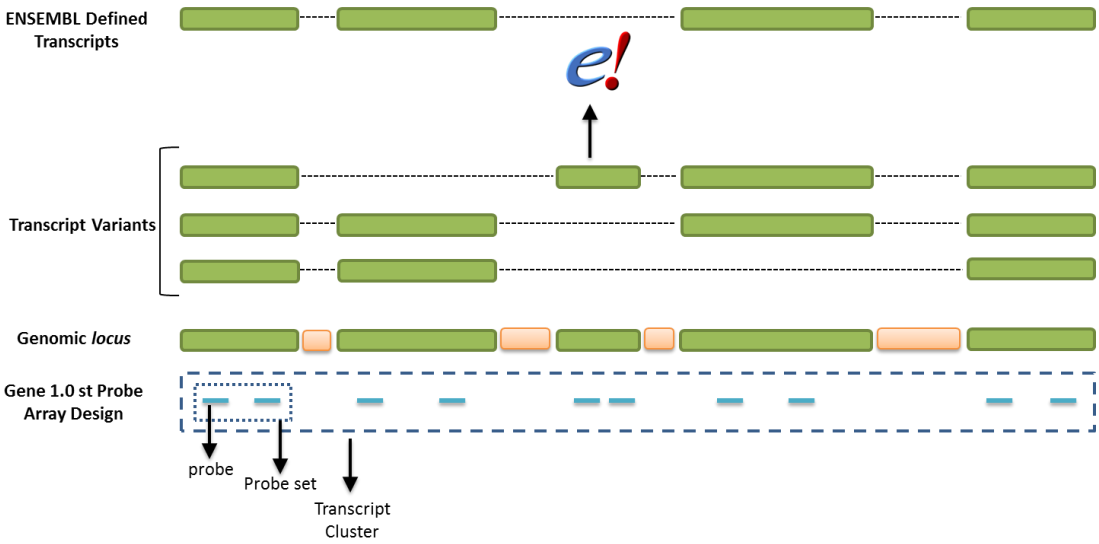


Figure 2. 4 Affymetrix Human Gene 1.0 st probe array design.
Probes are located along the entire length of the gene. On average, 2 probes map to each exon and the expression signals are gathered together into a probe set signal. Subsequently, probe sets can be put together into transcript clusters to give the expression signal of all the transcripts which are encoded by the same gene (gene expression value). In this example, 1 gene, containing 5 different exons, will originate 3 different transcripts following alternative splicing which ultimately will generate different proteins. Only transcripts (spliced isoforms) annotated by ENSEMBL (e!) will be studied. Green regions are exons whereas brown regions represent introns that are removed during splicing.

Table 2. 14 Human Array Gene 1.0 ST specifications

Oligonucleotide probe length	25-mer
Number of probes	764,885
Strand interrogated	Sense
Average number of probes per gene	26
Estimated number of genes	28,869
Gene-level probe sets with Ensembl support	28,132
Gene-level probe sets with putative full-length transcript support (GenBank and RefSeq)	19,734
Nucleic acid hybridised	cDNA

Total RNA samples were sent to the CRUK Manchester Institute molecular biology core facility to generate double stranded cDNA using the Ovation Pico WTA System V2 (NuGen). The whole explanatory process is shown in Figure 2.5. First strand cDNA was synthesised using DNA/RNA chimeric primers, which contains a mixture of random and oligodT primers so that whole transcriptome coverage is achieved, using the reverse transcriptase enzyme. Reverse transcription begins to synthesise DNA from the 3' end of each primer producing a cDNA/mRNA hybrid molecule holding a RNA tag at the 5' end of the cDNA. To generate the second strand of cDNA the old mRNA strand is degraded which allows the DNA polymerase to function, generating a complementary sequence that results in a double-stranded cDNA with a small portion of DNA/RNA heteroduplex due to the use of chimeric primers.

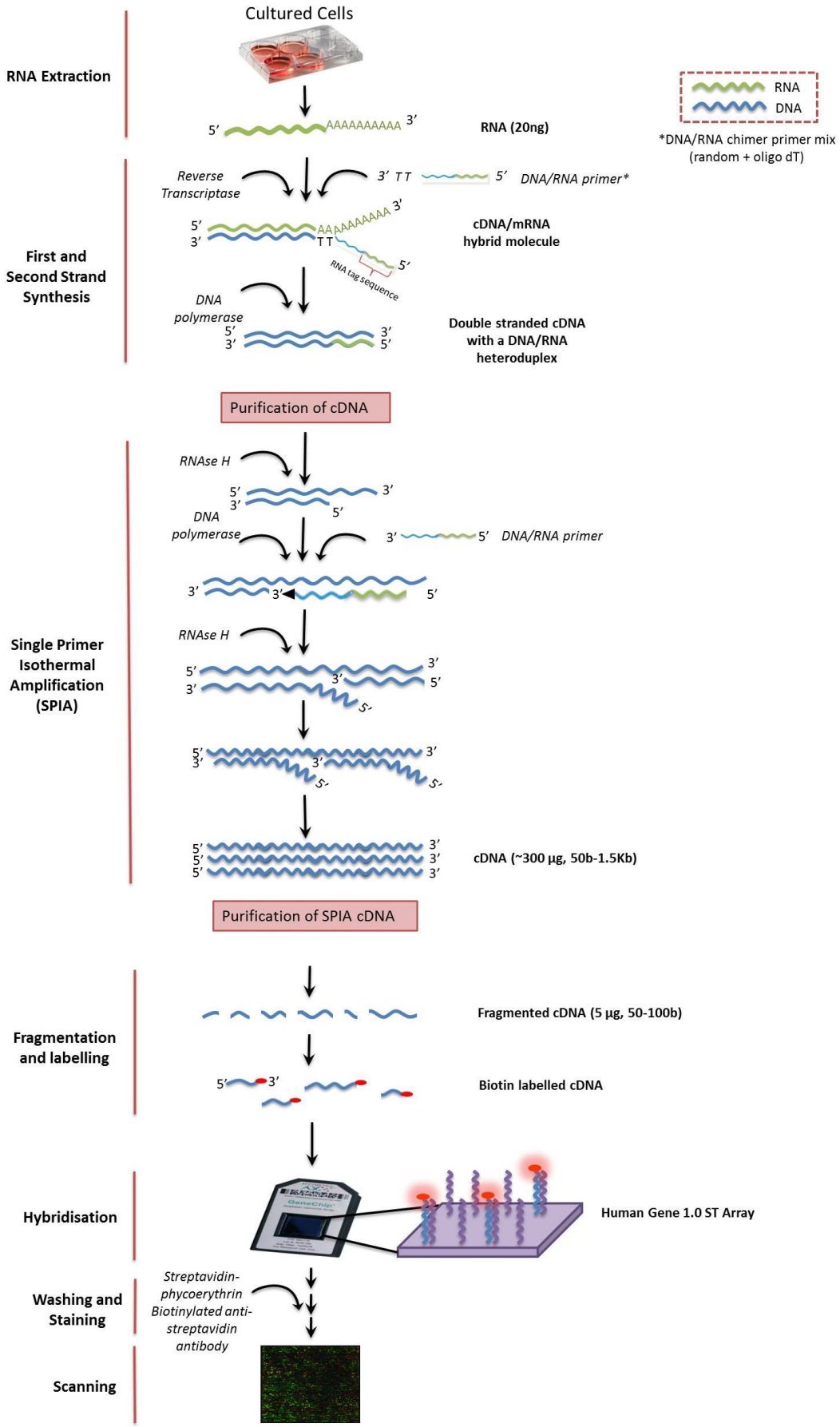


Figure 2. 5 Flowchart of the whole Affymetrix Arrays procedure.

Single Primer Isothermal Amplification (SPIA) process allowed the production of ~300 µg of cDNA from 20 ng of starting material. RNase H was used to get rid of the RNA portion of the DNA/RNA heteroduplex. Chimeric primers were added and DNA polymerase began to create cDNA starting at the 3' end of the primer displacing the existing forward strand. Chimeric DNA/RNA primers were used again so that the heteroduplex carrying the RNA tag could serve as a new substrate for RNase H and the initiation of the next cycle of cDNA synthesis.

2.2.5.2.1 Fragmentation and Labelling

cDNA was fragmented and biotin labeled to generate labeled targets suitable for hybridisation onto the array using the Encore Biotin Module (NuGen) and following the manufacturer's guidelines. Five µg of cDNA were processed by a mixture of enzymatic and chemical fragmentation, which allowed the generation of cDNA products between 50 to 100 base pairs long. cDNA was combined with 5 µl of the fragmentation buffer and 2 µl of the fragmentation enzyme and samples were then placed in a thermal cycler and run for 30 min at 37 ºC and 2 min at 95 ºC to generate small portions of cDNA. Subsequently, biotin-labeled cDNA was generated by mixing the fragmented cDNA along with the labeling buffer mix, biotin reagent and the labeling enzyme mix. Tubes were then placed in the thermal cycle programmed to run for 60 min at 37 ºC and 10 min at 70 ºC. Samples were then ready for array hybridisation.

2.2.5.2.2 Hybridisation and signal detection

Prior hybridisation, the GeneChip array was allowed to come to room temperature. The hybridization cocktail was prepared as indicated in the table below, heated at 99 ºC for 5 min and cooled down to 45 ºC for 5 min.

Table 2. 15 Hybridisation cocktail composition

Component	Volume for one 49/64 Format Array	Final Concentration
Fragmented and labeled DNA	~60 µl	25 ng/µl
Control oligonucleotide B2 (3nM)	3.7 µl	50 pM
20x Eukaryotic hybridization Controls (bioB, bioC, bioD, cre)	11 µl	1.5, 5, 25 and 100 pM, respectively
2x Hybridisation mix	110 µl	1x
DMSO	15.4 µl	7%
Nuclease-Free Water	Up to 220 µl	
Total Volume	220 µl	

The hybridisation cocktail (200 µl) was injected into the array through one of the septa (Figure 2.3). The GeneChip array was placed at 45 ºC and 60 rpm for 17h ± 1 h into the hybridisation oven (GeneChip hybridization Oven 640, Affymetrix) until hybridisation was completed.

Immediately following hybridisation, the GeneChip array was washed and stained using the Fluidics Station protocol FS450_0007 and the Affymetrix GeneChip Command Console Software (Affymetrix) following manufacturer's guidelines. Briefly, the bound target molecules were stained with a fluorescent-streptavidin-phycoerythrine (SAPE) conjugate, which is in turn bound to the biotin. Subsequently, the signal is amplified by washing with a solution biotinylated anti-streptavidin goat antibody, followed by a final staining with SAPE (Figure 2.6).

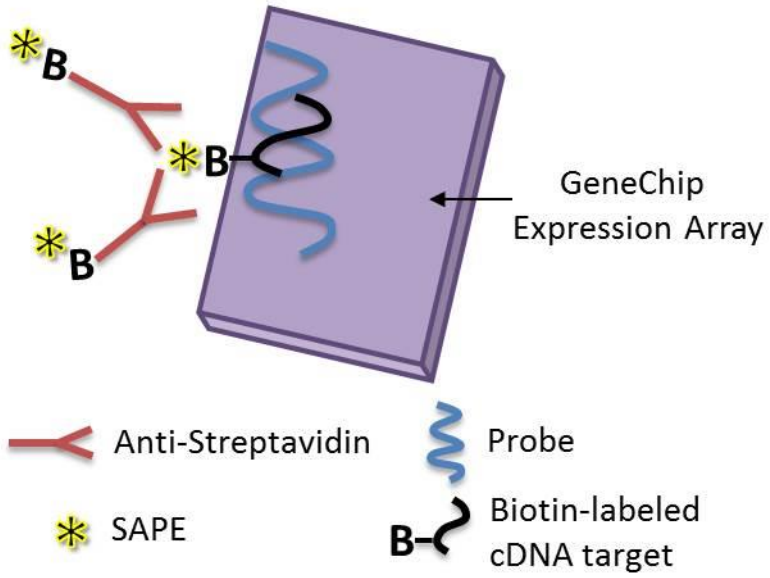


Figure 2. 6 Microarray hybridisation process.

Each spot in the microarray image needs to be matched to a specific spot in the physical experiment slide, which is shown for each particular gene. However, multiple errors can occur during microarray processing leading to irregular spot shapes or different spatial distribution of rows and columns. Oligo B2 controls address these problems and allow improved microarray image analysis. Oligo B2 controls hybridise to specific areas of the array providing fluorescent signals for the Affymetrix Microarray Suite Software to carry out automatic grid alignment during image analysis.

The GeneChip array was scanned using the AffymetrixGeneChip Scanner 3000 system with autoloader (Affymetrix).

2.2.6 Single cell gene expression analysis

The C1 system from Fluidigm is an automated single cell technology that isolates, extracts RNA, synthesises and amplifies cDNA from single cells. Medium C1 chips were used to interrogate gene expression in single ALDH pos MCF-7 cells since it allows capture of cells between 10-17 μm of size. C1 Chip was primed following manufacturer's instructions. Consequently, the mixture containing the C1 suspension

reagent and the cells of interest was loaded onto the chip for running the Cell Load Script (1782x1783x). Subsequently, lysis, reverse transcription and pre-amplification were performed by running the PreAmp script (1782x/1783x) on the C1 system, following manufacturer's guidelines. Outputs were harvested and stored at -20°C.

2.2.6.1 High-throughput qPCR (Biomark)

The Biomark qPCR system is a high-throughput device that enables researchers to interrogate expression of multiple targets in multiple samples at the same time. There are different available Integrated Fluidics Circuit (IFC) chips to use. The 96.96 Biomark IFC chip (Figure 2.7) is the technology employed to study the expression of 96 genes in 96 single cells that had been previously captured using the C1 system. On the other hand, the Flex Six IFC chip allows the study of 12 genes in 12 samples at a time and contains 6 different partitionings that can be used separately.

In order to investigate CSC heterogeneity, the gene expression patterns of 96 genes (Table 2.5) were interrogated in ALDH pos MCF-7 cells using the 96.96 IFC Biomark chip (Fluidigm) (Figure 2.7) and following the Gene Expression 96.96 IFC Standard TaqMan Assays protocol (PN 68000130 E1; Fluidigm). This system helped us to interrogate 9216 data points at once. Two independent sorting experiments and one technical replicate (same cDNA but used in different high-throughput qPCRs) resulted in 517 single ALDH pos cells (across the three treatments: control, Tam and Fulv) being analysed in 6 different Biomark qPCR.

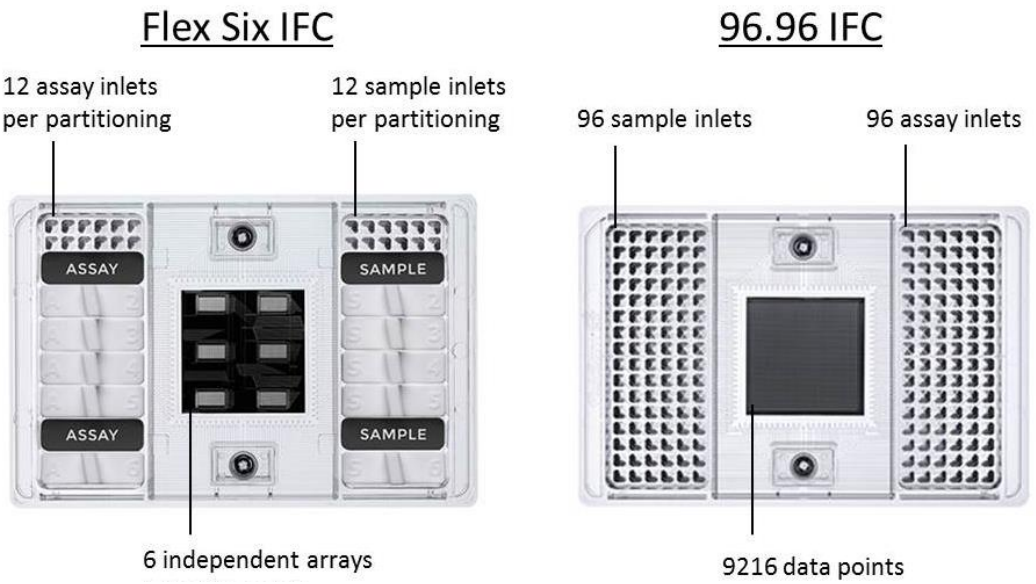


Figure 2. 7 Flex Six IFC vs. 96.96 IFC gene expression chips (Biomark).

2.2.6.2 Cell doublet detection within the C1 medium chips

Due to a malfunctioning of the medium C1 chips (Fluidigm), a large proportion of cell doublets were captured within the C1 chambers. In order to test the exact percentage of doublets captured using MCF-7 cells, $1x10^6$ cells were split into two groups, half of the cells were stained with the blue-fluorescent DNA dye Hoechst 33342 at 10 µg/ml for 20 min at 37 ºC and the other half was stained with the red-fluorescent DNA dye DRAQ5 at 20 µg/ml for 20 min at room temperature. Cells were also stained with the cell viability staining solution 7AAD to discard the dead cells, followed by FACS sorting. MCF-7 sorted cells were then counted using Trypan blue and the exact same number of blue and red-labelled cells were mixed together. Subsequently, cells were loaded into the C1 medium chip using the C1 system and chambers were visualised using a Leica Widefield Low Light microscope to record the number of double-coloured doublets, singlets and empty chambers. Finally, due to the difficulty of visualising single-coloured doublets, it was estimated that the number of red/red and blue/blue doublets would be the same as the double-coloured doublets.

2.2.14 Statistics and bioinformatics analysis

Microarray data was processed using the package Affy in R (Gautier et al. 2004). Data was quantile-normalised and Log2 transformed. Significance Analysis of Microarrays (SAM) was applied using the R package Siggenes (Schwender 2012) in order to identify differentially expressed genes. The total number of differentially expressed genes observed in SAM analysis depends on the designated threshold, known as Delta value. Rank products was applied using the R package RankProd as an alternative approach to detect differentially expressed genes (Hong et al. 2006)

Single cell transcriptomics data were analysed as shown in Figure 6.4. Briefly, data generated by the Biomark (Fluidigm) were converted into Log2 expression values. Data was filtered to remove all values under the limit of detection, which was set to threshold cycles (Ct) greater than 28, and outliers were removed using the function *identifyOutliers* implemented in the R package FluidigmSC (Fluidigm Corporation 2014). Missing values were inputted using the R package MICE (Van Buuren et al. 2008) and further filtering included the removal of genes that were expressed only in 3 or less cells. Batch effect was corrected using Combat (Johnson et al. 2007). Cell clusters were estimated via finite Gaussian mixture modelling using the function Mclust within the R package Mclust (Scrucca et al. 2016) resulting in 5, 7 and 8 clusters for Control, Tam and Fulv. Ward hierarchical clustering per treatment was used to merge very small clusters in larger ones (see Figure 5.9) resulting in 2, 2 and 3 clusters for Control, Tam and Fulv. Evaluation of the similarities and differences of clusters was done via (a) Ward hierarchical clustering and (b) Discriminant Analysis of Principal components resulting in the identification of treatment specific clusters.

Statistical significance for protein and RNA expression cell line data was assessed using t-test parametric testing. P values less that 0.05 were considered significant. $P<0.05$ (*), $P<0.01$(**), $P<0.001$(***).

Results are presented as the mean of at least 3 independent experiments ± Standard Error of the Mean (SEM) or Standard Error (SE)

3. Chapter 3: Study of the ALDH pos population in breast cancer cells

3.1 Introduction

Normal breast tissue shows a hierarchical organisation maintained by stem cells, which are thought to be responsible to drive breast growth during the morphological changes the mammary gland undergoes (Visvader 2009). It is believed that this hierarchy is maintained in cancer, where CSCs, which can be enriched based on the expression of different markers, are able to form heterogeneous tumours in mice as well as be more resistant to standard therapy (Visvader et al. 2008). Previous work from our lab, has shown that treatment of ER+ tumours with the therapy of choice Tam and Fulv, do not target breast CSCs, which helps explain the high rates of relapse in ER+ patients after anti-estrogen therapy (Simões et al. 2015).

Numerous cell surface markers have been identified to enrich for BCSCs in different types of cancers, although to date there is no method capable of isolating a pure population of CSCs. ALDH1 is an enzyme responsible for the oxidation of intracellular aldehydes and its activity has been shown to mark a population of BCSCs with self-renewal and tumour-initiating capabilities (Ginestier et al. 2007). Likewise, Michael Clarke and collaborators showed that human breast cancers contain a small population of tumourigenic breast CSCs expressing the cell surface marker $CD44^+/CD24^-$ (Al-Hajj et al. 2003). ALDH and $CD44^+/CD24^-$ mark two different breast CSC populations, an epithelial-like and a mesenchymal-like, however there is also a small population (0.05%) of cells that co-express both markers (Liu et al. 2014). Moreover, it has recently been suggested that intracellular autofluorescence of tumour cells can be used to mark cells with CSC features (Miranda-Lorenzo et al. 2014).

The mammosphere assay, developed by Dontu et al (Dontu et al. 2003) from the neurosphere assay (Reynolds et al. 1996), is an *in vitro* technique which has been widely used to culture breast stem cells in non-adherent conditions in both normal and tumorous mammary tissue (Ponti

et al. 2005). Differentiated cells undergo anoikis, a programmed cell death which is triggered in anchorage-dependent cells after detaching from the surrounding extracellular matrix (Frisch et al. 2001) (Shaw et al. 2012), when cultured in suspension; whereas undifferentiated stem cells survive low-adherent conditions forming clonal colonies, known as mammospheres. Our group has previously reported that anoikis resistant cells show increased capabilities of forming tumours *in vivo* (Harrison et al. 2010).

Given the putative role of the breast CSC population at driving resistance to Tam and Fulv treatment, this chapter aims at investigating the effects of both drugs on the breast CSC compartment by performing flow cytometry, mammosphere, knock-down and *in vivo* tumour-initiating assays using two common ER+ cell lines, MCF-7 and T47D cells and ER+ patient-derived samples (PDS).

3.2 Results

3.2.1 Effects of Tamoxifen and Fulvestrant on cellular viability

MCF-7 cells have been used for most of the experiments shown in this report since it represents, together with T47D cells, the only available ER+/PR+/HER2- breast cancer cell lines. In addition, MCF-7 is the most commonly used breast cancer cell line in the world and has become a powerful tool for cancer research. However, care must be taken when interpreting these data since cell lines do not always accurately resemble the original characteristics of breast tumours.

A common symptom during late stage breast cancer is the collection of fluid in the pleura of the lung (pleural effusion) or in the abdomen (ascitic fluids) which indicate metastatic spread. Fluid drainage is a standard practise in the clinic to improve overall wellbeing of the metastatic patient. Epithelial tumour cells isolated from these fluids may be used in the lab as these samples are normally large and easily accessible. Moreover, the possibility of obtaining repeated samples of the same patients make these fluids an attractive model for cancer research (Grannis 2011). Therefore, frozen pleural or ascitic patient-derived fluid

samples were also used in this report as an alternative model to cell lines in order to recapitulate the biology of breast tumours. Of note, samples originated from primary tumours may have been a superior model to study the anti-estrogen resistance mechanisms in CSCs, however the limited availability and amount of cells isolated from these biopsies makes it too challenging to use them for our experiments.

It has been reported that Tam and Fulv inhibit cell proliferation by causing cellular arrest. Figure 3.1 shows the effects that Tam (green) and Fulv (blue) have in cell density in MCF-7 cells, compared to ethanol control (pink) for 72 h, based on the measurement of cellular protein content using the colorimetric SRB proliferation assay. Treatment with Tam and Fulv reduces cell growth by 13.2% and 23% respectively after 24h, and both drugs are most effective at inhibiting cell proliferation after 72h when cell growth dropped by 16% (Tam) and 34% (Fulv) (both $p<0.001$).

In order to mimic early stages in the development of acquired resistance to anti-estrogen therapy and to maximise effects, a 6-day *in vitro* treatment approach was undertaken (Figure 3.2) as used previously (Simões et al. 2015).

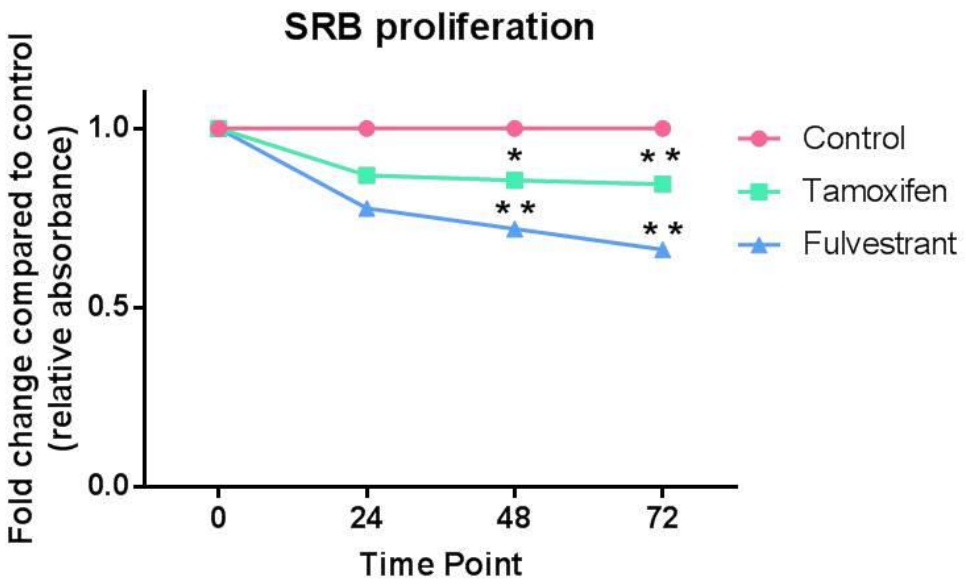


Figure 3. 1 Anti-estrogen drugs inhibit MCF-7 cell proliferation measured using SRB assay.
5,000 cells per well were seeded into 96 well plates treated with either control (ethanol; pink), Tam (green) or Fulv (blue) for 0, 24, 48 or 72 h. Optical density was measured at 550 nm in a microplate reader. Figure illustrates representative experiment and average from three technical replicates. SE is shown. *P≤0.05, **P≤0.01 (T-test).

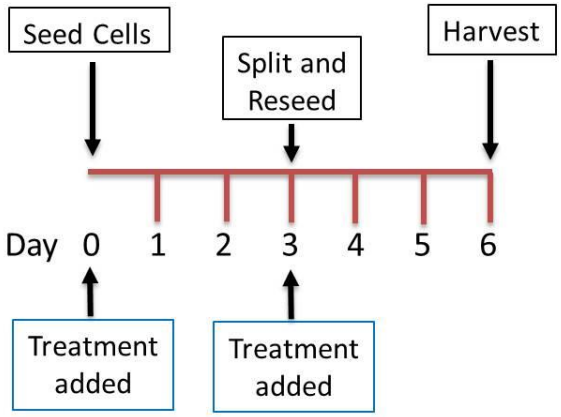


Figure 3. 2 Six-day *in vitro* treatment time-line.
MCF-7 cells were seeded in complete media containing ethanol control, 1 μM Tam or 0.1 μM Fulv and left to grow for 3 days until 80% confluency was reached. Cells were then harvested, syringed to create single cell suspension, and counted using trypan blue. Subsequently, cells were seeded with the treatments, at the same cell density as day 1, for another 3 days prior to cell harvesting.

3.2.2 Enrichment of ALDH positive cells after anti-estrogen treatment

Our group has reported previously that treatment with the anti-estrogen agent Tam increases mammosphere formation and self-renewal (secondary mammospheres) of patient-derived samples (Simões et al. 2015). In agreement with those findings, an increase in the MFE (primary mammosphere forming efficiency) of MCF-7 cells was demonstrated after 6 days *in vitro* treatment with Tam or Fulv (2.6 and 2 fold respectively) in the present study (Figure 3.3). These data show that treatment with anti-estrogen drugs enrich for stem cell activity.

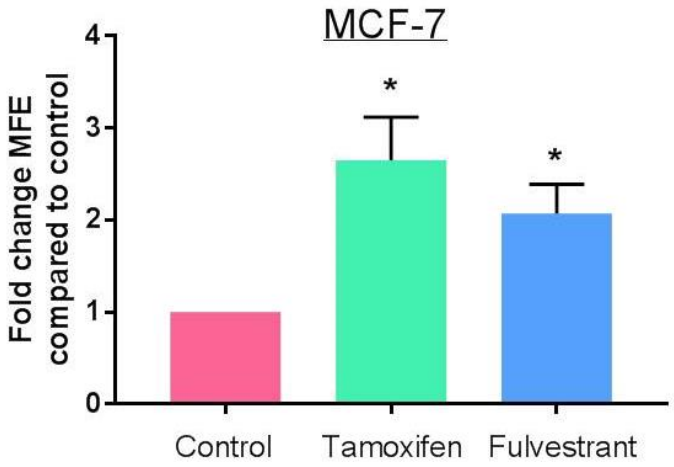


Figure 3. 3 MFE fold change of MCF-7 cells after 6-days Tam and Fulv treatment in monolayer.
MCF-7 cells were counted, syringed with a 25 G needle and resuspended with the adequate amount of mammosphere media. 3,000 cells per well were seeded in polyHema-coated 6 well plates and incubated at 37 °C and 5% CO_2 for 7 days. MFE was assessed after 7 days of culture by counting the number of mammospheres greater than 50 µm of diameter at 40X magnification using a microscope. Data shows average and SEM of three independent experiments. Data is represented as fold change (FC) compared to ethanol control *P≤0.05

ALDH enzyme activity enriches for breast cancer progenitor cells (Ginestier et al. 2007). We therefore tested the effect of Tam and Fulv on ALDH activity and found these two drugs significantly increase the percentage of ALDH pos cells by 2.7 (Tam) and 4 fold (Fulv) compared to ethanol control in MCF-7 and by 2.6 (Tam) and 5 (Fulv) fold in T47D cells (Figure 3.4).

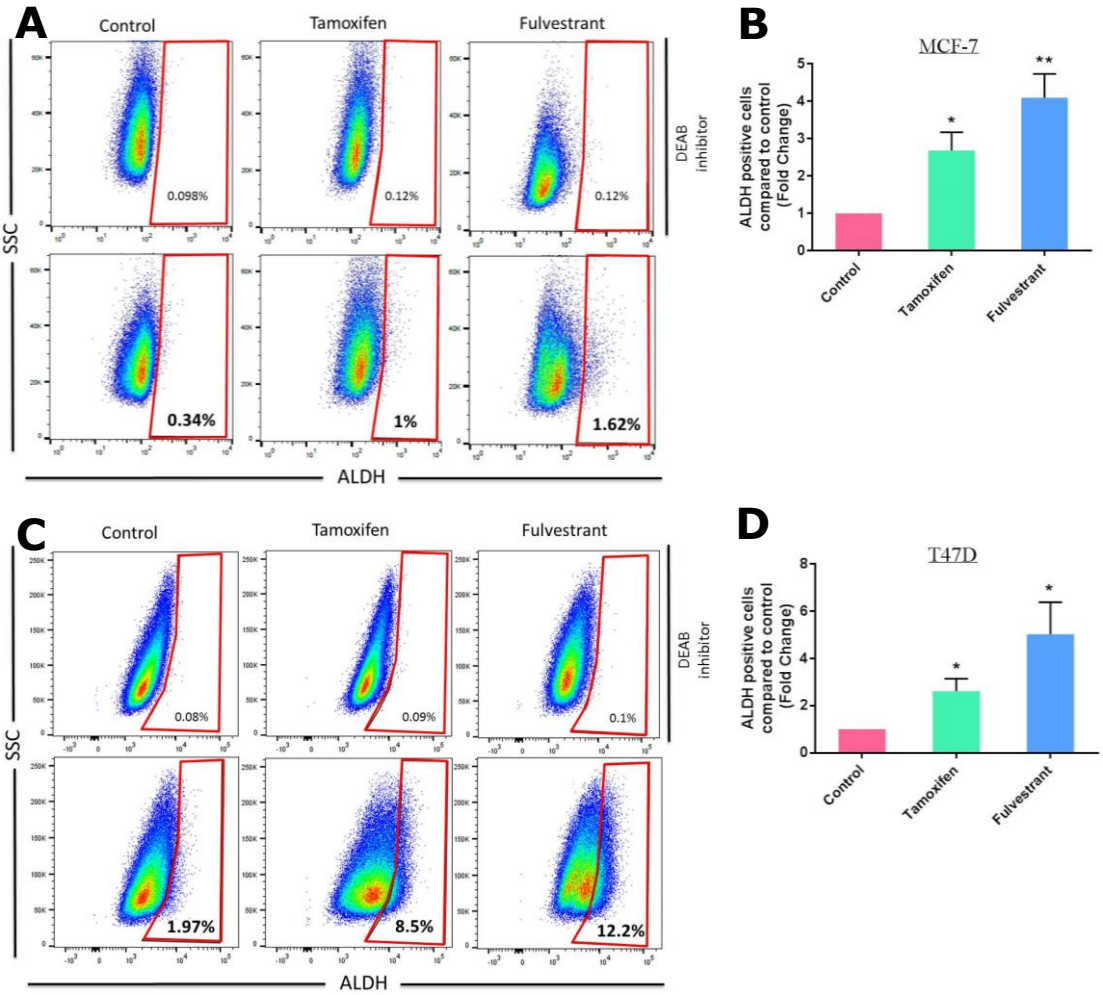


Figure 3. 4 Percentage of ALDH pos cells after Tam and Fulv treatment, measured by the Aldefluor assay.
A) Representative FACS plot of MCF-7 cells in the presence of control, Tam (1 μM) or Fulv (0.1 μM) for 6 days. ALDH pos cells (outlined in red) were discriminated from ALDH neg cells (cells outside the red box) using the ALDH inhibitor DEAB (top panel). **B)** Bar chart represents FC of the percentage of ALDH pos cells versus control-treated ALDH neg cells in MCF-7 cells following control (pink), Tam (green) and Fulv (blue) treatment. Figure illustrates averaged data from eight independent experiments and standard errors. **C)** Representative FACS plot of T47D cells after control, Tam (1 μM) or Fulv (0.1 μM) treatment. **D)** Bar chart represents FC of the percentage of ALDH pos cells versus control-treated ALDH neg cells for each condition in T47D cells. Figure shows average and SEM from five independent experiments. * P≤0.05, **P≤0.01 (T-test).

The effects of Tam and Fulv on the mesenchymal $CD44^+/CD24^-$ population were also assessed in the two ER+ cell lines used in this study, however the percentage of $CD44^+/CD24^-$ cells dropped following anti-estrogen treatment (Figure 3.5).

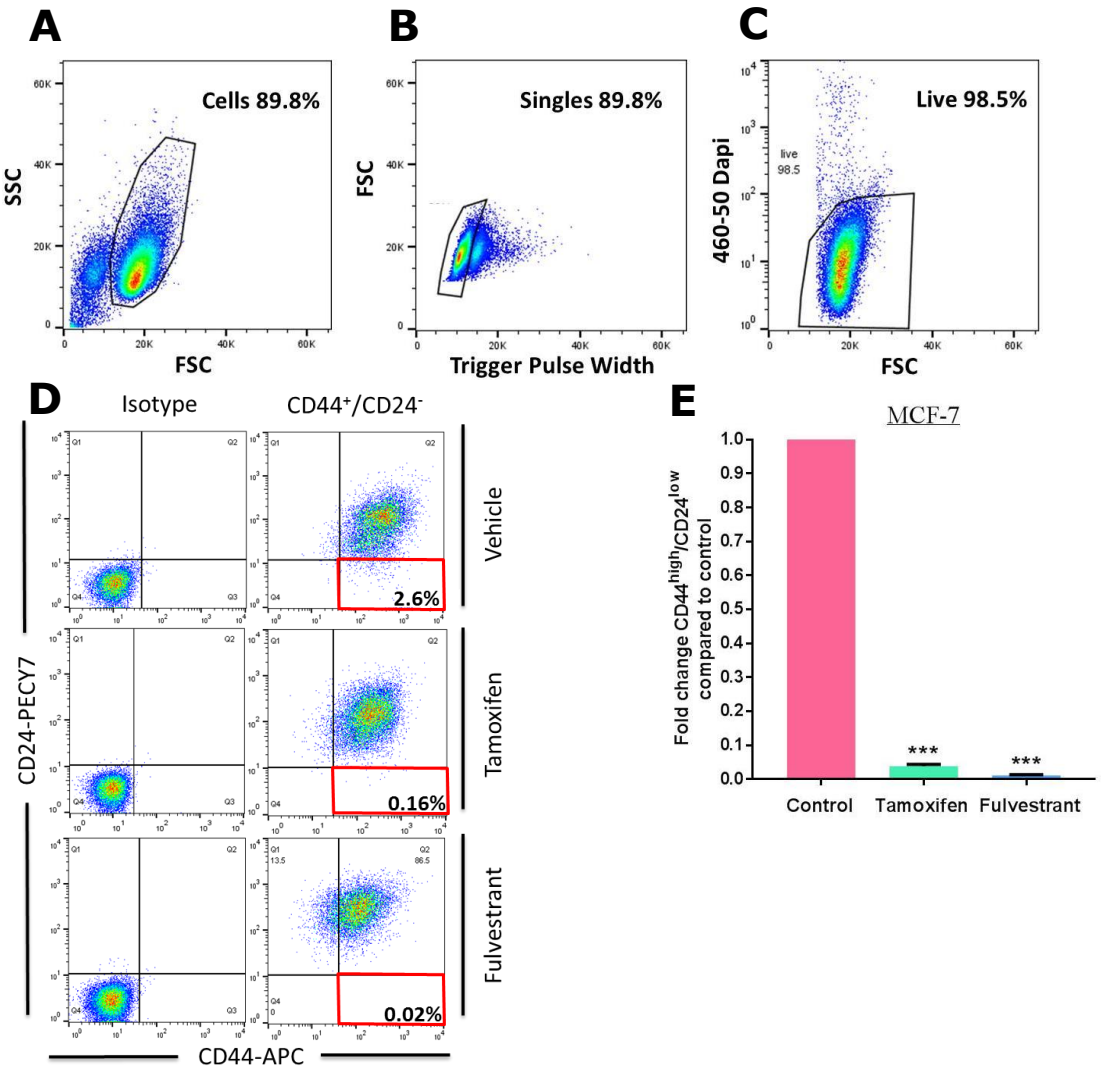


Figure 3. 5 Percentage of $CD44^{+}/CD24^{-}$ cells after anti-estrogen treatment.
MCF-7 representative FACS plots show A) conventional 488 nm side scatter and forward scatter profile to exclude debris, B) forward scatter and a trigger pulse width to remove doublets and C) Dapi and forward scatter to exclude dead cells. D) Representative FACS plot illustrates the $CD44^{+}/CD24^{-}$ population (red box) in MCF-7 cells in response to the vehicle control (top), Tam (middle) and Fulv (bottom) treatment for 6 days. MCF-7 cells were dual labelled with pre-conjugated antibodies to CD44 (APC) and CD24 (PE-CY7) and with the viability dye DAPI. Isotype control, represented on the left, was used to correct for non-specific binding. E) Bar chart shows FC of $CD44^{+}/CD24^{-}$ cells after Tam (green) and Fulv (blue) treatment compared to ethanol control (pink). ***P≤0.001

A recent publication showed that cells able to accumulate the fluorescent-vitamin riboflavin in intracellular vesicles are enriched for

CSCs in several solid tumours (Miranda-Lorenzo et al. 2014). Therefore, we implemented this technique to assess the effects of anti-estrogen therapy in the autofluorescent CSC population in breast cancer cell lines. Anti-estrogen treatment increased the autofluorescent CSC content by two fold compared to ethanol control in MCF-7 cells, whereas no enrichment was found in the other cell line tested, T47D cells (Figure 3.6).

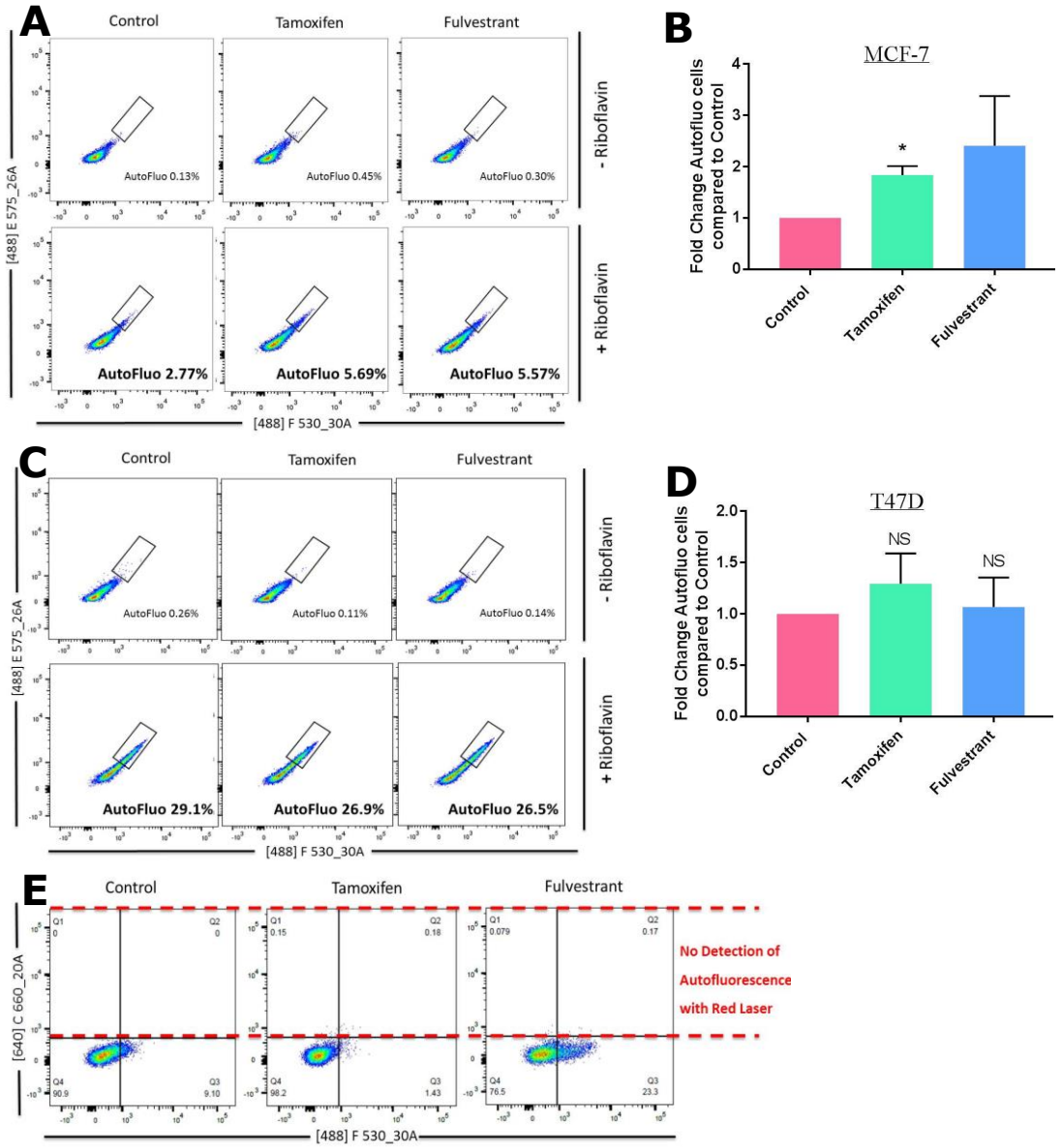


Figure 3. 6 Percentage of autofluorescent cells (AutoFluo) after anti-estrogen treatment.

A) FACS analysis of MCF-7 cells after 6 day treatment with control (left), Tam (middle) or Fulv (right) and supplemented with (bottom panel) or without (top panel) riboflavin (30 µM) for 2 h prior FACS analysis. Cells were gated to 0.1% in control group without Riboflavin. Autofluorescent cells were excited using a 488 nm blue laser source and 530/30 filters. **B)** Bar chart shows FC of autofluorescent cell content in Tam (blue) and Fulv (green) treated cells compared to ethanol control (pink) in MCF-7 cells. Figure shows data from six independent experiments. **C)** FACS analysis of T47D cells after 6 day treatment with control (left), Tam (middle) or Fulv (right) **D)** Bar chart shows FC of autofluorescent cell content in Tam (green) and Fulv (blue) treated cells compared to ethanol control (pink). Figure shows data from 9 independent experiments. **E)** Representative cytometry plot showing no excitation of autofluorescence with 640 nm red laser. * P≤0.05 (T-test).

Taking into account that ALDH pos cells were the only population increased following Tam and Fulv treatment in both cell lines, the expression of the two isoforms reported to be important in breast cancer, ALDH1A1 and ALDH1A3, was also measured. The mRNA expression of the ALDH1A3 isoform was studied in the two ER+ cell lines, MCF-7 and T47D cells using qPCR. As shown in Figure 3.7, ALDH1A3 expression in MCF-7 cells rose notably by 4.2 and 12.5 fold after Tam and Fulv treatment respectively, compared to ethanol control. A similar increase was observed in T47D cells, where ALDH1A3 expression significantly increased by 5.2 fold, with Tam, and 5 fold after Fulv treatment, in comparison to control. The expression of the other well-known isoform, ALDH1A1, was also tested in MCF-7 cells, however no expression was found (data not shown).

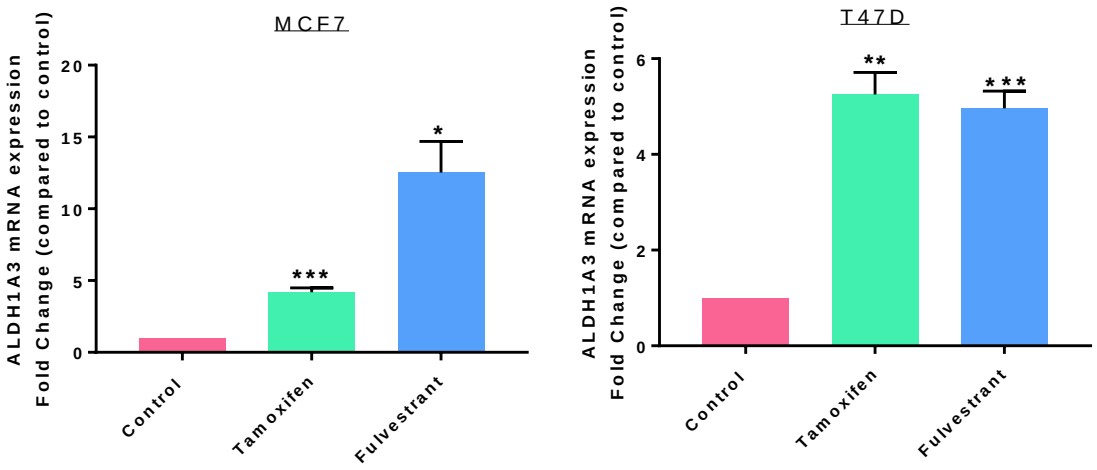


Figure 3. 7 ALDH1A3 mRNA expression after anti-estrogen treatment in MCF-7 and T47D cells.
Cells were treated in monolayer for 6 days with either ethanol control (pink), Tam (green) or Fulv (blue). mRNA was then extracted as indicated in chapter 2 of materials and methods. Control expression was set at 1 to represent the FC. Gene expression was normalised to the averaged expression of the two housekeeping genes SDHA and PGK1. Figure shows average and SEM from three independent experiments. * P<0.05 **P<0.01 ***P<0.001

3.2.3 Stable ALDH1A3 Knockdown by lentiviral shRNA

The ALDH1A3 isoform has been shown to be important for the activity of breast CSCs (Marcato et al. 2011). Moreover, as shown above, ALDH1A3 mRNA expression increases after Tam and Fulv therapy in bulk MCF-7 and T47D cells. Therefore, we sought to further establish the importance of ALDH1A3 in the anti-estrogen-mediated enrichment of ALDH pos cells in breast cancer cell lines. This was achieved by using lentiviral-mediated stable transduction to inducibly knockdown ALDH1A3 (ALDH1A3KD).

The plasmid construct used to knock down ALDH1A3 (refer to Appendix 8.1), employs a tetracycline-inducible promoter (TRE), which in the presence of DOX, drives the expression of a Turbo-RFP reporter. Three different shRNA constructs were used (V1, V4 and V5) and an empty TRIPZ vector was also utilised to allow for accurate interpretation of knockdown results. Lentiviral production was carried out with the packaging cell line HEK 293-T cells and viral particles were used to infect MCF-7 and also T47D cells (data not shown). After Puromycin selection, cells were incubated with DOX for 72h followed by microscopic examination of turbo-RFP expression to assess transduction efficiency.

Figure 3.8 shows images of turbo-RFP positive cells and flow cytometry plots illustrating the proportion of turbo-RFP positive cells for one of the shRNA constructs, V5. Overall, the virus-mediated gene transfer was successful and the transduction efficiency for V5 (measured as percentage of turbo-RFP positive cells) was 24.9%.

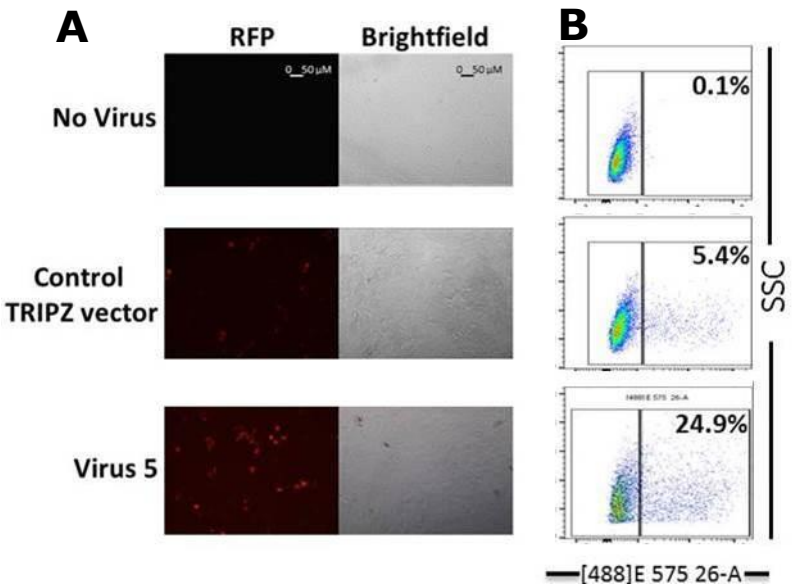


Figure 3. 8 Generation of a MCF-7 stable cell line knocking down ALDH1A3.
293T cells were transfected with viral plasmids and with second generation TRIPZ Inducible Lentiviral shRNA containing a turbo-RFP reporter to create lentiviral particles, which were used to infect MCF-7 cells to reduce ALDH1A3 expression. **A)** After Puromycin selection and upon addition of DOX for 72h, MCF-7 cells were photographed with the bright field (right panel) and with the UV lamp to visualise the turbo-RFP positive cells (left panel). **B)** Flow cytometry plots showing percentage of RFP-positive cells in untransfected (top) control TRIPZ empty vector (middle) and V5 shRNA constructs (bottom) after 72h of DOX exposure.

In order to have a pure turbo-RFP positive population, infected MCF-7 cells were sorted after Puromycin selection and DOX treatment for 14 days (Figure 3.9 A). Cells cultured after sorting were confirmed to express the red protein under the microscope (Figure 3.9 B).

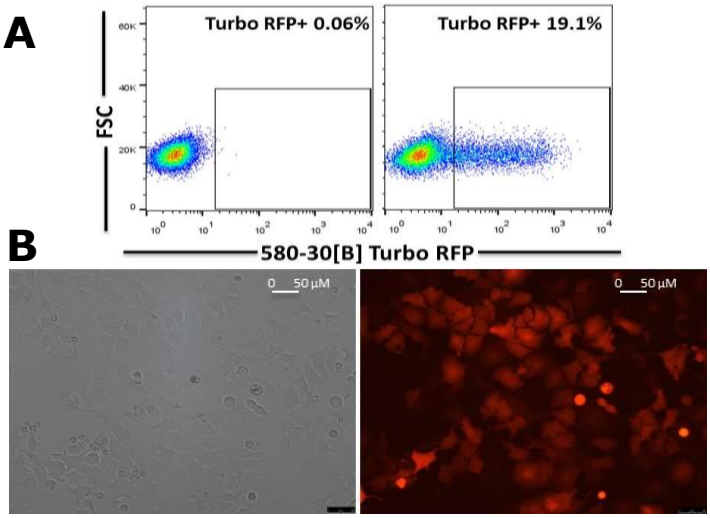


Figure 3. 9 Sorting Turbo-RFP positive cells.
Representative photo of MCF-7 cells transduced with V5. After Puromycin selection of infected MCF-7 cells, turbo-RFP positive cells were sorted following DOX administration for 14 days. **A)** Flow cytometry plots of ALDH1A3KD cells without DOX (left) and with DOX (right) exposure. **B**) Images show turbo-RFP positive cells in brightfield and red fluorescence after cell sorting.

The efficient downregulation of ALDH1A3 mRNA levels was assessed by qPCR (represented in pink in Figure 3.10). ALDH1A3 expression was reduced by 57.4 % in MCF-7 cells transduced with V5 and treated with DOX. Several attempts were performed to confirm downregulation of the target gene at the protein level, however they were unsuccessful, due to poor quality of the antibody used and/or low expression of the ALDH1A3 protein in the non-transfected MCF-7 cells.

In order to assess whether the anti-estrogen-mediated ALDH1A3 mRNA increase was obliterated in the knock-down cells, ALDH1A3KD cells were treated for 6 days with anti-estrogen therapy, followed by qPCR analysis of the ALDH1A3 expression (Figure 3.10). ALDH1A3 mRNA expression significantly increased after Tam (green) and Fulv (blue) treatments in the absence of DOX, which supports the results shown previously in figure 3.5. However, after doxycycline addition (+DOX), ALDH1A3 expression grew 2.6 and 3.1 fold less following Tam and Fulv treatment, compared to no-DOX cells (Figure 3.10).

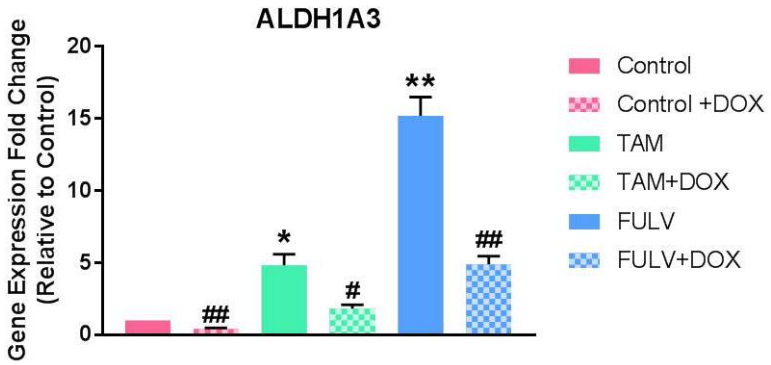


Figure 3. 10 ALDH1A3 expression in ALDH1A3KD cells after 6 days treatment.
Bar chart shows mRNA expression by qPCR of the ALDH1A3 isoform following control (pink), Tam (green) and Fulv (blue) treatment for 6 days in MCF-7 cells. Data was normalised with the expression of two housekeeping genes. Average and SEM from three independent experiments is shown. *P<0.05 **P<0.01 ***P<0.001 (no DOX); ♯ P<0.05 ♯♯P<0.01 (+ DOX).

3.2.4 ALDH1A3 isoform mediates the stem cell activity enrichment observed after anti-oestrogen treatment in MCF-7 cells.

Next, a 6-day *in vitro* treatment with the anti-estrogen drugs Tam and Fulv was performed on the ALDH1A3KD cells to assess ALDH pos cell content, as shown in figure 3.11. In the ALDH1A3KD cells, which were not exposed to DOX (- DOX), percentage of ALDH pos cells increased by 3.4 and 6.8 fold after Tam and Fulv therapy respectively. These results are in line with the ones shown in figure 3.7. However, in the presence of doxycycline (+DOX), ALDH1A3 downregulation prevented an increase in the percentage of ALDH pos cells after anti-estrogen treatment. These results highlight the importance of the ALDH1A3 isoform mediating the enrichment of the ALDH pos cells after anti-estrogen therapy in MCF-7 cells. Nonetheless, the ALDH pos population was not completely depleted, which could be due to a compensatory effect from the other relevant isoforms. The expression of ALDH1A1, ALDH1A2 and ALDH8A1 was then tested by qPCR and no expression was found before or after DOX exposure (data not shown).

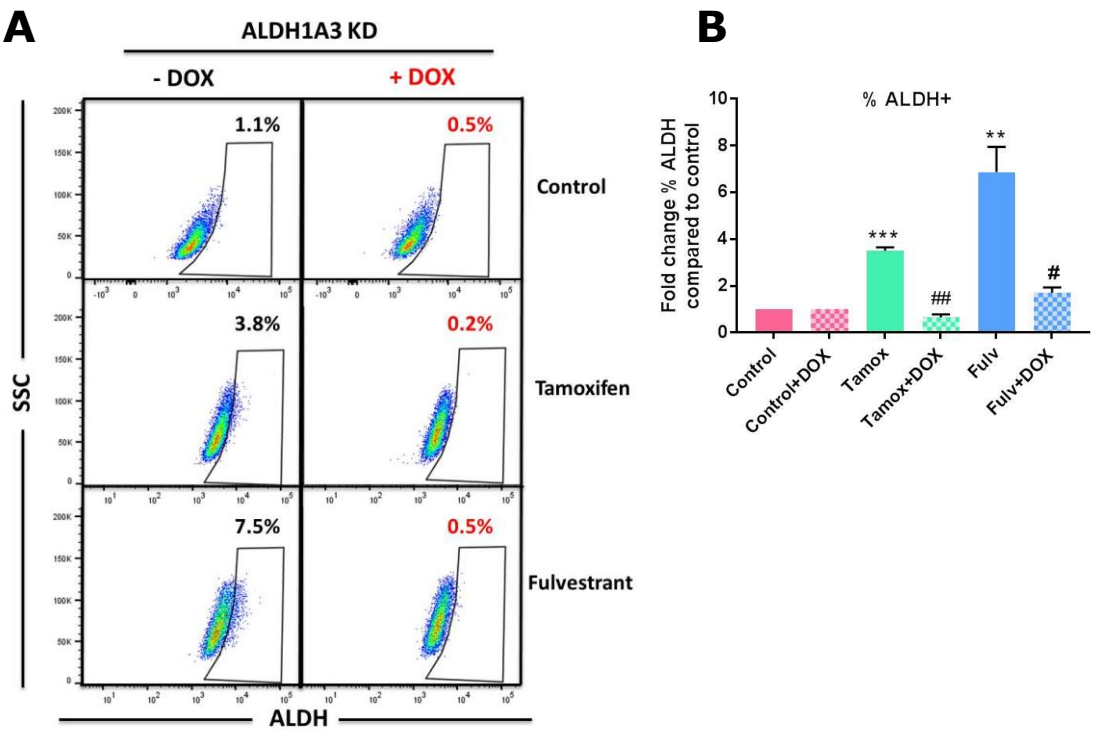


Figure 3. 11 Aldefluor assay in DOX-inducible ALDH1A3KD cells after 6-days in vitro treatment.

A) Representative FACS plots showing percentage of ALDH pos cells in the ALDH1A3KD cells. Stable ALDH1A3 shRNA expressing MCF-7 cells were not treated (left panel) or treated with DOX (right panel), alongside with control (top), Tam (middle) or Fulv (bottom) for 6 days followed by Flow cytometric analysis. Graph shows data from MCF-7 cells transfected with virus 5 only. **B**) ALDH pos FC of Tam (green) and Fulv-treated (blue) cells, compared to control (pink) with (squared fill) or without DOX (solid fill) treatment. Bar chart shows average and SEM of four independent experiments **P<0.01 ***P<0.001 (no DOX); ♯ P<0.05 ♯♯P<0.01 (+ DOX).

3.2.5 Characterising anti-estrogen resistant ALDH positive cells

ALDH pos cells are the only CSC population (out of the three tested in the current study) that is consistently enriched after anti-estrogen treatment in the ER+ cell lines MCF-7 and T47D. It has been published that ALDH pos cells, isolated from breast tumours, display stem cell properties, such as higher MFE and greater *in vivo* tumour formation compared to ALDH neg cells. These experiments were performed on ER+ and ER- Patient Derived Xenografts (PDXs) (Ginestier et al. 2007) as well as in triple negative cell lines (Charafe-Jauffret et al. 2009) (Charafe-Jauffret et al. 2010), however experiments using ER+ cell lines, such as MCF-7 or T47D are missing from the literature. Therefore, the MFE

between ALDH pos and ALDH neg cells was tested using MCF-7 cells. Figure 3.12 shows the gating strategy used to sort both populations following the Aldefluor assay. The DEAB inhibitor (left) is used to inhibit the enzymatic reaction and help us to account for background fluorescence and to perform the right gating for the ALDH pos cells. In order to sort a pure non-CSC population, the bottom 30% of the ALDH neg cells (purple box) was sorted, as indicated in the figure below.

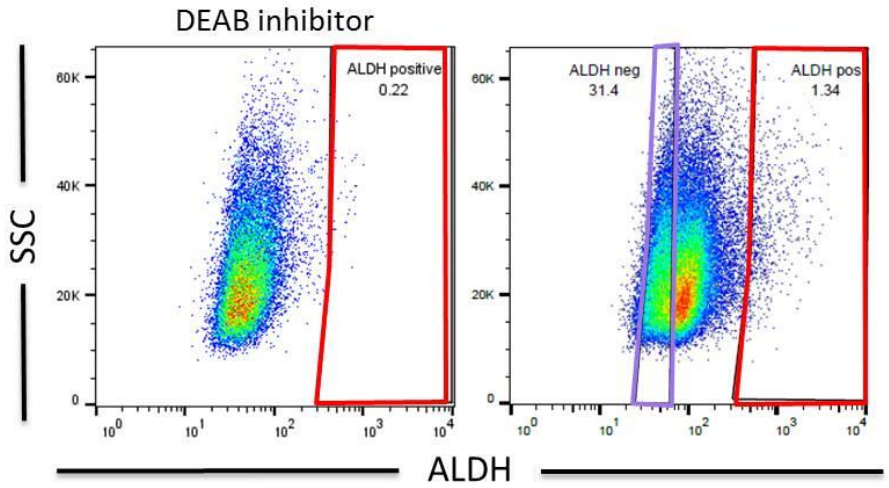


Figure 3. 12 ALDEFLUOR gating strategy.
FACS plot shows how gating was performed to isolate ALDH neg and pos cells in MCF-7 and PDS. Cells outlined by the red box show the ALDH pos cells, which were gated based on the DEAB inhibitor. Purple box highlights the bottom 30% of the ALDH neg population.

3.2.5.1 MFE of ALDH pos and ALDH neg cells in MCF7 cells

The MFE between ALDH pos and neg cells was tested using MCF-7 cells. Findings, illustrated in Figure 3.13, show that the MFE of ALDH pos cells is slightly greater than the one observed in ALDH neg cells, however these results are not statistically significant.

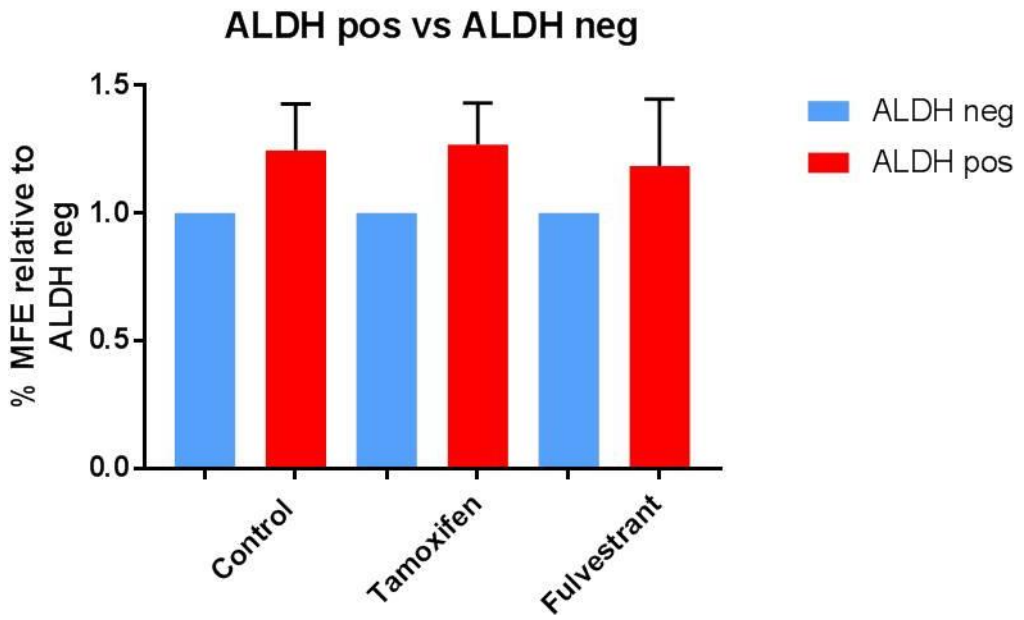


Figure 3. 13 MFE of ALDH pos and ALDH neg cells in MCF-7 cells. Cells were treated in monolayer for 6 days with the anti-estrogen drugs Tam and Fulv, as well as with the ethanol control. Treated ALDH pos and neg cells were then sorted, based on the ALDH activity, into eppendorfs containing 50 µl 2% FBS in HBSS. Sorted cells were counted using Trypan blue and 5,000 cells were seeded per well into low-attachment plates containing mammosphere media for 7 days. Mammospheres were then counted using a microscope. Bar chart shows averaged data from six independent experiments and SEM.

3.2.5.2 In vivo tumour formation capacity of ALDH pos and ALDH neg MCF7 cells.

The next question we wanted to address was whether ALDH pos cells derived from the ER+ cell line, MCF-7, showed higher *in vivo* tumour formation capabilities than the ALDH neg cells. Moreover, the tumour development capacity of ALDH pos vs. ALDH neg cells following anti-estrogen treatment was also tested in the *in vivo* experiment. NSG mice (n=4 per condition) were injected with 1000, 100 or 10 ALDH neg (left flank) or ALDH pos (right flank) MCF-7 cells, which had been treated *in vitro* in monolayer with ethanol control, Tam or Fulv for 6 days prior cell sorting. Cell doses were chosen based on our group's previous publication, where MCF-7-derived tumour growth was detected with as few as 100 cells (Simões et al. 2015). ALDH pos and ALDH neg cells were inoculated in different flanks of the same mouse to minimise variability of

the tumour growth rate due to engraftment in different animals. Mice were implanted with E_2 pellets (0.18 mg E_2 90 days slow release) for 7 days before cell implantation. After 90 days, to maintain E_2 levels, E_2 was administered to the drinking water (8 µg/ml). Tumours derived from 100 sorted MCF-7 cells (>200mm^3) grew by week 17 in all three conditions (Figure 3.14) (data for 10 and 1000 cells not shown). In the control group, ALDH pos and neg cell-derived tumours show comparable growth rates until week 18, when ALDH pos cells started forming slightly bigger tumours. In contrast, tumours from ALDH pos Tam-treated cells grew significantly larger compared to tumours from ALDH neg Tam-treated cells week 16. Similarly, differences between Fulv-treated ALDH neg and pos cells were noticeable from early stages of the *in vivo* experiment as shown in the figure below, where ALDH pos cells grew significantly bigger tumours from week 13 post-implantation. At week 20, ALDH pos cells grew bigger tumours compared to ALDH neg cells in the three conditions tested, however only in the Tam and Fulv treated groups were such differences statistically significant. Injection of 1000, 100 and 10 ALDH pos cells allowed the development of tumours in a greater number of mice, compared to the same cell dilutions of the ALDH neg cell population in all three treatment conditions tested.

The Extreme limiting dilution analysis (ELDA), an online application which enables limiting dilution analysis in stem cell assays (Hu et al. 2009), was used to assess the tumour initiating cell frequency between ALDH pos and ALDH neg cells after control, Tam and Fulv treatment, as indicated in Table 3.1. These results show a 3.8, 2.9 and 5.7-fold enrichment in tumour initiating cells in the ALDH pos population compared to the ALDH neg, in the control, Tam and Fulv group respectively. A chi-squared test (X^2) revealed statistical significance between the tumour initiating cell frequency of ALDH pos and ALDH neg cells in the control and Fulv group ($p \leq 0.05$).

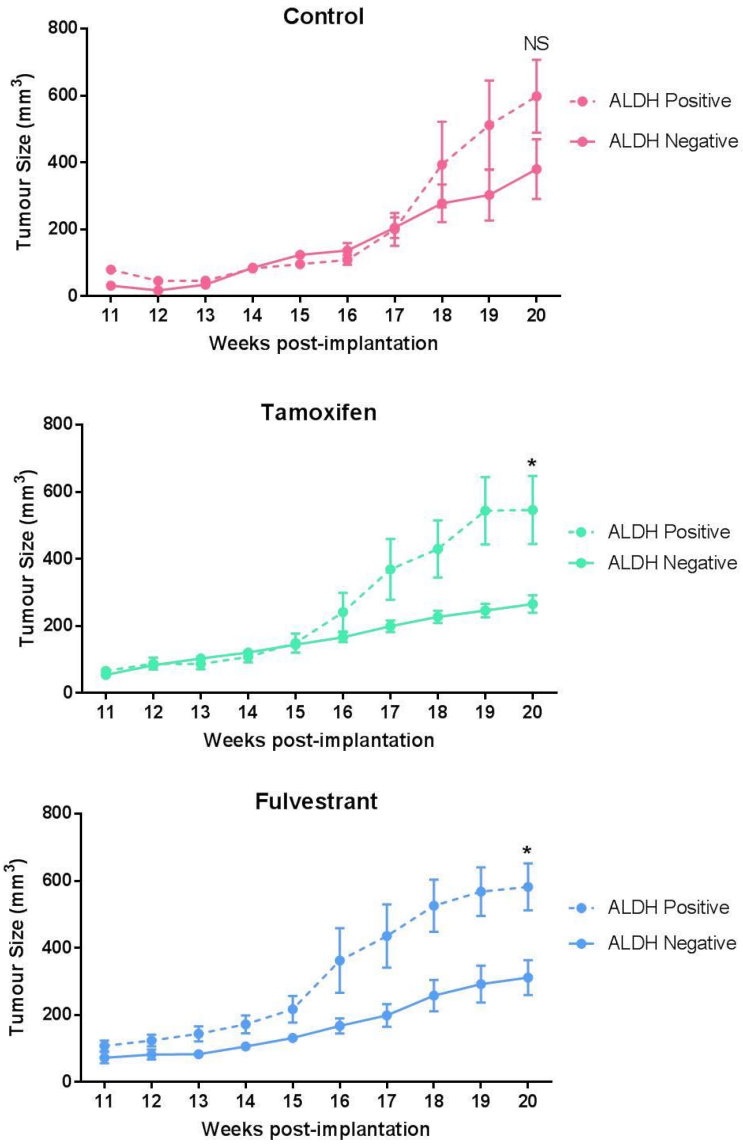


Figure 3. 14 *In vivo* tumour growth from 100 ALDH neg vs. ALDH pos MCF-7 cells after control, Tam and Fulv treatment.

MCF-7 cells were *in vitro* pre-treated for 6 days with control (ethanol), Tam (1 μM) or Fulv (0.1 μM) followed by the Aldefluor assay. ALDH neg and ALDH pos cells were FACS sorted and then counted using Trypan blue. 100 ALDH neg and ALDH pos cells were engrafted into the left and right flank of the same mice, respectively. Averaged tumour growths from control (pink; top panel), Tam (green; middle panel) or Fulv-treated (blue; bottom panel) ALDH neg (straight line) and ALDH pos (dash-dotted line) cells across three different weekly timepoints are represented. Female IL2γR NSG mice carried 90-day slow-release 0.18mg 17β-estradiol pellets. After 90 days, mice were given E_2 in the drinking water (8 μg/ml). Mice tumours were measured three times a week. * $P \leq 0.05$ (two tail, two sample equal variance T-test). Number of mice=4 apart from Control-treated mice (n=3 all time points due to the mouse dying at very early stages of the experiment). Data shows SEM.

Table 3. 1 Extreme Limiting Dilution Analysis (ELDA) of ALDH pos vs. ALDH neg cells following anti-estrogen treatment.

		MCF-7 cells pre-treatment groups					
		Control (Ethanol)		Tam		Fulv	
		ALDH neg	ALDH pos	ALDH neg	ALDH pos	ALDH neg	ALDH pos
Cell Number	$1x10^3$	2/4	3/4	2/4	3/4	0/3	2/3
	$1x10^2$	1/3	3/3	1/4	2/4	2/4	4/4
	10	0/4	1/4	0/4	1/4	0/4	0/4
Tumour Initiating Cell Frequency (95% CI)		1:1,044	**1:274**	1:1,069	**1:365**	1:1,670	**1:291**
P value compared to ALDH neg (X^2)		N/A	**0.05**	N/A	0.14	N/A	**0.02**

Only tumours bigger than 300 mm^3 were considered for this analysis. Data shown at week 20. N=3 in some groups due to some unexpected mouse deaths at early stages of the experiment.

3.2.5.3 MFE of ALDH pos vs. ALDH neg cells in Patient-derived samples

Next, the MFE between ALDH pos and ALDH neg cells was investigated using metastatic pleural effusion patient samples collected from the Michigan Biobank, USA, and the Manchester Biobank, UK. Samples were not further treated *in vitro* with Tam or Fulv, as patients had received previous chemo and hormonal therapy (refer to materials and methods Tables 2.7 and 2.8 for clinic-pathological information), to preserve their original characteristics. Figure 3.15 illustrates the distinct percentage of ALDH pos cells for each patient sample, which can range from as little as 0.32% in patient 66 to 27.3% in patient 56, highlighting inter-patient variability. As expected, MFE from ALDH pos cells was greater than the ALDH neg cells in 8 out of 9 patient samples (Figure 3.16 A). Overall, ALDH pos cells formed almost 4 times more mammospheres than the ALDH neg cells, which was highly statistical significant ($p<0.001$) (Figure 3.16 B). The morphology and the size of the mammospheres differed from patient to patient, as illustrated in the two

representative examples of patient-derived mammospheres shown in Figure 3.16 C.

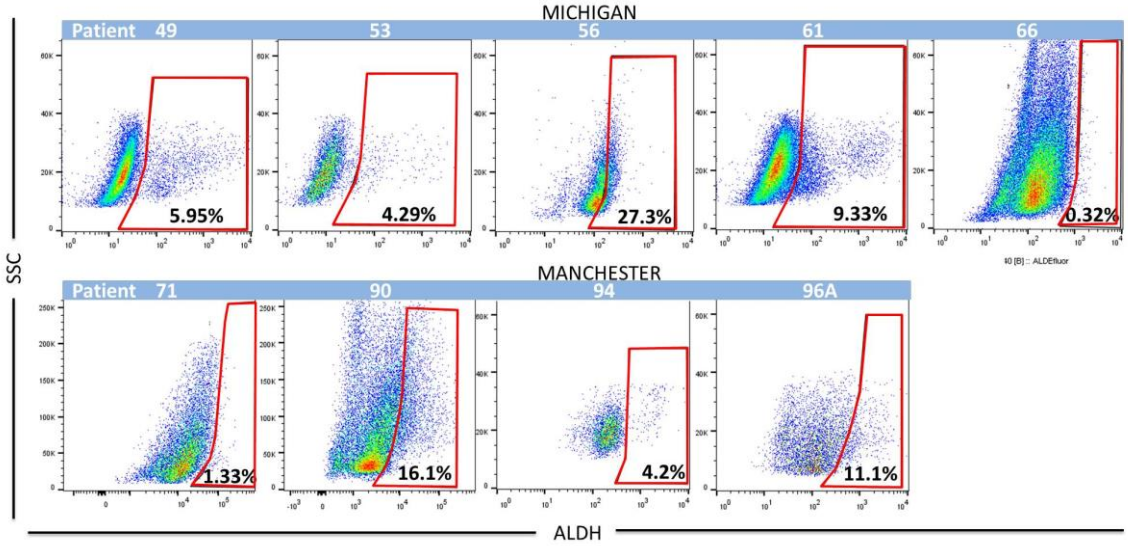


Figure 3. 15 FACS plots showing percentage of ALDH pos cells, measured by the Aldefluor assay, in metastatic patient samples. Cells were stained according to the manufacturer's guidelines. ALDH pos cells (red box) were discriminated from the ALDH neg cells using the DEAB inhibitor (DEAB plot not shown). PDS from Michigan's biobank (top) and from Manchester's biobank (bottom) are shown.

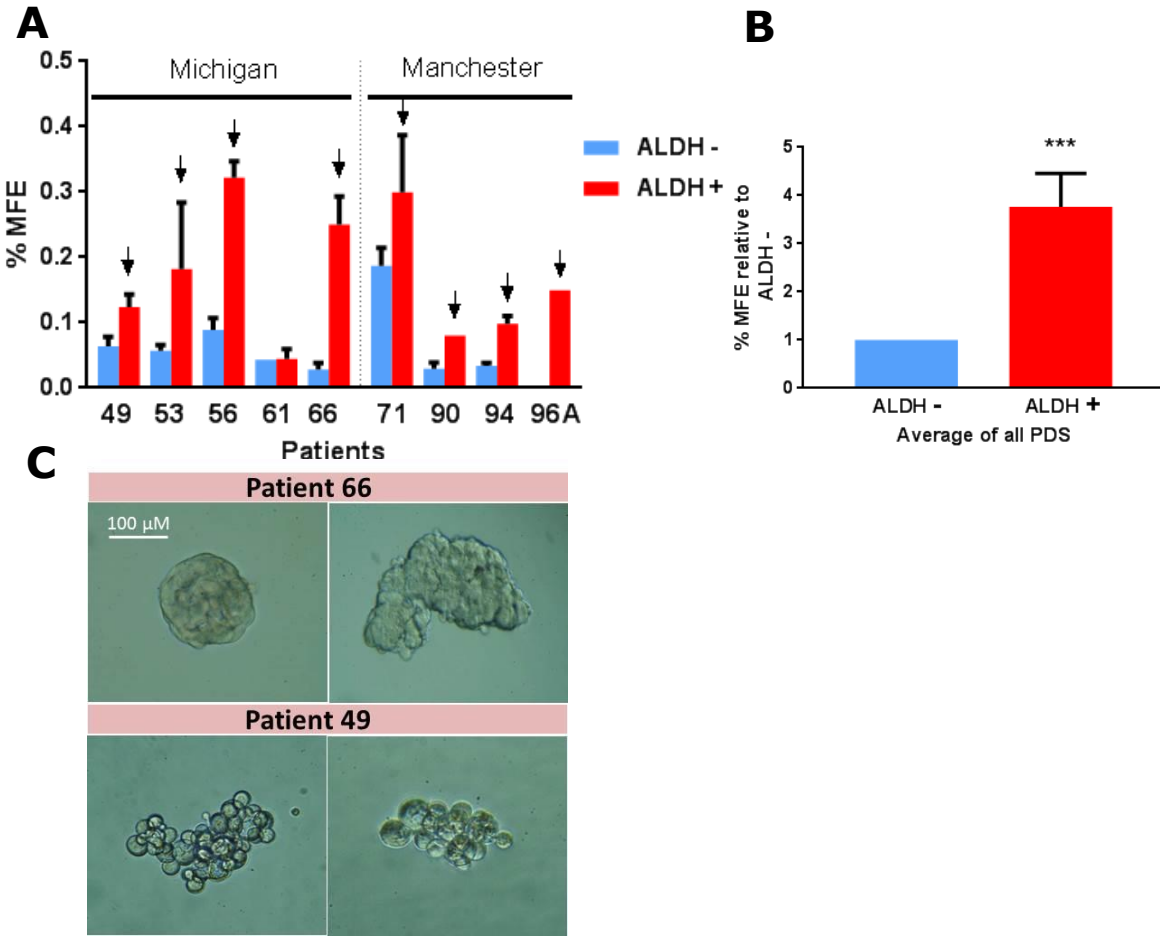


Figure 3. 16 MFE between ALDH pos and ALDH neg cells from PDS.
A) Metastatic pleural effusion samples from ER+ breast cancer patients were processed using density gradient to exclude leukocyte contamination. ALDH pos and ALDH neg cells, isolated with the Aldefluor assay, were sorted, counted and seeded into low-attachment 6 well plates containing mammosphere media at a cell density of 5,000 cells per well. Number of mammospheres was counted after 7 days using a microscope. **B)** FC of the percentage of the MFE between ALDH pos and ALDH neg cells across 9 different PDS. *** $P<0.001$. Bar chart shows SEM. **C)** Representative mammosphere photos from patient 66 and 49 using a microscope. Scale bar = 100 µM.

3.3 Discussion

It was observed that anti-estrogen therapy enriches for stem cell activity as 6-day *in vitro* treatment with Tam and Fulv leads to an increase in the MFE of MCF-7 cells (Figure 3.3; Tam: p=0.024, Fulv p=0.027). Different markers were interrogated to assess whether Tam and Fulv also enrich for CSCs, as measured by flow cytometry. It was found that anti-estrogen treatments mediate a modest increase in the autofluorescent cell content (Figure 3.6) and a drastic loss of the $CD44^{+}/CD24^{-}$ population (Figure 3.5). Tam and Fulv significantly raised the ALDH pos cell compartment, which according to the literature is a proliferative population with an epithelial-like phenotype (Liu et al. 2014). One potential explanation for the opposing effects exerted by anti-estrogen treatment in the three different CSC populations tested might be due to differences in cell confluency before and after anti-estrogen treatment. It has previously been reported that cell density affects the proportion of CSCs in colon cancer cell lines (Opdenaker et al. 2015). Moreover, it has also been shown that anti-estrogen treatment leads to an increase in the expression of the CD24 surface protein which may be the reason why the proportion of CD44+/CD24- cells drop after 6 days treatment with Tam and Fulv (Leung et al. 2017).

ALDH1A3 has been reported as the isoform responsible for ALDH activity in breast cancer cell lines (Marcato et al. 2011). Moreover, in the present work it was observed that anti-estrogen treatment increases ALDH1A3 mRNA levels in MCF-7 and T47D cells, therefore ALDH1A3 expression was genetically downregulated in MCF-7 cells to further explore the contribution of this isoform to the increase of ALDH pos cells after anti-estrogen treatment. Firstly, siRNA was used to transiently decrease ALDH1A3 expression in MCF-7 cells, however our 6-day anti-estrogen *in vitro* treatment approach made it too challenging to maintain low expression of the isoform for such long period of time (data not shown). Therefore, ALDH1A3 expression was stably downregulated with a lentiviral doxycycline-inducible shRNA system. Successful ALDH1A3 knockdown was achieved, however 42% ALDH1A3 mRNA expression was retained in MCF-7 cells transduced with V5. Therefore, an alternative

approach, which completely abolished target gene expression, such as CRISPR/Cas9, would likely be a better choice. We were unable to verify downregulation of ALDH1A3 at the protein level due to low expressing levels and poor antibody quality. Taking advantage of the ALDH1A3KD cells, it was possible to validate the importance of the ALDH1A3 isoform in driving the increase in the ALDH pos population after anti-estrogen treatment in MCF-7 cells, since such increase did not take place in the presence of DOX. This indicates that ALDH1A3 protein levels were reduced by shRNA knock-down.

Aldefluor-positive cell sorted MCF-7 cells were tested for their MFE and compared to that of ALDH neg cells (Figure 3.13). ALDH pos cells showed a clear trend with increased capacity to form more mammospheres than ALDH neg cells but results were not statistically significant. It is possible that the apparent lack of correlation is due to the dynamic state of ALDH pos cells, which are ready to transition to ALDH neg cells and vice versa depending on external stimuli and the microenvironment. For instance, when ALDH neg cells are sorted and then plated for 6 days to form mammospheres, these cells might de-differentiate to a cancer stem-like state to maintain a balanced equilibrium between non-CSCs and CSCs and to sustain cancer cell growth. This hypothesis was confirmed when testing whether the increase in the ALDH pos population, mediated by anti-estrogen treatment, was due to enrichment, by selectively killing the ALDH neg cells, or induction (data not shown). ALDH neg MCF-7 cells were sorted and treated in monolayer for 6 days with control, Tam and Fulv, followed by Flow cytometry analysis of the ALDH pos content (data not shown). However, even in the control-treated ALDH neg group, a small but consistent population of ALDH pos cells could be detected. These results suggest that Tam and Fulv might activate previously dormant ALDH pos cells, in the absence of CSCs in culture, or that these anti-estrogen agents induce a CSC phenotype increasing the proportion of ALDH pos cells in MCF-7 cells. Growing secondary mammospheres instead of primary mammospheres, as a way to assess ALDH pos self-renewal capabilities, might have shown a clearer correlation with the Aldefluor activity.

The ability of ALDH pos breast cancer cells to grow *in vivo* has been found to be much greater than the one shown by ALDH neg cells, however published experiments have focused only on triple negative subtypes (Charafe-Jauffret et al. 2010). This is due to triple negative breast cancer cells forming tumours with much greater capacity than ER+ tumours. Therefore, the present study conducted an *in vivo* limiting dilution transplantation of MCF-7 cells and found that 100 ALDH pos cells have greater capacity to form larger tumours *in vivo*, when compared to 100 ALDH neg cells (Figure 3.14). These data supports previously published results where ALDH pos cells from normal breast and triple negative breast tumours showed enhanced tumour initiating capabilities.

ALDH pos cells pre-treated with Tam and Fulv developed much bigger tumours than ALDH neg cells, which were also treated with the same drugs prior to cell sorting. These results indicate that anti-estrogen treatments enhance the *in vivo* cell proliferation capabilities of the ALDH pos population. Surprisingly, 10 ALDH pos Fulv-treated cells did not form any tumours, whereas 10 ALDH pos control-treated and 10 ALDH pos Tam-treated cells did form tumours. This might be due to the fact that Fulv exerts its anti-proliferative effects on tumour cells, which delays tumour growth considerably and so the *in vivo* experiment was terminated before any growth was seen. Another unanticipated finding was the fact that 1,000 ALDH neg Fulv-treated cells did not develop any tumours when injected in mice, whereas 100 ALDH neg Fulv-treated cells formed tumours in 2 out of 4 mice (Table 3.1). The reason for this rather contradictory result is still not entirely understood, however it is possible to argue that more mice per condition were needed in order to find out whether that was a meaningful biological result or an artefact instead.

ALDH pos cells from the control group formed bigger tumours than ALDH neg cells, however these results were not statistically significant (Figure 3.14). This is due to large variation in the data between mice of the same group, which might blur any real biological differences. Delivery of the cells via subcutaneous injection might have introduced bias into data collection. Filtration of part of the 0.2 ml of the cell-suspending medium into the surrounding external tissue might have account for differences in the injected volume between the four different mice. Flow

cytometry cell sorting is not 100% efficient as it can be seen in the purity check FACS plot shown in Appendix figure 8.2. Therefore, given the small amount of ALDH pos cells in the control group (0.34%) and the close proximity to the ALDH neg population in the Flow plots, it might be that some ALDH neg cells were sorted as pos, diluting out the tumorigenic potential of the latter. ELDA analysis revealed a 3.8, 2.9 and 5.6-fold enrichment in cell tumour-initiating cell frequency in ALDH pos cells following control, Tam and Fulv treatment respectively ($p=0.05$, $p=0.14$, $p=0.02$), when compared to ALDH neg cells. These results suggest that short-term Fulv treatment might selectively enrich for more tumorigenic ALDH pos MCF-7 cells.

In order to better apply our findings to the clinic, patient samples were used in our ALDH pos cells characterisation study and found a great inter-patient variability in the percentage of ALDH pos cells, identified with the aldefluor assay. Flow cytometric analysis revealed an ALDH pos population ranging from as little as 0.3% in patient 66 to 27.3% in patient 56 (Figure 3.15). ALDH pos cells had significantly higher MFE than ALDH neg cells in 8 out of 9 patient samples, which is in line with the results published previously in normal mammary tissue and triple negative breast tumours.

Overall, the MCF-7, T47D and PDS results support the use of ALDH as a suitable marker for enrichment of cells with higher MFE and more tumourigenic potential.

4. Chapter 4: Gene expression analysis of ALDH pos cells

4.1 Introduction

As shown in chapter 3, measurement of ALDH activity has been successfully used to isolate cells with CSC activity of MFE and high tumour-initiating cell frequency. ALDH pos cells have higher mammosphere formation rates and are more tumorigenic *in vivo*. Liu S. et al., (2014) found that ALDH pos cells, derived from primary human breast cancers, have a different gene expression pattern, when compared to the $CD44^+/CD24^-$ population. It was found that genes associated with an epithelial-like state were highly enriched in ALDH pos cells, compared to ALDH neg cells, whereas EMT-associated genes were significantly enriched in the $CD44^+/CD24^-$ population (Liu et al. 2014). In another relevant publication, the gene expression pattern between ALDH pos and the differentiated ALDH neg cells was interrogated in several breast cancer cell lines using Affymetrix whole-genome oligonucleotide microarray. It was found that ALDH pos cells showed overexpression of genes known to play a role in stem cells such as *CXCR1*, *NOTCH2* or *NFYA* and downregulation of other genes that are involved in cell differentiation and apoptosis such as *NACA*, *PDCD5* and *PDCD10* (Charafe-Jauffret et al. 2009). However, this study only used triple negative and HER2+ cell lines, leaving the most common breast cancer subtype, the ER+ subtype, out of this analysis. Therefore, in this chapter we wanted to interrogate the gene expression differences between ALDH pos and ALDH neg cells upon anti-estrogen treatment with Tam and Fulv in ER+ cell lines and PDS through Affymetrix microarray. The aim of this chapter is to identify biomarkers or activated signaling pathways that enable us to target CSC in endocrine resistant breast cancer.

4.2 Results

4.2.1 Analysis of the ALDH pos and ALDH neg populations in MCF7 cells by Affymetrix Arrays

RNA is susceptible to degradation by ribonucleases (RNases) present within the sample and in the environment. Downstream applications such as microarrays require good RNA quality; therefore evaluation of RNA integrity (RIN) prior to use is imperative especially after cell sorting using FACS, where RNA may be vulnerable to degradation. The electropherograms in Figure 4.1 show the RIN values of all the samples used for the microarray analysis. RNA samples from FACS-sorted ALDH pos and ALDH neg cells, used to interrogate gene expression pattern, had a RIN value ≥ 8.6 (1 being the most degraded profile and 10 being the most intact RNA). The control ALDH neg sample (top left corner of Figure 4.1) had no given RIN value, but well defined peaks for 18S and 28S rRNA confirmed RNA integrity.

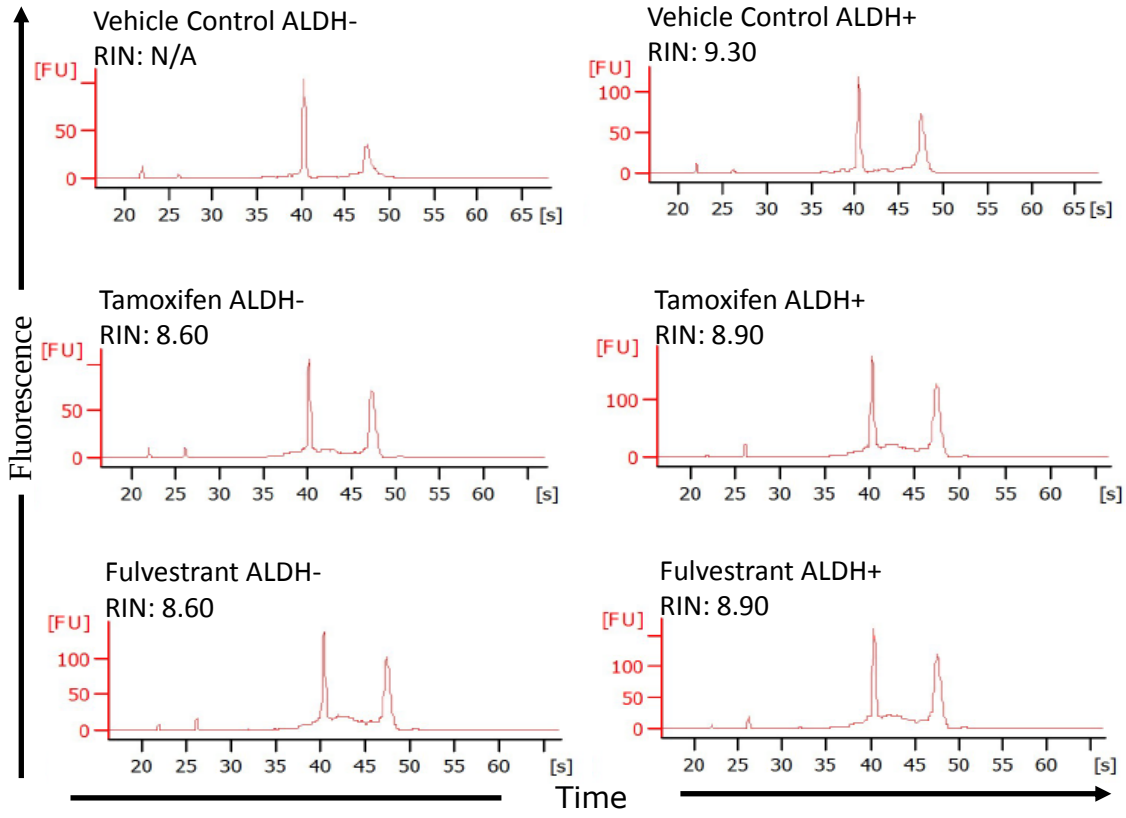


Figure 4. 1 RNA Quality control of ALDH neg and ALDH pos MCF-7 samples prior to Affymetrix microarray analysis.
Electropherogram of the 18S (left peak) and 28S (right peak) ribosomal subunits from MCF-7 cells treated with control (ethanol), 1μM Tam or 0.1 μM Fulv for 6 days prior to FACS sorting based on the ALDH enzymatic activity. RNA from ≤10,000 sorted cells was extracted using the MicroKit (Qiagen) as explained in materials and methods. Vertical axis represents fluorescence units (FU) and horizontal axis shows time in seconds.

The Affymetrix microarray technology used here (Human gene 1.0 ST array) enabled us to interrogate the whole transcriptome using a technology that involves 25-mer oligonucleotide probe arrays. Probes are not localized at the 3 prime end of the gene, which avoids relying on priming from a transcript's poly-A tail, and instead probes are along the full length of the gene. Random hexamers (rather than polyd(T)) are used to reverse transcribe mRNA to cDNA providing template distributed throughout the mRNA rather than with a bias to 3 prime untranslated region.

Before microarray data is interrogated for differential expression, several steps must be undertaken to eliminate as many undesirable

sources of variation as possible. After the biological experiments are performed, the first steps in microarray data analysis involve imaging scanning. Initial visualisation of the raw image data was carried out to assess the overall quality of the hybridisation process. By visually inspecting the image files, it is possible to find out whether there were any flaws when the hybridisation cocktail was injected into the array. In Figure 4.2, the intensity signals of the individual probes can be seen uniformly spread as bright spots onto the array, which highlights the hybridization was performed successfully.

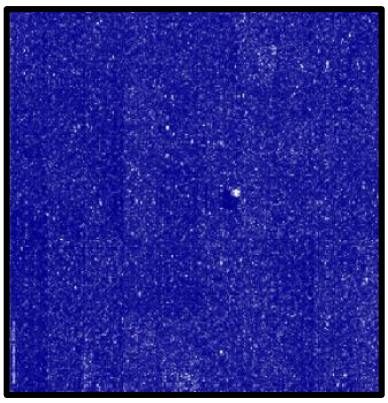


Figure 4. 2 Representative scanned image of the Human Gene 1.0 st array.
25-mer oligonucleotide probes placed on the quartz chip array are bound to labeled cDNA (sample of interest) and laser scanned. The image is divided into grids and the amount of brightness produced by each spot quantified. When assessing the image visually it is possible to assess the performance of the hybridisation process indicated by dull patches.

Microarray analysis generates .CEL files that contain the expression levels of the 764,885 individual probes. Robust Multiarray Average (RMA) normalisation was carried out to account for technical variation between the different arrays and to convert probe level data to expression values (Irizarry et al. 2003). Post-normalisation samples are represented in Figure 4.3, showing equal distribution of intensities between the arrays.

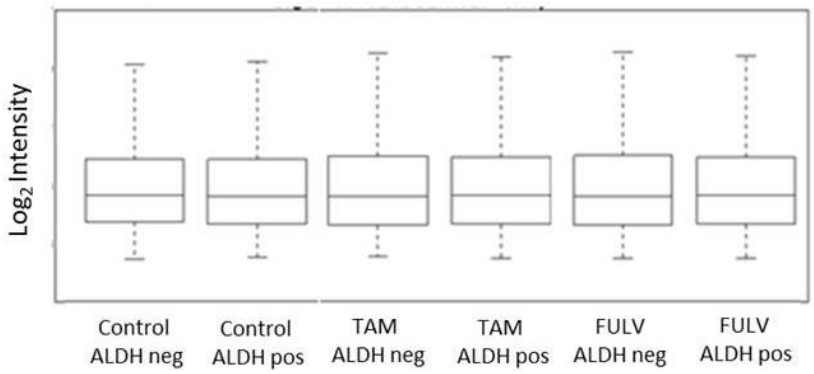


Figure 4. 3 Box Plots showing the distribution of intensities after microarray normalisation in MCF-7 cells using RMA.
Boxplots show distribution of Log2 intensities as a central box bracketed by horizontal lines known as whiskers, which represent the extreme values of the distribution. The line through the centre of the box shows the mean of the distribution and the box itself represents the SD.

To decode the data, and decipher which gene IDs on the ENSEMBL database link to each probe set used on the array, mapping files were downloaded from brainarray.com. The total number of genes expressed and mapped to the probe sets was 22,097.

Unsorted MCF-7 cells treated with control, Tam or Fulv for 6 days were also interrogated by Affymetrix microrrays to observe Tam and Fulv-induced gene expression changes and validate the *in vitro* treatment results (Figure 4.4). A supervised paired analysis was performed and the need for two controls in order to perform the paired SAM analysis meant we included the ALDH neg as a second control. Both control vehicle and control ALDH neg are highly correlated as indicated by the small branch length of the dendrogram when both samples were compared (data not shown). Therefore, the unsorted ethanol control and ALDH neg control samples were compared against the unsorted, treated populations (Tam and Fulv) using a two class paired SAM analysis to identify differentially expressed genes before and after treatments. This analysis found 168 genes (represented as green circles) that were differentially expressed between control and treated samples using Delta=1 (Figure 4.5).

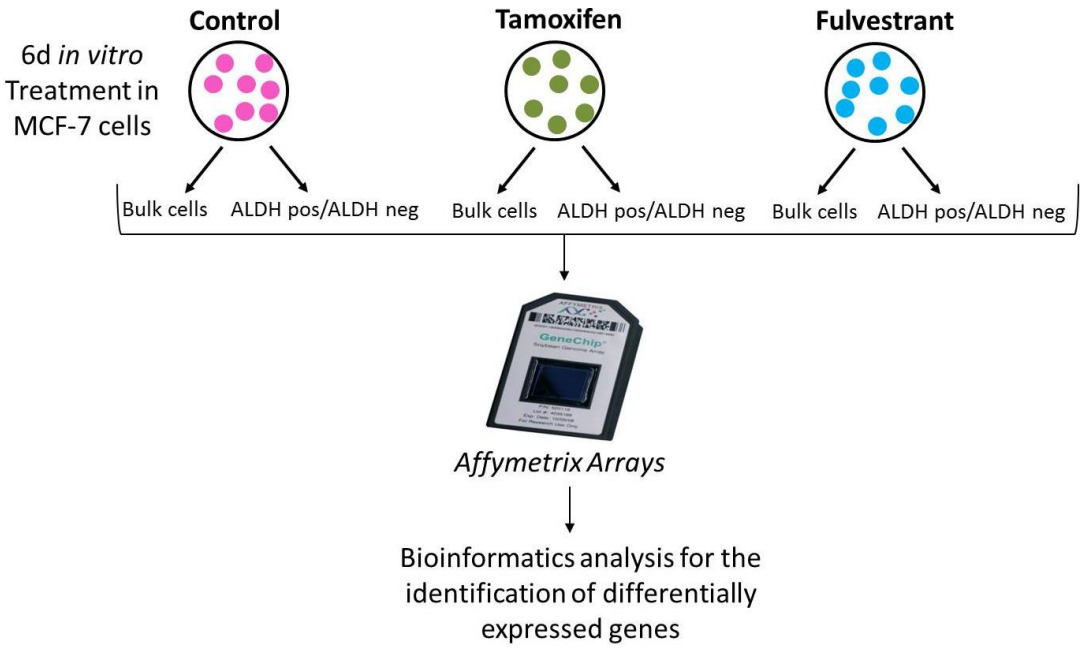


Figure 4. 4 Flow diagram showing the gene expression analysis performed in MCF-7 cells.

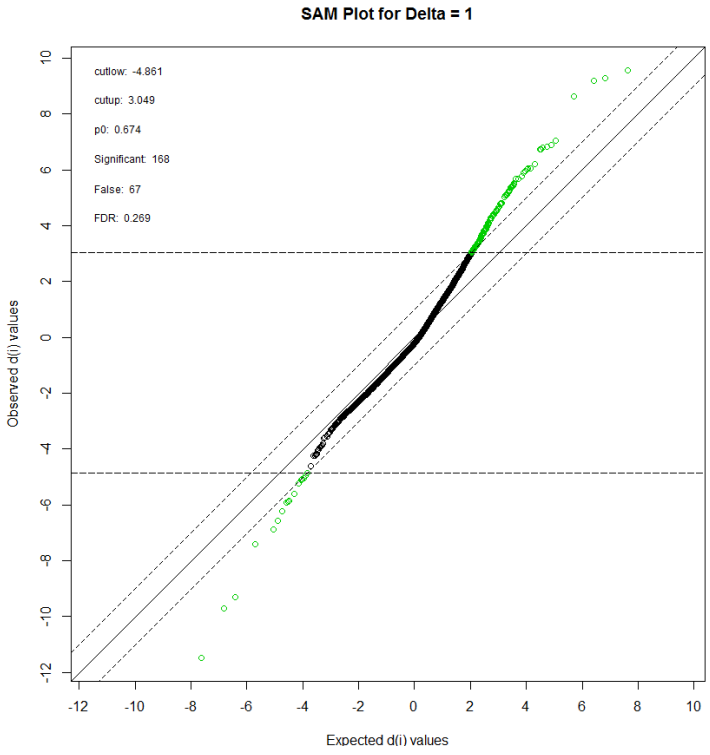


Figure 4. 5 SAM plot showing differentially expressed genes between controls vs. Tam and Fulv-treated unsorted MCF-7 cells using Delta=1.

Unsorted control and control-treated ALDH neg cells were compared against Tam and Fulv-treated unsorted cells. 168 up and downregulated genes, represented in green, were detected by performing SAM analysis between the control samples vs. Tam and Fulv treatments with a false discovery rate (FDR)=0.269.

28 of the most relevant differentially expressed genes, between control and tam or fulv-treated unsorted MCF-7 cells, identified through SAM analyses were plotted in a heatmap for identification, and better visualisation of the differences in expression. The clustering of the samples (Figure 4.6), illustrated at the top of the heatmap using a dendrogram, shows a clear separation of the control and the treated samples reflecting the biological differences between them. *SNAI2* (Slug) was one of the genes overexpressed after Fulv treatment, which is in line with results published previously (Patani et al. 2014). Likewise, Cyclin-dependent Kinase 6 (*CDK6)* was upregulated in unsorted MCF-7 cells after Fulv treatment. It has been published that *CDK6* is overexpressed in Fulv-resistant cells, which allows these resistant cells to overcome the Fulv-mediated apoptosis signaling (Alves et al. 2016). Of note, *CDK6* is the target of Palbociclib and similar drugs, which are currently achieving much success in combination with Fulv. Conversely, expression of the estrogen responsive genes *GREB1* (Ghosh et al. 2000) and *EGR3* (Inoue et al. 2004) and *PGR* were downregulated following Tam and Fulv treatment.

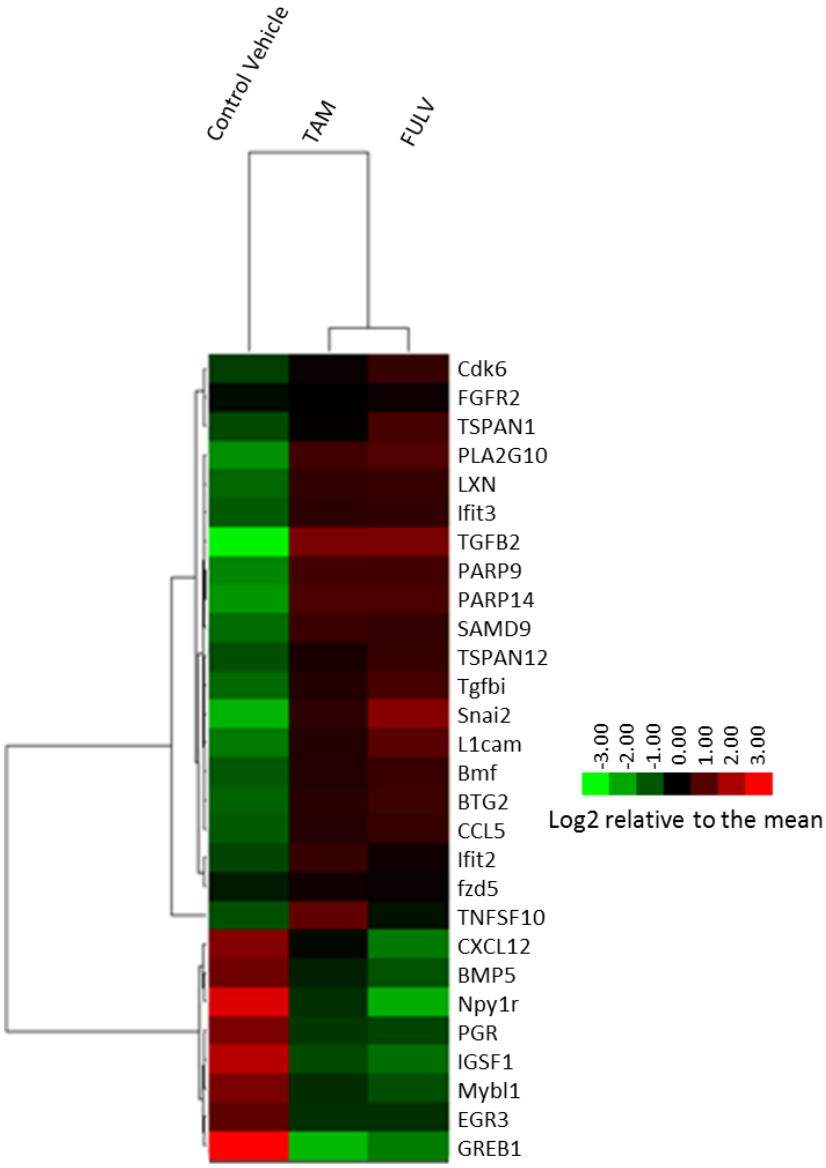


Figure 4. 6 Graphic display of 28 of the most representative up and downregulated genes after SAM analysis between controls and Tam and Fulv-treated unsorted MCF-7 cells.
Heatmap shows 28 of the differentially expressed genes out of the 168 genes identified using SAM analysis. The colours show Log2 gene expression values relative to the mean of individual genes across all 3 samples, and then subjected to cluster analysis for generating the heatmap. Red represents overexpression while green represents underexpression.

The two-class, paired SAM analysis was next applied to the sorted samples: ALDH pos (control, Tam and Fulv) against ALDH neg (control, Tam and Fulv). In total, 22 upregulated genes were identified in ALDH pos compared to ALDH neg cells with a False Discovery Rate (FDR) of 0.171 (Figure 4.7). The delta value chosen for this analysis was 1

(Delta=1), however applying a different Delta value results in a different number of differentially regulated genes as well as different FDR values as shown in Table 4.1. The 22 genes identified in the SAM analysis are shown in Table 4.2 and have been ranked according to control FC. The Deiodinase Iodothyronine Type II (*DIO2)* was the gene with the greatest FC between ALDH pos and ALDH neg in all three treatments. The expression of the ALDH1A3 isoform was much higher in ALDH pos compared to ALDH neg cells, which validates the cell sorting. No downregulated genes were identified by this analysis.

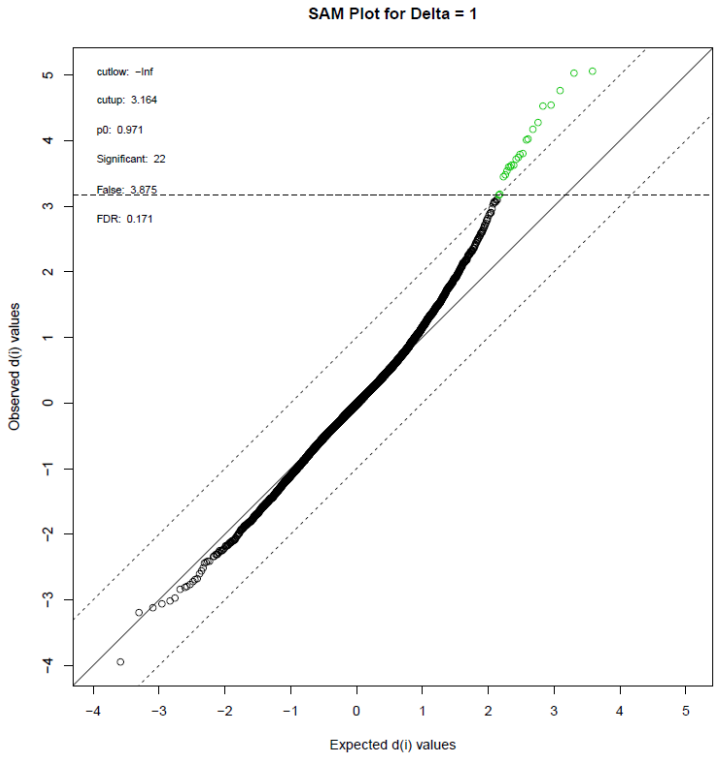


Figure 4. 7 SAM plot for the two class, paired data using Delta=1 showing differentially expressed genes between ALDH pos and ALDH neg in MCF-7.
22 differentially expressed genes were identified from processed data using SAM analysis between ALDH neg (control, Tam and Fulv) and ALDH pos (control, Tam and Fulv) with a FDR of 0.17.

Table 4. 1 Number of differentially expressed genes between ALDH pos and ALDH neg applying different Delta values.

Delta	False	False Discovery Rate (FDR)	Number of Significant Genes
0.4	76.25	0.465	164
0.5	14.25	0.291	49
0.6	27.125	0.331	82
0.7	6.75	0.225	30
0.8	11.5	0.267	43
0.9	3.375	0.178	19
1	3.875	0.171	22
1.1	2	0.2	10
1.2	0.75	0.15	5
1.3	0	0	0

Table 4. 2 Representation of the 22 differentially expressed genes identified with SAM analysis Delta=1. Values are expressed as FC between ALDH pos and ALDH neg for each group (control, Tam and Fulv) and ranked according to the control-treated group.

	FC ALDH pos/ALDH neg		
Official Gene Symbol	**Control**	**Tam**	**Fulv**
DIO2	10.97	7.08	2.82
ALDH1A3	6.50	2.61	3.81
IGFBP5	5.43	3.24	3.41
CST6	4.18	2.23	3.29
PLA2G10	3.69	2.82	2.68
MFGE8	2.38	2.69	3.41
PPARG	2.20	3.29	2.70
H19	2.37	2.29	2.75
S100P	2.05	2.11	2.81
MAP1B	1.96	2.11	2.81
GPRC5A	2.17	2.42	1.85
ENTPD3	1.84	2.47	2.08
GSTM4	2.41	1.85	1.91
HLA-DQB1	1.96	1.85	1.99
MTTY	1.98	1.80	1.93
RDX	1.79	1.86	2.03
CYP1A1	1.72	1.86	1.96
TMPRSS4	2.04	1.61	1.88
CES1	1.67	1.94	1.83
RGL1	1.91	1.76	1.78
C16ORF45	1.59	1.64	1.61
SLCO2A1	1.68	1.61	1.53

Next, ALDH1A3 expression was interrogated by qPCR in ALDH pos and ALDH neg cells after treatment with Tam and Fulv to validate the Affymetrix microarray results. Anti-estrogen treatments significantly increased the gene expression levels of the ALDH1A3 isoform in the ALDH neg population, relative to control. Moreover, ALDH1A3 expression was dramatically higher in ALDH pos cells (squared fill pattern), compare to ALDH neg (solid fill), in all three conditions tested, which is in line with the Affymetrix micorarrays results shown above. ALDH1A3 mRNA expression levels were comparable across the three conditions tested, control, Tam and Fulv (Figure 4.8).

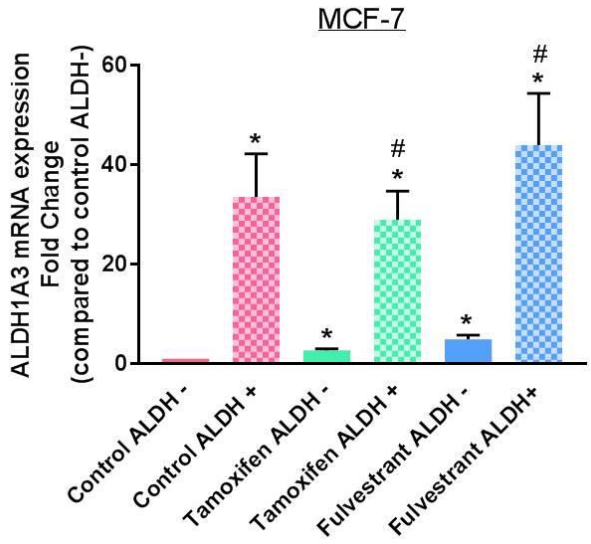


Figure 4. 8 ALDH1A3 mRNA expression in ALDH pos and neg populations in MCF-7 cells by qPCR.
MCF-7 cells were treated in monolayer for 6 days with either ethanol control (pink), Tam (green) or Fulv (blue), stained with the Aldefluor kit, as indicated in materials and methods, and subsequently ALDH pos (squared fill pattern) and ALDH neg cells (solid fill) were sorted into RNA lysis buffer to extract the mRNA. ALDH1A3 mRNA expression levels were studied using The Flex Six IFC Biomark chip. ALDH1A3 expression was set to 1 in control ALDH neg cells to show FC. Data shows average and SEM of three independent experiments. *P<0.05 (compared to control ALDH neg); ♯ P<0.05 (compared to ALDH neg from same treatment condition).

4.2.2 Analysis of ALDH pos and ALDH neg populations in patient-derived samples by Affymetrix Arrays

Next, metastatic patient samples were interrogated by Affymetrix Arrays in order to further study the gene expression differences between ALDH pos and ALDH neg cells (Figure 4.9). RMA normalisation was performed on gene expression values from all 9 patient-derived samples to account for technical variation between the different arrays. Figure 4.10 shows boxplots representing equal distribution of intensities between the 18 different arrays.

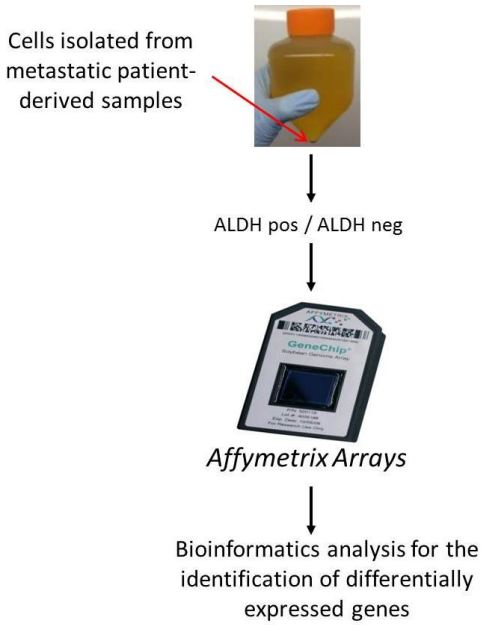


Figure 4. 9 Flow diagram showing the gene expression analysis performed in patient samples.

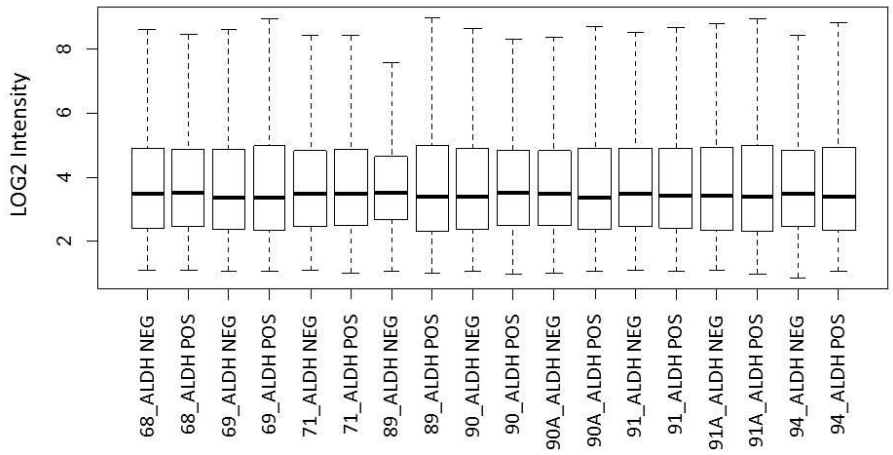


Figure 4. 10 Box Plot showing distribution of intensities after microarray normalisation in patient-derived samples using RMA. Boxplots represent distribution of Log2 intensities as a central box bracketed by horizontal lines known as whiskers, which show the extreme values of the distribution. The line through the centre of the box represents the mean of the distribution and the box itself represents the SD.

On this occasion, we implemented a different technique for identifying differential expressed genes in our microarray data set, Rank Products (RP), which is a non-parametric test similar to SAM analysis. A paired rank products analysis between ALDH pos and ALDH neg samples was carried out as shown in Figure 4.11. A total of 447 upregulated genes and 152 downregulated genes in ALDH pos cells were found in this analysis using a percentage of false positive (PFP) of 0.05. The horizontal dendrogram displayed at the top of the heatmap shows how patients who received chemo and hormonal therapy cluster together, whereas patient BB3RC94, who is treatment naïve, clusters separately (refer to Table 2.8 for clinic-pathological information about the patient samples). As expected, samples BB3RC91 and BB3RC91A, longitudinal samples from the same patient, grouped together within the branches of the dendrogram, which shows consistent gene expression pattern. The additional treatment sample BB3RC91A received, compared to BB3RC91, which were three extra weeks of Exemestane plus Everolimus and 5 days of Tam treatment (Figure 4.12), did not lead to a different gene expression pattern. Conversely, samples BB3RC90 and BB3RC90A (treatment scheme not shown), which also came from the same patient

but received the same treatment, showed a lesser degree of homology. Sample BB3RC90 shared more commonly differentially expressed genes with patient sample BB3RC89 than with its own biological replicate, as shown by the dendrogram branching (Figure 4.11).

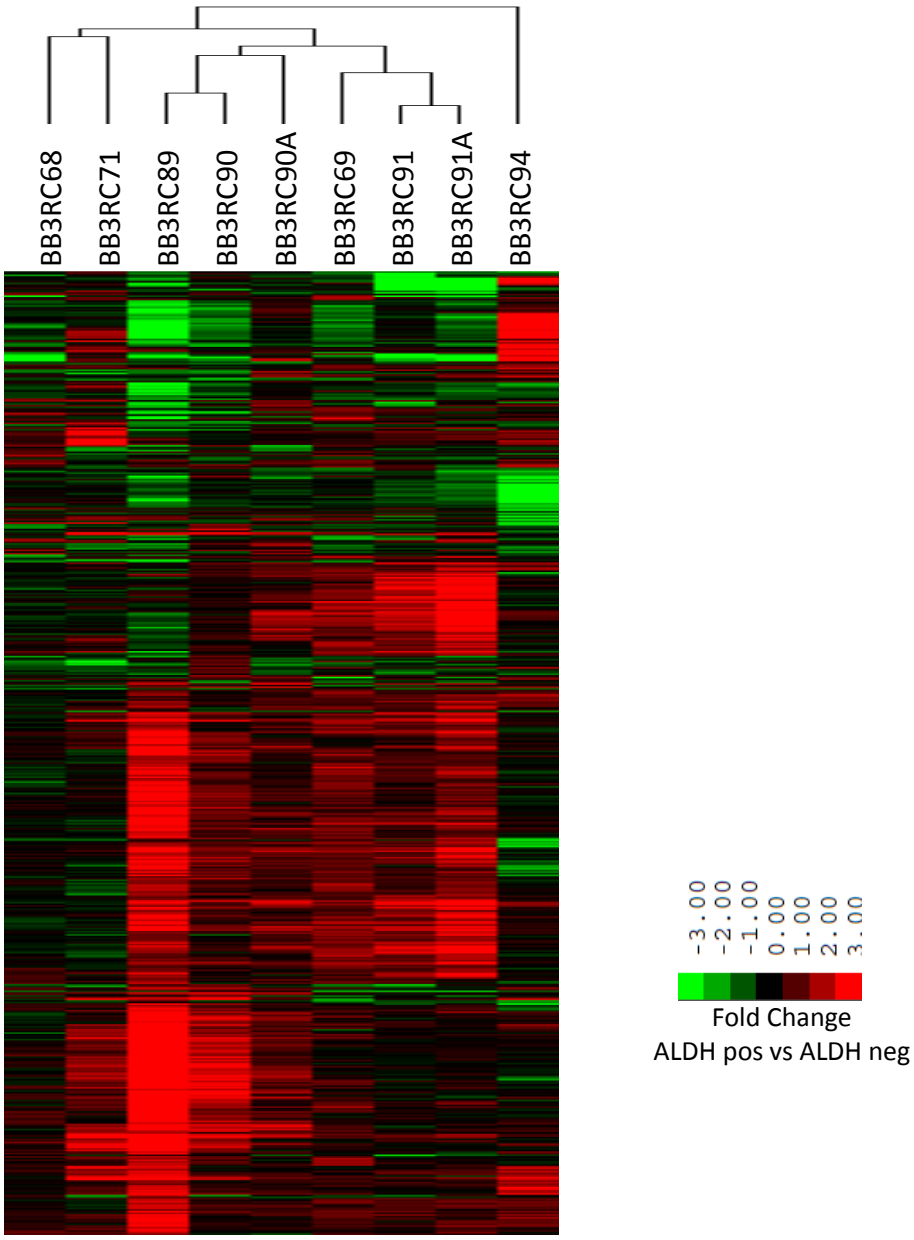


Figure 4. 11 Heatmap of Rank Products illustrating the top 599 differentially expressed genes between ALDH neg and ALDH pos cells.

Samples BB3RC90 and BB3RC90A, as well as BB3RC91 and BB3RC91A belong to the same patient respectively but were collected at different times. All samples are metastatic ER+ and HER2-, except patient 71, which is ER+, HER2+. For clinical information refer to Table 2.8. Rank product test was the non-parametric statistical method used for the detection of the top 599 differentially expressed genes illustrated in this heatmap. Red and green colours represent upregulated and downregulated genes respectively in the ALDH pos relative to ALDH neg cells.

Figure 4. 12 Schematic representation of the clinical information for patients BB3RC91 and BB3RC91A.
Yellow lines highlight the time when the patient sample was collected and processed.

Through using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, it was possible to identify the relevant biological pathways found in the ALDH pos compared to ALDH neg populations, as shown in Table 4.3. Cell adhesion molecules and cytokine-cytokine receptor interaction were two of the most enriched pathways in ALDH pos cells from ER+ PDS. Other pathways found in this analysis relevant for our study are the ECM-receptor interaction, focal adhesion, PI3K-AKT and hippo-signaling pathway, amongst others.

Table 4. 3 Gene ontology from paired rank products between ALDH pos and ALDH neg cells in 9 different patient-derived samples.

KEGG Pathway	Genes	P value	FDR
Cell adhesion molecules (CAMs)	*F11R, MPZL1, OCLN, CLDN4, SELL, VTCN1, ICAM3, CDH1, NEO1, SDC4, CDH3, VCAM1, ITGB8, CLDN1, VCAN, CD28*	2.01E-05	0.025504
Cytokine-cytokine receptor interaction	*TNFRSF21, IL6, IL1R1, TNFRSF12A, OSMR, IL6ST, TGFB3, KITLG, TGFB2, VEGFC, TNFRSF11B, CCR7, CNTF, PRLR, CXCR4, CCR2, CX3CR1, IFNG*	5.04E-04	0.637442
Hematopoietic cell lineage	*CD9, IL1R1, IL6, CD3G, CD3E, CD59, CD1C, KITLG, ITGA3, CD1A*	9.35E-04	1.180125
Extracellular Matrix (ECM)-receptor interaction	*LAMB2, ITGB8, ITGB6, COL1A2, ITGA3, SDC4, LAMB1, THBS2*	0.015103	17.5635
Glutathione metabolism	*MGST3, GSTM3, GSTA4, GSTM4, GPX3, MGST1*	0.017665	20.24444
Focal adhesion	*PAK6, VEGFC, CAV1, CCND1, LAMB2, ITGB8, ITGB6, COL1A2, ITGA3, LAMB1, THBS2, PARVA, MYL9*	0.020614	23.2305
Tight junction	*F11R, TJP1, CTTN, OCLN, CLDN4, CLDN1, PRKCH, RAB13, TJP2, MYL9*	0.021495	24.10242
Leukocyte transendothelial migration	*VCAM1, F11R, OCLN, CLDN4, CXCR4, CLDN1, MMP2, MYL9, THY1*	0.024902	27.38933
PI3K-Akt signaling pathway	*IL6, PPP2R3A, EFNA1, OSMR, KITLG, ITGA3, GNG12, VEGFC, CCND1, LAMB2, PRLR, ITGB8, ITGB6, COL1A2, CREB3L1, EFNA5, LAMB1, THBS2*	0.029245	31.3886
Hippo signaling pathway	*WNT5A, CCND1, CTGF, SERPINE1, TGFB3, TEAD1, TEAD2, CDH1, YAP1, TGFB2*	0.037134	38.13838
Pathways in cancer	*WNT5A, IL6, EPAS1, EGLN3, TGFB3, KITLG, CDH1, ITGA3, GNG12, ARHGEF12, MMP2, GLI3, TGFB2, JUP, VEGFC, CCND1, LAMB2, CXCR4, LAMB1*	0.046814	45.58421

Next, we wanted to find genes consistently differentially expressed between ALDH pos and ALDH neg cells across 5 or more patient samples as a stringent way to identify the biologically relevant expression changes. As shown in Figure 4.13, mRNA expression of Gap Junction protein Alpha 1 (*GJA1*; also known as *Connexin43*), Phospholipase A2 group IIA (*PLA2G2A*), Cortactin (*CTTN*), Pre-B-Cell Leukaemia Transcription Factor 1 (*PBX1*), Carcinoembryonic antigen related cell adhesion molecule 6 (*CEACAM6*) and some of the S100A genes were highly upregulated in ALDH pos cells relative to ALDH neg. A small number of genes were found to be downregulated in ALDH pos cells in our PDS microarray dataset, such as C-C Motif chemokine Receptor 2 (*CCR2*) and C-Type Lectin Domain Family 5 Member A (*CLEC5A*).

Figure 4.14 shows the gene expression values of 17 out of the 19 ALDH isoforms, since ALDH4A1 and ALDH3B1 were not detected in the Affymetrix data set. ALDH1A1, ALDH1A2, ALDH1A3 and ALDH2 were the isoforms that showed the greatest FC between ALDH pos and ALDH neg cells. The expression of these 4 isoforms for each individual patient samples is shown in Figure 4.15. Patient BB3RC89 shows the biggest FC between ALDH neg and ALDH pos cells for ALDH1A1, ALDH1A2 and ALDH1A3 isoforms. Conversely, patient BB3RC94 (patient who did not receive any treatment) shows the greatest FC for the ALDH2 isoform. Overall, ALDH1A1 is the isoform expressed at higher levels in ALDH pos compared to ALDH neg cells in all 9 patient samples.

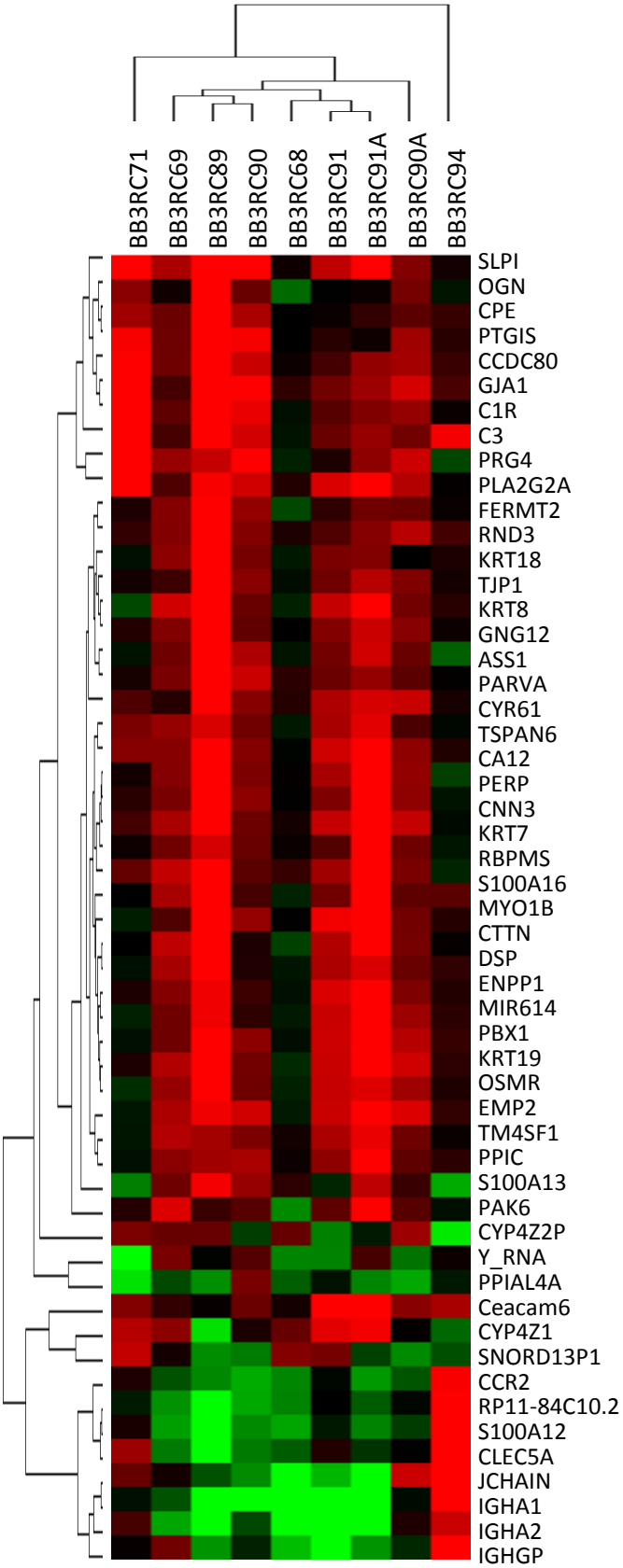


Figure 4. 13 Heatmap showing the largest FC between ALDH pos and ALDH neg cells in at least 5 or more PDS.

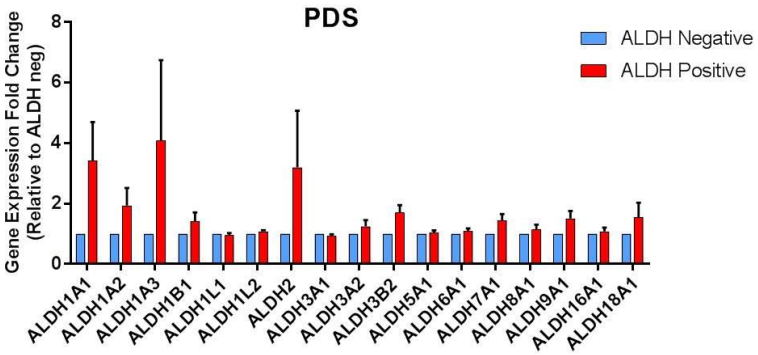


Figure 4. 14 Expression of 17 ALDH isoform across 9 patient samples.
Data from Affymetrix array represent FC, between ALDH pos (red) and ALDH neg (blue) cells across 9 different patient samples, and SEM. Expression of the two remaining ALDH isoforms, ALDH4A1 ALDH3B1, was not detected in the Affymetrix array. Although not statistically significant, P values of the four isoforms with the highest FC between ALDH pos and ALDH neg: ALDH1A1 p value=0.08, ALDH1A3 p value=0.25

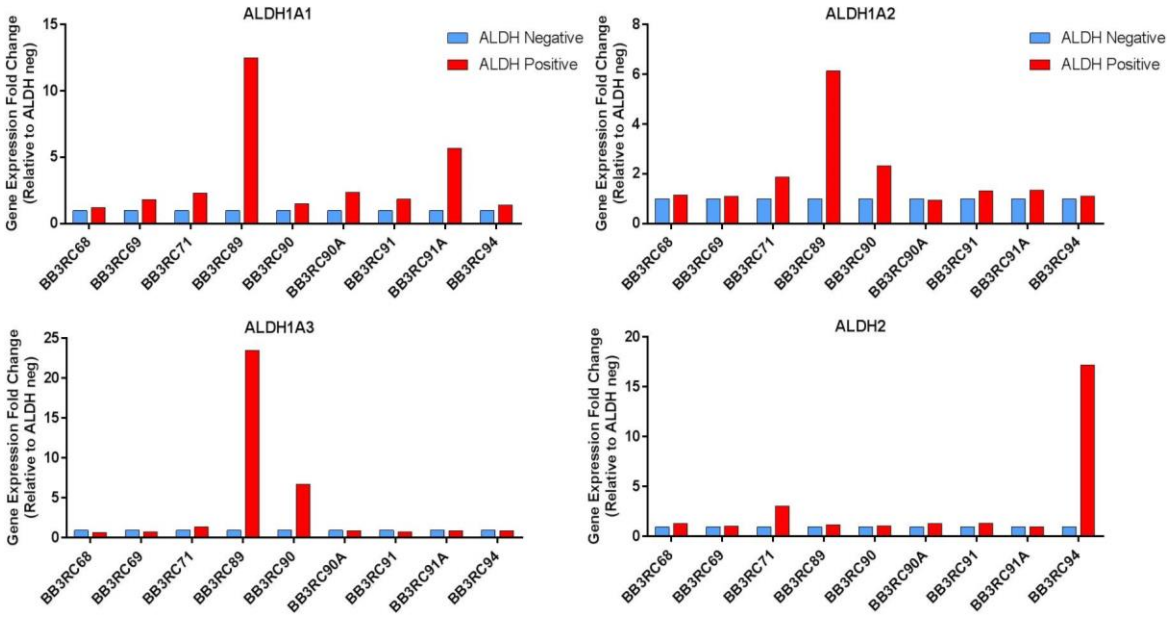


Figure 4. 15 Gene expression of the most representative isoforms in the ALDH pos population from patient samples.
Data from Affymetrix array represent FC, between ALDH pos (red) and neg (blue) cells in 9 different patient samples for ALDH1A1, ALDH1A2, ALDH1A3 and ALDH2 isoforms.

qPCR was used to further validate the ALDH gene expression results obtained by Affymetrix arrays because of its high sensitivity and large measurement range. ALDH1A1 and ALDH1A3 data validation from

the 9 patient samples is shown in Figure 4.16. Overall, qPCR data show much greater FC than Affymetrix data, which demonstrates a tendency for the Affymetrix platform to underestimate differential expression. ALDH1A3 was overexpressed in ALDH pos cells with greater FC than ALDH1A1, however the latter was upregulated in all patient samples whereas ALDH1A3 expression was not detected in patients BB3RC91 and BB3RC94.

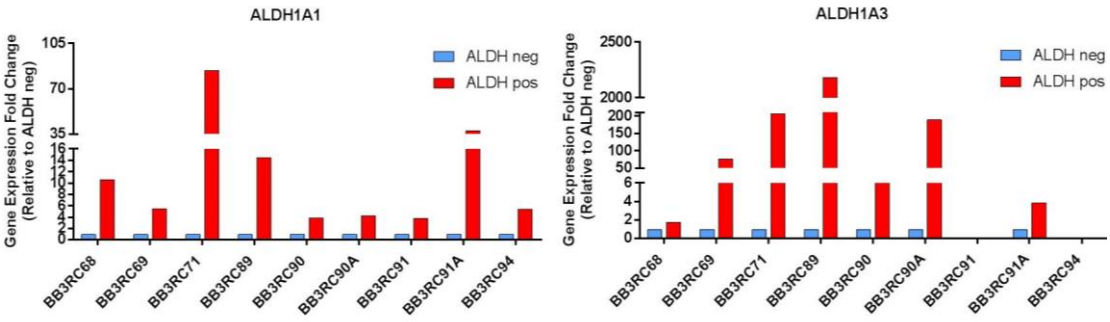


Figure 4. 16 ALDH1A1 and ALDH1A3 gene expression in PDS using qPCR.
Data from qPCR represent FC, between ALDH pos (red) and neg (blue) cells in 9 different patient samples.

4.2.3 Meta-analysis of MCF7 and patient-derived samples gene expression dataset

The Venn diagram (Figure 4.17) comprises the common genes shared between cell line and the PDS datasets. 7 genes were differentially expressed between ALDH pos and ALDH neg cells in MCF-7 and PDS in all treatments (Table 4.4). We were also able to identify 4 genes commonly shared between Tam and Fulv-treated MCF-7 cells and PDS, which had received chemo and anti-estrogen therapy. The genes associated to anti-estrogen resistance are listed in Table 4.5.

Note that ALDH1A3 was not included within these 7 genes because even though the overall Log2 FC expression of this isoform between ALDH pos and ALDH neg was 0.707 in PDS (higher than the set-up threshold of 0.599 Log2 FC), the p-value was not statistically significant (p=0.266). This can be explained because the isoform ALDH1A1 is more consistently responsible for breast CSCs activity in patient samples (1.31 Log2 FC).

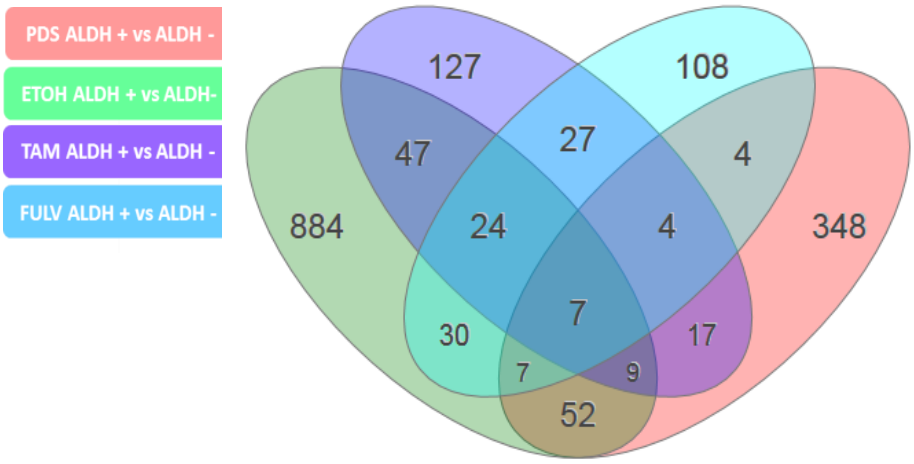


Figure 4. 17 Venn diagram illustrates meta-analysis of the cell line data between control, Tam, Fulv groups and the patient data. The iPathway guide software tool was used to plot the diagrams. The Log2 FC cutoff applied to the ALDH pos vs. ALDH neg cells obtained from the meta-analysis data was 0.599.

Table 4. 4 List of the 7 commonly differentially expressed genes between ALDH pos and ALDH neg cells in control, Tam and Fulv-treated MCF7 cells and PDS.
Red shows upregulation, green shows downregulation. Data shows Log2 FC ALDH pos vs. ALDH neg. NA: gene ID could not be mapped to any official gene symbol.

Gene symbol	Entrez ID	PDS	Control	Fulvestrant	Tamoxifen
APOD	347	**1.008**	**2.737**	**0.938**	**1.156**
GPRC5A	9052	**1.456**	**1.121**	**0.890**	**1.277**
H19	283120	**1.230**	**1.247**	**1.460**	**1.194**
ITGB6	3694	**0.690**	**4.014**	**1.502**	**2.123**
MAP1B	4131	**1.158**	**0.973**	**1.488**	**1.078**
NA	4663	**-1.260**	**1.226**	**-1.136**	**-0.936**
SUMO1P1	391257	**-0.704**	**-0.803**	**-1.580**	**-0.810**

Table 4. 5 List of the 4 genes differentially expressed between ALDH pos and ALDH neg cells shared in samples treated with anti-estrogen drugs.
Red shows upregulation, green shows downregulation, grey shows FC between ALDH pos and neg when FC is less than the set up threshold of 0.599. Data shows Log2 FC ALDH pos vs. ALDH neg.

Gene symbol	Entrez ID	PDS	Control	Fulvestrant	Tamoxifen
A2M	2	**1.101**	-0.283	**0.866**	**0.867**
CP	1356	**1.165**	0.152	**1.189**	**1.075**
PRSS23	11098	**0.846**	0.446	**0.840**	**0.841**
TIMP1	7076	**1.013**	0.591	**0.866**	**0.831**

4.3 Discussion

The main aim of this chapter was to identify up or downregulated markers in the ALDH pos population, compared to ALDH neg, in MCF-7 and PDS. By doing this it was expected to identify important markers for the enriched ALDH pos population after Tam and Fulv treatment, which could potentially be therapeutically targeted to overcome ER+ breast cancer anti-estrogen resistance. The use of microarray technology to perform gene expression profile of thousands of transcripts simultaneously has been successfully implemented in cancer research since the late 1990s. Therefore we employed this technique to assess gene expression differences between ALDH pos and ALDH neg populations. One source of variation for microarray data is the hybridisation procedure, therefore all Affymetrix chips were visually inspected to ensure that probes hybridised properly to the RNA target. Figure 4.3 shows homogenously distribution of the intensity signals across the chip array.

Following RMA normalisation of the microarray dataset, unsorted MCF-7 samples treated with either control, Tam or Fulv were plotted in a heatmap to observe the gene expression differences associated to the anti-estrogen treatment (Figure 4.6). EMT-associated genes such as *SNAI2* (Jiang et al. 2014), pro-apoptotic regulators such as *BCL2* (*BMF*) (Lam et al. 2009), members of the transforming growth factor beta family of cytokines (TGFB2) (Brandt et al. 2003) or chemokines such as chemokine c-c motif ligand 5 (CCL5 also known as RANTES) were upregulated upon anti-estrogen treatment (Yi et al. 2013), as previously described in the literature. Conversely, many estrogen-responsive genes were found to be down-regulated after Tam and Fulv treatments such as Early Growth Response 3 (EGR3), Growth Regulation by Estrogen in Breast Cancer 1 (GREB1) and the Progesterone Receptor (PgR) in line with previous research (Inoue et al. 2004) (Rae et al. 2005). These gene expression results validate the 6 day-anti-estrogen treatments performed on the unsorted MCF-7 cells.

SAM analysis was applied to the ALDH-sorted MCF-7 dataset as a tool to identify differentially regulated genes between ALDH pos and their

differentiated progeny ALDH neg cells. 22 genes were differentially expressed between ALDH pos and ALDH neg cells when applying a delta value of 1 (Figure 4.7 and Table 4.2). Delta value was adjusted to 1 since it was the value that provided us with a significant number of genes and an acceptable FDR. Different delta values result in different number of genes and different FDR, as shown in Table 4.1. Surprisingly, no downregulated genes were found in the MCF-7 array data set, however even in the PDS dataset, which identified many more upregulated genes than the 22 observed in the cell line data, very few downregulated genes were found in ALDH pos cells. *ALDH1A3* isoform was amongst those 22 genes and its expression was up-regulated in the ALDH pos population, by approximately 6.5 in the control and 2.5 and 3.2-fold in Tam and Fulv groups respectively. *ALDH1A3* has been shown to be a good breast CSC marker, as *ALDH1A3* expression has been shown to correlate better with tumour grade, metastasis and cancer stage than ALDH1A1 (Marcato et al. 2011). *ALDH1A3* activity has been shown to account for the majority of the total ALDH enzyme activity in breast cancer, and knockdown studies showed a decreased BCSC activity when its expression is reduced (Marcato et al. 2011). The results shown in Chapter 3 of the present study also support these findings. Therefore, the finding that *ALDH1A3* expression was increased in all three groups validates the sorting of ALDH pos in MCF-7 cells. Different members of the secretory phospholipase A_2 ($sPLA_2$) have been shown to generate biologically active lipid mediators such as arachidonic acid (AA), which can then be converted into prostaglandins via cyclooxygenase (Cox) isoenzymes. Prostaglandins will in turn exert their inflammatory effects by activating rhodopsin-like seven transmembrane spanning G protein-coupled receptors (GPCRs) (Ricciotti et al. 2011). Moreover, the arachidonate metabolite 15-deoxy-$\Delta^{12,14}$-prostaglandin J_2 has been found to bind the RA receptor PPAR-γ and induce adipogenesis (Raman et al. 2011). The expression of phospholipase A2 group X (*PLA2G10*) was also greater in the ALDH pos population after control, Tam and Fulv treatment. PLA2G10 has been identified in a census as one of the genes likely to play a prominent role in the development of breast cancer (Santarius et al. 2010) and the $sPLA_2$ have been shown to play an important role at

mediating cell growth and invasive potential of lung CSCs (Bennett et al. 2014). Insulin-like Growth Factor Binding Protein (*IGFBP5*) and G protein-coupled receptor Class C Group 5 Member A (*GPRC5A*) were also identified amongst the 22 genes observed in the SAM analysis. IGFBP5 has different functions in normal and tumorous breast. On the one hand, it has been argued that IGFBP5 plays an important role in inducing apoptosis during involution of the mammary gland after pregnancy (Tonner et al. 1995). However, IGFBP5 has also shown to have IGF-independent effects such as inducing anti-apoptotic effects or promoting metastasis in breast cancer (Akkiprik et al. 2008). GPRC5A has been shown to be induced by RA and play a role as tumour suppressor in lung cancer through functioning as a negative modulator of EGFR in mice (Zhong et al. 2015). It has also been reported that GPRC5A can activate non-canonical Wnt signalling through biding the Frizzled receptors in zebra fish embryos (Harada et al. 2007). Another gene identified through SAM analysis was PPAR-γ, which is one of the nuclear receptors RA binds to and has been linked to insulin resistance, lipid metabolism and anti-inflammatory responses. Firstly, PPAR-γ forms a heterodimer with the retinoic x receptor (RXR) and secondly binds to ligands such as unsaturated fatty acid (i.e. arachidonic acid) or eicosanoids, which modulate various cellular functions. IL4 has been found to induce 12/15 lipoxygenase and PPAR-γ production in macrophages, which establishes a link between cytokines and the nuclear receptor PPAR-γ (Huang et al. 1999). In another relevant publication it was found that the PPAR-α agonist Wy14643 enhances mammosphere formation through upregulating the NF-KB/IL-6 axis (Papi et al. 2012). Moreover, PPAR-γ has been shown to play a key role in ALDH pos cells of HER2+ breast cancers (Wang et al. 2013). Activation of PPAR-δ, another member of the PPAR family, leads to an increase in the proliferation of the hematopoietic stem cell compartment, whereas deleting this receptor results in inhibition of fatty acid oxidation (Ito et al. 2012).

RP was applied to the PDS microarray dataset and it was found that the ALDH1A1 isoform, but not ALDH1A3, was upregulated in ALDH pos cells compared to ALDH neg (Figure 4.11). These results indicate that ALDH pos cells from MCF-7 cells may rely on different isoforms to sustain

ALDH activity. The PDS dataset showed much greater gene expression differences than the cell line data, highlighting the importance of using patient samples for generating reliable and robust data that reflects real breast tumours. *GJA1*, also known as Connexin43, was found to be upregulated in ALDH pos cells compared to ALDH neg in all metastatic patient samples investigated. *GJA1* is a component of GAP junctions, which are channels on the cell membrane made up of connexion proteins. GAP junctions allow tumour cells to communicate with endothelial cells and therefore play a key role in cancer cell extravasation and metastasis. Zibara K. et al., used the triple negative breast cancer cell line, MDA-MB-231, to show that gap junction inhibition with Oleamide induced cell arrest, reduced migration and invasion *in vitro* and inhibited metastasis to lung and liver *in vivo* (Zibara et al. 2015). However, other studies have found *GJA1* has tumour suppressor capabilities in breast cancers, via a mechanism that is independent of gap junctional intercellular communication (Qin et al. 2002). *PLA2G2A* was also upregulated in ALDH pos cells in all nine patient samples interrogated. A recent publication has shown that *PLA2G2A* works as a ligand for EGFR family and subsequently enhances CSC properties via HER/ERBB-elicited signalling (Lu et al. 2017). *PLA2G2A* expression is increased upon inflammation (Lambeau et al. 2008) and interestingly, its expression stimulates intestinal stem cells as it triggers the release of AA, which is then converted by Cox into prostaglandin E2 to synergistically activate the Wnt signalling pathway. Schewe and colleagues found that both *PLA2G2A* and *PLA2G10* play regulatory roles as stem cell niche factors in the intestine (Schewe et al. 2016).

Likewise, CTTN, a substrate of Src-related tyrosine kinases, was also amongst the top 599 most differentially expressed genes between ALDH pos and ALDH neg cells since it was upregulated in 7 out of 9 samples. It has been published that *CTTN* promotes metastasis by boosting the interaction between tumour cells and endothelial cells (Li et al. 2001). *PBX1*, which also contributes to the development of metastasis in ER+ breast cancers, was overexpressed in 7 out of 9 PDS (Magnani et al. 2015). Several members of the S100A family were also found amongst the top 599 genes. The S100 protein family regulate cellular

responses acting as calcium sensors and they can contribute to metastasis, angiogenesis and immune evasion (Bresnick et al. 2015). Moreover, different genes involved in calmodulin binding, an intracellular target of the secondary messenger calcium, such as Calponin-3 (*CNN3*) and Myosin 1B (*MYO1B*) were upregulated in the ALDH pos population in virtually all patients. Calmodulin plays an important role in lipid metabolism by lowering blood calcium levels, which activates G protein cascades leading to the generation of *cAMP* (Alberts B et al. 2002). A recent publication has shown that calmodulin promotes cancer cell proliferation through *KRAS* by binding Phosphatidylinositide-3-kinase alpha (*PI3K-α*).

Some of the downregulated genes identified in the PDS dataset such as IGHA or CLEC5A are involved in immune response (Figure 4.13). However, some of the pathways identified in the KEGG analysis such as the hematopoietic cell lineage suggest inefficient leukocyte contamination when processing the PDS. These results show that applying the gradient lymphoprep for white blood cell exclusion is not sufficient and in addition, CD45+ cell depletion should have been performed. Cell adhesion molecules, cytokine-cytokine receptor interaction and ECM-receptor interaction were some of the cellular processes found to be significantly differentially expressed between ALDH pos and ALDH neg cells derived from metastatic patients. These findings highlight the interaction between CSC and their surrounding extracellular matrix as a required step before cancer cells leave the primary tumour and travel to form metastasis.

Next, we wanted to elucidate which genes were common to both cell line and the PDS datasets. We did this by performing a meta-analysis, which identified 7 genes to be upregulated in ALDH pos population (Table 4.4), which included the retinoic acid-inducible gene *GPRC5A* and the integrin *ITGB6*. The apolipoprotein D (*APOD*) was also amongst those 7 genes upregulated in ALDH pos cells compared to ALDH neg in MCF-7 and PDS. A similar lipoprotein, *APOE,* had been previously found to be significantly upregulated in mammosphere-derived cells from normal breast tissue compared to differentiated cells (Dontu et al. 2003). When looking at the differentially expressed genes between ALDH pos and ALDH neg cells after Tam or Fulv treatment compared to control, 4

genes were found. Ceruloplasmin (CP), which was amongst the 4 genes identified in the meta-analysis, is a metalloprotein shown to be regulated by the transcription factor FOXO1 in hepatic carcinoma, which enhances the antioxidant pathway in response to IL-6 signaling (Sidhu et al. 2011).

Overall, a large number of genes differentially expressed between ALDH pos and ALDH neg cells have been found in both data sets, which support the hypothesis that ALDH pos cells are transcriptionally different compared to ALDH neg cells in MCF-7 and ER+ patient samples.

5. Chapter 5: An investigation into the cellular diversity of the resistant ALDH pos population

5.1 Introduction

Tumours are composed of a heterogeneous population of cells, not only between different individuals with the same type of cancer, but also between different stages of tumour progression, and also between different regions of the same tumour, and this heterogeneity occurs at multiple levels: mutations, gene and protein expression. Breast cancer is an especially heterogeneous disease, as shown by the distinct subtypes containing different molecular signatures, however most studies to date have used bulk cells to explore the molecular biology underpinning breast cancer, which hides the intrinsic characteristics of individual cells. Differences in gene expression define cell progression and fate throughout development, therefore it is essential to study genetic changes and mutations at the single cell level to understand tumour heterogeneity, which might influence therapeutic strategies and improve breast cancer outcomes.

The CSC compartment is a heterogeneous population of cells demonstrated by the different marker profiles, which have been employed to identify and isolate breast CSCs in the different breast cancer subtypes. Epigenetic plasticity allows interconversion between CSCs and their differentiated progeny and also between different CSC populations, which can be promoted by the microenvironment at different stages of the disease (Brooks et al. 2015). It has been suggested that this phenomenon might be regulated by the expression of certain microRNAs (Liu et al. 2012). Therefore, taking into account the complex cellular heterogeneity present in breast tumours, we wanted to assess whether there is diversity within the ALDH pos population in response to anti-estrogen therapy by looking at the gene expression profile of this therapy resistant population at the single cell level. In order to assess changes in gene expression levels upon Tam and Fulv treatment in single ALDH pos MCF-7 cells, the C1 system and Biomark HD technology (Fluidigm) was used. This microfluidic technology enables automated

single cell isolation and interrogation of 96 genes. In our study, the 96 genes were selected based on important signalling pathways for CSC regulation, EMT and CSC markers. The MCF-7 and PDS array data, shown in chapter 4, was used to identify many of the 96 genes selected to study individual cell response to anti-estrogen treatment of single ALDH pos MCF-7 cells.

5.2 Results

5.2.1 Designing a 96 gene panel to study ALDH pos heterogeneity after anti-estrogen treatment

The 96 genes selected to interrogate ALDH pos cell diversity were based on the literature and the MCF-7 and PDS microarray data presented in chapter 4. The heatmap illustrated in Figure 5.1 shows the level of expression of each of these 96 genes in our MCF-7 and PDS microarray dataset (Figure 5.1A and 5.1B respectively) and it tests for statistical significance using a paired t-test. As shown previously in chapter 4, the expression of the *ALDH1A3* isoform is significantly higher in ALDH pos cells, compared to ALDH neg ($p=0.03$) in MCF-7 cells, whereas the *ALDH1A1* gene expression levels did not differ between the two populations investigated ($p=0.68$). Conversely, in the PDS dataset ALDH1A1 expression was significantly higher in ALDH pos cells ($p=0.007$), whereas *ALDH1A3* was not ($P=0.26$). *GPRC5A* and *TSPAN6* were the only two genes consistently upregulated in ALDH pos cells across treatments in both MCF-7 ($p= 0.01$, $p=0.05$ respectively) and PDS ($p=0.008$ $p=0.004$ respectively) dataset. In addition to *GPRC5A* and *TSPAN6*, *PLA2G2A* and *PLA2G10* were also included into the 96 gene list as their expression is significantly upregulated in ALDH pos cells compared to ALDH neg and also due to their role in inflammation and stem cells. *CDH3*, also known as p-Cadherin, has been linked to the expression of the ALDH1 marker in breast tumours of the basal-like molecular subtype and its expression also correlates with bad prognosis (Vieira et al. 2012). Given the important role that CDH3 plays in breast cancer and the fact that its expression was significantly higher in ALDH

pos compared to ALDH neg patient-derived cells, CDH3 was also included in the gene list to investigate CSC heterogeneity at the single cell level.

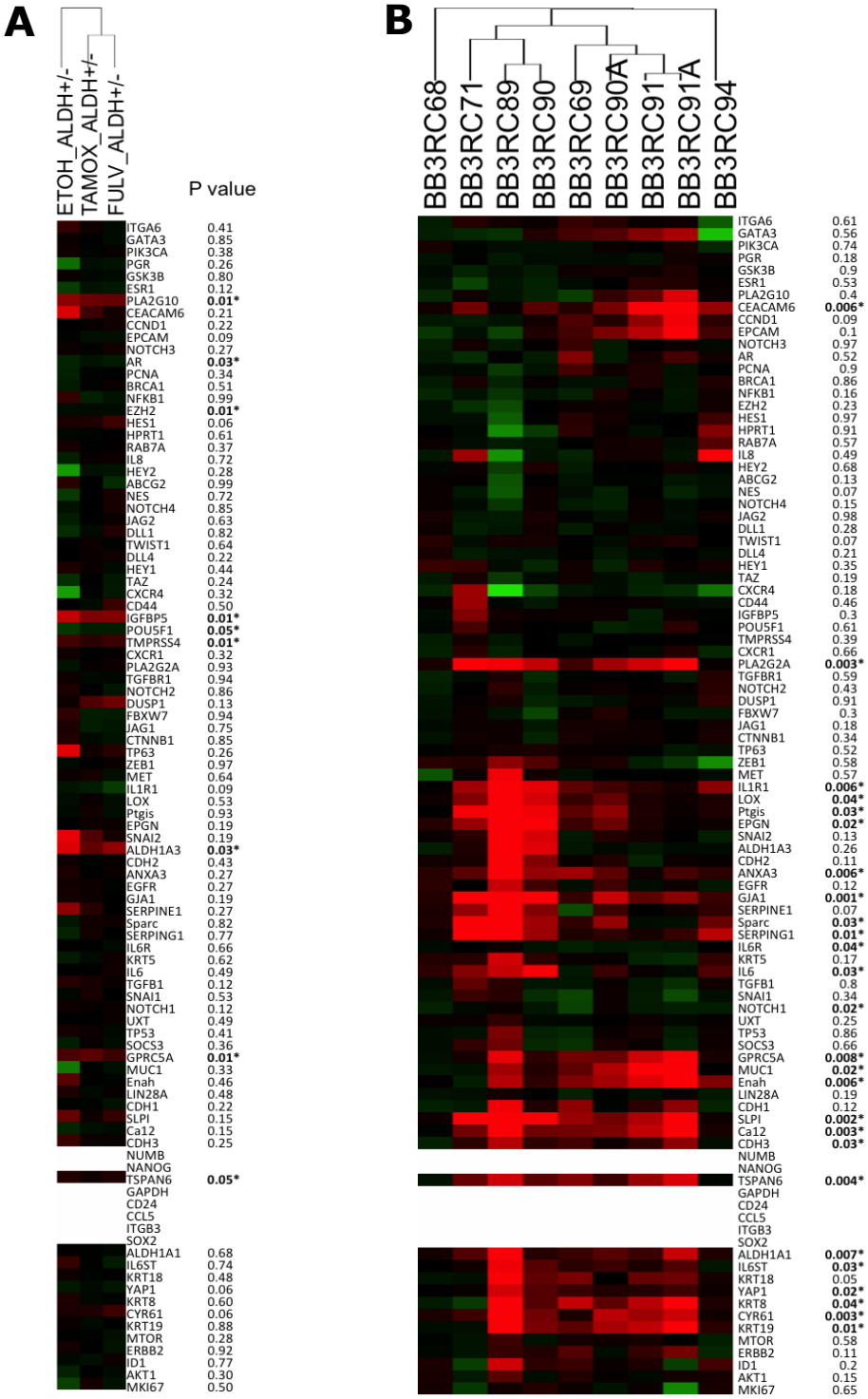


Figure 5. 1 Heatmap showing expression fold change of the selected 96 genes between ALDH pos and ALDH neg cells in MCF-7 cells (A) and PDS (B).

Paired t-test was applied to the microarray dataset using bioconductor R. Official gene symbol is represented on the right hand side of the heatmap. Red shows upregulation and green shows downregulation of gene expression in ALDH pos cells. Significant p-values are represented with an asterisk. When colour code was absent for a particular gene means that gene is not represented in the dataset.

5.2.2 Single cell technology and gating strategy

In order to assess diversity of the ALDH pos population, it was necessary to exclude the small percentage of cells, which express both CSC markers, ALDH and $CD44^{+}/CD24^{-}$ (Liu et al. 2014). The gating strategy undertaken is represented in Figure 5.2 and it shows how once the ALDH pos cells are gated, according to the DEAB inhibitor, a further gating is performed to exclude CD44-APC^{+}/CD24-PE-$CY7^{-}$.

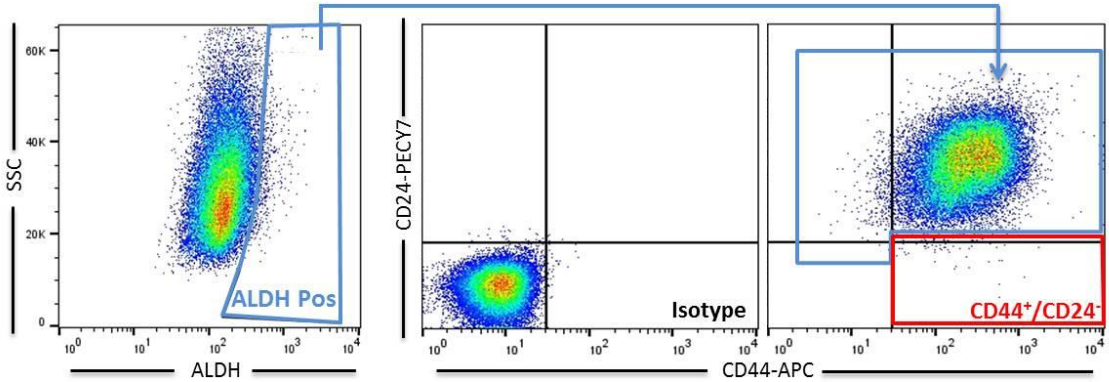


Figure 5. 2 Flow Cytometry gating strategy to assess ALDH pos heterogeneity at the single cell level.
MCF-7 cells were assayed for the Aldefluor activity using the Aldefluor assay, dual stained with the pre-conjugated antibodies CD44-APC and CD24-PE-CY7 and with the DNA-binding dye DAPI. Isotype control (middle plot) was used to gate the $CD44^{+}/CD24^{-}$ and to correct for background staining. $CD44^{+}/CD24^{-}$ (red gate) cells were excluded from the ALDH pos (blue gate) populations as shown in the gating strategy.

Once the ALDH pos cells have been sorted, cells are loaded into the C1 microfluidic chip and all 96 chambers are manually inspected using a microscope to record the number of chambers that contained one single cell, multiple cells, debris, as well as those that were empty. Cells or debris placed outside the chamber but within the capture zone (Figure 5.3 B), were also recorded since the lysis buffer would drag them into the reverse transcription zone. Figure 5.3 C shows the multiple scenarios that can be seen in the C1 system chips: single cells perfectly located within the chambers, single cell placed outside the chamber captured site but within the capture zone, debris, multiple cells and empty chambers. Only those chambers containing visible single cells, like the photographic

examples C1 or C2, were considered for further experimental approaches.

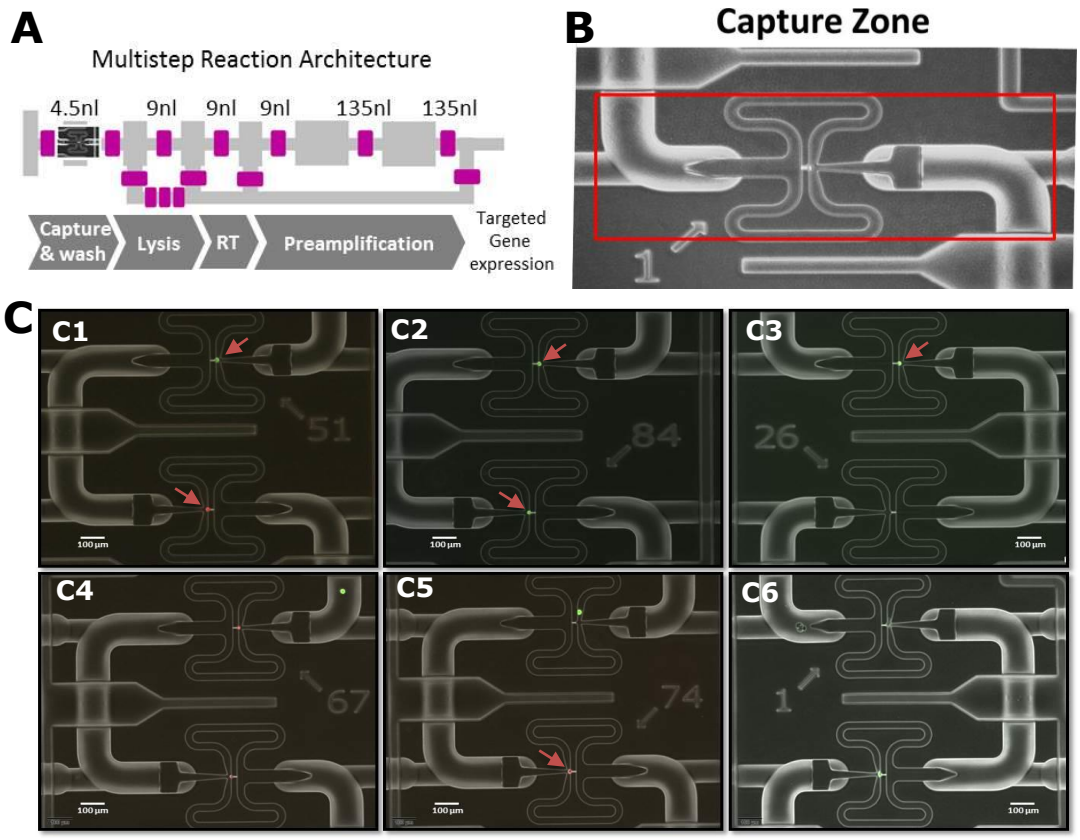


Figure 5. 3 C1 system workflow, architecture and photographic examples of captured cells within the C1 chambers.
A) Cell capture module shows the five-reaction step. Cells are first of all loaded onto the C1 chip and each single cell is captured within an individual chamber in a 4.5 nl volume reaction. Cells are washed and lysed in a 9 nl volume reaction and mRNA is then reverse transcribed into cDNA. Subsequently, PCR preamplification of all 96 genes in each single cell occurs. Finally, Preamplified cDNA gene expression is interrogated using the Biomark HD technology. Pink bars represent valves that separate different chambers in the device. Figure taken from Fluidigm. **B**) The capture zone includes an individual chamber, where the cell gets captured in and also part of the pipes, which will transport the required reagents to fill each chamber. Figure taken from Fluidigm. **C**) After cell loading and prior cell washing, cells within the chambers were photographed and images were scored for single (C1, C2, C3-top chamber and C5) empty/debris (C3-bottom chamber and C6-top chamber) and multiple cells (C4-top chamber). Arrows indicate single viable cells.

Once the pre-amplification step was finished, cDNA from each single cell was harvested, diluted and used to interrogate expression of the 96 genes using the Biomark HD technology. Gene expression data were extracted and inputted in the statistical software R (R Development Core Team 2008) for further analyses. A schematic description of the analysis procedure is shown in figure 5.4 but in summary: firstly all samples were converted into Log2 expression values via equation shown below for which a limit of detection (LOD) of 28 Ct was used.

$$Log2(expression) = LOD - Ct$$

Next, all samples with a Ct value higher than the LOD were discarded. Further quality controls involved the elimination from the study of all genes that were only expressed in 3 or less cells per Biomark plate and the removal of outliers (via Fluidigm® SINGuLAR™ Analysis Toolset 2.0, which calls the princomp R package (http://stat.ethz.ch/R-manual/R-patched/library/stats/html/princomp.html) (Fluidigm Corporation 2014). After data processing, a total of 69 genes were used out of the 96 genes that were studied initially. Finally all missing completely at random values were inputted using the package MICE (Van Buuren et al. 2008) that uses chained equations to estimate missing values. Preliminary exploratory analysis via Principal Component Analysis (PCA) showed that there was a batch effect that could be addressed using the ComBat package in R (M Wang et al. 2017) resulting in a large reduction of batch-related variance (Figure 5.5).

Figure 5. 4 Flow chart showing single cell data processing.

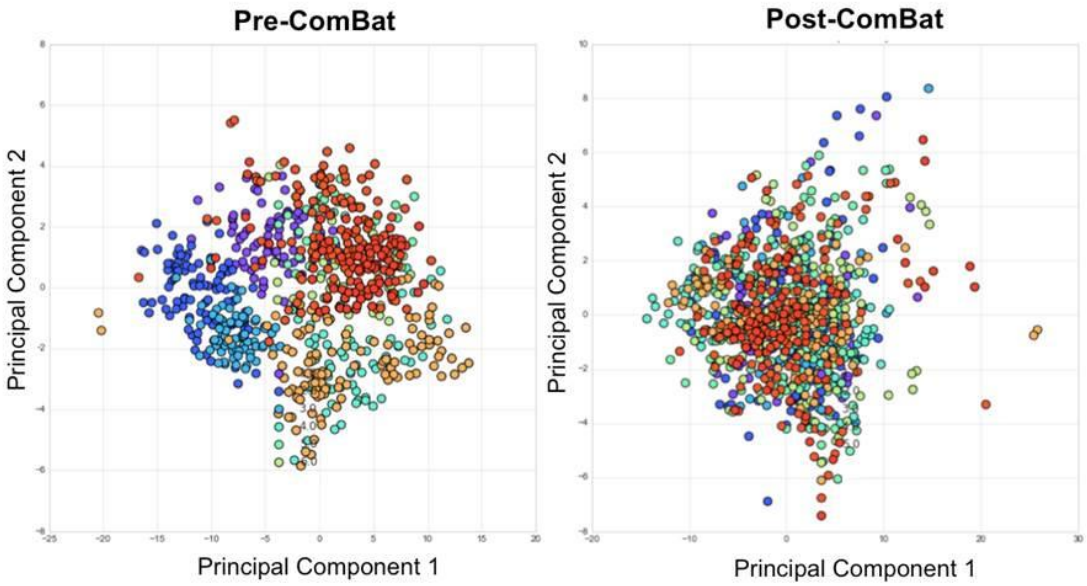


Figure 5. 5 PCA showing cell distribution before and after applying ComBat to the single cell data set.
PCA shows 9 clusters of cells, each cluster representing an independent Biomark experiment. Pre-combat PCA (left) show distinct clusters separated from each other, which highlights the existing batch effect between the different Biomark runs. Post-ComBat PCA (right) shows all clusters overlapping each other, which emphasises the batch effect removal.

5.2.3 Identification of cell doublets

Following collection of data, Fluidigm published a white paper informing users of their C1 systems that due to a manufacturing problem there was up to 40% chance to capture doublets within the chips (Fluidigm Corporation 2016). However, there was no information regarding our cell line of choice, MCF-7 cells. Therefore, sorted MCF-7 cells were injected into a C1 chip and bright-field z-stack images were taken to quantify the percentage of doublets. It was found that bright-field images were only powerful enough to identify doublets in a very small number of cases, normally when two or more cells were captured next to each other and their nuclei could be seen (Figure 5.6 A). However this method failed to identify doublets when cells were on top of each other, in the z-plane. Therefore, in order to find out the exact percentage of doublets in our cell line of interest, MCF-7 cells were stained with either the blue-fluorescent DNA dye Hoechst 33342 or the red-fluorescent DNA dye DRAQ5, as explained in materials and methods, and doublets

were assessed applying z-stacks using a Leica Widefield Low light microscope. Figure 5.6 B shows a double coloured cell doublet, which is not possible to identify in bright-field but can be detected when applying z-stacks. RGB light source enabled us to visualise cell doublets using the two colours, blue and red. Overall, 27 double-coloured doublets were identified out of the 96 captured cells. However, when the two cells that were part of the doublet showed the same colour, doublet identification was arduous. Therefore, it was estimated that the number of red/red or blue/blue doublets would be a similar number. The results from this single experiment indicate that the number of doublet MCF-7 cells being captured could be greater than 50%.

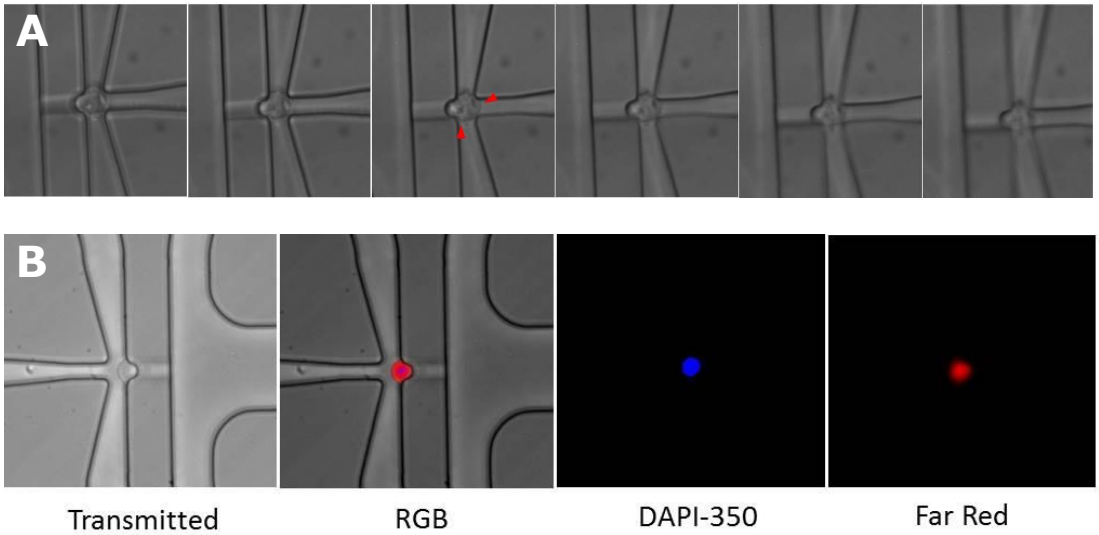


Figure 5. 6 Photographic images of the C1 chambers to assess cell doublet ratio.
A) Bright-field z-stack images of a doublet captured with the C1 system. Red arrows point at the two nuclei seen at a specific focal distance. **B)** Z-stack images show a doublet with the transmitted light (left; bright field), the RGB light source (red and blue colour), the DAPI filter (blue colour only) and Far Red filter (right; red colour only).

In order to filter our already collected datasets to ensure the analysis of only single cells, a statistical approach was undertaken. Using the package Mclust (Fraley et al. 2017) in R, Gaussian mixture models were fitted to identify cell clusters within each treatment. These models indicated the existence of two very well defined cell clusters in each condition. The nature of these clusters was further investigated by

plotting the average Log2 expression per gene in both clusters. The results indicated a much higher expression of all 96 genes investigated in one of the clusters, compared to the other cell cluster (Figure 5.7). Henceforth, it was proposed that the cluster with lower expression pattern corresponded to cell singlets and we discarded all the cells assigned to the cluster with the highest gene expression pattern for the analysis, as they were considered to be doublets. Figure 5.7 shows how Control Cluster 1, represented in green, had lower averaged gene expression values in all 69 genes interrogated, compared to Control Cluster 2 (orange), therefore the former was considered to represent singlets and the latter doublets. Similarly, Tam Cluster 2 and Fulv Cluster 2 (singlets) show reduced gene expression levels of all genes, compared to Tam Cluster 1 and Fulv Cluster 1 (doublets). Table 5.1 shows the percentage of singlets and doublets identified in each cluster and each treatment.

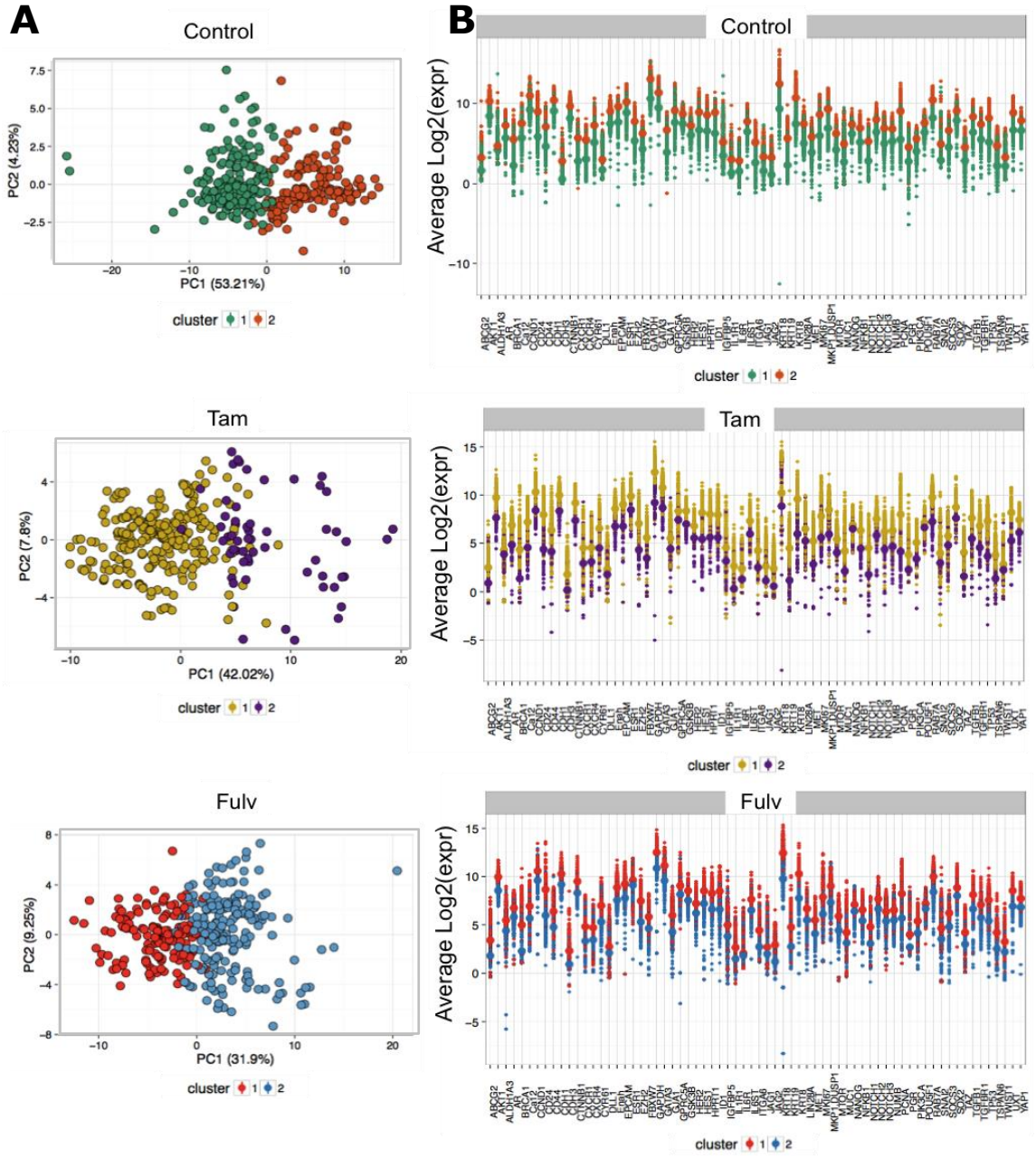


Figure 5. 7 Identification of cell doublets using Mclust in ALDH pos MCF-7 cells.
A) Figure shows PCA using the gene expression data from 69 genes in single ALDH pos cells. PCA of ALDH pos cells shows two distinct cell clusters in each treatment group (top: control; middle: Tam; bottom: Fulv). **B)** Log2 gene expression values of each investigated gene were averaged across all cells for each cluster and each treatment in order to identify singlets and doublets. Top graph: Control cluster 1 (green; singlets), Control cluster 2 (orange; doublets) Middle graph: Tam cluster 1 (yellow; doublets), Tam cluster 2 (purple; singlets). Bottom graph: Fulv cluster 1 (red; doublets), Fulv cluster 2 (blue; singlets).

Table 5. 1 Percentage of ALDH pos cells in each singlet-cluster and doublet-cluster.
Single cell clusters are represented in red.

	Cluster 1	Cluster 2
Control	56.1%	43.9%
Tam	81.2%	18.8%
Fulv	45.4%	54.6%

5.2.4 Identification of ALDH pos subclusters with Mclust and refinement via hierarchical clustering

Upon cell doublet exclusion, we wanted to identify cell subpopulations within the ALDH pos cells across the three conditions tested. Control, Tam and Fulv-treated ALDH pos cells were subjected to clustering using the package Mclust in R (Fraley et al. 2017). This resulted in 5 clusters for control, 7 clusters for Tam and 8 clusters for Fulv. The proportion of cells in each cluster is shown in Table 5.2. As our main objective is to compare representative clusters between treatments, it was decided to merge small related clusters in order to reduce their number. To do so in an unbiased manner, we used the hierarchical clustering with bootstrapping resampling using the functions provided in the R package Pvclust (Suzuki et al. 2015) to group the smaller clusters into large ones. Pvclust analysis and results are represented in Figure 5.8. The control cells, represented in the top dendrogram, are divided into 5 different subpopulations (Ctrl 1-5). Cell populations Ctrl 1-4 showed very high Approximately Unbiased (AU) *p*-value (%) and Bootstrap Probability (BP) values (%), which indicates how strong these clusters are supported by the data. AU is considered as a more reliable p-value. For a cluster with AU *p*-value greater than 0.95, the null hypothesis that "the cluster does not exist" is rejected with significance level 0.05. Cluster Ctrl-5 contains only two cells, which is not a large enough bootstrap sample, therefore it does not group with the rest of the control clusters identified. Tam-treated ALDH pos cells were initially divided into 7 different cell clusters. Tam-1, Tam-2 and Tam-6 cell clusters showed a great degree of homology amongst them, compared to

the other Tam clusters, as shown by an AU value of 90. Likewise, data supports the clustering seen between Tam-3 and Tam-4, as it shows an AU value of 100. Tam-7 only contains two ALDH pos cells, which is not large enough to perform bootstrap analysis. Clustering of the Fulv-treated ALDH pos cells identified eight distinct subpopulations of cells, although AU values were not as strong as in Ctrl and Tam. These results suggest that the clustering performed in the Fulv ALDH pos cells is less supported by the data than the clustering seen in Control and Tam-treated cells.

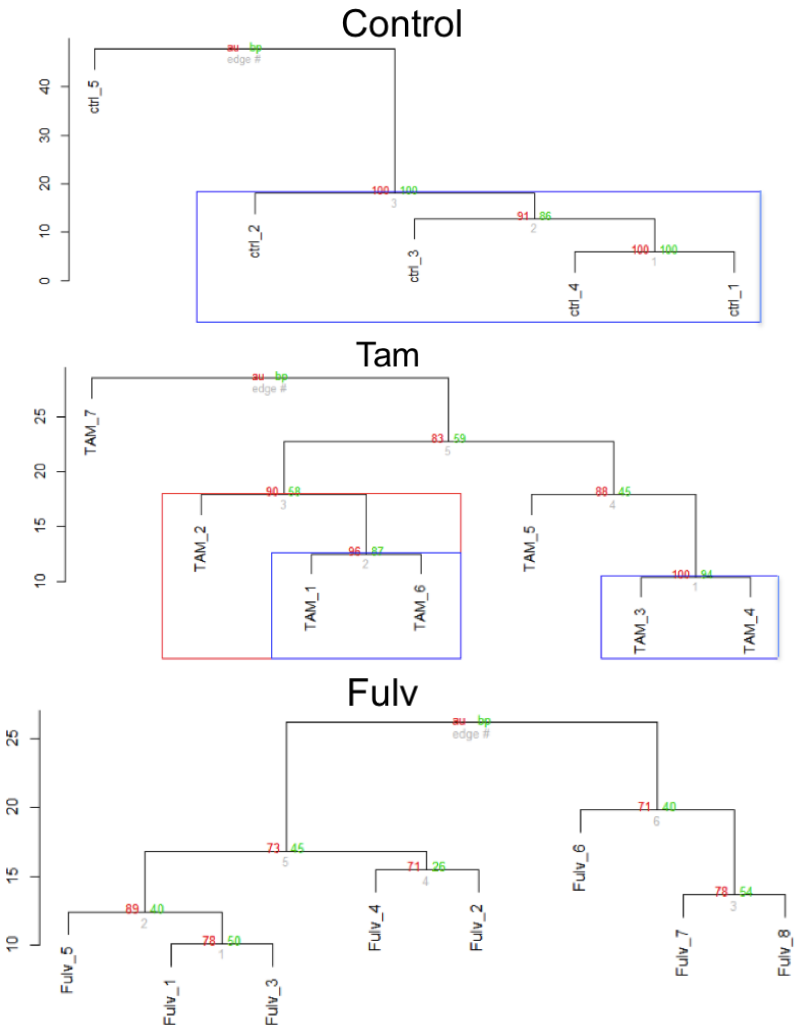


Figure 5. 8 Hierarchical clustering of single ALDH pos cells using Mclust.
Single ALDH pos cells were clustered using Mclust for each treatment. Control (top), Tam (middle), Fulv (Bottom). AU (red) and BP (green) numbers in branches correspond to p-values that provide information on how strong the cluster (branching) is supported by the data. Red and blue boxes represent AU p-value of at least 90% and 95% confidence interval respectively.

Next, to ensure the cell clusters identified using Mclust were not experimental artefacts, but were genuine biological differences, the data was separated into the individual experiments (Table 5.2). This allowed us to identify the contribution of each experiment to generating the clusters, and revealed that cells from the same experiment belonged to different cell clusters, rejecting the idea that a batch effect helped defining the subpopulations.

Table 5. 2 Number of ALDH pos cells that belong to each cluster and each experiment.

Control

Experiment N°	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4	CLUSTER 5	Total N° cells
1	7	1	4	23	0	35
2	53	4	5	34	0	96
3	22	2	3	38	2	67

Tam

Experiment N°	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4	CLUSTER 5	CLUSTER 6	CLUSTER 7	Total N° cells
1	2	5	2	0	0	0	0	9
2	6	0	10	6	2	4	1	29
3	0	0	6	4	2	8	1	21

Fulv

Experiment N°	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4	CLUSTER 5	CLUSTER 6	CLUSTER 7	CLUSTER 8	Total N° cells
1	46	0	0	3	0	0	0	0	49
2	51	2	6	13	2	2	2	2	80
3	30	2	6	16	2	2	2	2	62

Some of the clusters identified above (Figure 5.8) are very similar to each other and consist of too few cells. Therefore, in order to simplify the data and analyse only the relevant subpopulations of cells, some of the clusters were merged and others excluded (Ctrl-5 and Tam-7) from the analysis. Taking into account the hierarchical clustering and the AU/BP values, the merging of the clusters was carried out as follows:

Cluster 1 (**Ctrl-1** from here onwards): Ctrl-1 + Ctrl-4
Cluster 2 (**Ctrl-2** from here onwards): Ctrl-2 + Ctrl-3
Cluster 6 (**Tam-6** from here onwards): Tam-1 + Tam-2 + Tam-6
Cluster 7 (**Tam-7** from here onwards): Tam-3 + Tam-4 + Tam-5
Cluster 3 (**Fulv-3** from here onwards): Fulv-1 + Fulv-3 + Fulv-5
Cluster 4 (**Fulv-4** from here onwards): Fulv-2 + Fulv-4
Cluster 5 (**Fulv-5** from here onwards): Fulv-6 + Fulv-7 + Fulv-8

5.2.5 Investigation of the relationship between clusters from different treatments

The next step in the single cell data analysis was to investigate what the similarities and differences between control and Tam and Fulv cell clusters were. Hierarchical clustering using Mclust was applied in the newly merged clusters, as shown in Figure 5.9. The dendrogram illustrates the final clusters obtained after single cell data processing, which are 2 control clusters, 2 Tam clusters and 3 Fulv clusters. Cellular clusters were further grouped into clades, which can be defined here as a group of highly related branches of a dendrogram. Ctrl-2 and Fulv-4 clusters (Clade A) group together and are different from the rest of the clusters, as supported by an AU value greater than 95. On the right hand side of the dendrogram Tam-7, Ctrl-1 and Fulv-3 are also clustered together in a separate group (Clade B) (AU:100%, BP:100%). Tam-6 and Fulv-5 belong to different branches of the dendrogram, which are different from Clades A and B. The exact number of ALDH pos cells per cluster is shown in table 5.3.

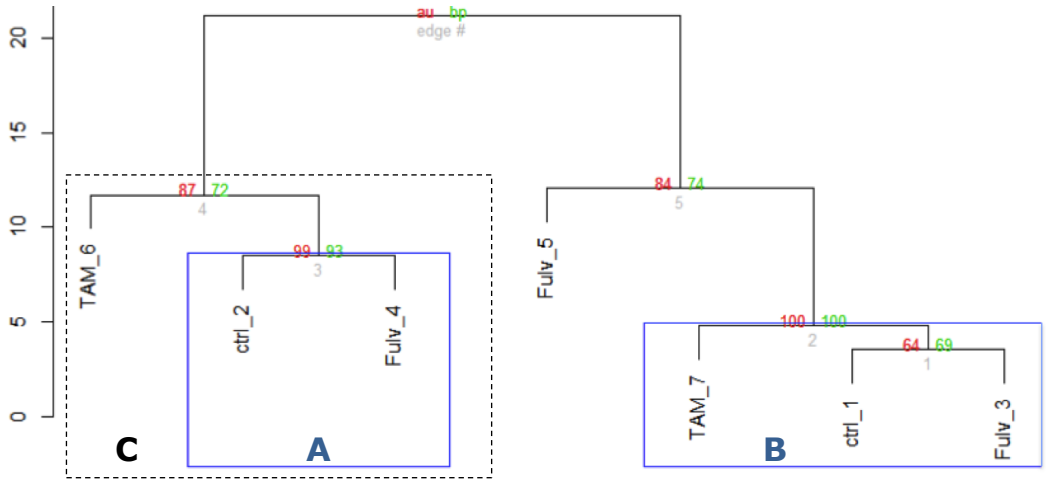


Figure 5. 9 Hierarchical clustering of ALDH pos cells using Mclust after merging the clusters.
Following the merge of different clusters, explained in section 6.2.4, the newly created clusters were plotted in a dendrogram to visualise how they relate to each other. AU (red) and BP (green) numbers in branches correspond to p-values that provide information on how strong the cluster (branching) is supported by the data. Blue boxes represent AU p-value of at least 95% confidence interval. Ctrl-2/Fulv-4 cluster group together as Clade A and Tam-7/Ctrl-1/Fulv-3 as Clade B. Further Clade C (dotted line), based on DAPC analysis in Figure 5.11, is discussed in page 165.

Table 5. 3 Number of ALDH pos cells per cluster after merging using Mclust.

	CLUSTER 1	CLUSTER 2
Control	177 (90.3%)	19 (9.7%)

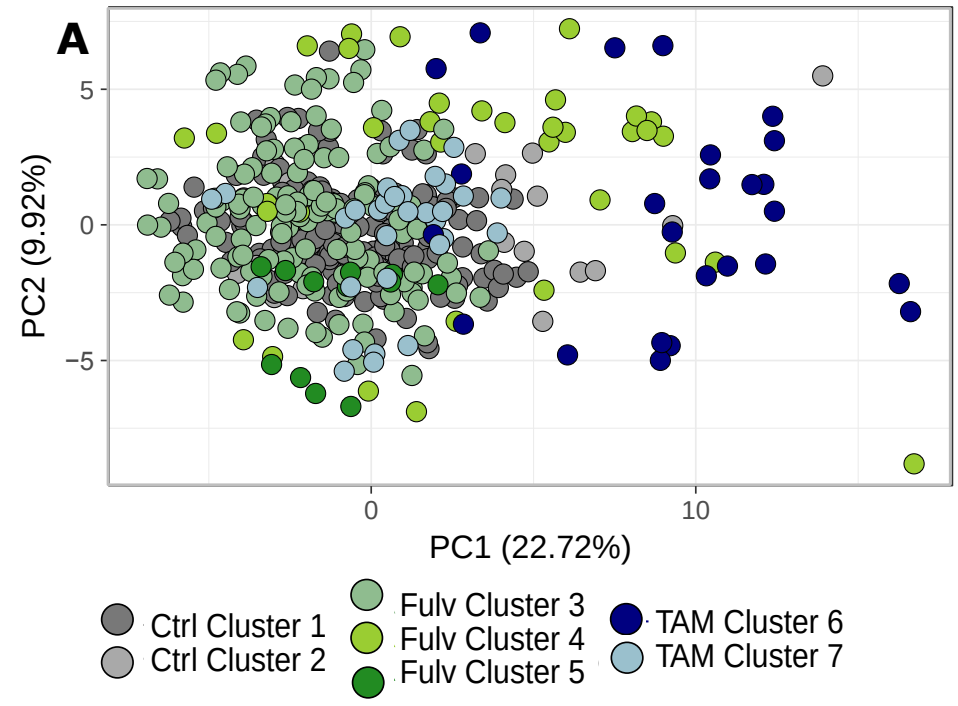
	CLUSTER 6	CLUSTER 7
Tam	25 (43.9%)	32 (56.1%)

	CLUSTER 3	CLUSTER 4	CLUSTER 5
Fulv	143 (74.9%)	36 (18.8%)	12 (6.3%)

Next, PCA was applied to the newly merged clusters and the PCA score and loadings plots are represented in Figure 5.10 A & B respectively. PCA is a data transformation that produces new uncorrelated variables called Principal Components (PCs), that capture the maximum amount of variance in the data in decreasing order (the first PC captures the most variance whereas the last PC captures the least). This effectively "compresses" the data allowing most of the information to be presented in just a few variables, which can be easily visualised. The distance between individual ALDH pos cells, represented by coloured circles in the PCA, illustrates their similarity. The 7 clusters observed in the hierarchical clustering shown above cannot be easily discriminated in the PCA below, which represents two PCs only, however data shows a trend for the appearance of some cellular clusters. PCA score results show that Tam-6 (dark blue) and Fulv-5 (dark green) are significantly separated from each other and from the rest of the clusters, which highlights their gene expression differences. Some of the cells that belong to Ctrl-2 (light grey) overlap with some of the cells from Tam-6, which indicates that a small number of control cells are more similar to a population of Tam-treated cells than to its own control group. Fulv-4 (bright green) contains cells scattered throughout the entire PCA, however some of them are also related to Tam-6 and to Ctrl-2. Tam-7 (light blue), Ctrl-1 (dark grey) and Fulv-3 (light green) overlapped each other completely, which indicates high similarity amongst them. Overall, these data support the results observed in the above dendrogram,

however using two PCs is insufficient to recapitulate all the cellular heterogeneity observed when applying hierarchical clustering using Mclust (Figure 5.9).

PCA loadings (Figure 5.10 B) indicate how each factor, in this particular case each gene, is associated with the observable variables, in this case gene expression, for this particular analysis. When looking at the PCA loadings, a subset of 8 genes circled in red can be seen at the top right of the graph. The enclosed genes present high correlation between them and at the same time they are inversely correlated with the rest of the genes. This cluster of genes include *LIN28A*, *NANOG* and *POU5F1* (also known as *OCT4*), which were identified by Yu and colleagues as three of the four factors necessary for human somatic cells to reprogram into pluripotent stem cells, retaining the essential characteristics of embryonic stem cells (Yu et al. 2007). CXCR1, another gene identified in the PCA loadings plot as potentially relevant for the cellular heterogeneity observed in our study, has been shown to be important for CSC maintenance. Ginestier et al., used antibodies or chemical inhibitors to target the *CXCR1* receptor in breast cancer cell lines and reported a delay in tumour growth and a reduction in metastasis (Ginestier et al. 2010). Likewise, another chemokine receptor, *CXCR4*, was also amongst the 8 genes identified. Our group has observed that CSCs, isolated as anoikis resistant cells, had increased *CXCR4* mRNA and protein levels, compared to non-CSCs (Ablett et al. 2014). Moreover, *TWIST1* has been found to contribute to metastasis by downregulating the cell-cell adhesion marker E-cadherin and promoting an epithelial-to-mesenchymal transition (Yang et al. 2004).



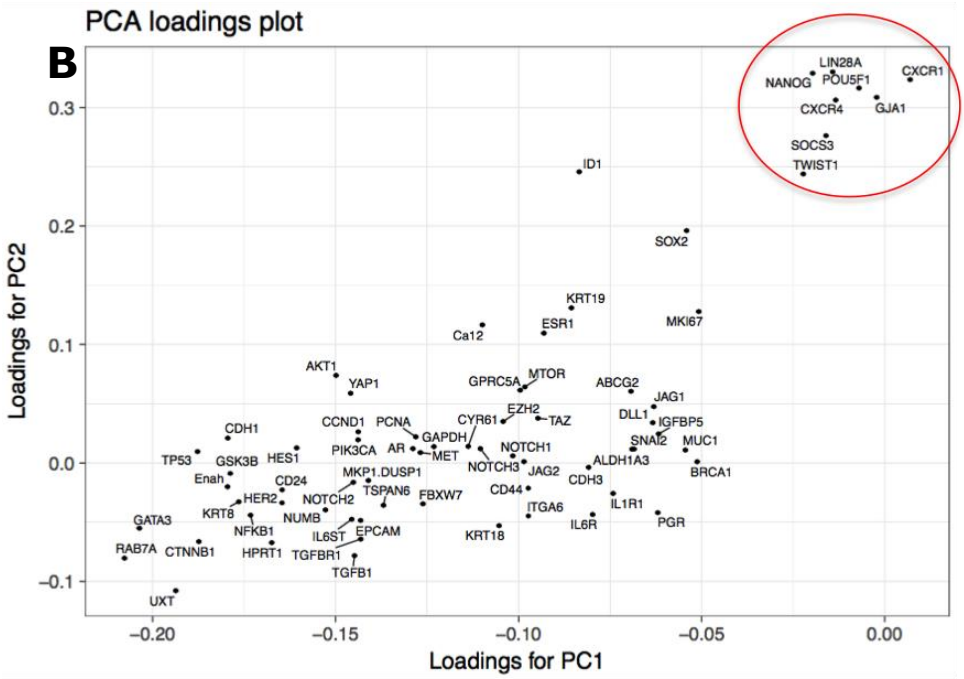


Figure 5. 10 PCA of the merged ALDH pos cell clusters.
A) Figure shows the PCA score plot of the two first principal components what accounts for almost 30% variance of the data. Each point represents a single ALDH pos cell. Points are coloured by the cluster they were assigned to, with control cells coloured in grey, Fulv treated in green and Tam in blue. **B)** PCA loading plot shows the contribution and relation of each gene for each principal component shown in Figure 5.10 A.

In order to further study the genes that are significantly different between clusters, an analysis of variance (ANOVA) test was carried out followed by a Tukey's Honestly Significant Difference (HSD) Post Hoc test. Gene expression differences between Clades A and B in Figure 5.9 were investigated and the significantly differentially expressed genes are listed in table 5.4. Anova data revealed 40 genes differentially expressed between Clades A and B. Overall, most genes show higher gene expression values in Clade B compared to Clade A. Some of these genes are key regulators of the Notch signalling pathways such as *DLL1* and *HES1*, the Wnt signalling pathway such as *CTNNB1* (β-catenin) and *CDH1* (E-cadherin), the Hippo pathway such as *TAZ* and *YAP1* or cytokine signalling such as *IL6ST* or *CXCR1*. Luminal and epithelial markers such as KRT8, KRT18, GATA3, CD24 and MUC1 were also found to be upregulated in Clade B compared to A. Clade B appears to be a more proliferative cell population as indicated by higher mRNA expression of the proliferating cell nuclear antigen (*PCNA*) and the proliferative marker *MKI67*.

Table 5. 4 List of genes differentially expressed between Clade A and Clade B identified using Mclust.
The two main cellular clades circled in blue boxes (Clade A and B; AU p-value ≥ 95%) found in figure 5.9 were compared to identify differentially expressed genes. The statistically significant genes found after performing Anova and a Tukey´s HSD Post-hoc test between the two clades are shown below. When p-value is very close to 0 then p-value=0.

Gene name	P-value	Cellular Process
AKT1	0	PI3K pathway
PIK3CA	0.0000194	
CCND1	0	Cell cycle
GAPDH	0	Metabolism
HES1	0	Notch signaling
DLL1	0.002685438	
FBXW7	0.005447442	
HPRT1	0	Purine metabolism
KRT18	0	Luminal-associated cytokeratin. Apoptosis and survival caspase cascade
KRT8	0	
RAB7A	0	Regulators of vesicular transport
TP53	0	Chromatin binding, actin filament binding
UXT	0	Transcription factor activity
GATA3	9.49E-10	DNA binding and transcription factor binding
NFKB1	6.08E-09	Cytokine signalling, development IGF-1 receptor signaling
EPCAM	0.00000003	Cell adhesion
IGFBP5	0.000000165	IGF-1/AKT signaling
Enah	0.000000535	Actin binding, cytoskeleton remodelling
CTNNB1	0.000000613	Wnt signaling
HER2	0.00000489	EGFR signalling, GPCR pathway
IL6ST	0.00000799	Cytokine signalling, MAPK1/3 activation

TGFB1	0.0000241	TGFB/SMAD signaling
TGFBR1	0.000327989	
POU5F1	0.000139475	Embryonic development and stem cell Pluripotency
LIN28A	0.030878998	
CD24	0.000291429	Developmental biology
CDH1	0.000352286	Calcium-dependent cell-cell adhesion
YAP1	0.0004595	Hippo signalling pathway
TAZ	0.002511528	
Ca12	0.002395943	Nitrogen metabolism
PGR	0.003797786	Progesterone signalling
MET	0.003920032	Tyrosine kinase signalling, IGF-1 receptor signalling
GPRC5A	0.004437088	G-protein signalling, RA signaling
CXCR1	0.004438069	Cytokine signalling, G-protein signalling
PCNA	0.006282843	DNA replication, IGF-1 signalling
CD44	0.032433079	Degradation of the extracellular matrix, migration
CYR61 (IGFBP10)	0.034363712	Cell proliferation, cell adhesion
MUC1	0.038328673	MAPK, PI3K/AKT
MKI67	0.045511902	Cell proliferation, DNA damage
MKP1.DUSP1	0.046439773	MAPK

As a strategy to further separate the clusters identified through PCA, the Discriminant analysis of principal components (DAPC) was applied to centre and unit variance scaled data. DAPC can be used to infer the number of clusters of genetically related individual cells. Contrary to traditional methods such as PCA, which focus on the entire genetic variation, DAPC yields linear combinations of the original variables, which maximize differences between clusters while minimizing variation within clusters. DAPC was performed using the 'DAPC' function in the 'adegenet' package (version 1.4-2) (Jombart et al. 2016). Cross-

validation was performed to identify the number of principle components needed to retain the linear discriminant analysis (LDA) model. For cross-validation the data was separated into two sub-data sets: the training dataset (80% of the data) and the validation dataset (20% of the data). The training dataset was used to produce an LDA model with varying numbers of principal components; these models were then used to classify the validation set. The most appropriate model was identified as that which provides the maximum correct prediction with the lowest number of principle components. This process was repeated iteratively 30 times for each model. The final model was built using 40 principal components and 8 linear discriminants (LD). Figure 5.11 shows a DAPC from the control (grey), Fulv-treated (green) and Tam (blue) ALDH pos cells, representing the most important LDs (LD1 and LD2). DAPC reveals greater differences between the previously identified clusters, compared to the PCA shown above. Fulv-4 (bright green) and Tam-6 (dark blue) separate clearly along LD1 from the rest of the clusters, whereas Fulv-5 (dark green) is different from the rest of the clusters mainly along LD2. The results obtained from the DAPC support the ALDH pos hierarchical clustering seen after applying Mclust (Figure 5.9). Overall, this data identified two major clades (clade A and clade B) that represent two main cellular clusters of related ALDH pos cells in control and after anti-estrogen treatments.

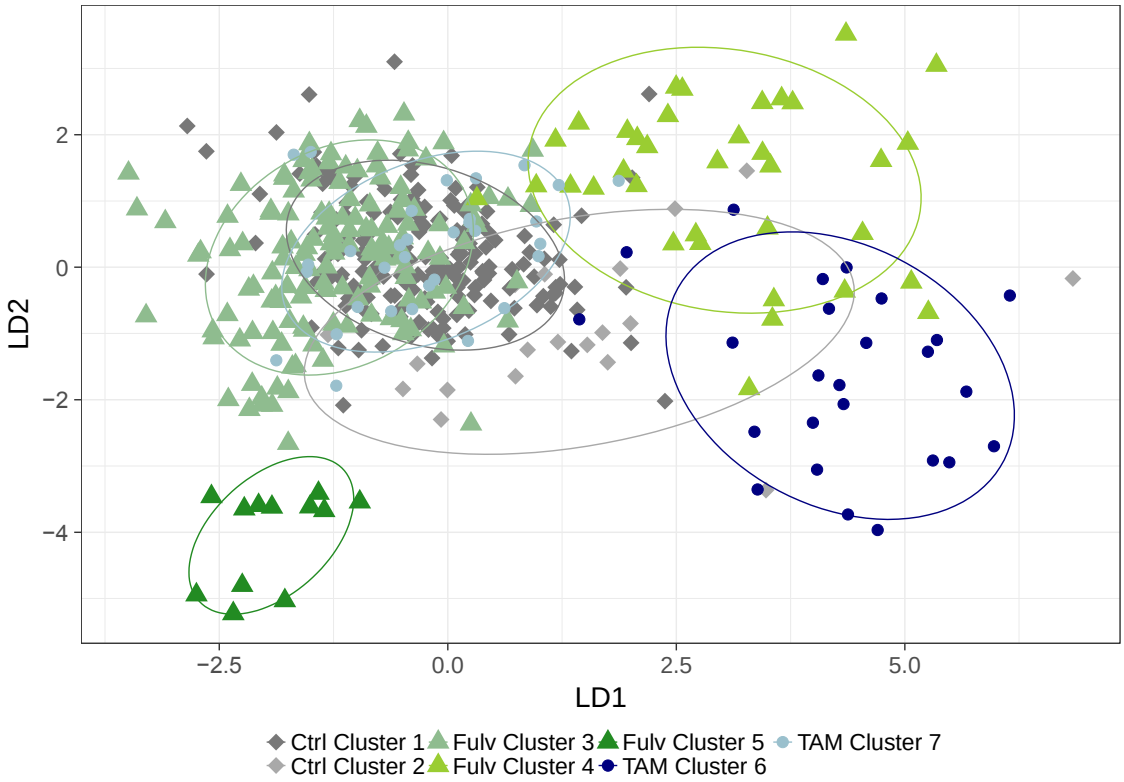


Figure 5. 11 Scatter plot of the DAPC analysis for single ALDH pos MCF-7 cells after anti-estrogen treatment.
The scatter plot shows the cluster of individual ALDH pos cells (rhomboid: control group; triangle: Fulv group; circle: Tam group). Control ALDH pos cells (grey) clustered within two groups (Cluster 1 & 2), Fulv-treated ALDH pos cells (green) clustered within three groups (Cluster 3, 4 & 5) and Tam-treated ALDH pos cells (blue) clustered within two groups (Cluster 6 & 7).

Figure 5.12 shows the PCA loading of the DAPC, which illustrates the genes that contribute the most to the separation observed in LD1 and LD2. The top 3 genes listed are *IGFBP5*, the Carbonic anhydrase 12 (*CA12*) and the Delta-like Notch ligand, *DLL1*. Interestingly, CA12 has recently been identified as a novel prognostic biomarker in HER2+ breast cancers and is associated with breast CSCs. The EMT marker *TWIST1* that clustered with the embryonic genes shown in figure 5.10 B is also part of the top 25% genes that contribute to the separation observed in the DAPC (Figure 5.12). These results suggest that the genes from the 8 embryonic gene signature are not relevant for the segregation of the clusters observed in the hierarchical clustering and the DAPC. Instead, those 8 genes might have clustered together because they show a similar

trend regarding their gene expression pattern but without any relevance in regards with the cell clusters.

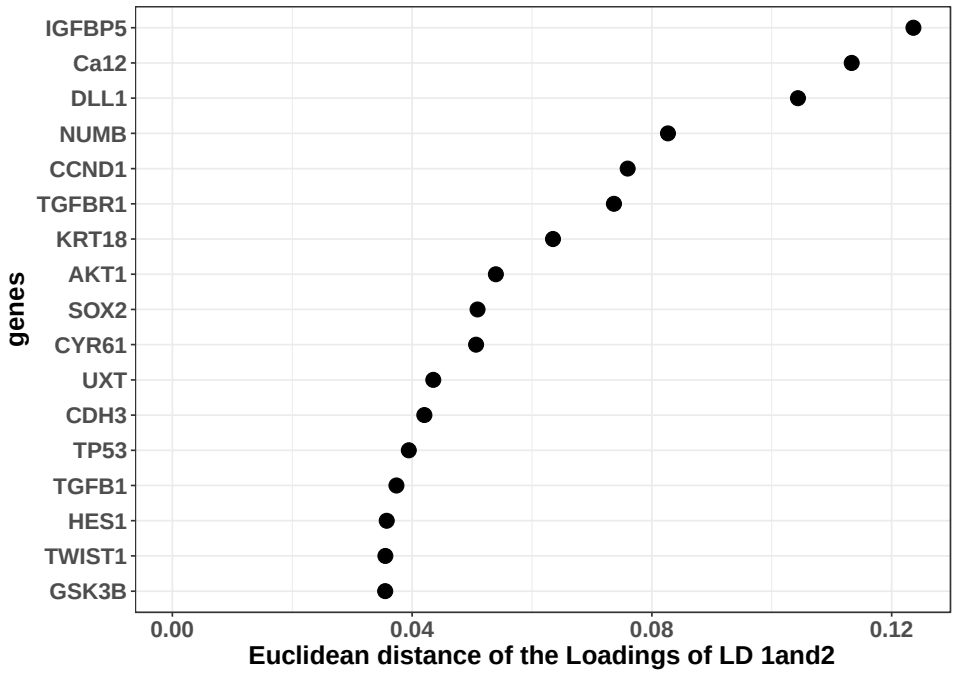


Figure 5. 12 PCA loading of the DAPC.
Only 25% of the most influent genes are shown.

Finally, in order to observe significant gene expression differences amongst ALDH pos cells treated with the same drug, clusters that belonged to the same treatment condition but were located in different clades of the dendrogram were also compared. According to the DAPC (Figure 5.11) clusters Tam-6, Ctrl-2 and Fulv-4 were more similar between each other than to the rest of the cellular clusters and were near neighbours in the dendrogram (Figure 5.9). Therefore, these three clusters are grouped as Clade C. Cellular clusters that belonged to Clade C were individually compared to the clusters from clade B (Tam-7, Ctrl-1 and Fulv-3) that received the same treatment. The gene expression comparison was as follows: Ctrl-2 versus Ctrl-1, Tam-7 versus Tam-6 and Fulv-4 versus Fulv-3. The common significantly differentially expressed genes amongst control, Tam and Fulv group from clades C and B are represented in Table 5.5. These 12 genes are involved in a variety

of cellular processes such as the PI3K/AKT pathway (*AKT1*), the Wnt signalling pathway (*CTNNB1*) or cellular adhesion and metastasis (*EPCAM*). Overall, gene expression is higher in Clade B compared to Clade C, and for each of the individual genes.

Table 5. 5 Common significantly differentially expressed genes between Clades C and B.
The gene expression comparison was performed as follows: Ctrl-2 versus Ctrl-1, Tam-7 versus Tam6 and Fulv-4 versus Fulv-3. Only the common differentially expressed genes are shown.

Gene name
AKT1
CTNNB1
EPCAM
HPRT1
IL6ST
KRT18
KRT8
NFKB1
RAB7A
TGFBR1
TP53
UXT

5.3 Discussion

This chapter aimed at finding whether ALDH pos MCF-7 cells comprise a homogeneous or a heterogeneous population of cells and whether anti-estrogen treatment can drive cell diversity through the accumulation of changes in their gene expression pattern. By interrogating the gene expression levels of 96 genes associated with stemness and cancer in 444 single ALDH pos cells (following 6-day treatment with control, Tam or Fulv), it was possible to determine the diversity within this cellular compartment. The extensive bioinformatics analysis using hierarchical clustering and PCA that was applied to the single cell data set revealed the existence of distinct clusters of ALDH pos cells before and after anti-estrogen treatment. The present piece of work provides the first evidence of the existence of cellular diversity within the breast CSC population (ALDH pos cells) in MCF-7 cells.

The C1 system (Fluidigm) technology enabled us to assess the expression of up to 96 genes in single ALDH pos cells by performing cell loading, capture, washing, lysis, reverse transcription and amplification within the C1 chip, which increases the reproducibility of single cell data and reduces the possibility of variability or contamination. C1 plates come in three different sizes based upon cell type and size. The medium chip, which captures cells ranging from 10-17 μm was the one used for this study since the MCF-7 cell line has been reported to have a diameter of 11.95 μm (SD 3.22) (Nexcelom 2017), however MCF-7 cells are highly heterogeneous regarding cell size (as it can be seen using the forward/side scatter in a Flow cytometric plot) and also as it has been described in the literature (Frimat et al. 2011, Ding et al. 2014). Therefore, it is possible that our studies might be focusing on a subpopulation of cells with a particular cell size (10-17 μm), and that this approach might neglect smaller and bigger cells, which cannot be captured within the medium C1 chips. Nevertheless, despite this drawback, the C1 system was the best automated single cell technology available at the beginning of the present study.

The 96-gene list used to explore ALDH pos cell diversity (Figure 5.1) was generated based on the MCF-7 and PDS Affymetrix microarray

data shown in chapter 4 and on the available literature. Prior data analysis, single cell data was filtered in order to discard the genes that were expressed in less than 3 cells per Biomark plate and also those genes that did not show amplification during the qPCR. Therefore, only 69 genes were assessable out of the initial 96 genes tested. This reduction in the number of genes used to assess heterogeneity could have limited the power to detect cellular diversity in MCF-7 cells.

The present piece of work showed that anti-estrogen treatment enriches for ALDH pos cells and not for $CD44^+/CD24^-$ cells. Therefore, in order to isolate a pure epithelial ALDH pos population for gene expression analysis, it was necessary to exclude the small percentage of cells that co-express ALDH and $CD44^+/CD24^-$ markers (Figure 5.2). Following cell sorting, ALDH pos cells were loaded onto the C1 system and were also visually inspected under the microscope to record the cell capture efficiency. Our workflow revealed that not all 96 chambers captured cells and that many others contained debris or dead cells, which decreased the overall efficiency of the C1 system. Moreover, in 2016 Fluidigm reported that a large proportion of cell doublets were being captured within the C1 medium chips in the form of stacked doublets (two cells stacked one over the other in the capture site) as a consequence of the capture architecture of the medium C1 chips (Fluidigm Corporation 2016). However, only the mouse fibroblast NIH3T3 and the human embryonic kidney HEK293 cell lines were used in the Fluidigm report. In an attempt to define the exact percentage of cell doublets in our cell line of choice, MCF-7 cells were stained using a red and a blue nuclear dye and by microscopically inspecting the C1 chambers it was reported that the ratio of cell doublets in MCF-7 cells using the C1 system technology was around 58%. Two distinct groups of cells were identified through averaging Log2 expression values of each gene for all single cells studied in each condition via PCA (Figure 5.7). These results showed the existence of a cluster of cells with much higher expression of all genes investigated, which were considered doublets and subsequently discarded from the analysis, and another cluster of cells with reduced gene expression levels, which were defined as singlets. Although the doublet exclusion method was a necessary step in order to be able to study the

ALDH pos cell diversity, the PCA approach used here might not have been completely effective at identifying cell doublets and might have introduced bias into our cellular analysis. For instance, Tam Cluster 1, which was considered to contain doublets, constituted 81.6% of all Tam cells, which is much greater than the 58% of doublets observed in our MCF-7 pilot experiment. Therefore, these results suggest about 23.6% of the doublets excluded from Tam Cluster 1 might in fact be singlets (Figure 5.7). In addition, this approach reduced the number of single cells studied by more than half, which limits the scope of our findings, particularly for Tam treatment.

In order to study whether control, Tam and Fulv-treated ALDH pos cells comprised different subpopulations of cells, hierarchical clustering using Mclust was applied to the data set (Figure 5.8). This hierarchical clustering aims at partitioning the data into homogeneous groups, such that the within-group similarities are greater than the between-group similarities. 5, 7 and 8 clusters were identified in control, Tam and Fulv-treated ALDH pos cells, respectively, which pointed towards the existence of phenotypic differences within small clusters of cells. It is important to mention that some of the clusters contained very few cells. In particular, the clusters from before the merge Ctrl-5 and Tam-7 (Figure 5.8 and Table 5.2) only had two cells in each cluster, which puts into question their biological relevance. Therefore, those two clusters were excluded from the analysis and some others were merged into bigger clusters taking into account the hierarchical clustering and the probability values (p values) shown in Figure 5.8. Two main clades of cells were identified in the dendrogram of the newly merged clusters. Clade A contained clusters Ctrl-2 and Fulv-4 and Clade B consisted of clusters Ctrl-1, Tam-7, and Fulv-3 (Figure 5.9).

PCA was applied to the dataset to identify the patterns encoding the highest variance and facilitate data exploration and visualisation (Figure 5.10). However, overall PCA did not infer clear and distinct genetic clusters. Therefore, DAPC was used next in order to maximise the discrimination between groups seen in the PCA (Figure 5.11). DAPC plot revealed clear distinct clusters of related cells such as Ctrl-2, Fulv-4 and Tam-6, which correlate highly with the hierarchical clustering shown in

Figure 5.9. Using DAPC loadings, the top 25% of the genes responsible for the spatial distribution of the clusters seen in the DAPC were identified (Figure 5.12). This new gene list contained 17 genes and some of them have been shown to be important at facilitating EMT such as *CDH3* (P-cadherin) or *TWIST1*, regulating the Notch signalling pathway such as DLL1, NUMB or HES1 and also at regulating developmental processes such as IGFBP5.

The present study has identified so far different cellular clusters within the ALDH pos population from MCF-7 cells and a list of genes relevant for such cellular segregation. However, it was still necessary to specify which genes are important for the generation of each particular cell cluster. Therefore, an Anova test was performed on the dataset to compare the clusters followed by a Post Hoc test in order to identify the differentially expressed genes between Clade A and Clade B. A total of 40 genes were found to be differentially expressed between Clade A (Ctrl-2, Fulv-4) and Clade B (Ctrl-1, Tam-7 and Fulv-3) and overall, most genes were upregulated in the latter. Clade B showed higher expression of most genes studied in comparison to Clade A, including the proliferation markers *PCNA* and *MIK67*. These data suggest that Clade B may comprise highly proliferative cells that might be in S, G2 or M phase of the cell cycle; whereas Clade A may contain a group of cells that have entered in G0 or G1 phase of the cell cycle, which can be described as a state of quiescence.

Next, since clusters Tam-6, Ctrl-2 and Fulv-4 were more similar to each other than to the rest of the cellular clusters, according to the DAPC, they were grouped in what was known as Clade C (Clade A plus Tam-6). In an attempt to investigate the gene expression differences between opposed cellular clusters (based on the hierarchical clustering) but that had received the exact same treatment, an Anova test was performed on the dataset followed by a Post Hoc test. Ctrl-2 was compared to Ctrl-1, Tam6 to Tam-7 and Fulv-4 to Fulv-3. Only those differentially expressed genes that were shared across the three treatment comparisons were listed (Table 5.5). Epithelial markers such as *EPCAM* and cytokeratins such as *KRT8* and *KRT18* were amongst the 12 differentially expressed genes found in this analysis. Altogether, these

results suggest that those cells that belong to Clade B may show a more epithelial and proliferative phenotype.

Single cell data analysis is being widely used to investigate cellular heterogeneity. There are 2 new relevant studies in breast cancer that aim to define stem cells, Akrap et al (2016) and Lawson et al (2015). Akrap and colleagues applied single cell analyses to compare CSCs from MCF-7 cells enriched with three different functional strategies: mammospheres, growth in hypoxia and PKH26 retention. Researchers reported a hierarchical cellular organisation model in ER+ and ER- tumours supported by single cell data from primary breast tumours (Akrap et al. 2016). Lawson et al., applied the Biomark technology that we have used to resolve cellular diversity at the single cell level during breast cancer metastasis by investigating 116 genes involved in stemness, EMT and dormancy amongst others cellular processes, which further support our investigation of cellular heterogeneity present in breast cancers (Lawson et al. 2015). In our study, we reveal for the first time the existence of distinct cellular clusters within the ALDH pos anti-estrogen resistant stem cell population of ER+ MCF-7 cells. This study extends the previous work of Simões et al by demonstrating that anti-estrogen treatment enriches for 3 different ALDH pos populations that have might have functional significance, possibly a proliferative and a quiescent population and one of unknown function.

In summary, two major cellular clusters (Clade A and Clade B) have been identified in control and treated ALDH pos cells based on the expression of 69 genes associated with stemness and cancer. Based on the higher gene expression levels of markers of proliferation such as MKI67 and PCNA, it can be hypothesised that Clade B comprises a more proliferative group of cellular clusters. At the same time, these clusters of cells might show a more epithelial phenotype, as indicated by higher gene expression of EPCAM, KRT8 and KRT18 markers. Conversely, Clade A comprises two clusters that may exhibit a more quiescent phenotype with overall lower gene expression levels than Clade B. In addition, Fulv treatment led to the identification of a third cellular cluster (Fulv-5), which consists of 12 Fulv-treated ALDH pos cells. It remains to be discovered if this also exists after TAM treatment since four-fold less cells

were analysed, compared to control and Fulv-treated ALDH pos cells. Based on the ALDH pos MCF-7 single cell data presented in this study, a model representing the cellular diversity of this population before and after anti-estrogen therapy can be drawn. This testable model (Figure 5.13) illustrates how anti-estrogen treatment with Tam and Fulv might enrich for a more quiescent group of cells (Clade A). According to this hypothesis, highly proliferative clusters of cells, which represent the majority of the ALDH pos cells, might be targeted by anti-estrogen treatment, which may lead to an enrichment of a quiescent population of cells (Clade A). Our data also points towards the existence of a small group of Fulv-treated ALDH pos cells (Fulv-5) that emerge only following anti-estrogen treatment.

These findings will inform how to better target anti-estrogen resistant cells with additional therapies.

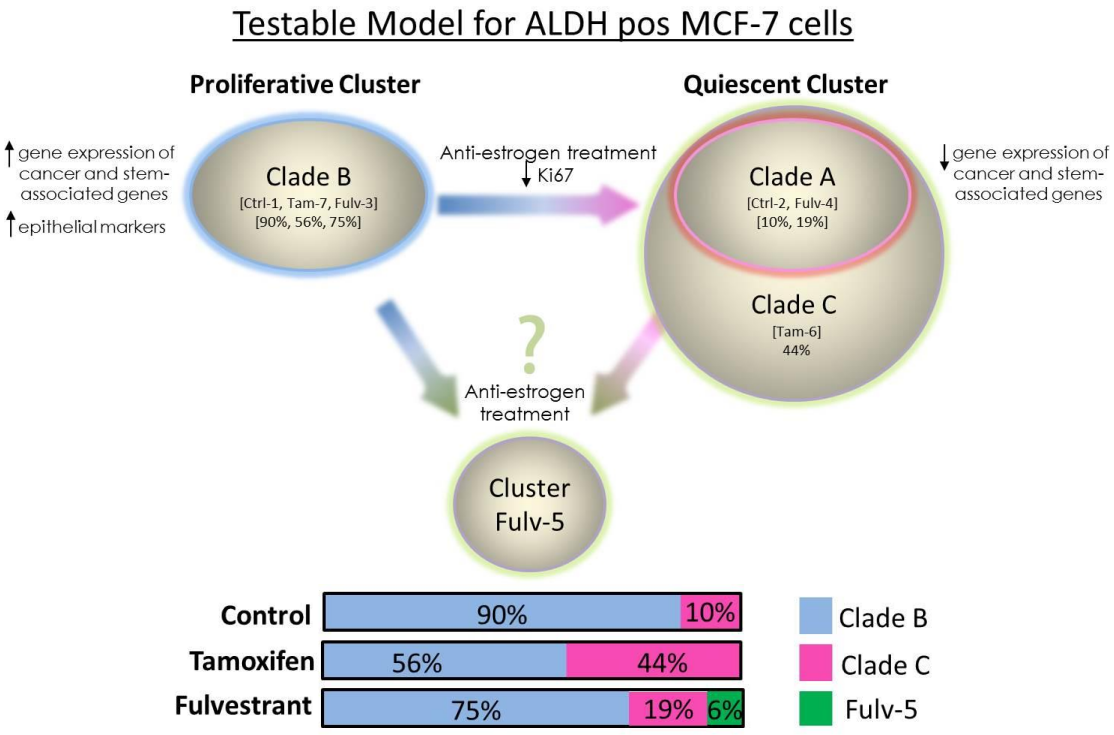


Figure 5. 13 Working model for ALDH pos cellular heterogeneity before and after anti-estrogen treatment in MCF-7 cells.
Top figure shows a schematic representation of the three clades and the cluster identified in control, Tam and Fulv-treated ALDH pos MCF-7 cells. Bottom figure represents the percentage of ALDH pos cells that belong to each clade/cluster for each treatment.

6. Chapter 6: Discussion and future perspectives

Breast cancer is the most commonly diagnosed cancer in women and it is estimated that 508,000 women died from this disease worldwide in 2011, a number that averages to approximately one death every minute (World Health Organization 2017). Estrogen fuels the growth of the largest molecular subtype of breast cancer, the ER+ subtype, which comprises around 75% of all breast cancer cases. Anti-estrogen drugs such as Tam and Fulv impair the estrogen signalling in these tumours and are commonly used as standard-of-care to treat premenopausal ER+ breast cancer patients. Tam has been shown to reduce by almost one-half the 10-year risk of recurrence and by around one-third the risk of death (Clarke et al. 2015) (Early Breast Cancer Trialists' Collaborative 2011). Nevertheless, more women die from ER+ breast cancers than from any other breast cancer subtype. In addition, late recurrences (several decades after completion of the adjuvant therapy) are very common in ER+ breast tumours, whereas tumours with worse prognosis (ER- tumours) will normally develop secondary tumours within the first 3-5 years after diagnosis. It has been argued that breast CSCs might play an important role in this late relapse. CSCs show anti-apoptotic capabilities that would enable them to halt their growth in a hostile microenvironment (i.e. hypoxia) until conditions are favourable to drive formation of a secondary tumour or to develop metastasis (Clarke et al. 2015) (Patel et al. 2012). Therefore, it is crucial to understand the molecular processes important for CSCs to drive anti-estrogen resistance in breast cancers. CSCs can be enriched using different combinations of cell surface markers or assays that test ALDH enzymatic activity, followed by FACS sorting. However, it is established that not all the sorted cells have CSC properties such as the capabilities to form tumours *in vivo* or grow spheres under low adherent conditions. To date, there are not general markers that can isolate to purity the truly malignant CSCs. Therefore, sorted, putative CSCs may actually be a mixture of differentiated, progenitor and CSCs instead, highlighting the need for studying these cells as individual entities. In addition, cellular plasticity

and bidirectional conversion between CSC and non-CSC leads to poor therapeutic response and highlights the need for combinational approaches that target all different cells present in the tumour. Single cell technology is an optimal approach to investigate cellular heterogeneity, which will provide researchers with valuable information to better understand this resistant population.

The present study aimed at providing an insight into the diversity of ALDH pos cells from breast tumours that survive anti-estrogen therapy. Based on the data presented here, several conclusions can be drawn:

1) ALDH pos cells are enriched after anti-estrogen treatment with Tam and Fulv and although the exact mechanism remains elusive, the ALDH1A3 isoform appears to be important at driving this enrichment in MCF-7 cells and ALDH1A1 is important in patient-derived samples (PDS).

2) ALDH pos cells from MCF-7 and from PDS have a distinct gene expression pattern when compared to the non-CSC ALDH neg cells; although it is important to note that patient samples show a much larger list of differentially expressed genes between ALDH pos and neg cells than the MCF-7 cell line.

3) The ALDH pos population in MCF-7 cells is not a homogeneous cellular compartment but is instead divided in at least 2-3 smaller subpopulations of cells with some particular genes contributing more than others to such diversity.

It has been published that CSCs are ER-/low and resistant to the direct effects of anti-estrogen therapy (Simoes et al. 2011). Data from the present study and previous published work from our group have shown that treatment with Tam and Fulv enrich for stem cell activity in breast cancer cells as measured by an increase in the MFE and an enlargement in the percentage of ALDH pos cells (Simões et al. 2015). In addition, knock down studies in MCF-7 cells, shown in chapter 3, revealed that the ALDH1A3 isoform is important at mediating the enrichment of the ALDH pos population. These results support the data published previously by Marcato and colleagues who identified ALDH1A3 as the isoform important for ALDH activity in breast cancer cell lines. Once ALDH pos cells were identified as the population resistant to the anti-estrogen

therapy, the next step was to characterise them and confirm their stem cell properties. It was shown that ALDH pos cells form more mammospheres compared to the ALDH neg cells in ER+ breast cancer patient samples, and also that ALDH pos MCF-7 cells form bigger tumours in mice and have higher tumour-initiating cell frequency by ELDA, than ALDH neg cells. These results support the data from ALDH pos cells published previously by Ginestier and colleagues in normal and cancerous primary breast tissue (Ginestier et al. 2007). In order to identify biomarkers that would allow us to target anti-estrogen-resistant CSCs, ALDH pos cells from MCF-7 and PDS were subjected to microarray analysis. MCF-7 cells showed a smaller number of differentially expressed genes between ALDH pos and ALDH neg cells, compared to the PDS dataset. The PDS dataset found hundreds of genes that were consistently differentially expressed between CSCs and non-CSCs, however inter-patient heterogeneity was also visible especially due to patient samples 68, 71 and 94. Patient sample 71 was the only sample showing HER2 amplification and patient sample 94 was treatment-free, which might explain the genetic differences. Gene ontology enrichment analysis identified the cell adhesion molecules and the cytokine-cytokine receptor interaction pathways being significantly differentially expressed between ALDH pos and ALDH neg cells. When the MCF-7 and PDS datasets were compared, it was found that 7 genes were commonly differentially expressed between ALDH pos and ALDH neg cells. ITGB6 was one of the commonly upregulated genes in MCF-7 and PDS data set and its expression is upregulated in breast cancers according to the literature. In addition, co-upregulation of ITGB6 and HER2 identified one of the worse prognostic sub-groups of breast cancer (Moore et al. 2014). Overall, this approach helped us understand the gene expression profile of ALDH pos cells in the metastatic ER+ cell line, MCF-7 cells and in metastatic patient samples.

Taking into account that the main aim of the present study is elucidating the cellular processes behind the development of anti-estrogen resistance, primary breast cancer samples before and after treatment with anti-estrogen drugs could have been better as opposed to the metastatic samples used in this study. Cells present in metastatic

fluids grow in a non-attached (suspension) form, which is characteristic for less-differentiated cells. Therefore, pleural or ascitic fluids might be enriched for cells with a more cancer stem-like phenotype compared to cells from solid tumours, which highlights the need for using primary samples as well.In addition, sequential tumour samples from the same patient would allow us to compare the changes in the gene expression pattern from ALDH pos cells over time and provide data on selection and clonal evolution during endocrine treatment. These should be addressed in future studies if samples are made available.

It was possible to identify important markers in the acquisition of anti-estrogen resistance by interrogating the gene expression profile of ALDH pos and ALDH neg populations in metastastic ER+ breast cancer cells. This list of genes, together with the available literature, enabled us to create a 96 gene panel with which it was possible to study and assesses the cellular diversity present in the resistant ALDH pos cells. Single cell data from MCF-7 cells before and after anti-estrogen therapy, pointed towards the existence of distinct cellular clusters within the ALDH pos population. These results highlight the relevance of tumour heterogeneity and its implications in cancer therapeutics. Resistant populations must be studied individually in order to identify the right targetable biomarkers to eradicate all different kinds of cells within a tumour and avoid tumour relapse in subsequent years. Nowadays, researchers are applying single cell analysis to refine methods that can be used to elucidate tumour heterogeneity and improve targeted therapy at individual patient level (Anjanappa et al. 2017). Chung and colleagues have recently used single cell RNA-seq to explore breast cancer transcriptome and study the immune cells that are in the surrounding tissue. However, the present research is the first study to report that anti-estrogen treatment may select for CSCs in a more dormant state. Further research will assess the relevance of this work, but it is anticipated that this study will contribute to the growing knowledge of CSC and cellular diversity.

A question that remains unanswered is whether Tam and Fulv are enhancing the formation of ALDH pos cells to sustain tumour regrowth or whether these therapeutic agents are selectively killing the ALDH neg

cells leaving the ALDH pos population unimpaired. Therefore, future strategies such as the use of an ALDH1A3 reporter to test these two hypotheses will give us a better insight into how CSCs survive anti-estrogen treatment in ER+ breast tumours.

For our current on-going work, the number of genes used to investigate ALDH pos heterogeneity has been reduced from 96 to 24, based on the single cell line data presented here. The expression of these 24 genes has been interrogated in 576 single ALDH pos cells using the newly designed C1 chips, which avoid capturing cell doublets and at the moment data is awaiting analysis. This will allow us to further explore ALDH pos cellular diversity before and after anti-estrogen treatment. In addition, we are planning on extending our investigation by using patient-derived samples (PDS) to study the cellular diversity present within the ALDH pos compartment in a clinically relevant population who have had prior anti-estrogen treatments. In order to do this, we will apply the novel microfluidic platform ICELL8 System from Takara. This system enables us to isolate thousands of single cells from up to 8 different samples at once followed by RNA-seq, which will increase the strength of our research significantly and reduce the batch effect. In addition, this technology can capture cells of any size to achieve unbiased cellular isolation. Alternatively, the 10X Genomics system could be implemented in our study as it has been recently reported as a successful novel technology to perform transcriptional profiling of thousands of single cells (Zheng et al. 2017). Novel RNA-seq data analysis tools, such as T-SNE and pseudotime or branching inference will be applied on the new dataset and will help us in the identification of developmental trajectories and rare types of cells in the anti-estrogen resistance breast cancer setting (Rostom et al. 2017).

7. Chapter 7: References

Ablett, M. P., C. S. O'Brien, A. H. Sims, G. Farnie and R. B. Clarke (2014). "A differential role for CXCR4 in the regulation of normal versus malignant breast stem cell activity." Oncotarget **5**(3): 599-612.

Akkiprik, M., Y. Feng, H. Wang, K. Chen, L. Hu, A. Sahin, S. Krishnamurthy, A. Ozer, X. Hao and W. Zhang (2008). "Multifunctional roles of insulin-like growth factor binding protein 5 in breast cancer." Breast Cancer Res **10**(4): 212.

Akrap, N., D. Andersson, E. Bom, P. Gregersson, A. Stahlberg and G. Landberg (2016). "Identification of Distinct Breast Cancer Stem Cell Populations Based on Single-Cell Analyses of Functionally Enriched Stem and Progenitor Pools." Stem Cell Reports **6**(1): 121-136.

Al-Hajj, M., M. S. Wicha, A. Benito-Hernandez, S. J. Morrison and M. F. Clarke (2003). "Prospective identification of tumorigenic breast cancer cells." Proc Natl Acad Sci U S A **100**(7): 3983-3988.

Alberts B, J. A., Lewis J, et al. and (2002). "Signaling through G-Protein-Linked Cell-Surface Receptors." Molecular Biology of the Cell(4th edition (https://www.ncbi.nlm.nih.gov/books/NBK26912/)).

Alves, C. L., D. Elias, M. Lyng, M. Bak, T. Kirkegaard, A. E. Lykkesfeldt and H. J. Ditzel (2016). "High CDK6 Protects Cells from Fulvestrant-Mediated Apoptosis and is a Predictor of Resistance to Fulvestrant in Estrogen Receptor-Positive Metastatic Breast Cancer." Clin Cancer Res **22**(22): 5514-5526.

Anderson, E. (2002). "Progesterone receptors - animal models and cell signaling in breast cancer - The role of oestrogen and progesterone receptors in human mammary development and tumorigenesis." Breast Cancer Research **4**(5): 197-201.

Anjanappa, M., A. Cardoso, L. Cheng, S. Mohamad, A. Gunawan, S. Rice, Y. Dong, L. Li, G. E. Sandusky, E. F. Srour and H. Nakshatri (2017). "Individualized Breast Cancer Characterization through Single-Cell Analysis of Tumor and Adjacent Normal Cells." Cancer Res **77**(10): 2759-2769.

Asselin-Labat, M. L., K. D. Sutherland, H. Barker, R. Thomas, M. Shackleton, N. C. Forrest, L. Hartley, L. Robb, F. G. Grosveld, J. van der Wees, G. J. Lindeman and J. E. Visvader (2007). "Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation." Nat Cell Biol **9**(2): 201-209.

Bachelot, T., C. Bourgier, C. Cropet, I. Ray-Coquard, J. M. Ferrero, G. Freyer, S. Abadie-Lacourtoisie, J. C. Eymard, M. Debled, D. Spaeth, E. Legouffe, D. Allouache, C. El Kouri and E. Pujade-Lauraine (2012). "Randomized phase II trial of everolimus in combination with tamoxifen in patients with hormone receptor-positive, human epidermal growth factor receptor 2-negative metastatic breast cancer with prior exposure to aromatase inhibitors: a GINECO study." J Clin Oncol **30**(22): 2718-2724.

Baum, M., D. M. Brinkley, J. A. Dossett, K. McPherson, J. S. Patterson, R. D. Rubens, F. G. Smiddy, B. A. Stoll, A. Wilson, J. C. Lea, D. Richards and S. H. Ellis (1983). "Improved survival among patients treated with adjuvant tamoxifen after mastectomy for early breast cancer." Lancet **2**(8347): 450.

Beatson, G. (1896). "On the treatment of inoperable cases of carcinoma of the mamma: suggestions for a new method of treatment, with illustrative cases." The Lancet **148**(3803): 162 - 165.

Beleut, M., R. D. Rajaram, M. Caikovski, A. Ayyanan, D. Germano, Y. Choi, P. Schneider and C. Brisken (2010). "Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland." Proc Natl Acad Sci U S A **107**(7): 2989-2994.

Bennett, D. T., X. S. Deng, J. A. Yu, M. T. Bell, D. C. Mauchley, X. Meng, T. B. Reece, D. A. Fullerton and M. J. Weyant (2014). "Cancer Stem Cell Phenotype Is Supported by Secretory Phospholipase A(2) in Human Lung Cancer Cells." Ann Thorac Surg **98**(2): 439-446.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-242.

Bonnet, D. and J. E. Dick (1997). "Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell." Nat Med **3**(7): 730-737.

Boveri, T. (1929). "The origin of malignant tumours." Ann. Scien. Natur. **16**: 277-336.

Boyd, S. (1899). "Remarks on oophorectomy in the treatment of cancer of the breast " BMJ **1**: 257-262.

Brandt, S., A. Kopp, B. Grage and C. Knabbe (2003). "Effects of tamoxifen on transcriptional level of transforming growth factor beta (TGF-beta) isoforms 1 and 2 in tumor tissue during primary treatment of patients with breast cancer." Anticancer Res **23**(1A): 223-229.

Braun, T., E. Bober, S. Singh, D. P. Agarwal and H. W. Goedde (1987). "Evidence for a signal peptide at the amino-terminal end of human mitochondrial aldehyde dehydrogenase." FEBS Lett **215**(2): 233-236.

Bresnick, A. R., D. J. Weber and D. B. Zimmer (2015). "S100 proteins in cancer." Nat Rev Cancer **15**(2): 96-109.

Brisken, C. and B. O'Malley (2010). "Hormone action in the mammary gland." Cold Spring Harb Perspect Biol **2**(12): a003178.

Brodie, A., Q. Lu and J. Nakamura (1997). "Aromatase in the normal breast and breast cancer." J Steroid Biochem Mol Biol **61**(3-6): 281-286.

Brooks, M. D., M. L. Burness and M. S. Wicha (2015). "Therapeutic Implications of Cellular Heterogeneity and Plasticity in Breast Cancer." Cell Stem Cell **17**(3): 260-271.

Brown, K. (2009). "Is tamoxifen a genotoxic carcinogen in women?" Mutagenesis **24**(5): 391-404.

Brzozowski, A. M., A. C. Pike, Z. Dauter, R. E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G. L. Greene, J. A. Gustafsson and M. Carlquist (1997). "Molecular basis of agonism and antagonism in the oestrogen receptor." Nature **389**(6652): 753-758.

Bulun, S. E., Z. Lin, G. Imir, S. Amin, M. Demura, B. Yilmaz, R. Martin, H. Utsunomiya, S. Thung, B. Gurates, M. Tamura, D. Langoi and S. Deb (2005). "Regulation of aromatase expression in estrogen-responsive breast and uterine disease: from bench to treatment." Pharmacol Rev **57**(3): 359-383.

Busch, S., L. Ryden, O. Stal, K. Jirstrom and G. Landberg (2012). "Low ERK phosphorylation in cancer-associated fibroblasts is associated with tamoxifen resistance in pre-menopausal breast cancer." PLoS One **7**(9): e45669.

Butenandt, A. (1931). "Uber die chemische Untersuchung der Sexualhormone." Zeitschrift fur angewandte Chernie 905-916.

Cancer Research UK. (2017). "Breast Cancer Statistics." Retrieved 9 August, 2017, from http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero.

Carroll, J. S., X. S. Liu, A. S. Brodsky, W. Li, C. A. Meyer, A. J. Szary, J. Eeckhoute, W. Shao, E. V. Hestermann, T. R. Geistlinger, E. A. Fox, P. A. Silver and M. Brown (2005). "Chromosome-wide mapping of estrogen

receptor binding reveals long-range regulation requiring the forkhead protein FoxA1." Cell **122**(1): 33-43.

Charafe-Jauffret, E., C. Ginestier, F. Iovino, C. Tarpin, M. Diebel, B. Esterni, G. Houvenaeghel, J. M. Extra, F. Bertucci, J. Jacquemier, L. Xerri, G. Dontu, G. Stassi, Y. Xiao, S. H. Barsky, D. Birnbaum, P. Viens and M. S. Wicha (2010). "Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer." Clin Cancer Res **16**(1): 45-55.

Charafe-Jauffret, E., C. Ginestier, F. Iovino, J. Wicinski, N. Cervera, P. Finetti, M. H. Hur, M. E. Diebel, F. Monville, J. Dutcher, M. Brown, P. Viens, L. Xerri, F. Bertucci, G. Stassi, G. Dontu, D. Birnbaum and M. S. Wicha (2009). "Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature." Cancer Res **69**(4): 1302-1313.

Chen, D., Q. C. Cui, H. Yang and Q. P. Dou (2006). "Disulfiram, a clinically used anti-alcoholism drug and copper-binding agent, induces apoptotic cell death in breast cancer cultures and xenografts via inhibition of the proteasome activity." Cancer Res **66**(21): 10425-10433.

Ciarloni, L., S. Mallepell and C. Brisken (2007). "Amphiregulin is an essential mediator of estrogen receptor alpha function in mammary gland development." Proc Natl Acad Sci U S A **104**(13): 5455-5460.

Clarke, M. F., J. E. Dick, P. B. Dirks, C. J. Eaves, C. H. Jamieson, D. L. Jones, J. Visvader, I. L. Weissman and G. M. Wahl (2006). "Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells." Cancer Res **66**(19): 9339-9344.

Clarke, R., F. Leonessa, J. N. Welch and T. C. Skaar (2001). "Cellular and molecular pharmacology of antiestrogen action and resistance." Pharmacol Rev **53**(1): 25-71.

Clarke, R., J. J. Tyson and J. M. Dixon (2015). "Endocrine resistance in breast cancer – an overview and update." Mol Cell Endocrinol **418**(0 3): 220-234.

Clarke, R. B., A. Howell and E. Anderson (1997). "Estrogen sensitivity of normal human breast tissue in vivo and implanted into athymic nude mice: analysis of the relationship between estrogen-induced proliferation and progesterone receptor expression." Breast Cancer Res Treat **45**(2): 121-133.

Clarke, R. B., A. Howell, C. S. Potten and E. Anderson (1997). "Dissociation between steroid receptor expression and cell proliferation in the human breast." Cancer Res **57**(22): 4987-4991.

Clarke, R. B., K. Spence, E. Anderson, A. Howell, H. Okano and C. S. Potten (2005). "A putative human breast stem cell population is enriched for steroid receptor-positive cells." Dev Biol **277**(2): 443-456.

Clayton, H., I. Titley and M. Vivanco (2004). "Growth and differentiation of progenitor/stem cells derived from the human mammary gland." Exp Cell Res **297**(2): 444-460.

Cole, M. P., C. T. Jones and I. D. Todd (1971). "A new anti-oestrogenic agent in late breast cancer. An early clinical appraisal of ICI46474." Br J Cancer **25**(2): 270-275.

Collaborative Group on Hormonal Factors in Breast Cancer (2012). "Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies." Lancet Oncol **13**(11): 1141-1151.

Cristofanilli, M., N. C. Turner, I. Bondarenko, J. Ro, S. A. Im, N. Masuda, M. Colleoni, A. DeMichele, S. Loi, S. Verma, H. Iwata, N. Harbeck, K. Zhang, K. P. Theall, Y. Jiang, C. H. Bartlett, M. Koehler and D. Slamon (2016). "Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial." Lancet Oncol **17**(4): 425-439.

Croker, A. K. and A. L. Allan (2012). "Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDHhiCD44(+) human breast cancer cells." Breast Cancer Res Treat **133**(1): 75-87.

Curtis, C., S. P. Shah, S. F. Chin, G. Turashvili, O. M. Rueda, M. J. Dunning, D. Speed, A. G. Lynch, S. Samarajiwa, Y. Yuan, S. Gräf, G. Ha, G. Haffari, A. Bashashati, R. Russell, S. McKinney, A. Langerød, A. Green, E. Provenzano, G. Wishart, S. Pinder, P. Watson, F. Markowetz, L. Murphy, I. Ellis, A. Purushotham, A. L. Børresen-Dale, J. D. Brenton, S. Tavaré, C. Caldas and S. Aparicio (2012). "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups." Nature **486**(7403): 346-352.

Cuzick, J., J. Forbes, R. Edwards, M. Baum, S. Cawthorn, A. Coates, A. Hamed, A. Howell and T. Powles (2002). "First results from the International Breast Cancer Intervention Study (IBIS-I): a randomised prevention trial." Lancet **360**(9336): 817-824.

Dai, X., T. Li, Z. Bai, Y. Yang, X. Liu, J. Zhan and B. Shi (2015). "Breast cancer intrinsic subtype classification, clinical use and future trends." Am J Cancer Res **5**(10): 2929-2943.

Davies, C., J. Godwin, R. Gray, M. Clarke, D. Cutter, S. Darby, P. McGale, H. C. Pan, C. Taylor, Y. C. Wang, M. Dowsett, J. Ingle and R. Peto (2011). "Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials." Lancet **378**(9793): 771-784.

Davies, C., H. Pan, J. Godwin, R. Gray, R. Arriagada, V. Raina, M. Abraham, V. H. Medeiros Alencar, A. Badran, X. Bonfill, J. Bradbury, M. Clarke, R. Collins, S. R. Davis, A. Delmestri, J. F. Forbes, P. Haddad, M. F. Hou, M. Inbar, H. Khaled, J. Kielanowska, W. H. Kwan, B. S. Mathew, I. Mittra, B. Muller, A. Nicolucci, O. Peralta, F. Pernas, L. Petruzelka, T. Pienkowski, R. Radhika, B. Rajan, M. T. Rubach, S. Tort, G. Urrutia, M. Valentini, Y. Wang and R. Peto (2013). "Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial." Lancet **381**(9869): 805-816.

Deome, K. B., L. J. Faulkin, Jr., H. A. Bern and P. B. Blair (1959). "Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice." Cancer Res **19**(5): 515-520.

Devin-Leclerc, J., X. Meng, F. Delahaye, P. Leclerc, E. E. Baulieu and M. G. Catelli (1998). "Interaction and dissociation by ligands of estrogen receptor and Hsp90: the antiestrogen RU 58668 induces a protein synthesis-dependent clustering of the receptor in the cytoplasm." Mol Endocrinol **12**(6): 842-854.

DeVita, V. T., Jr. and S. A. Rosenberg (2012). "Two hundred years of cancer research." N Engl J Med **366**(23): 2207-2214.

Ding, X., Z. Peng, S. C. S. Lin, M. Geri, S. Li, P. Li, Y. Chen, M. Dao, S. Suresh and T. J. Huang (2014). "Cell separation using tilted-angle standing surface acoustic waves." Proc Natl Acad Sci U S A **111**(36): 12992-12997.

Dontu, G., W. M. Abdallah, J. M. Foley, K. W. Jackson, M. F. Clarke, M. J. Kawamura and M. S. Wicha (2003). "In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells." Genes Dev **17**(10): 1253-1270.

Dowsett, M., R. I. Nicholson and R. J. Pietras (2005). "Biological characteristics of the pure antiestrogen fulvestrant: overcoming endocrine resistance." Breast Cancer Res Treat **93 Suppl 1**: S11-18.

Early Breast Cancer Trialists' Collaborative, G. (2011). "Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials." Lancet **378**(9793): 771-784.

Eirew, P., J. Stingl, A. Raouf, G. Turashvili, S. Aparicio, J. T. Emerman and C. J. Eaves (2008). "A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability." Nat Med **14**(12): 1384-1389.

Etienne, M. C., G. Milano, J. L. Fischel, M. Frenay, E. Francois, J. L. Formento, J. Gioanni and M. Namer (1989). "Tamoxifen metabolism: pharmacokinetic and in vitro study." Br J Cancer **60**(1): 30-35.

Fan, P., J. Wang, R. J. Santen and W. Yue (2007). "Long-term treatment with tamoxifen facilitates translocation of estrogen receptor alpha out of the nucleus and enhances its interaction with EGFR in MCF-7 breast cancer cells." Cancer Res **67**(3): 1352-1360.

Fisher, B., J. P. Costantino, D. L. Wickerham, C. K. Redmond, M. Kavanah, W. M. Cronin, V. Vogel, A. Robidoux, N. Dimitrov, J. Atkins, M. Daly, S. Wieand, E. Tan-Chiu, L. Ford and N. Wolmark (1998). "Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study." J Natl Cancer Inst **90**(18): 1371-1388.

Fluidigm Corporation (2014). "Fluidigm® SINGuLAR™ Analysis Toolset 2.0 R package." RELEASE NOTES, Fluidigm Media **3.0.3**(PN 100-7692 C1): 1, http://stat.ethz.ch/R-manual/R-patched/library/stats/html/princomp.html.

Fluidigm Corporation (2016). "Doublet Rate and Detection on the C1 IFCs." White Paper, Fluidigm Media(PN 101-2711 A1): 1-12, https://mailbox.zimbra.upenn.edu/home/ngsc@zimbra.upenn.edu/Briefcase/PublicFiles/Equipment/Fluidigm/C11/C11_Doublet_wp_101-2711A2711_20160106.pdf.

Foulkes, W. D., I. M. Stefansson, P. O. Chappuis, L. R. Begin, J. R. Goffin, N. Wong, M. Trudel and L. A. Akslen (2003). "Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer." J Natl Cancer Inst **95**(19): 1482-1485.

Fraley, C., A. Raftery, L. Scrucca, T. Brendan and M. F. Murphy (2017). "Gaussian Mixture Modelling for Model-Based Clustering, Classification and Density Estimation." R Package, R Foundation for Statistical Computing (https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-M).

Frimat, J. P., M. Becker, Y. Y. Chiang, U. Marggraf, D. Janasek, J. G. Hengstler, J. Franzke and J. West (2011). "A microfluidic array with cellular valving for single cell co-culture." Lab Chip **11**(2): 231-237.

Frisch, S. M. and R. A. Screaton (2001). "Anoikis mechanisms." Curr Opin Cell Biol **13**(5): 555-562.

Fu, G. K., J. Hu, P. H. Wang and S. P. Fodor (2011). "Counting individual DNA molecules by the stochastic attachment of diverse labels." Proc Natl Acad Sci U S A **108**(22): 9026-9031.

Gautier, L., L. Cope, B. M. Bolstad and R. A. Irizarry (2004). "affy--analysis of Affymetrix GeneChip data at the probe level." R Package, R Foundation for Statistical Computing **20**(https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-M): 307-315.

Geisler, J. (2003). "Breast cancer tissue estrogens and their manipulation with aromatase inhibitors and inactivators." J Steroid Biochem Mol Biol **86**(3-5): 245-253.

Ghosh, M. G., D. A. Thompson and R. J. Weigel (2000). "PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer." Cancer Res **60**(22): 6367-6375.

Giesen, C., H. A. Wang, D. Schapiro, N. Zivanovic, A. Jacobs, B. Hattendorf, P. J. Schuffler, D. Grolimund, J. M. Buhmann, S. Brandt, Z. Varga, P. J. Wild, D. Gunther and B. Bodenmiller (2014). "Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry." Nat Methods **11**(4): 417-422.

Ginestier, C., M. H. Hur, E. Charafe-Jauffret, F. Monville, J. Dutcher, M. Brown, J. Jacquemier, P. Viens, C. G. Kleer, S. Liu, A. Schott, D. Hayes, D. Birnbaum, M. S. Wicha and G. Dontu (2007). "ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome." Cell Stem Cell **1**(5): 555-567.

Ginestier, C., S. Liu, M. E. Diebel, H. Korkaya, M. Luo, M. Brown, J. Wicinski, O. Cabaud, E. Charafe-Jauffret, D. Birnbaum, J. L. Guan, G. Dontu and M. S. Wicha (2010). "CXCR1 blockade selectively targets human breast cancer stem cells in vitro and in xenografts." J Clin Invest **120**(2): 485-497.

Giraddi, R. R., M. Shehata, M. Gallardo, M. A. Blasco, B. D. Simons and J. Stingl (2015). "Stem and progenitor cell division kinetics during postnatal mouse mammary gland development." Nat Commun **6**: 8487.

Goldhirsch, A., J. N. Ingle, R. D. Gelber, A. S. Coates, B. Thurlimann and H. J. Senn (2009). "Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009." Ann Oncol **20**(8): 1319-1329.

Gong, C., H. Yao, Q. Liu, J. Chen, J. Shi, F. Su and E. Song (2010). "Markers of tumor-initiating cells predict chemoresistance in breast cancer." PLoS One **5**(12): e15630.

Grannis, F. W., Jr. (2011). "Use of discard pleural fluid in molecular research." Nat Rev Clin Oncol **9**(1): 64; author reply 64.

Gronemeyer, H., J. A. Gustafsson and V. Laudet (2004). "Principles for modulation of the nuclear receptor superfamily." Nat Rev Drug Discov **3**(11): 950-964.

Group, E. B. C. T. C. (1988). "Effects of Adjuvant Tamoxifen and of Cytotoxic Therapy on Mortality in Early Breast Cancer." New England Journal of Medicine **319**(26): 1681-1692.

Gudjonsson, T., M. C. Adriance, M. D. Sternlicht, O. W. Petersen and M. J. Bissell (2005). "Myoepithelial Cells: Their Origin and Function in Breast Morphogenesis and Neoplasia." J Mammary Gland Biol Neoplasia **10**(3): 261-272.

Gudjonsson, T., R. Villadsen, H. L. Nielsen, L. Ronnov-Jessen, M. J. Bissell and O. W. Petersen (2002). "Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties." Genes Dev **16**(6): 693-706.

Gusterson, B. A., P. Monaghan, R. Mahendran, J. Ellis and M. J. O'Hare (1986). "Identification of myoepithelial cells in human and rat breasts by anti-common acute lymphoblastic leukemia antigen antibody A12." J Natl Cancer Inst **77**(2): 343-349.

Gusterson, B. A. and T. Stein (2000). "Human breast development." Seminars in Cell & Developmental Biology **23**(5): 567-573.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.

Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.

Harada, Y., C. Yokota, R. Habas, D. C. Slusarski and X. He (2007). "Retinoic Acid-Inducible G Protein-Coupled Receptors Bind to Frizzled Receptors and May Activate Non-canonical Wnt Signaling." Biochem Biophys Res Commun **358**(4): 968-975.

Harrison, H., G. Farnie, S. J. Howell, R. E. Rock, S. Stylianou, K. R. Brennan, N. J. Bundred and R. B. Clarke (2010). "Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor." Cancer Res **70**(2): 709-718.

Harrison, H., B. M. Simoes, L. Rogerson, S. J. Howell, G. Landberg and R. B. Clarke (2013). "Oestrogen increases the activity of oestrogen receptor negative breast cancer stem cells through paracrine EGFR and Notch signalling." Breast Cancer Res **15**(2): R21.

Hodges, L. C., J. D. Cook, E. K. Lobenhofer, L. Li, L. Bennett, P. R. Bushel, C. M. Aldaz, C. A. Afshari and C. L. Walker (2003). "Tamoxifen functions as a molecular agonist inducing cell cycle-associated genes in breast cancer cells." Mol Cancer Res **1**(4): 300-311.

Hong, F., R. Breitling, C. W. McEntee, B. S. Wittner, J. L. Nemhauser and J. Chory (2006). "RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis." Bioinformatics **22**(22): 2825-2827.

Hoskins, J. M., L. A. Carey and H. L. McLeod (2009). "CYP2D6 and tamoxifen: DNA matters in breast cancer." Nat Rev Cancer **9**(8): 576-586.

Howell, A. (2006). "Fulvestrant ('Faslodex'): current and future role in breast cancer management." Crit Rev Oncol Hematol **57**(3): 265-273.

Hu, X., W. Huang and M. Fan (2017). "Emerging therapies for breast cancer." J Hematol Oncol **10**.

Hu, Y. and G. K. Smyth (2009). "ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays." J Immunol Methods **347**(1-2): 70-78.

Hua, S., R. Kittler and K. P. White (2009). "Genomic antagonism between retinoic acid and estrogen signaling in breast cancer." Cell **137**(7): 1259-1271.

Huang, J. T., J. S. Welch, M. Ricote, C. J. Binder, T. M. Willson, C. Kelly, J. L. Witztum, C. D. Funk, D. Conrad and C. K. Glass (1999). "Interleukin-4-dependent production of PPAR-gamma ligands in macrophages by 12/15-lipoxygenase." Nature **400**(6742): 378-382.

Huls, M., F. G. Russel and R. Masereeuw (2009). "The role of ATP binding cassette transporters in tissue defense and organ regeneration." J Pharmacol Exp Ther **328**(1): 3-9.

Inman, J. L., C. Robertson, J. D. Mott and M. J. Bissell (2015). "Mammary gland development: cell fate specification, stem cells and the microenvironment." Development **142**(6): 1028-1042.

Inoue, A., Y. Omoto, Y. Yamaguchi, R. Kiyama and S. I. Hayashi (2004). "Transcription factor EGR3 is involved in the estrogen-signaling pathway in breast cancer cells." J Mol Endocrinol **32**(3): 649-661.

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

Ishizawa, K., Z. A. Rasheed, R. Karisch, Q. Wang, J. Kowalski, E. Susky, K. Pereira, C. Karamboulas, N. Moghal, N. Rajeshkumar, M. Hidalgo, M. Tsao, L. Ailles, T. Waddell, A. Maitra, B. G. Neel and W. Matsui (2010). "Tumor-initiating cells are rare in many human tumors." Cell Stem Cell **7**(3): 279-282.

Ito, K., A. Carracedo, D. Weiss, F. Arai, U. Ala, D. E. Avigan, Z. T. Schafer, R. M. Evans, T. Suda, C. H. Lee and P. P. Pandolfi (2012). "A PML-PPARδ pathway for fatty acid oxidation regulates haematopoietic stem cell maintenance." Nat Med **18**(9): 1350-1358.

Jensen, E. V., T. Suzuki, T. Kawashima, W. E. Stumpf, P. W. Jungblut and E. R. DeSombre (1968). "A two-step mechanism for the interaction of estradiol with rat uterus." Proc Natl Acad Sci U S A **59**(2): 632-638.

Jiang, Y., X. T. Zhao, Q. Xiao, Q. B. Liu, K. S. Ding, F. Yu, R. Zhang, T. Zhu and G. X. Ge (2014). "Snail and Slug mediate tamoxifen resistance in breast cancer cells through activation of EGFR-ERK independent of epithelial-mesenchymal transition." Journal of Molecular Cell Biology **6**(4): 352-354.

Johansson, H. J., B. C. Sanchez, F. Mundt, J. Forshed, A. Kovacs, E. Panizza, L. Hultin-Rosenberg, B. Lundgren, U. Martens, G. Mathe, Z. Yakhini, K. Helou, K. Krawiec, L. Kanter, A. Hjerpe, O. Stal, B. K. Linderholm and J. Lehtio (2013). "Retinoic acid receptor alpha is associated with tamoxifen resistance in breast cancer." Nat Commun **4**: 2175.

Johnson, W. E., C. Li and A. Rabinovic (2007). "Adjusting batch effects in microarray expression data using empirical Bayes methods." Biostatistics **8**(1): 118-127.

Jombart, T., Z. Kamvar, R. Lustrik, C. Collins, M. Beugin, B. Knaus, P. Solymos, K. Schliep, I. Ahmed, A. Cori and F. Calboli (2016). "Exploratory Analysis of Genetic and Genomic Data." R Package, R Foundation for Statistical Computing (https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-M).

Jordan, V. C. (2006). "Tamoxifen (ICI46,474) as a targeted therapy to treat and prevent breast cancer." Br J Pharmacol **147 Suppl 1**: S269-276.

Knight, W. A., 3rd, C. K. Osborne, M. G. Yochmowitz and W. L. McGuire (1980). "Steroid hormone receptors in the management of human breast cancer." Ann Clin Res **12**(5): 202-207.

Koboldt, D. C. (2012). "Comprehensive molecular portraits of human breast tumours." Nature **490**(7418): 61-70.

Koppaka, V., D. C. Thompson, Y. Chen, M. Ellermann, K. C. Nicolaou, R. O. Juvonen, D. Petersen, R. A. Deitrich, T. D. Hurley and V. Vasiliou (2012). "Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application." Pharmacol Rev **64**(3): 520-539.

Kordon, E. C. and G. H. Smith (1998). "An entire functional mammary gland may comprise the progeny from a single cell." Development **125**(10): 1921-1930.

Krivtsov, A. V., D. Twomey, Z. Feng, M. C. Stubbs, Y. Wang, J. Faber, J. E. Levine, J. Wang, W. C. Hahn, D. G. Gilliland, T. R. Golub and S. A. Armstrong (2006). "Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9." Nature **442**(7104): 818-822.

Kumar, R., M. N. Zakharov, S. H. Khan, R. Miki, H. Jang, G. Toraldo, R. Singh, S. Bhasin and R. Jasuja (2011). "The dynamic structure of the estrogen receptor." J Amino Acids **2011**: 812540.

LaBarge, M. A., O. W. Petersen and M. J. Bissell (2007). "Of microenvironments and mammary stem cells." Stem Cell Rev **3**(2): 137-146.

Lacassagne (1932). "Apparition de cancers de la mamelle chez, la souris mâle soumise à des injections de folliculine." Compt Rend Acad Sci **195**: 630-632.

Lam, L., X. Hu, Z. Aktary, D. W. Andrews and M. Pasdar (2009). "Tamoxifen and ICI 182,780 increase Bcl-2 levels and inhibit growth of breast carcinoma cells by modulating PI3K/AKT, ERK and IGF-1R pathways independent of ERalpha." Breast Cancer Res Treat **118**(3): 605-621.

Lamb, R., S. Lehn, L. Rogerson, R. B. Clarke and G. Landberg (2013). "Cell cycle regulators cyclin D1 and CDK4/6 have estrogen receptor-dependent divergent functions in breast cancer migration and stem cell-like activity." Cell Cycle **12**(15): 2384-2394.

Lambeau, G. and M. H. Gelb (2008). "Biochemistry and physiology of mammalian secreted phospholipases A2." Annu Rev Biochem **77**: 495-520.

Lapidot, T., C. Sirard, J. Vormoor, B. Murdoch, T. Hoang, J. Caceres-Cortes, M. Minden, B. Paterson, M. A. Caligiuri and J. E. Dick (1994). "A cell initiating human acute myeloid leukaemia after transplantation into SCID mice." Nature **367**(6464): 645-648.

Laplane, L. (2016). "Cancer Stem Cells." Harvard University Press.

Larsen, S. L., A. V. Laenkholm, A. K. Duun-Henriksen, M. Bak, A. E. Lykkesfeldt and T. Kirkegaard (2015). "SRC drives growth of antiestrogen resistant breast cancer cell lines and is a marker for reduced benefit of tamoxifen treatment." PLoS One **10**(2): e0118346.

Lawson, D. A., N. R. Bhakta, K. Kessenbrock, K. D. Prummel, Y. Yu, K. Takai, A. Zhou, H. Eyob, S. Balakrishnan, C. Y. Wang, P. Yaswen, A. Goga and Z. Werb (2015). "Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells." Nature **526**(7571): 131-135.

Leake, R., D. Barnes, S. Pinder, I. Ellis, L. Anderson, T. Anderson, R. Adamson, T. Rhodes, K. Miller and R. Walker (2000). "Immunohistochemical detection of steroid receptors in breast cancer: a working protocol." Journal of Clinical Pathology **53**(8): 634-635.

Lee, Y. T. (1983). "Breast carcinoma: pattern of metastasis at autopsy." J Surg Oncol **23**(3): 175-180.

Leung, E. Y., M. E. Askarian-Amiri, D. Sarkar, C. Ferraro-Peyret, W. R. Joseph, G. J. Finlay and B. C. Baguley (2017). "Endocrine Therapy of Estrogen Receptor-Positive Breast Cancer Cells: Early Differential Effects on Stem Cell Markers." Front Oncol **7**.

Li, X., M. T. Lewis, J. Huang, C. Gutierrez, C. K. Osborne, M. F. Wu, S. G. Hilsenbeck, A. Pavlick, X. Zhang, G. C. Chamness, H. Wong, J. Rosen and J. C. Chang (2008). "Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy." J Natl Cancer Inst **100**(9): 672-679.

Li, Y. S., M. Tondravi, J. L. Liu, E. Smith, C. C. Haudenschild, M. Kaczmarek and X. Zhan (2001). "Cortactin potentiates bone metastasis of breast cancer cells." Cancer Research **61**(18): 6906-6911.

Lim, E., F. Vaillant, D. Wu, N. C. Forrest, B. Pal, A. H. Hart, M. L. Asselin-Labat, D. E. Gyorki, T. Ward, A. Partanen, F. Feleppa, L. I. Huschtscha, H. J. Thorne, S. B. Fox, M. Yan, J. D. French, M. A. Brown, G. K. Smyth, J. E. Visvader and G. J. Lindeman (2009). "Aberrant luminal progenitors

as the candidate target population for basal tumor development in BRCA1 mutation carriers." Nat Med **15**(8): 907-913.

Liu, S., S. G. Clouthier and M. S. Wicha (2012). "Role of microRNAs in the regulation of breast cancer stem cells." J Mammary Gland Biol Neoplasia **17**(1): 15-21.

Liu, S., Y. Cong, D. Wang, Y. Sun, L. Deng, Y. Liu, R. Martin-Trevino, L. Shang, S. P. McDermott, M. D. Landis, S. Hong, A. Adams, R. D'Angelo, C. Ginestier, E. Charafe-Jauffret, S. G. Clouthier, D. Birnbaum, S. T. Wong, M. Zhan, J. C. Chang and M. S. Wicha (2014). "Breast Cancer Stem Cells Transition between Epithelial and Mesenchymal States Reflective of their Normal Counterparts." Stem Cell Reports **2**(1): 78-91.

Liu, Y., D. Lv, J. Duan, S. Xu, J. Zhang, X. Yang, X. Zhang, Y. Cui, X. Bian and S. Yu (2014). "ALDH1A1 expression correlates with clinicopathologic features and poor prognosis of breast cancer patients: a systematic review and meta-analysis." BMC Cancer **14**: 444.

Lu, S. and Z. Dong (2017). "Overexpression of secretory phospholipase A2-IIa supports cancer stem cell phenotype via HER/ERBB-elicited signaling in lung and prostate cancer cells." Int J Oncol **50**(6): 2113-2122.

Lupulescu, A. (1981). "Hormonal regulation of epidermal tumor development." J Invest Dermatol **77**(2): 186-195.

M Wang, J Huang, JB Potash and S. Han (2017). "A combined Association Test for Genes using Summary Statistics " R Package, R Foundation for Statistical Computing(https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-M).

Ma, R., G. M. Karthik, J. Lovrot, F. Haglund, G. Rosin, A. Katchy, X. Zhang, L. Viberg, J. Frisell, C. Williams, S. Linder, I. Fredriksson and J. Hartman (2017). "Estrogen Receptor beta as a Therapeutic Target in Breast Cancer Stem Cells." J Natl Cancer Inst **109**(3): 1-14.

Macconaill, L. E. and L. A. Garraway (2010). "Clinical implications of the cancer genome." J Clin Oncol **28**(35): 5219-5228.

Maehle, A. H. (2011). "Ambiguous Cells: The Emergence of the Stem Cell Concept in the Nineteenth and Twentieth Centuries." Notes Rec R Soc Lond **65**(4): 359-378.

Magnani, L., D. K. Patten, V. T. Nguyen, S. P. Hong, J. H. Steel, N. Patel, Y. Lombardo, M. Faronato, A. R. Gomes, L. Woodley, K. Page, D. Guttery, L. Primrose, D. Fernandez Garcia, J. Shaw, P. Viola, A. Green, C. Nolan,

I. O. Ellis, E. A. Rakha, S. Shousha, E. W. Lam, B. Gyorffy, M. Lupien and R. C. Coombes (2015). "The pioneer factor PBX1 is a novel driver of metastatic progression in ERalpha-positive breast cancer." Oncotarget **6**(26): 21878-21891.

Maley, C. C., P. C. Galipeau, J. C. Finley, V. J. Wongsurawat, X. Li, C. A. Sanchez, T. G. Paulson, P. L. Blount, R. A. Risques, P. S. Rabinovitch and B. J. Reid (2006). "Genetic clonal diversity predicts progression to esophageal adenocarcinoma." Nat Genet **38**(4): 468-473.

Marcato, P., C. A. Dean, D. Pan, R. Araslanova, M. Gillis, M. Joshi, L. Helyer, L. Pan, A. Leidal, S. Gujar, C. A. Giacomantonio and P. W. Lee (2011). "Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis." Stem Cells **29**(1): 32-45.

Meacham, C. E. and S. J. Morrison (2013). "Tumour heterogeneity and cancer cell plasticity." Nature **501**(7467): 328-337.

Miranda-Lorenzo, I., J. Dorado, E. Lonardo, S. Alcala, A. G. Serrano, J. Clausell-Tormos, M. Cioffi, D. Megias, S. Zagorac, A. Balic, M. Hidalgo, M. Erkan, J. Kleeff, A. Scarpa, B. Sainz, Jr. and C. Heeschen (2014). "Intracellular autofluorescence: a biomarker for epithelial cancer stem cells." Nat Methods **11**(11): 1161-1169.

Moore, K. M., G. J. Thomas, S. W. Duffy, J. Warwick, R. Gabe, P. Chou, I. O. Ellis, A. R. Green, S. Haider, K. Brouilette, A. Saha, S. Vallath, R. Bowen, C. Chelala, D. Eccles, W. J. Tapper, A. M. Thompson, P. Quinlan, L. Jordan, C. Gillett, A. Brentnall, S. Violette, P. H. Weinreb, J. Kendrew, S. T. Barry, I. R. Hart, J. L. Jones and J. F. Marshall (2014). "Therapeutic targeting of integrin alphavbeta6 in breast cancer." J Natl Cancer Inst **106**(8).

Moore, S. A., H. M. Baker, T. J. Blythe, K. E. Kitson, T. M. Kitson and E. N. Baker (1998). "Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases." Structure **6**(12): 1541-1551.

Moretti, A., J. Li, S. Donini, R. W. Sobol, M. Rizzi and S. Garavaglia (2016). "Crystal structure of human aldehyde dehydrogenase 1A3 complexed with NAD(+) and retinoic acid." Sci Rep **6**.

Morgan, C. A. and T. D. Hurley (2015). "Characterization of two distinct structural classes of selective aldehyde dehydrogenase 1A1 inhibitors." J Med Chem **58**(4): 1964-1975.

Musgrove, E. A. and R. L. Sutherland (2009). "Biological determinants of endocrine resistance in breast cancer." Nat Rev Cancer **9**(9): 631-643.

Nathanson, I. T. (1946). "The effect of stilbestrol on advanced cancer of the breast." Cancer Res **6**.

Nelson, L. R. and S. E. Bulun (2001). "Estrogen production and action." Journal of the American Academy of Dermatology **45**(3, Supplement): S116-S124.

Nexcelom. (2017). "Cell size distribution for all 60 NCI cancer cell lines." Retrieved 17 August, 2017, from http://www.nexcelom.com/landing/NCI-60-cell-table-thank-you.php.

Nik-Zainal, S., H. Davies, J. Staaf, M. Ramakrishna, D. Glodzik, X. Zou, I. Martincorena, L. B. Alexandrov, S. Martin, D. C. Wedge, P. Van Loo, Y. S. Ju, M. Smid, A. B. Brinkman, S. Morganella, M. R. Aure, O. C. Lingjaerde, A. Langerod, M. Ringner, S. M. Ahn, S. Boyault, J. E. Brock, A. Broeks, A. Butler, C. Desmedt, L. Dirix, S. Dronov, A. Fatima, J. A. Foekens, M. Gerstung, G. K. Hooijer, S. J. Jang, D. R. Jones, H. Y. Kim, T. A. King, S. Krishnamurthy, H. J. Lee, J. Y. Lee, Y. Li, S. McLaren, A. Menzies, V. Mustonen, S. O'Meara, I. Pauporte, X. Pivot, C. A. Purdie, K. Raine, K. Ramakrishnan, F. G. Rodriguez-Gonzalez, G. Romieu, A. M. Sieuwerts, P. T. Simpson, R. Shepherd, L. Stebbings, O. A. Stefansson, J. Teague, S. Tommasi, I. Treilleux, G. G. Van den Eynden, P. Vermeulen, A. Vincent-Salomon, L. Yates, C. Caldas, L. van't Veer, A. Tutt, S. Knappskog, B. K. Tan, J. Jonkers, A. Borg, N. T. Ueno, C. Sotiriou, A. Viari, P. A. Futreal, P. J. Campbell, P. N. Span, S. Van Laere, S. R. Lakhani, J. E. Eyfjord, A. M. Thompson, E. Birney, H. G. Stunnenberg, M. J. van de Vijver, J. W. Martens, A. L. Borresen-Dale, A. L. Richardson, G. Kong, G. Thomas and M. R. Stratton (2016). "Landscape of somatic mutations in 560 breast cancer whole-genome sequences." Nature **534**(7605): 47-54.

Nolan, E., F. Vaillant, D. Branstetter, B. Pal, G. Giner, L. Whitehead, S. W. Lok, G. B. Mann, K. Rohrbach, L. Y. Huang, R. Soriano, G. K. Smyth, W. C. Dougall, J. E. Visvader and G. J. Lindeman (2016). "RANK ligand as a potential target for breast cancer prevention in BRCA1-mutation carriers." Nat Med **22**(8): 933-939.

Nowell, P. C. (1976). "The clonal evolution of tumor cell populations." Science **194**(4260): 23-28.

O'Brien, C. S., S. J. Howell, G. Farnie and R. B. Clarke (2009). "Resistance to endocrine therapy: are breast cancer stem cells the culprits?" J Mammary Gland Biol Neoplasia **14**(1): 45-54.

Opdenaker, L. M., S. R. Modarai and B. M. Boman (2015). "The Proportion of ALDEFLUOR-Positive Cancer Stem Cells Changes with Cell Culture Density Due to the Expression of Different ALDH Isoforms." Cancer Stud Mol Med **2**(2): 87-95.

Osborne, C. K., J. Pippen, S. E. Jones, L. M. Parker, M. Ellis, S. Come, S. Z. Gertler, J. T. May, G. Burton, I. Dimery, A. Webster, C. Morris, R. Elledge and A. Buzdar (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 **20**(16): 3386-3395.

Osborne, C. K. and R. Schiff (2011). "Mechanisms of endocrine resistance in breast cancer." Annu Rev Med **62**: 233-247.

Papi, A., T. Guarnieri, G. Storci, D. Santini, C. Ceccarelli, M. Taffurelli, S. De Carolis, N. Avenia, A. Sanguinetti, A. Sidoni, M. Orlandi and M. Bonafe (2012). "Nuclear receptors agonists exert opposing effects on the inflammation dependent survival of breast cancer stem cells." Cell Death Differ **19**(7): 1208-1219.

Pasqualini, J. R., C. Sumida and N. Giambiagi (1988). "Pharmacodynamic and biological effects of anti-estrogens in different models." J Steroid Biochem **31**(4B): 613-643.

Patani, N., A. K. Dunbier, H. Anderson, Z. Ghazoui, R. Ribas, E. Anderson, Q. Gao, R. A'Hern, A. Mackay, J. Lindemann, R. Wellings, J. Walker, I. Kuter, L. A. Martin and M. Dowsett (2014). "Differences in the transcriptional response to fulvestrant and estrogen deprivation in ER-positive breast cancer." Clin Cancer Res **20**(15): 3962-3973.

Patel, A. P., I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suvà, A. Regev and B. E. Bernstein (2014). "Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma." Science **344**(6190): 1396-1401.

Patel, P. and E. I. Chen (2012). "Cancer stem cells, tumor dormancy, and metastasis." Front Endocrinol (Lausanne) **3**.

Paterson, R. and M. H. Russell (1959). "Clinical trials in malignant disease: Part III—Breast cancer: Evaluation of post-operative radiotherapy." Journal of the Faculty of Radiologists **10**(4): 175-180.

Pavlik, E. J., K. Nelson, S. Srinivasan, D. E. Powell, D. E. Kenady, P. D. DePriest, H. H. Gallion and J. R. van Nagell, Jr. (1992). "Resistance to tamoxifen with persisting sensitivity to estrogen: possible mediation by excessive antiestrogen binding site activity." Cancer Res **52**(15): 4106-4112.

Pece, S., D. Tosoni, S. Confalonieri, G. Mazzarol, M. Vecchi, S. Ronzoni, L. Bernard, G. Viale, P. G. Pelicci and P. P. Di Fiore (2010). "Biological

and molecular heterogeneity of breast cancers correlates with their cancer stem cell content." Cell **140**(1): 62-73.

Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein (2000). "Molecular portraits of human breast tumours." Nature **406**(6797): 747-752.

Piccart, M., G. N. Hortobagyi, M. Campone, K. I. Pritchard, F. Lebrun, Y. Ito, S. Noguchi, A. Perez, H. S. Rugo, I. Deleu, H. A. Burris, 3rd, L. Provencher, P. Neven, M. Gnant, M. Shtivelband, C. Wu, J. Fan, W. Feng, T. Taran and J. Baselga (2014). "Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2dagger." Ann Oncol **25**(12): 2357-2362.

Piva, M., G. Domenici, O. Iriondo, M. Rábano, B. M. Simões, V. Comaills, I. Barredo, J. A. López-Ruiz, I. Zabalza, R. Kypta and M. d. M. Vivanco (2014). "Sox2 promotes tamoxifen resistance in breast cancer cells." EMBO Molecular Medicine **6**(1): 66-79.

Ponti, D., A. Costa, N. Zaffaroni, G. Pratesi, G. Petrangolini, D. Coradini, S. Pilotti, M. A. Pierotti and M. G. Daidone (2005). "Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties." Cancer Res **65**(13): 5506-5511.

Qin, H., Q. Shao, H. Curtis, J. Galipeau, D. J. Belliveau, T. Q. Wang, M. A. Alaoui-Jamali and D. W. Laird (2002). "Retroviral delivery of connexin genes to human breast tumor cells inhibits in vivo tumor growth by a mechanism that is independent of significant gap junctional intercellular communication." Journal of Biological Chemistry **277**(32): 29132-29138.

Qiu, Y., T. Pu, L. Li, F. Cheng, C. Lu, L. Sun, X. Teng, F. Ye and H. Bu (2014). "The Expression of Aldehyde Dehydrogenase Family in Breast Cancer." J Breast Cancer **17**(1): 54-60.

Quintana, E., M. Shackleton, M. S. Sabel, D. R. Fullen, T. M. Johnson and S. J. Morrison (2008). "Efficient tumour formation by single human melanoma cells." Nature **456**(7222): 593-598.

R Development Core Team (2008). "R: A language and environment for statistical computing. R Foundation for Statistical Computing."

Rae, J. M., M. D. Johnson, J. O. Scheys, K. E. Cordero, J. M. Larios and M. E. Lippman (2005). "GREB 1 is a critical regulator of hormone dependent breast cancer growth." Breast Cancer Res Treat **92**(2): 141-149.

Raica, M., A. M. Cimpean, R. A. Ceausu, V. Fulga, C. Nica, L. Rudico and L. Saptefrati (2014). "Hormone receptors and HER2 expression in primary breast carcinoma and corresponding lymph node metastasis: do we need both?" Anticancer Res **34**(3): 1435-1440.

Raman, P., B. L. F. Kaplan, J. T. Thompson, J. P. Vanden Heuvel and N. E. Kaminski (2011). "15-Deoxy-Δ(12,14)-prostaglandin J(2)-Glycerol Ester, a Putative Metabolite of 2-Arachidonyl Glycerol, Activates Peroxisome Proliferator Activated Receptor γ." Mol Pharmacol **80**(1): 201-209.

Reya, T., S. J. Morrison, M. F. Clarke and I. L. Weissman (2001). "Stem cells, cancer, and cancer stem cells." Nature **414**(6859): 105-111.

Reynolds, B. A. and S. Weiss (1996). "Clonal and population analyses demonstrate that an EGF-responsive mammalian embryonic CNS precursor is a stem cell." Dev Biol **175**(1): 1-13.

Ricciotti, E. and G. A. FitzGerald (2011). "Prostaglandins and inflammation." Arterioscler Thromb Vasc Biol **31**(5): 986-1000.

Rios, A. C., N. Y. Fu, G. J. Lindeman and J. E. Visvader (2014). "In situ identification of bipotent stem cells in the mammary gland." Nature **506**(7488): 322-327.

Robertson, J., R. Barr, N. Shulman, B. Forte and N. Magrini (2016). "Essential medicines for cancer: WHO recommendations and national priorities." Bulletin of the World Health Organization.

Robinson, D. R., Y. M. Wu, P. Vats, F. Su, R. J. Lonigro, X. Cao, S. Kalyana-Sundaram, R. Wang, Y. Ning, L. Hodges, A. Gursky, J. Siddiqui, S. A. Tomlins, S. Roychowdhury, K. J. Pienta, S. Y. Kim, J. S. Roberts, J. M. Rae, C. H. Van Poznak, D. F. Hayes, R. Chugh, L. P. Kunju, M. Talpaz, A. F. Schott and A. M. Chinnaiyan (2013). "Activating ESR1 mutations in hormone-resistant metastatic breast cancer." Nat Genet **45**(12): 1446-1451.

Roger, P., M. E. Sahla, S. Makela, J. A. Gustafsson, P. Baldet and H. Rochefort (2001). "Decreased expression of estrogen receptor beta protein in proliferative preinvasive mammary tumors." Cancer Res **61**(6): 2537-2541.

Rostom, R., V. Svensson, S. A. Teichmann and G. Kar (2017). "Computational approaches for interpreting scRNA-seq data." FEBS Lett **591**(15): 2213-2225.

Rutqvist, L. E., B. Cedermark, U. Glas, A. Mattsson, L. Skoog, A. Somell, T. Theve, N. Wilking, J. Askergren, M. L. Hjalmar and et al. (1991).

"Contralateral primary tumors in breast cancer patients in a randomized trial of adjuvant tamoxifen therapy." J Natl Cancer Inst **83**(18): 1299-1306.

Santanam, N., R. Shern-Brewer, R. McClatchey, P. Z. Castellano, A. A. Murphy, S. Voelkel and S. Parthasarathy (1998). "Estradiol as an antioxidant: incompatible with its physiological concentrations and function." J Lipid Res **39**(11): 2111-2118.

Santarius, T., J. Shipley, D. Brewer, M. R. Stratton and C. S. Cooper (2010). "A census of amplified and overexpressed human cancer genes." Nat Rev Cancer **10**(1): 59-64.

Schewe, M., P. F. Franken, A. Sacchetti, M. Schmitt, R. Joosten, R. Bottcher, M. E. van Royen, L. Jeammet, C. Payre, P. M. Scott, N. R. Webb, M. Gelb, R. T. Cormier, G. Lambeau and R. Fodde (2016). "Secreted Phospholipases A2 Are Intestinal Stem Cell Niche Factors with Distinct Roles in Homeostasis, Inflammation, and Cancer." Cell Stem Cell **19**(1): 38-51.

Schug, T. T., D. C. Berry, N. S. Shaw, S. N. Travis and N. Noy (2007). "Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors." Cell **129**(4): 723-733.

Schwender, H. (2012). "Siggenes: Multiple testing using SAM and Efron's empirical Bayes approaches." R Package, R Foundation for Statistical Computing (https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-M).

Scrucca, L., M. Fop, T. B. Murphy and A. E. Raftery (2016). "mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models." R Package, R Foundation for Statistical Computing (https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-M).

Shackleton, M., F. Vaillant, K. J. Simpson, J. Stingl, G. K. Smyth, M. L. Asselin-Labat, L. Wu, G. J. Lindeman and J. E. Visvader (2006). "Generation of a functional mammary gland from a single stem cell." Nature **439**(7072): 84-88.

Shaw, F. L., H. Harrison, K. Spence, M. P. Ablett, B. M. Simoes, G. Farnie and R. B. Clarke (2012). "A detailed mammosphere assay protocol for the quantification of breast stem cell activity." J Mammary Gland Biol Neoplasia **17**(2): 111-117.

Shiau, A. K., D. Barstad, P. M. Loria, L. Cheng, P. J. Kushner, D. A. Agard and G. L. Greene (1998). "The structural basis of estrogen

receptor/coactivator recognition and the antagonism of this interaction by tamoxifen." Cell **95**(7): 927-937.

Shoker, B. S., C. Jarvis, R. B. Clarke, E. Anderson, J. Hewlett, M. P. A. Davies, D. R. Sibson and J. P. Sloane (1999). "Estrogen Receptor-Positive Proliferating Cells in the Normal and Precancerous Breast." Am J Pathol **155**(6): 1811-1815.

Sidhu, A., P. J. Miller and A. D. Hollenbach (2011). "FOXO1 stimulates ceruloplasmin promoter activity in human hepatoma cells treated with IL-6." Biochem Biophys Res Commun **404**(4): 963-967.

Simões, B. M., C. S. O'Brien, R. Eyre, A. Silva, L. Yu, A. Sarmiento-Castro, D. G. Alférez, K. Spence, A. Santiago-Gómez, F. Chemi, A. Acar, A. Gandhi, A. Howell, K. Brennan, L. Rydén, S. Catalano, S. Andó, J. Gee, A. Ucar, A. H. Sims, E. Marangoni, G. Farnie, G. Landberg, S. J. Howell and R. B. Clarke (2015). "Anti-estrogen Resistance in Human Breast Tumors Is Driven by JAG1-NOTCH4-Dependent Cancer Stem Cell Activity." Cell Reports **12**: 1968–1977.

Simoes, B. M., M. Piva, O. Iriondo, V. Comaills, J. A. Lopez-Ruiz, I. Zabalza, J. A. Mieza, O. Acinas and M. D. Vivanco (2011). "Effects of estrogen on the proportion of stem cells in the breast." Breast Cancer Res Treat **129**(1): 23-35.

Simpson, E., G. Rubin, C. Clyne, K. Robertson, L. O'Donnell, M. Jones and S. Davis (2000). "The Role of Local Estrogen Biosynthesis in Males and Females." Trends in Endocrinology & Metabolism **11**(5): 184-188.

Sims, A. H., A. Howell, S. J. Howell and R. B. Clarke (2007). "Origins of breast cancer subtypes and therapeutic implications." Nat Clin Pract Oncol **4**(9): 516-525.

Slamon, D. J., G. M. Clark, S. G. Wong, W. J. Levin, A. Ullrich and W. L. McGuire (1987). "Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene." Science **235**(4785): 177-182.

Sleeman, K. E., H. Kendrick, A. Ashworth, C. M. Isacke and M. J. Smalley (2006). "CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells." Breast Cancer Res **8**(1): R7.

Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning and A. L. Borresen-Dale (2001). "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." Proc Natl Acad Sci U S A **98**(19): 10869-10874.

Sreerama, L. and N. E. Sladek (1997). "Cellular levels of class 1 and class 3 aldehyde dehydrogenases and certain other drug-metabolizing enzymes in human breast malignancies." Clin Cancer Res **3**(11): 1901-1914.

Stingl, J., C. J. Eaves, U. Kuusk and J. T. Emerman (1998). "Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast." Differentiation **63**(4): 201-213.

Stingl, J., C. J. Eaves, I. Zandieh and J. T. Emerman (2001). "Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue." Breast Cancer Res Treat **67**(2): 93-109.

Stingl, J., P. Eirew, I. Ricketson, M. Shackleton, F. Vaillant, D. Choi, H. I. Li and C. J. Eaves (2006). "Purification and unique properties of mammary epithelial stem cells." Nature **439**(7079): 993-997.

Storms, R. W., A. P. Trujillo, J. B. Springer, L. Shah, O. M. Colvin, S. M. Ludeman and C. Smith (1999). "Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity." Proc Natl Acad Sci U S A **96**(16): 9118-9123.

Suzuki, R. and H. Shimodaira (2015). "Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling." R Package, R Foundation for Statistical Computing (https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-M).

Takebe, N., L. Miele, P. J. Harris, W. Jeong, H. Bando, M. Kahn, S. X. Yang and S. P. Ivy (2015). "Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: clinical update." Nat Rev Clin Oncol **12**(8): 445-464.

Tallman, M. S. (1996). "Differentiating therapy with all-trans retinoic acid in acute myeloid leukemia." Leukemia **10 Suppl 1**: S12-15.

Thomas, M. L., R. de Antueno, K. M. Coyle, M. Sultan, B. M. Cruickshank, M. A. Giacomantonio, C. A. Giacomantonio, R. Duncan and P. Marcato (2016). "Citral reduces breast tumor growth by inhibiting the cancer stem cell marker ALDH1A3." Mol Oncol **10**(9): 1485-1496.

Tonner, E., L. Quarrie, M. Travers, M. Barber, A. Logan, C. Wilde and D. Flint (1995). "Does an IGF-binding protein (IGFBP) present in involuting rat mammary gland regulate apoptosis?" Prog Growth Factor Res **6**(2-4): 409-414.

Toy, W., Y. Shen, H. Won, B. Green, R. A. Sakr, M. Will, Z. Li, K. Gala, S. Fanning, T. A. King, C. Hudis, D. Chen, T. Taran, G. Hortobagyi, G. Greene, M. Berger, J. Baselga and S. Chandarlapaty (2013). "ESR1 ligand-binding domain mutations in hormone-resistant breast cancer." Nat Genet **45**(12): 1439-1445.

Tsai, Y. C., Y. Lu, P. W. Nichols, G. Zlotnikov, P. A. Jones and H. S. Smith (1996). "Contiguous patches of normal human mammary epithelium derived from a single stem cell: implications for breast carcinogenesis." Cancer Res **56**(2): 402-404.

Van Amerongen, R., A. N. Bowman and R. Nusse (2012). "Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland." Cell Stem Cell **11**(3): 387-400.

Van Buuren, S. and K. Groothuis-Oudshoorn (2008). "mice: Multivariate Imputation by Chained Equations in R." R Package, R Foundation for Statistical Computing (https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-M).

Van Keymeulen, A., A. S. Rocha, M. Ousset, B. Beck, G. Bouvencourt, J. Rock, N. Sharma, S. Dekoninck and C. Blanpain (2011). "Distinct stem cells contribute to mammary gland development and maintenance." Nature **479**(7372): 189-193.

Vasiliou, V., A. Pappa and D. R. Petersen (2000). "Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism." Chem Biol Interact **129**(1-2): 1-19.

Verma, A. K., E. A. Conrad and R. K. Boutwell (1982). "Differential effects of retinoic acid and 7,8-benzoflavone on the induction of mouse skin tumors by the complete carcinogenesis process and by the initiation-promotion regimen." Cancer Res **42**(9): 3519-3525.

Veronesi, U., L. Mariani, A. Decensi, F. Formelli, T. Camerini, R. Miceli, M. G. Di Mauro, A. Costa, E. Marubini, M. B. Sporn and G. De Palo (2006). "Fifteen-year results of a randomized phase III trial of fenretinide to prevent second breast cancer." Ann Oncol **17**(7): 1065-1071.

Vieira, A. F., S. Ricardo, M. P. Ablett, M. R. Dionisio, N. Mendes, A. Albergaria, G. Farnie, R. Gerhard, J. F. Cameselle-Teijeiro, R. Seruca, F. Schmitt, R. B. Clarke and J. Paredes (2012). "P-cadherin is coexpressed with CD44 and CD49f and mediates stem cell properties in basal-like breast cancer." Stem Cells **30**(5): 854-864.

Villadsen, R., A. J. Fridriksdottir, L. Ronnov-Jessen, T. Gudjonsson, F. Rank, M. A. LaBarge, M. J. Bissell and O. W. Petersen (2007). "Evidence

for a stem cell hierarchy in the adult human breast." J Cell Biol **177**(1): 87-101.

Visvader, J. E. (2009). "Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis." Genes Dev **23**(22): 2563-2577.

Visvader, J. E. and G. J. Lindeman (2008). "Cancer stem cells in solid tumours: accumulating evidence and unresolved questions." Nat Rev Cancer **8**(10): 755-768.

Vogel, V. G. (2009). "The NSABP Study of Tamoxifen and Raloxifene (STAR) trial." Expert Rev Anticancer Ther **9**(1): 51-60.

Wang, D., C. Cai, X. Dong, Q. C. Yu, X. O. Zhang, L. Yang and Y. A. Zeng (2015). "Identification of multipotent mammary stem cells by protein C receptor expression." Nature **517**(7532): 81-84.

Wang, X., Y. Sun, J. Wong and D. S. Conklin (2013). "PPARgamma maintains ERBB2-positive breast cancer stem cells." Oncogene **32**(49): 5512-5521.

Watson, C. J. (2006). "Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ." Breast Cancer Res **8**(2): 203.

Weigelt, B., J. L. Peterse and L. J. van 't Veer (2005). "Breast cancer metastasis: markers and models." Nat Rev Cancer **5**(8): 591-602.

Welm, B. E., S. B. Tepera, T. Venezia, T. A. Graubert, J. M. Rosen and M. A. Goodell (2002). "Sca-1(pos) cells in the mouse mammary gland represent an enriched progenitor cell population." Dev Biol **245**(1): 42-56.

Welshons, W. V., M. E. Lieberman and J. Gorski (1984). "Nuclear localization of unoccupied oestrogen receptors." Nature **307**(5953): 747-749.

World Health Organization. (2017). "Breast Cancer Prevention and Control." Retrieved 17 August, 2017, from http://www.who.int/cancer/detection/breastcancer/en/index1.html.

Wu, X., J. R. Hawse, M. Subramaniam, M. P. Goetz, J. N. Ingle and T. C. Spelsberg (2009). "The tamoxifen metabolite, endoxifen, is a potent antiestrogen that targets estrogen receptor alpha for degradation in breast cancer cells." Cancer Res **69**(5): 1722-1727.

Wuidart, A., M. Ousset, S. Rulands, B. D. Simons, A. Van Keymeulen and C. Blanpain (2016). "Quantitative lineage tracing strategies to resolve

multipotency in tissue-specific stem cells." Genes Dev **30**(11): 1261-1277.

Xin, Y., J. Kim, M. Ni, Y. Wei, H. Okamoto, J. Lee, C. Adler, K. Cavino, A. J. Murphy, G. D. Yancopoulos, H. C. Lin and J. Gromada (2016). "Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells." Proc Natl Acad Sci U S A **113**(12): 3293-3298.

Yang, J., S. A. Mani, J. L. Donaher, S. Ramaswamy, R. A. Itzykson, C. Come, P. Savagner, I. Gitelman, A. Richardson and R. A. Weinberg (2004). "Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis." Cell **117**(7): 927-939.

Yi, E. H., C. S. Lee, J.-K. Lee, Y. J. Lee, M. K. Shin, C.-H. Cho, K. W. Kang, J. W. Lee, W. Han, D.-Y. Noh, Y.-N. Kim, I.-H. Cho and S.-k. Ye (2013). "STAT3-RANTES Autocrine Signaling Is Essential for Tamoxifen Resistance in Human Breast Cancer Cells." Molecular Cancer Research **11**(1): 31-42.

Yu, J., M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, Slukvin, II and J. A. Thomson (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-1920.

Zhang, X., L. Ding, L. Kang and Z. Y. Wang (2012). "Estrogen receptor-alpha 36 mediates mitogenic antiestrogen signaling in ER-negative breast cancer cells." PLoS One **7**(1): e30174.

Zhang, X. and Z. Y. Wang (2013). "Estrogen receptor-alpha variant, ER-alpha36, is involved in tamoxifen resistance and estrogen hypersensitivity." Endocrinology **154**(6): 1990-1998.

Zheng, G. X., J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson and J. H. Bielas (2017). "Massively parallel digital transcriptional profiling of single cells." Nat Commun **8**: 14049.

Zhong, S., H. Yin, Y. Liao, F. Yao, Q. Li, J. Zhang, H. Jiao, Y. Zhao, D. Xu, S. Liu, H. Song, Y. Gao, J. Liu, L. Ma, Z. Pang, R. Yang, C. Ding, B. Sun, X. Lin, X. Ye, W. Guo, B. Han, B. P. Zhou, Y. E. Chin and J. Deng (2015). "Lung Tumor Suppressor GPRC5A Binds EGFR and Restrains Its Effector Signaling." Cancer Res **75**(9): 1801-1814.

Zibara, K., Z. Awada, L. Dib, J. El-Saghir, S. Al-Ghadban, A. Ibrik, N. El-Zein and M. El-Sabban (2015). "Anti-angiogenesis therapy and gap

junction inhibition reduce MDA-MB-231 breast cancer cell invasion and metastasis in vitro and in vivo." Scientific Reports **5**.

8. Chapter 8: Appendices

8.1 ALDH1A3 shRNA Knock-down

Table 8. 1 Identification of the Aldehyde Dehydrogenase 1 family member A3 (ALDH1A3) shRNA sequences in the *Homo sapiens* ALDH1A3 mRNA NCBI sequence (NM_000693.3).
The ALDH1A3 mRNA sequence shown below contains 13 exons, which are represented in black and blue colours. Three different ALDH1A3 shRNAs (V1, V4, V5) that bind inversely and complementary to ALDH1A3 are highlighted with different colours in the sequence below. **V1** shRNA sequence: TGTTGATAAATATCTTGGT is located between Exon 1 and Exon 2 and it has been highlighted in yellow in the NCBI sequence. Note that shRNA from Virus 1 and Virus 5 share part of the mRNA sequence and this portion has been highlighted in grey. **V4** shRNA sequence: TGGTTGAAGAACACTCCCT is located in Exon 9, and it has been highlighted in green in the NCBI sequence. **V5** shRNA sequence: ATCTTGGTGAACTTGACCT is located between Exon 1 and Exon 2, and it has been highlighted in red in the NCBI sequence. Note that shRNA from Virus 1 and Virus 5 share part of the mRNA sequence and this portion has been highlighted in grey.

```
1          gaggcgggcc  ggccgccctc  ccctctcctc  cctcccgcag  cccgccctcc
cttcctccgg          61   tcccgcagcc aattaggcgg ccccctcggg cgggaggcgt
ggggcgctgc ataaagcggc         121   ggggagctgc caccccggga gcgggctgcg
cagtgtccgg gccgagccgg tgcgccgcag         181   actagggcgc ctcgggccag
ggagcgcgga ggagccatgg ccaccgctaa cggggccgtg         241   gaaaacgggc
agccggacag gaagccgccg gccctgccgc gccccatccg caacctggag          301
gtcaagttca ccaagatatt tatcaacaat gaatggcacg aatccaagag tgggaaaaag
361      tttgctacat  gtaacccttc  aactcgggag  caaatatgtg  aagtggaaga
aggagataag          421   cccgacgtgg acaaggctgt ggaggctgca caggttgcct
tccagagggg ctcgccatgg         481   cgccggctgg atgccctgag tcgtgggcgg
ctgctgcacc agctggctga cctggtggag         541   agggaccgcg ccaccttggc
cgccctggag acgatggata cagggaagcc atttcttcat         601   gctttttca
tcgacctgga gggctgtatt agaaccctca gatactttgc agggtgggca          661
gacaaaatcc agggcaagac catccccaca gatgacaacg tcgtgtgctt caccaggcat
721      gagcccattg  gtgtctgtgg  ggccatcact  ccatggaact  tccccctgct
gatgctggtg          781   tggaagctgg cacccgccct ctgctgtggg aacaccatgg
tcctgaagcc tgcggagcag         841   acacctctca ccgcccttta tctcggctct
ctgatcaaag aggccgggtt ccctccagga         901   gtggtgaaca ttgtgccagg
attcgggccc acagtgggag cagcaatttc ttctcaccct         961   cagatcaaca
agatcgcctt caccggctcc acagaggttg gaaaactggt taaagaagct         1021
gcgtcccgga gcaatctgaa gcgggtgacg ctggagctgg gggggaagaa cccctgcatc
1081   gtgtgtgcgg   acgctgactt  ggacttggca  gtggagtgtg  cccatcaggg
agtgttcttc          1141  aaccaaggcc agtgttgcac ggcagcctcc agggtgttcg
tggaggagca ggtctactct         1201  gagtttgtca ggcggagcgt ggagtatgcc
aagaaacggc ccgtgggaga ccccttcgat         1261  gtcaaaacag aacaggggcc
tcagattgat caaaagcagt tcgacaaaat cttagagctg         1321  atcgagagtg
ggaagaagga aggggccaag ctggaatgcg ggggctcagc catggaagac         1381
aaggggctct tcatcaaacc cactgtcttc tcagaagtca cagacaacat gcggattgcc
1441   aaagaggaga  ttttcgggcc  agtgcaacca  atactgaagt  tcaaaagtat
cgaagaagtg          1501  ataaaaagag cgaatagcac cgactatgga ctcacagcag
ccgtgttcac aaaaaatctc         1561  gacaaagccc tgaagttggc ttctgcctta
gagtctggaa cggtctggat caactgctac         1621  aacgccctct atgcacaggc
tccatttggt ggctttaaaa tgtcaggaaa tggcagagaa         1681  ctaggtgaat
acgctttggc cgaatacaca gaagtgaaaa ctgtcaccat caaacttggc         1741
gacaagaacc cctgaaggaa aggcggggct ccttcctcaa acatcggacg gcggaatgtg
1801   gcagatgaaa  tgtgctggag  gaaaaaaatg  acatttctga  ccttcccggg
acacattctt          1861  ctggaggctt tacatctact ggagttgaat gattgctgtt
ttcctctcac tctcctgttt         1921  attcaccaga ctggggatgc ctataggttg
```

```
tctgtgaaat cgcagtcctg cctggggagg        1981 gagctgttgg ccatttctgt
gtttcccttt aaaccagatc ctggagacag tgagatactc        2041 agggcgttgt
taacagggag tggtatttga agtgtccagc agttgcttga aatgctttgc        2101
cgaatctgac tccagtaaga atgtgggaaa acccccctgtg tgttctgcaa gcagggctct
2161 tgcaccagcg gtctcctcag ggtggacctg cttacagagc aagccacgcc
tctttccgag        2221 gtgaaggtgg gaccattcct tgggaaagga ttcacagtaa
ggtttttgg ttttgtttt        2281 ttgttttctt gtttttaaaa aaaggatttc
acagtgagaa agttttggtt agtgcatacc        2341 gtggaagggc gccagggtct
ttgtggattg catgttgaca ttgaccgtga gattcggctt        2401 caaaccaata
ctgcctttgg aatatgacag aatcaatagc ccagagagct tagtcaaaga        2461
cgatatcacg gtctacctta accaaggcac tttcttaagc agaaaatatt gttgaggtta
2521 cctttgctgc taaagatcca atcttctaac gccacaacag catagcaaat
cctaggataa        2581 ttcacctcct catttgacaa atcagagctg taattcgctt
taacaaatta cgcatttcta        2641 tcacgttcac taacagctta tgataagtct
gtgtagtctt ccttttctcc agttctgtta        2701 cccaatttag attagtaaag
cgtacacaac tggaaagact gctgtaataa cacagccttg        2761 ttatttttaa
gtcctatttt gatattaatt tctgattagt tagtaaataa cacctggatt        2821
ctatggagga cctcggtctt catccaagtg gcctgagtat ttcactggca ggttgtgaat
2881 ttttcttttc ctctttgggg atccaaatga tgatgtgcaa tttcatgttt
taacttggga        2941 aactgaaagt gttcccatat agcttcaaaa acaaaaacaa
atgtgttatc cgacggatac        3001 ttttatggtt actaactagt actttcctaa
ttgggaaagt agtgcttaag tttgcaaatt        3061 aagttgggga gggcaataat
aaaatgaggg cccgtaacag aaccagtgtg tgtataacga        3121 aaaccatgta
taaaatgggc ctatcaccct tgtcagagat ataaattacc acatttgcct        3181
tcccttcatc agctaacact tatcacttat actaccaata acttgttaaa tcaggatttg
3241 gcttcataca ctgaattttc agtattttat ctcaagtaga tatagacact
aaccttgata        3301 gtgatacgtt agagggttcc tattcttcca ttgtacgata
atgtctttaa tatgaaatgc        3361 tacattattt ataattggta gagttattgt
atctttttat agttgtaagt acacagaggt        3421 ggtatattta aacttctgta
atatactgta tttagaaatg gaaatatata tagtgttagg        3481 tttcacttct
tttaaggttt acccctgtgg tgtggtttaa aaatctatag gcctgggaat        3541
tccgatccta gctgcagatc gcatcccaca atgcgagaat gataaaataa aattggatat
3601 ttgagaaaaa aaaaaaaaaa aa
```

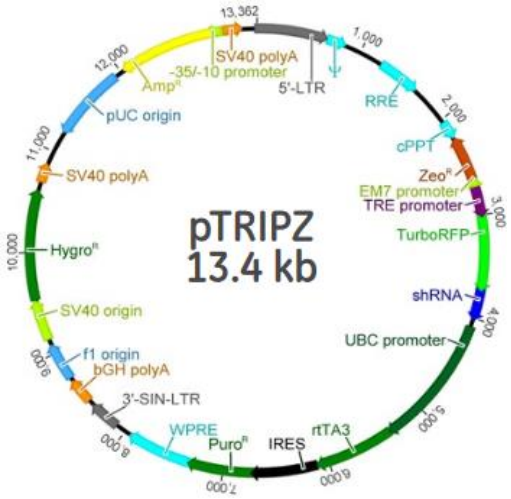


Figure 8. 1 Detailed vector map of the pTRIPZ lentiviral vector. This second-generation TRIPZ vector is engineered to be Tet-On. DOX activates the synthetic factor rtTA3, which binds the tetracycline response element (TRE). TRE drives expression of turbo-RFP and shRNA in the same transcript. Puromycin was used for antibiotic selection of infected cells, as it is constitutively driven by the UBC promoter, independently of DOX. Image taken from http://dharmacon.gelifesciences.com/uploadedFiles/Resources/ptripz-inducible-lentiviral-manual.pdf.

8.2 Flow cytometry purity check

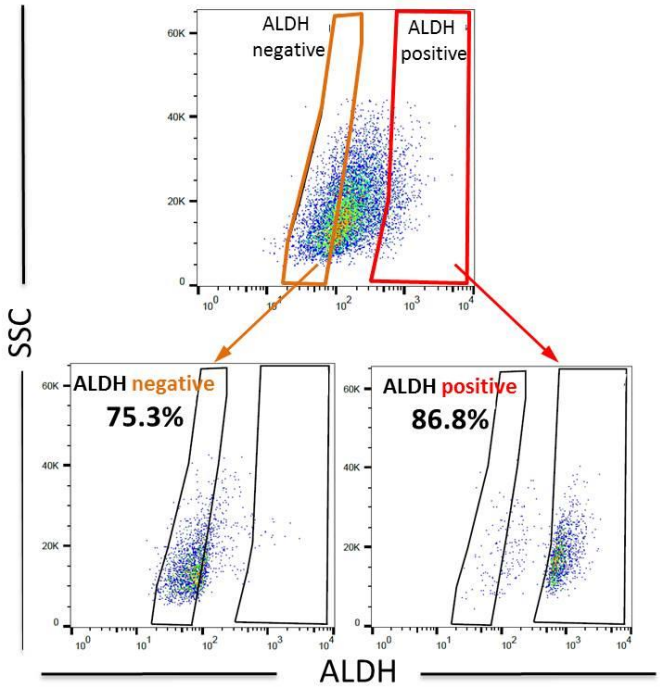


Figure 8. 2 Assessment of purity of sorted cells.
A small fraction of the sorted ALDH neg and ALDH pos cells was re-run through the flow cytometer analyser to evaluate percentage of non-target cells within the gates. Cells were sorted into Aldefluor buffer containing DAPI to avoid loss of the substrate and to account for dead cells.

8.3 Co-authored papers

Anti-estrogen resistance in human breast tumours is driven by JAG-1-NOTCH4-Dependent cancer stem cell activity

Citation: Simões BM, O'Brien CS, Eyre R, et al. Anti-estrogen Resistance in Human Breast Tumors Is Driven by JAG1-NOTCH4-Dependent Cancer Stem Cell Activity. *Cell Reports*. 2015;12(12):1968-1977. doi:10.1016/j.celrep.2015.08.050.

Patient-derived mammosphere and xenograft tumour initiation correlates with progression to metastasis

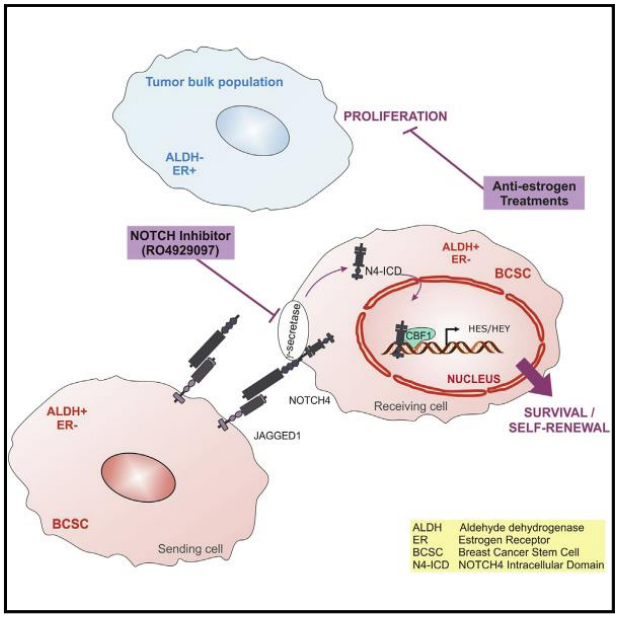
Citation: Eyre R, Alférez DG, Spence K, et al. Patient-derived Mammosphere and Xenograft Tumour Initiation Correlates with Progression to Metastasis. *Journal of Mammary Gland Biology and Neoplasia*. 2016;21(3):99-109. doi:10.1007/s10911-016-9361-8.

Report

Cell Reports

Anti-estrogen Resistance in Human Breast Tumors Is Driven by JAG1-NOTCH4-Dependent Cancer Stem Cell Activity

Graphical Abstract



Authors

Bruno M. Simões, Ciara S. O'Brien, Rachel Eyre, ..., Göran Landberg, Sacha J. Howell, Robert B. Clarke

Correspondence

sacha.howell@christie.nhs.uk (S.J.H.), robert.clarke@manchester.ac.uk (R.B.C.)

In Brief

Breast cancers frequently develop resistance to anti-estrogen treatment, which makes it imperative to understanding how therapy resistance develops. Here, Simões et al. show that combining standard anti-estrogen therapies with anti-Notch4 drugs targeting breast cancer stem cells should improve treatment of ER+ breast cancer patients by preventing relapse due to therapy resistance.

Highlights

- Anti-estrogen therapies selectively enrich for BCSCs and activate Notch signaling
- Notch pathway activation and ALDH1 predict for anti-estrogen treatment failure
- Targeting of Notch4 reduces the population of BCSCs
- Notch inhibitors might prevent relapse or overcome resistance in ER+ tumors

Simões et al., 2015, Cell Reports *12*, 1968–1977
September 29, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.celrep.2015.08.050

OPEN ACCESS CellPress

Cell Reports
Report

Anti-estrogen Resistance in Human Breast Tumors Is Driven by JAG1-NOTCH4-Dependent Cancer Stem Cell Activity

Bruno M. Simões,[1,9] Ciara S. O'Brien,[1,9] Rachel Eyre,[1] Andreia Silva,[1] Ling Yu,[1] Aida Sarmiento-Castro,[1] Denis G. Alférez,[1] Kath Spence,[1] Angélica Santiago-Gómez,[1] Francesca Chemi,[1,2] Ahmet Acar,[3] Ashu Gandhi,[4] Anthony Howell,[1] Keith Brennan,[3] Lisa Rydén,[5] Stefania Catalano,[2] Sebastiano Andó,[2] Julia Gee,[6] Ahmet Ucar,[1,3] Andrew H. Sims,[7] Elisabetta Marangoni,[8] Gillian Farnie,[1] Göran Landberg,[1] Sacha J. Howell,[1,*] and Robert B. Clarke[1,*]

[1]Breast Cancer Now Research Unit, Institute of Cancer Sciences, University of Manchester, Wilmslow Road, Manchester M20 4BX, UK
[2]Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, 87036 Arcavacata di Rende, Cosenza, Italy
[3]Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, UK
[4]Manchester Academic Health Science Centre, University Hospital of South Manchester NHS Foundation Trust, Southmoor Road, Manchester M23 9LT, UK
[5]Department of Surgery, Clinical Sciences, Lund University, Skåne University Hospital, 21428 Malmö, Sweden
[6]Cardiff School of Pharmacy and Pharmaceutical Sciences, University of Cardiff, Cardiff, Wales CF10 3NB, UK
[7]Applied Bioinformatics of Cancer Group, Systems Medicine Building, Western General Hospital, University of Edinburgh, Edinburgh EH4 2XU, UK
[8]Laboratoire d'Investigation Préclinique, Institut Curie, 26 rue d'Ulm 75248 Paris Cedex 05, France
[9]Co-first author
*Correspondence: sacha.howell@christie.nhs.uk (S.J.H.), robert.clarke@manchester.ac.uk (R.B.C.)
http://dx.doi.org/10.1016/j.celrep.2015.08.050

SUMMARY

Breast cancers (BCs) typically express estrogen receptors (ERs) but frequently exhibit de novo or acquired resistance to hormonal therapies. Here, we show that short-term treatment with the anti-estrogens tamoxifen or fulvestrant decrease cell proliferation but increase BC stem cell (BCSC) activity through JAG1-NOTCH4 receptor activation both in patient-derived samples and xenograft (PDX) tumors. In support of this mechanism, we demonstrate that high ALDH1 predicts resistance in women treated with tamoxifen and that a NOTCH4/HES/HEY gene signature predicts for a poor response/prognosis in 2 ER+ patient cohorts. Targeting of NOTCH4 reverses the increase in Notch and BCSC activity induced by anti-estrogens. Importantly, in PDX tumors with acquired tamoxifen resistance, NOTCH4 inhibition reduced BCSC activity. Thus, we establish that BCSC and NOTCH4 activities predict both de novo and acquired tamoxifen resistance and that combining endocrine therapy with targeting JAG1-NOTCH4 overcomes resistance in human breast cancers.

INTRODUCTION

Resistance to endocrine therapies such as selective estrogen receptor (ER) modulators (SERMs; e.g., tamoxifen), selective ER downregulators (SERDs; e.g., fulvestrant), and the aromatase inhibitors is seen in 50%–60% of early breast cancer (BC) cases and develops in almost all patients with advanced disease (Davies et al., 2011; Palmieri et al., 2014).

Evidence suggests that tumor-initiating or cancer stem-like cells (CSCs) are responsible for tumor recurrence after chemo- and endocrine therapy (Li et al., 2008; Creighton et al., 2009). Al-Hajj et al. (2003) were the first to show that tumor-initiating cells were capable of recapitulating the original tumor phenotype when transplanted into immunodeficient mice. In vitro functional assays for BC stem cell (BCSC) activity include aldehyde dehydrogenase 1 (ALDH1) enzyme activity and the capacity to form clonogenic mammospheres in suspension culture (Ginestier et al., 2007). It has been demonstrated that the BCSC population is ER negative/low and resistant to the direct effects of endocrine therapy (Simões et al., 2011; Harrison et al., 2013; Piva et al., 2014).

We have shown that aberrant Notch activation transforms normal breast cells, is found in pre-invasive and invasive human BCs, and correlates with early recurrence (Stylianou et al., 2006; Farnie et al., 2007). Moreover, we reported that inhibition of Notch signaling, particularly NOTCH4 receptor, reduced BCSC activity (Harrison et al., 2010).

Here, using patient-derived ER+ BC samples and patient-derived xenografts (PDXs), we report that short-term treatment with endocrine therapies enriches for JAG1-NOTCH4-regulated BCSCs, suggesting that these effects are not through genetic selection. Furthermore, we show that ALDH1 expression and NOTCH4 activation in human primary tumors are predictive of resistance to endocrine treatments. Finally, we demonstrate that NOTCH inhibition in vivo reduces BCSC activity in long-term acquired resistant PDX tumors. Thus, we propose that inhibiting Notch signaling will help overcome endocrine therapy resistance and recurrence in ER+ BC.

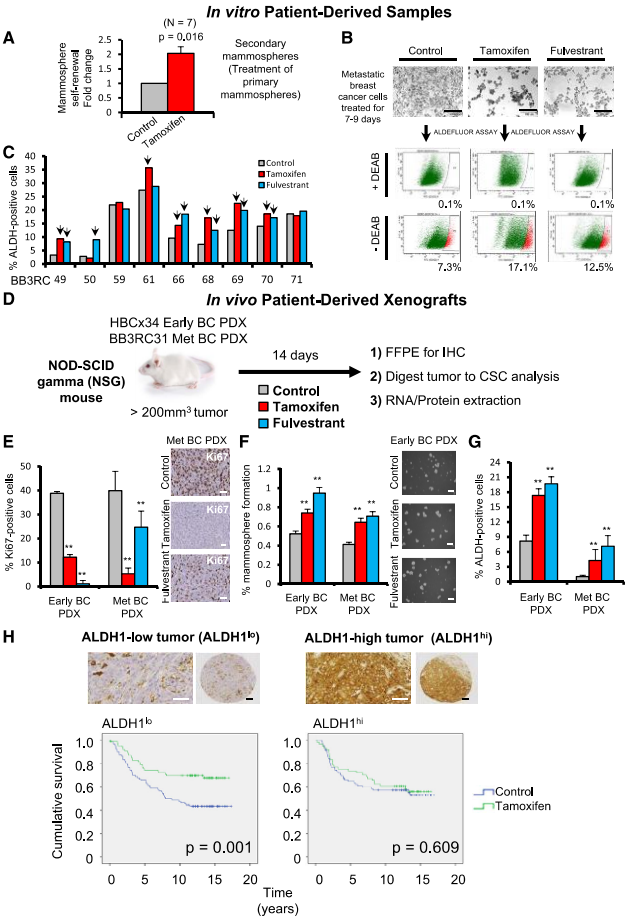


Figure 1. Tamoxifen or Fulvestrant Treatment of ER+ Patient-Derived Samples and PDXs Selectively Enriches for Cells with CSC Properties

High BCSC frequency is associated with worse outcomes for tamoxifen-treated BC patients.

(A) Mammosphere self-renewal of freshly isolated ER+ early and metastatic patient-derived samples. Primary mammospheres cultured in the presence of ethanol (Control) or 10^{-6} M 4-hydroxy-tamoxifen (Tamoxifen) were dissociated and re-plated in secondary mammosphere suspension culture for a further 7–9 days to measure self-renewal of mammosphere-initiating cells treated in the first generation. p value was calculated with Wilcoxon signed-rank test.

(B) Representative micrographs of metastatic BC cells before fluorescence-activated cell sorting (FACS) analysis of ALDH1 enzymatic activity (ALDEFLUOR assay). ALDH-positive cells were discriminated from ALDH-negative cells using the ALDH inhibitor DEAB.

(C) Percentage of ALDH-positive cells in nine ER+ metastatic BC patient-derived samples. Cells were grown in adherence with ethanol (Control), tamoxifen (10^{-6} M), or fulvestrant (10^{-7} M) for 7–9 days. Arrows indicate fold change greater than 20% compared to control.

(D–G) Early (HBCx34) and metastatic (BB3RC31) BC estrogen-dependent PDX tumors treated in vivo for 14 days with tamoxifen (10 mg/kg/day, oral gavage; red bars) or fulvestrant (200 mg/kg/week, subcutaneous injection; blue bars). Gray bars correspond to vehicle control. FFPE, formalin-fixed paraffin-embedded. (E) Representative micrographs and quantification of Ki67 expression determined by immunohistochemistry (IHC). (F) Percentage of MFE. (G) ALDH-positive cells (%) determined using the ALDEFLUOR assay.

(H) ALDH1 expression was assessed by immunohistochemistry in breast tumor epithelial cells, and the percentage of positive cells was scored. Representative micrographs of ALDH-high ($ALDH^{hi}$) and -low ($ALDH^{lo}$) epithelial expression are shown. Kaplan-Meier curves represent cumulative survival for the $ALDH^{lo}$ population and $ALDH^{hi}$ population of a cohort of 322 pre-menopausal ER+ BC patients who participated in a randomized trial of 2 years of adjuvant tamoxifen treatment versus no systemic treatment (control). Vertical bars on survival curves indicate censored cases. p values are based on a log-rank (Mantel-Cox) test of equality of survival distributions.

Scale bars, 100 μm. Data are represented as mean ± SEM. *$p < 0.05$; **$p < 0.01$.

See also Figure S1.

RESULTS

BCSC Activity Is Enriched by Tamoxifen and Fulvestrant

We tested the effect of the anti-estrogen tamoxifen on the mammosphere-forming efficiency (MFE) of patient-derived ER+ tumor cells and found that tamoxifen increases mammosphere self-renewal by about 2-fold (Figures 1A, S1A, and S1B). Next, we investigated ALDH activity, another functional assay for CSCs, in nine patient samples treated with tamoxifen or fulvestrant and showed significant increases in ALDH enzymatic activity in seven patients (Figures 1B and 1C). These data suggest that endocrine therapies, given for a period of a few days, enrich for stem cell activity.

Then, we tested the in vivo impact of endocrine therapies on stem cell activity in ER+ BC using PDXs grown subcutaneously in mice. We used both an early (treatment-naive; early BC) and a metastatic ER+ PDX tumor that both maintain biological characteristics (such as the expression of ER and estrogen

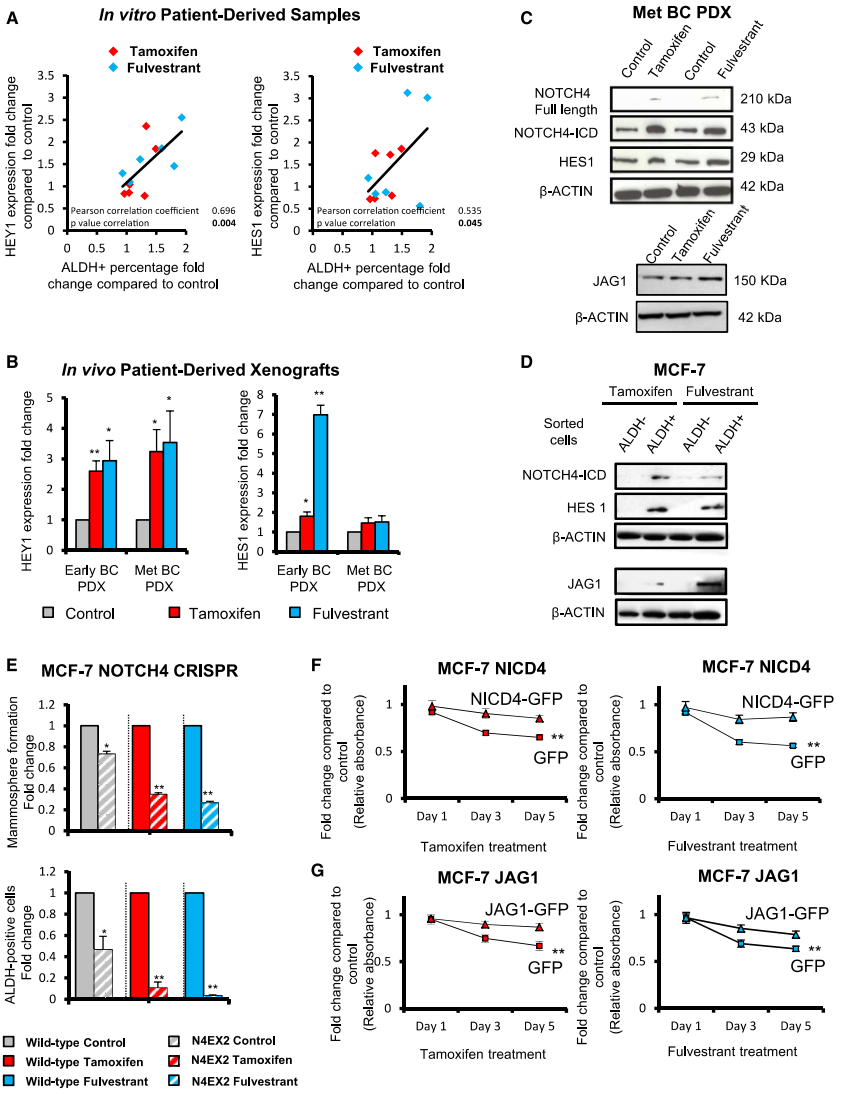


Figure 2. Tamoxifen or Fulvestrant Treatment Upregulates Notch Target Genes in Patient-Derived Samples and PDXs

JAG1-NOTCH4 receptor signaling in ALDH-positive cells drives Notch activity in endocrine-resistant BC.

(A and B) Expression of Notch target genes *HEY1* and *HES1* was assessed by real-time qPCR analysis and compared to control to determine fold change. (A) Metastatic BC patient-derived cells were treated for 7–9 days with ethanol (control), tamoxifen (10^{-6} M), or fulvestrant (10^{-7} M) and a correlation between fold change of expression of HEY1 and HES1 and fold change of percentage of ALDH-positive cells is shown. (B) Early (HBCx34) and metastatic (BB3RC31) BC PDXs: the effect of in vivo treatment for 14 days with tamoxifen (10 mg/kg/day, oral gavage) or fulvestrant (200 mg/kg/week, subcutaneous injection) on *HEY1* and *HES1*.

(legend continued on next page)

dependence) of the patient primary tumor from which they were derived (Figures S1D and S1E). The estrogen dependence of the HBCx34 PDX model (early BC) has been previously reported (Cottu et al., 2012). Using a 14-day in vivo "window" treatment (Figure 1D), we showed that both tamoxifen and fulvestrant treatment decrease proliferation (Figure 1E). However, there is an increase in MFE and ALDH enzymatic activity (Figures 1F and 1G), suggesting a mechanism for endocrine resistance driven by enrichment for a stem cell phenotype.

The mechanism for this enrichment by anti-estrogens may be partly explained by more than 90% of sorted ALDH-positive cells being ER negative (Figure S1C). Thus, we hypothesized that frequency of ALDH-positive cells would predict for response to tamoxifen treatment, and we analyzed ALDH1 in 322 ER+ BC samples taken prior to a randomized trial of tamoxifen versus no systemic treatment. ALDH1 percentage dichotomized at the median value predicted benefit from tamoxifen so that improvement in survival (i.e., a response to treatment) was only seen in women with low epithelial ALDH1 expression (Figure 1H; Table S3). We saw no significant difference in recurrence between control treated patients with high versus low ALDH1 expression ($p = 0.59$). These data, from a prospective randomized trial, establish for the first time that ALDH-positive cell frequency predicts response to tamoxifen treatment, suggesting that stem cell numbers may be responsible for de novo endocrine resistance.

Tamoxifen or Fulvestrant Treatment Upregulates Notch Target Genes

We analyzed the patient-derived BC cells that were treated with tamoxifen and fulvestrant in Figures 1B and 1C and found that increased numbers of ALDH-positive cells were strongly correlated to increased expression of Notch target genes (*HEY1* and *HES1*) (Figure 2A). In addition, the BC PDX tumors treated in vivo with tamoxifen or fulvestrant (Figure 1D) for 2 weeks showed increased *HEY1* and *HES1* expression (Figure 2B), supporting an increased role for the Notch signaling pathway after endocrine therapies.

In ER+ cell lines (MCF-7, T47D, and ZR-75-1) in vitro, treatment with tamoxifen or fulvestrant for 6 days preferentially increased expression of *HEY1* and *HES1* (Figure S2A). Similarly, in tamoxifen-resistant (TAMR) or fulvestrant-resistant (FULVR) MCF-7 models, which have acquired resistance after long-term tamoxifen or fulvestrant treatment, we found upregulation of Notch target genes and increased Notch transcriptional activity (Figure S4A).

JAG1 and NOTCH4 Receptor Signaling Drives Endocrine Resistance

Next, we assessed the expression of Notch receptors and ligands in parental, TAMR, and FULVR cell lines. NOTCH4 and its intracellular domain (ICD) were upregulated while NOTCH1, -2, and -3 were downregulated (Figure S4B) in the resistant versus parental cell lines. We found the Notch ligand JAG1 to be highly expressed in both resistant models (Figure S4B), while expression of the other four ligands was either unchanged (DLL1 and DLL4; Figure S4B) or absent (JAG2 and DLL3; data not shown). JAG1 and NOTCH4-ICD were also upregulated after a 14-day window treatment of PDXs in vivo, and after short-term treatment with tamoxifen or fulvestrant of MCF-7 cells in vitro, suggesting that activation of Notch signaling (demonstrated by increased HES1 expression) is an early event in the acquisition of endocrine resistance (Figures 2C and S2B). Importantly, JAG1, NOTCH4-ICD, and HES1 are expressed at higher levels in ALDH-positive cells, which suggests JAG1-NOTCH4 signaling between ALDH-positive cells (Figure 2D).

To further confirm the role of NOTCH4 activity in endocrine resistance and the stem cell phenotype, we analyzed loss-of- and gain-of-function phenotypes for NOTCH4-ICD in MCF-7 cells. Genomic disruption of exon 2 of *NOTCH4* by using a CRISPR approach led to loss of protein expression (Figures S2C–S2E) and a significant inhibition of MFE and ALDH-positive cells, especially after tamoxifen and fulvestrant treatments (Figure 2E). In contrast, overexpression of NOTCH4-ICD or JAG1 conferred tamoxifen and fulvestrant resistance in parental MCF-7 cells (Figures 2F and 2G).

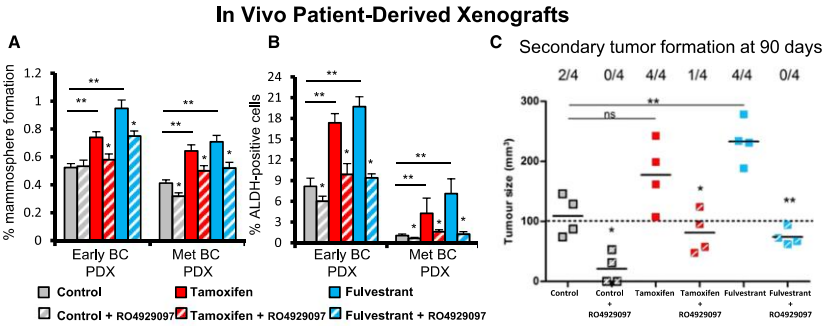
Overall, these results indicate that JAG1 ligand and cleavage of NOTCH4-ICD may be responsible for Notch signaling activation after endocrine treatment, which is in agreement with recent reports that NOTCH4 expression is increased in TAMR cell lines (Yun et al., 2013; Lombardo et al., 2014).

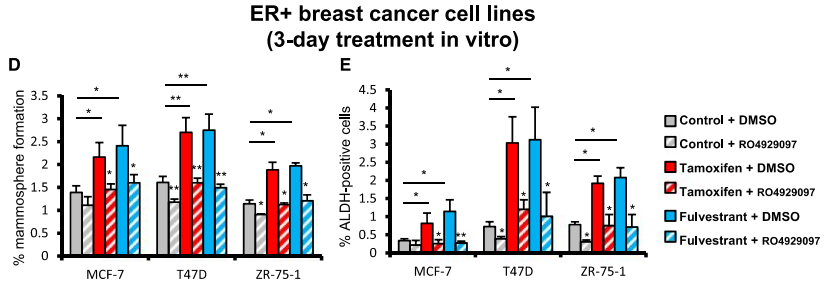
GSI RO4929097 Abrogates Tamoxifen- and Fulvestrant-Stimulated CSC Activity

In order to inhibit NOTCH4 signaling, we used the gamma-secretase inhibitor (GSI) RO4929097, which we found to be effective in reducing levels of the active NOTCH4 ICD in endocrine-resistant models (Figure S4C). RO4929097 inhibited *HEY1* and *HES1* expression, as well as CBF1-Notch transcriptional activity in TAMR and FULVR cell lines, but not in parental MCF-7 cells (Figure S4D). Therefore, we tested whether RO4929097 would abrogate increases in MFE and ALDH-positive cells induced in vivo by anti-estrogens administered in short-term window

(C) NOTCH4, HES1, and JAG1 protein expression levels determined by western blot in metastatic (Met) (BB3RC31) BC PDX. β-actin was used as a reference for the loading control.
(D) NOTCH4, HES1, and JAG1 protein expression levels were determined by western blot in MCF-7 ALDH-negative and ALDH-positive sorted cells. MCF-7 cells were treated with tamoxifen or fulvestrant for 6 days before ALDH sorting.
(E) Wild-type MCF-7 cells (filled bars) and a CRISPR clone containing a disruption of *NOTCH4* exon 2 (N4EX2 cells, hatched bars) treated in adherence with ethanol (Control, gray bars), 10^{-6} M tamoxifen (red bars), and 10^{-7} M fulvestrant (blue bars) for 6 days. N4EX2 cells' fold change of MFE and ALDH-positive cells after treatments was compared to that of the wild-type cells.
(F and G) NICD4 and JAG1 rescue tamoxifen- or fulvestrant-inhibited growth: cell number (using sulforhodamine B [SRB] assay, y axis) of MCF-7 overexpressing (F) NICD4-GFP, (G) JAG1-GFP, or GFP control incubated with tamoxifen or fulvestrant for 1, 3, and 5 days (x axis) compared to the respective cell line treated with control ethanol. p values are for the 5-day treatment.
Data are represented as mean ± SEM. $^{*}p < 0.05$; $^{**}p < 0.01$.
See also Figures S2 and S4.

OPEN ACCESS
CellPress

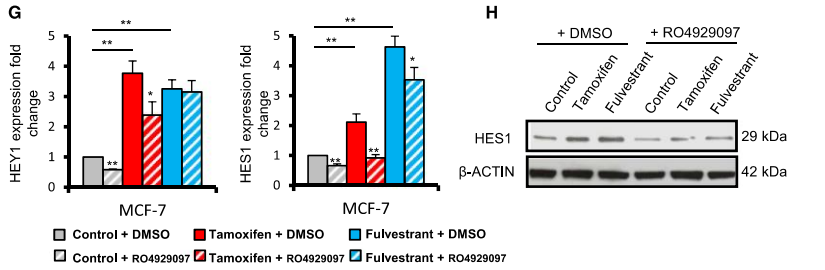




F

MCF-7
(6-day treatment in vitro)

MCF-7 cells 6-day pre-treatment groups		Control + DMSO	Control + RO4929097	Tamoxifen + DMSO	Tamoxifen + RO4929097	Fulvestrant + DMSO	Fulvestrant + RO4929097
Positive tumor growth from different cell numbers	10 000 cells	4/4	3/4	4/4	4/4	4/4	4/4
	1 000 cells	3/4	1/4	4/4	3/4	4/4	2/4
	100 cells	0/4	0/4	2/4	1/4	2/4	1/4
	10 cells	0/4	0/4	2/4	0/4	2/4	0/4
Tumor Initiating Cell Frequency (95% CI)		**1:888** (1:282-1:2801)	**1:6276** (1:2115-1:18619)	**1:77** (1:25-1:239)	**1:626** (1:211-1:1861)	**1:77** (1:25-1:239)	**1:1078** (1:325-1:3572)
P from Control + DMSO group (χ^2)		N/A	N/A	**0.0025**	N/A	**0.0025**	N/A
P from respective DMSO group (χ^2)		N/A	**0.0205**	N/A	**0.0070**	N/A	**0.0009**



(legend on next page)

treatments, using the same estrogen-dependent ER+ PDX tumors as in Figures 1D–1G. Tamoxifen and fulvestrant treatments reduced tumor growth and proliferation (Figures S3A and S3B) while increasing both MFE and ALDH activity (Figures 3A and 3B). RO4929097 had no impact on growth or proliferation (% Ki67; Figure S3B) but significantly inhibited endocrine-stimulated MFE and ALDH activity (Figures 3A and 3B). The gold standard for functionally determining tumor-initiating cells is xenograft formation in secondary mouse hosts, which we performed using dissociated cells from PDX tumors treated in vivo with anti-estrogens and/or RO4929097. Cells isolated from tumors treated in vivo with RO4929097 had significantly reduced tumor-initiating capacity 90 days post-implantation (Figure 3C). Furthermore, the stimulation of tumorigenicity following in vivo tamoxifen and fulvestrant treatment was completely reversed by RO4929097 (Figure 3C). Overall, these data suggest that BCSCs, measured by tumor-initiating activity and enriched by short-term anti-estrogen treatments, are dependent on NOTCH4 signaling that can be blocked by combination treatment with a NOTCH4 inhibitor.

To further substantiate this finding, we analyzed MFE and ALDH activity of MCF-7, T47D, and ZR-75-1 cells treated for 3 days with tamoxifen or fulvestrant in combination with RO4929097. In all cases, RO4929097 reduced MFE and ALDH-positive cells (Figures 3D and 3E). To confirm that RO4929097 reduced the tumor-initiating capacity, we conducted in vivo limiting dilution transplantation of MCF-7 cells. Extreme limiting dilution analysis (ELDA) revealed an 11-fold enrichment in tumor-initiating cell frequency following tamoxifen or fulvestrant pre-treatment, which was reversed by co-treatment with RO4929097 (Figure 3F). Inhibition of NOTCH4 cleavage/activation by RO4929097 was evidenced by decreased HEY1 and HES1 mRNA and protein levels (Figures 3G and 3H). Thus, we established, using PDX models and cell lines in tumor-initiating cell assays, that NOTCH4 inhibition reduces BCSC activity induced by anti-estrogen treatment.

NOTCH4 Inhibition Targets CSCs in TAMR PDX Models

The next question we asked was whether inhibiting NOTCH4 signaling to target BCSCs will overcome long-term acquired anti-estrogen resistance in ER+ BC patients. We investigated RO4929097 treatment in two established PDXs (HBCx22 and HBCx34) that have long-term acquired resistance to tamoxifen in vivo. Analysis of HES1 expression by immunohistochemistry revealed that these two TAMR PDXs displayed increased Notch signaling activation compared to the parental control (Figure 4A). Notably, the TAMR HBCx34 PDX model has a higher percentage of MFE and ALDH activity than the endocrine-sensitive HBCx34 PDX model (compare Figures 1F and 1G with Figures 4B and 4C). These data suggest that acquired tamoxifen resistance in PDX models involves enrichment for BCSC activity through Notch signaling. Treatment with RO4929097 for 14 days demonstrates that MFE and ALDH activity can be significantly reduced in TAMR PDX tumors in vivo (Figures 4B–4D).

NOTCH4/HES/HEY Gene Signature Predicts for Resistance to Tamoxifen Treatment and Prognosis in ER+ Tumors

Based on the aforementioned observations, we hypothesized that NOTCH4 activity, comprising a NOTCH4/HES/HEY gene signature, would predict for response to tamoxifen treatment. In gene expression data from 669 pre-treatment tumors from four published Affymetrix microarray datasets of ER+ patients who subsequently received adjuvant tamoxifen therapy, we found *NOTCH4*, *HES1*, *HEY1*, and *HEY2* to be co-expressed in some tumors, as demonstrated in the heatmap ordered from left to right by the sum of the four genes (Figure 5A). Importantly, elevated expression of these Notch genes before treatment was significantly associated with distant metastasis (Figure 5A) and with reduced overall survival in an independent cohort of 343 untreated ER+ patients (Figure 5B). Thus, *NOTCH4* gene expression and activity in tumors before treatment with endocrine therapy predicts sensitivity to treatment, indicating that this signaling pathway predicts de novo as well as acquired endocrine resistance. These data strengthen the case for therapies against NOTCH4 to target the endocrine-resistant ALDH-positive cells responsible for relapse of ER+ tumors following hormonal therapy (Figure 5C).

DISCUSSION

Here, we report that BCSC activity and frequency are increased in response to the common endocrine therapies tamoxifen and fulvestrant in ER+ patient samples and in early and metastatic PDXs. Our findings suggest that endocrine therapies do not target BCSCs, and this may explain how residual drug-resistant cells are responsible for the relapse of ER+ tumors following hormonal therapy. Although we observe increased BCSC frequency after endocrine treatments, we do not know whether absolute BCSC numbers remain the same and are selected for or whether

Figure 3. NOTCH4 Inhibition using RO4929097 Abrogates Tamoxifen and Fulvestrant Enrichment of CSC Activities

(A–C) Early (HBCx34) and metastatic (Met) (BB3RC31) PDX tumors treated in vivo for 14 days with tamoxifen (10 mg/kg/day, oral gavage) or fulvestrant (200 mg/kg/week, subcutaneous injection) in the presence or absence of the NOTCH4 inhibitor RO4929097 (3 mg/kg/day, oral gavage). (A) MFE (%). (B) Percentage of ALDH-positive cells. (C) Secondary tumor formation. 100,000 cells of metastatic (BB3RC31) PDX were re-implanted subcutaneously in NSG mice with 90-day slow-release estrogen pellets. Tumor growth (>100 mm^3) was determined at day 90 after cell injection.

(D and E) MCF-7, T47D, and ZR-75-1 cells were pre-treated in adherence with 10^{-6} M tamoxifen (red bars) and 10^{-7} M fulvestrant (blue bars) with RO4929097 (10 μM; hatched bars) or DMSO (filled bars) for 72 hr. (D) MFE and (E) percentage of ALDH-positive cells were assessed after pre-treatments.

(F–H) MCF-7 cells were pre-treated in adherence for 6 days in the presence of RO4929097 (10 μM; hatched bars) or DMSO (filled bars). (F) In vivo experiments were carried out in NSG mice with 90-day slow-release estrogen pellets. Tumor growth (>100 mm^3) was assessed at day 60 and is represented as mice positive for growth/mice tested for each cell number tested. ELDA of tumor-initiating cell frequency is shown. (G) Expression of *HEY1* and *HES1* by real-time PCR was compared to the control. (H) HES1 protein expression levels determined by western blot.

Data are represented as mean ± SEM. p values refer to hatched bars compared to filled control bars. *$p < 0.05$; **$p < 0.01$.

See also Figures S3 and S4.

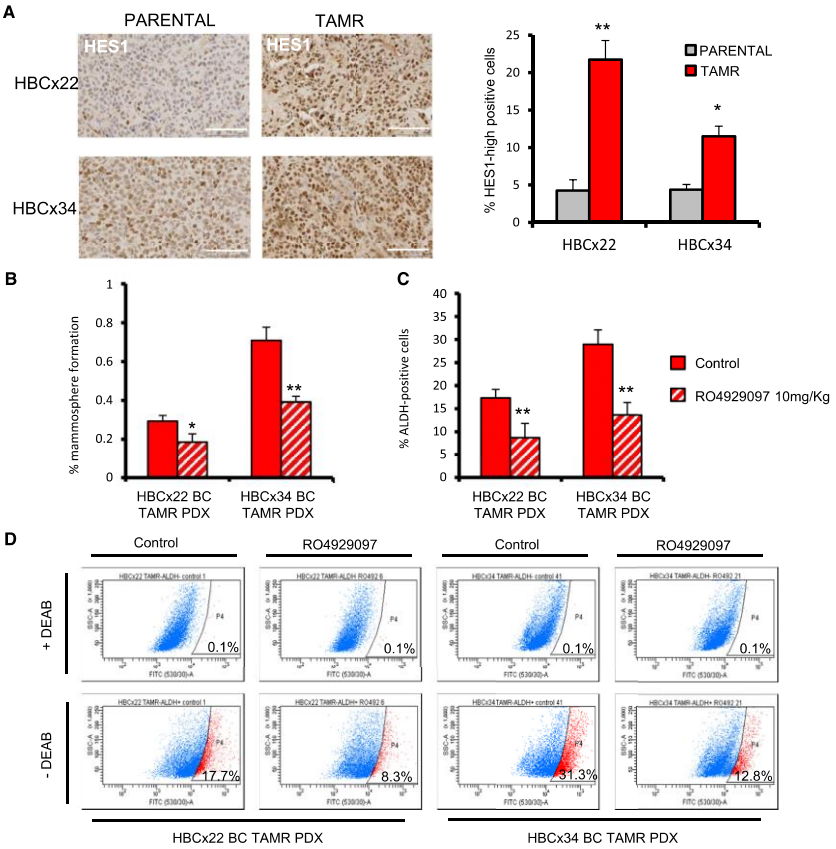


Figure 4. HBCx22 and HBCx34 TAMR PDXs Express High Levels of HES1

NOTCH4 inhibitor RO4929097 targets CSCs in tamoxifen-resistant (TAMR) PDXs.

(A) Representative micrographs and quantification of HES1 expression determined by immunohistochemistry. Scale bars, 100 μm.

(B–D) HBCx22 and HBCx34 TAMR PDXs treated in vivo for 14 days in the presence or absence of the GSI RO4929097 (10 mg/kg/day, oral gavage). (B) MFE (%). (C) Percentage of ALDH-positive cells. (D) Representative FACS plots of ALDEFLUOR assay. ALDH-positive cells were discriminated from ALDH-negative cells using the ALDH inhibitor DEAB.

Data are represented as mean ± SEM. *$p < 0.05$; **$p < 0.01$.

See also Figure S4.

they can be induced by anti-estrogen treatment. Tamoxifen and fulvestrant are clearly successful in reducing BC recurrence in some patients. In other patients with poorer outcome after endocrine therapies, we demonstrate that tumors have high pre-treatment levels of ALDH1 expression and NOTCH4 activation. Moreover, we found that treating ER+ BC cells with endocrine therapies specifically increases JAG1-NOTCH4 signaling and that combining endocrine therapies with a Notch pathway inhibitor can prevent BCSC enrichment induced by endocrine therapies. Thus, our findings in patient-derived BCSCs establish that JAG1 ligand signaling through the NOTCH4 receptor in ALDH-positive cell populations is a determining factor in the acquisition of endocrine resistance.

The best described strategy for inhibition of Notch signaling is the use of small-molecule GSIs, which prevent the release of Notch ICD (NICD). In our study, the GSI RO4929097 specifically targets NOTCH4 cleavage in anti-estrogen-treated cells and, thus, decreases BCSC activity in vitro (MFE and ALDH activity)

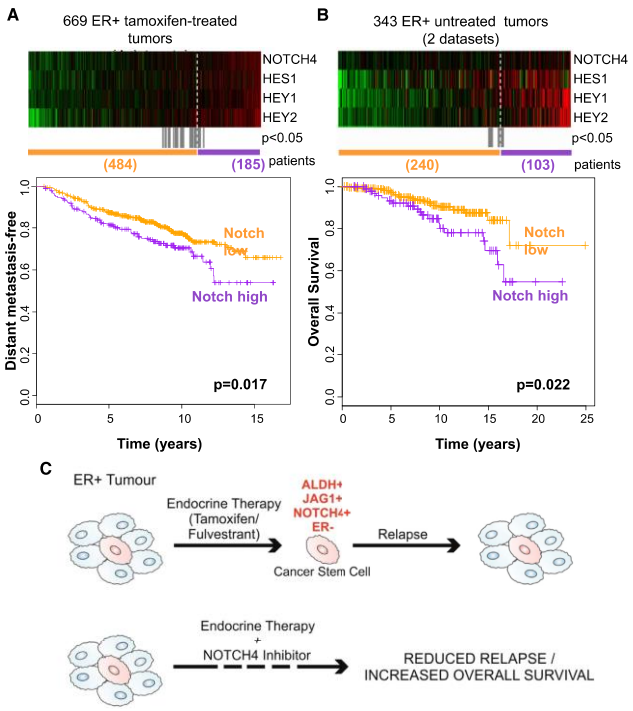


Figure 5. NOTCH4 Receptor Activity Predicts for Resistance to Tamoxifen Treatment and Prognosis in ER+ Tumors

(A and B) *NOTCH4*, *HES1*, *HEY1*, and *HEY2* genes in ER+ primary tumors from (A) tamoxifen-treated or (B) untreated patients are co-expressed in the heatmap ranked from left to right using the four-gene signature. Colors are $\log_2$ mean-centered values; red indicates high, and green indicates low. All significant cut-points ($p < 0.05$) are shown in gray. Kaplan-Meier analysis using the optimum cut-point (dashed white line) demonstrates that elevated expression of the Notch genes is significantly associated with an increased rate of (A) distant metastasis and (B) decreased overall survival. Vertical bars on survival curves indicate censored cases. p values are based on a log-rank (Mantel-Cox) test.

(C) Diagram suggesting that endocrine therapies do not target BCSCs and emphasizing the need of targeting residual drug-resistant cells to eliminate all cancer cells and prevent long-term recurrences of ER+ BC.

and tumor initiation in vivo. Our investigations in ER+ PDX tumors provide the rationale for the use of NOTCH4 inhibitors together with endocrine therapies in the adjuvant or advanced settings (Figure 5C). Significantly, we demonstrated the utility of RO4929097 to target BCSCs in pre-clinical models of TAMR patient tumors.

In conclusion, our data establish that tamoxifen and fulvestrant select for stem cell activity in short- and long-term-treated BC cells, as well as in early endocrine therapy naive and metastatic-endocrine-treated patient-derived samples and PDXs. Importantly, we report that low numbers of stem cells and low Notch signaling activation in patient tumors predict response to tamoxifen therapy and better survival. Overall, these results suggest that ER+ BC recurrence after endocrine therapies, which target the majority of cells (ER+ cells), will be reduced by targeting the JAG1+/NOTCH4+/ALDH1+/ER− BCSC population.

EXPERIMENTAL PROCEDURES

Patient-Derived Samples

Early BC samples were collected in RPMI (GIBCO), dissected into 1- to 2-mm^3 cubes and digested with the Human Tumor Dissociation Kit (Miltenyi Biotec) for 2 hr at 37°C. Digested tissue was filtered sequentially through 100- and 40-μm cell strainers, then centrifuged at 300 × g for 5 min and washed in PBS.

Metastatic samples (ascites or pleural effusions) were centrifuged at 1,000 × g for 10 min at 4°C. The cell pellets were diluted in PBS. Erythrocytes and leucocytes were removed using Lymphoprep (Axis-Shield) and CD45-negative magnetic sorting (Miltenyi Biotec), respectively. Cells were cultured in adherence for 7–9 days in DMEM/F-12 medium, GlutaMAX (GIBCO) with 10% fetal bovine serum (FBS; GIBCO), 10 μg/ml insulin (Sigma-Aldrich), 10 μg/ml hydrocortisone (Sigma-Aldrich), and 5 ng/ml epidermal growth factor (EGF; Sigma-Aldrich), in 10^{-6} M 4-OH tamoxifen (Sigma-Aldrich, H7904), 10^{-7} M fulvestrant (ICI 182,780, Tocris, 1047), or ethanol (control).

Clinico-pathological details of the samples are summarized in Tables S1 (primary BC) and S2 (metastatic BC).

Please refer to the Supplemental Experimental Procedures for further details.

PDXs and In Vivo Experiments

Mouse studies commenced in 8- to 12-week-old female mice and were conducted in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986, using NSG (NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ) mice. All in vivo work was performed with a minimum of n = 4 mice per condition.

Serial passaging of the PDX was carried out by implanting small fragments of the tumor subcutaneously into dorsal flanks of NSG mice. Early (HBCx34) and metastatic (BB3RC31) BC estrogen-dependent PDXs were administered with 8 μg/ml of 17-beta estradiol in drinking water at all times and were treated with drugs when tumors reached 200–300 mm^3. Experiments were performed using PDX tumors between passages 5 and 8. Animal weight and tumor size was measured bi-dimensionally using calipers twice a week.

Tamoxifen citrate (Sigma, T9262, 10 mg/kg/day) and RO4929097 (Cellagen Technology, 3 mg/kg/day) were administered by oral gavage (0.1 ml per dose) on a basis of 5 days out of 7 (weekends excluded) for 14 days. Tamoxifen citrate and RO4929097 were prepared in 1% carboxymethylcellulose (Sigma, C9481) dissolved in distilled water. Fulvestrant (kindly provided by AstraZeneca, 200 mg/kg/week) was administered by subcutaneous injection

(0.1 ml per dose) on a weekly basis for 14 days. The HBCx22 and HBCx34 TAMR PDXs were treated for 14 days in the presence or absence of the GSI RO4929097 (10 mg/kg/day, oral gavage). Xenografts were collected in ice-cold DMEM for live-cell assays, histological analysis, and RNA and protein extraction. PDX single-cell suspension was obtained using a collagenase-hyaluronidase mixture for digestion (Stem Cell Technologies).

Please refer to the Supplemental Experimental Procedures for further details.

Mammosphere Colony Assay

MFE was calculated by dividing the number of mammospheres formed ($\geq$50 μm) by the original number of single cells seeded (500 cells per square centimeter for primary cells) and is expressed as fold change normalized to control or as the mean percentage of MFE (Shaw et al., 2012).

Please refer to the Supplemental Experimental Procedures for further details.

Tamoxifen Trial Study

Premenopausal BC patients with invasive stage II disease were enrolled in SBII:2a, a Swedish clinical trial in which patients were randomly assigned to receive 2 years of adjuvant tamoxifen or no treatment (control) and followed up for recurrence-free and overall survival (Rydén et al., 2005). Our data represent cumulative survival for a cohort of 322 premenopausal ER+ BC patients stratified by ALDH-low (below median) and ALDH-high (above median) expression over time.

Notch Gene Expression Signature

The gene expression data on 669 ER+ tamoxifen-treated tumors (GSE6532, GSE9195, GSE17705, and GSE12093) and 343 ER+ untreated tumors (GSE2034 and GSE7390) are from published Affymetrix microarray datasets.

Please refer to the Supplemental Experimental Procedures for additional details.

Statistical Analysis

If not stated otherwise, a two-tailed Student's t test was performed for statistical analysis. A value of $p < 0.05$ was considered to be statistically significant. Error bars represent the SEM of at least three independent experiments. Data are shown as mean ± SEM.

SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes Supplemental Experimental Procedures, four figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.08.050.

AUTHOR CONTRIBUTIONS

B.M.S. and C.S.O.: conception and design, collection and/or assembly of data, data analysis and interpretation, and manuscript writing. R.E., A.S., L.Y., A.S.-C., D.G.A., K.S., A.S.-G., F.C., A.A., S.C., S.A., A.U., and G.F.: collection and/or assembly of data, data analysis and interpretation, and manuscript critique. A.G.: study recruitment, collection of BC tissue, and manuscript critique. A.H. and K.B.: conception, design, and manuscript critique. J.G.: generation of resistant cell lines and manuscript critique. L.R. and G.L.: Tamoxifen trial study and BC clinical pathology expertise and manuscript critique. A.H.S.: gene expression analysis and manuscript critique. E.M.: generation of PDX and manuscript critique. S.J.H. and R.B.C.: conception and design, data analysis and interpretation, manuscript writing, and final approval of manuscript.

ACKNOWLEDGMENTS

We are a Breast Cancer Now-funded research group (Institute of Cancer Sciences, University of Manchester). We are grateful for funding from the Cancer Research UK (C.S.O.), Medical Research Council (A.S.-C.), Fundación Alfonso Martín Escudero (A.S.-G.), and MultiFun EU FP7 grant #262943 (B.M.S. and K.S.). We thank Allan Jordan and Rebecca Newton (Drug Discovery Unit, Cancer Research UK Manchester Institute) for advice regarding small-molecule Notch inhibitors and Anne Lykkesfeldt for advice on the use of TAMR MCF-7 cells (Danish Cancer Society Research Center). We acknowledge the South-East Swedish Breast Cancer Group for providing access to clinical data and tumor tissue.

Received: March 17, 2015
Revised: July 15, 2015
Accepted: August 17, 2015
Published: September 17, 2015

REFERENCES

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. USA *100*, 3983–3988.

Cottu, P., Marangoni, E., Assayag, F., de Cremoux, P., Vincent-Salomon, A., Guyader, Ch., de Plater, L., Elbaz, C., Karboul, N., Fontaine, J.J., et al. (2012). Modeling of response to endocrine therapy in a panel of human luminal breast cancer xenografts. Breast Cancer Res. Treat. *133*, 595–606.

Creighton, C.J., Li, X., Landis, M., Dixon, J.M., Neumeister, V.M., Sjolund, A., Rimm, D.L., Wong, H., Rodriguez, A., Herschkowitz, J.I., et al. (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc. Natl. Acad. Sci. USA *106*, 13820–13825.

Davies, C., Godwin, J., Gray, R., Clarke, M., Cutter, D., Darby, S., McGale, P., Pan, H.C., Taylor, C., Wang, Y.C., et al.; Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2011). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet *378*, 771–784.

Farnie, G., Clarke, R.B., Spence, K., Pinnock, N., Brennan, K., Anderson, N.G., and Bundred, N.J. (2007). Novel cell culture technique for primary ductal carcinoma in situ: role of Notch and epidermal growth factor receptor signaling pathways. J. Natl. Cancer Inst. *99*, 616–627.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., et al. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1*, 555–567.

Harrison, H., Farnie, G., Howell, S.J., Rock, R.E., Stylianou, S., Brennan, K.R., Bundred, N.J., and Clarke, R.B. (2010). Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. Cancer Res. *70*, 709–718.

Harrison, H., Simoes, B.M., Rogerson, L., Howell, S.J., Landberg, G., and Clarke, R.B. (2013). Oestrogen increases the activity of oestrogen receptor negative breast cancer stem cells through paracrine EGFR and Notch signalling. Breast Cancer Res. *15*, R21.

Li, X., Lewis, M.T., Huang, J., Gutierrez, C., Osborne, C.K., Wu, M.F., Hilsenbeck, S.G., Pavlick, A., Zhang, X., Chamness, G.C., et al. (2008). Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J. Natl. Cancer Inst. *100*, 672–679.

Lombardo, Y., Faronato, M., Filipovic, A., Vircillo, V., Magnani, L., and Coombes, R.C. (2014). Nicastrin and Notch4 drive endocrine therapy resistance and epithelial to mesenchymal transition in MCF7 breast cancer cells. Breast Cancer Res. *16*, R62.

Palmieri, C., Patten, D.K., Januszewski, A., Zucchini, G., and Howell, S.J. (2014). Breast cancer: current and future endocrine therapies. Mol. Cell. Endocrinol. *382*, 695–723.

Piva, M., Domenici, G., Iriondo, O., Rábano, M., Simões, B.M., Comaills, V., Barredo, I., López-Ruiz, J.A., Zabalza, I., Kypta, R., and Vivanco, Md. (2014). Sox2 promotes tamoxifen resistance in breast cancer cells. EMBO Mol. Med. *6*, 66–79.

Rydén, L., Jönsson, P.E., Chebil, G., Dufmats, M., Fernö, M., Jirström, K., Källström, A.C., Landberg, G., Stål, O., Thorstenson, S., et al. (2005). Two years of adjuvant tamoxifen in premenopausal patients with breast cancer: a randomised, controlled trial with long-term follow-up. Eur. J. Cancer *41*, 256–264.

Shaw, F.L., Harrison, H., Spence, K., Ablett, M.P., Simões, B.M., Farnie, G., and Clarke, R.B. (2012). A detailed mammosphere assay protocol for the quantification of breast stem cell activity. J. Mammary Gland Biol. Neoplasia *17*, 111–117.

Simões, B.M., Piva, M., Iriondo, O., Comaills, V., López-Ruiz, J.A., Zabalza, I., Mieza, J.A., Acinas, O., and Vivanco, M.D. (2011). Effects of estrogen on the proportion of stem cells in the breast. Breast Cancer Res. Treat. *129*, 23–35.

Stylianou, S., Clarke, R.B., and Brennan, K. (2006). Aberrant activation of notch signaling in human breast cancer. Cancer Res. *66*, 1517–1525.

Yun, J., Pannuti, A., Espinoza, I., Zhu, H., Hicks, C., Zhu, X., Caskey, M., Rizzo, P., D'Souza, G., Backus, K., et al. (2013). Crosstalk between PKCα and Notch-4 in endocrine-resistant breast cancer cells. Oncogenesis *2*, e60.

J Mammary Gland Biol Neoplasia (2016) 21:99–109
DOI 10.1007/s10911-016-9361-8

Patient-derived Mammosphere and Xenograft Tumour Initiation Correlates with Progression to Metastasis

Rachel Eyre[1] · Denis G. Alférez[1] · Kath Spence[1] · Mohamed Kamal[1,2] · Frances L. Shaw[1] · Bruno M. Simões[1] · Angélica Santiago-Gómez[1] · Aida Sarmiento-Castro[1] · Maria Bramley[3] · Mohammed Absar[3] · Zahida Saad[4] · Sumohan Chatterjee[4] · Cliona Kirwan[5] · Ashu Gandhi[5] · Anne C. Armstrong[6] · Andrew M. Wardley[6] · Ciara S. O'Brien[6] · Gillian Farnie[7] · Sacha J. Howell[8,6] · Robert B. Clarke[1]

Received: 25 May 2016 / Accepted: 5 September 2016 / Published online: 28 September 2016

Abstract Breast cancer specific mortality results from tumour cell dissemination and metastatic colonisation. Identification of the cells and processes responsible for metastasis will enable better prevention and control of metastatic disease, thus reducing relapse and mortality. To better understand these processes, we prospectively collected 307 patient-derived breast cancer samples (n = 195 early breast cancers (EBC) and n = 112 metastatic samples (MBC)). We assessed colony-forming activity *in vitro* by growing isolated cells in both primary (formation) and secondary (self-renewal) mammosphere culture, and tumour initiating activity *in vivo* through subcutaneous transplantation of fragments or cells into mice. Metastatic samples formed primary mammosphere colonies significantly more frequently than early breast cancers and had significantly higher primary mammosphere colony formation efficiency (0.9 % vs. 0.6 %; $p < 0.0001$). Tumour initiation *in vivo* was significantly higher in metastatic than early breast cancer samples (63 % vs. 38 %,

The original version of this article was revised: Table 1 contained the following errors: "Grade" was incorrectly labelled as "Grade 1", "Nodal Involvement" was incorrectly labelled as "Hormone Receptor Status", the ER positive and ER negative labels were reversed, the Her2 positive was labelled as "Po" instead of "Pos", and "Endocrine therapy" was labelled as" "Endoctrine therapy". In Table 2, "Grade" was incorrectly labelled as "Grade 1", and "Hormone Receptor Status" was incorrectly labelled as "Nodal Involvement". In Table 3, "Pathology" was incorrectly labelled as "Pathlofy".

Rachel Eyre and Denis G. Alférez contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10911-016-9361-8) contains supplementary material, which is available to authorized users.

✉ Robert B. Clarke
robert.clarke@manchester.ac.uk

[1] Breast Biology Group, Breast Cancer Now Research Unit, Division of Molecular and Clinical Cancer Sciences, Manchester Cancer Research Centre, University of Manchester, Wilmslow Road, Manchester M20 4QL, UK

[2] Department of Zoology, Faculty of Science, University of Benha, Benha, Egypt

[3] Pennine Acute Hospitals NHS Trust, Manchester, UK

[4] Salford Royal NHS Foundation Trust, Manchester, UK

[5] University Hospitals of South Manchester NHS Foundation Trust, Manchester, UK

[6] The Christie NHS Foundation Trust, Wilmslow Road, Manchester M20 4BX, UK

[7] Cancer Stem Cell Research, Division of Molecular and Clinical Cancer Sciences, Manchester Cancer Research Centre, University of Manchester, Wilmslow Road, Manchester M20 4QL, UK

[8] Breast Cancer Now Research Unit, Division of Molecular and Clinical Cancer Sciences, Manchester Cancer Research Centre, University of Manchester, Wilmslow Road, Manchester M20 4QL, UK

$p = 0.04$). Of 144 breast cancer samples implanted in *vivo*, we established 20 stable patient-derived xenograft (PDX) models at passage 2 or greater. Lung metastases were detected in mice from 14 PDX models. Mammosphere colony formation *in vitro* significantly correlated with the ability of a tumour to metastasise to the lungs *in vivo* ($p = 0.05$), but not with subcutaneous tumour initiation. In summary, the breast cancer stem cell activities of colony formation and tumour initiation are increased in metastatic compared to early samples, and predict metastasis *in vivo*. These results suggest that breast stem cell activity will predict for poor outcome tumours, and therapy targeting this activity will improve outcomes for patients with metastatic disease.

Keywords Breast cancer · Patient-derived xenografts · Metastasis · Mammosphere · Stem cell activity

Introduction

Despite advances in breast cancer diagnoses and treatment, 20 % of patients will develop metastatic tumours at distant sites, eventually leading to death [1]. Metastasis is a multi-step process involving local invasion, intravasation, survival in the circulation, extravasation and colonisation at secondary sites [2]. Although much research has concentrated on understanding the metastatic process, little is known about the cells within a tumour which are able to successfully colonise distant organs. Understanding which cells drive the colonisation and growth of breast cancer cells at distant sites should lead to improvements in adjuvant therapies, designed to prevent the development of secondary tumours.

Much research has focussed on a subset of tumour cells termed cancer stem cells (CSCs), which are capable of self-renewal, and are responsible for tumour initiation [3, 4]. Several methods are used to isolate CSCs and assess their activity, including both functional assays and the expression of cellular markers. Tumour formation following transplantation *in vivo* is considered the gold standard assay to measure CSC activity, and the mammosphere colony forming assay is widely used *in vitro*. This assay was first developed to quantify neural stem cells [5], and was later demonstrated to isolate a stem cell population in mammary tissue [6]. It has since been used to measure cancer stem cell activity in both DCIS and invasive ductal carcinomas [7, 8].

There is evidence that breast cancer cells with a stem cell-like phenotype are metastasis-forming cells in breast cancer. Breast cancer stem-like cells isolated from cell lines are more frequently metastatic compared to non-CSC populations when injected into immunocompromised mice [9, 10]. Furthermore, tumour ALDH1 expression, a marker of breast CSCs, has been shown to be an independent predictive factor for early metastasis in patients [9]. Most recently, single cell analysis of metastatic cells in patient-derived xenograft models revealed that these possessed a stem cell-like gene signature [11]. However the role of CSCs in the metastatic behaviour of patient-derived breast cancer samples is yet to be determined.

Using 307 patient-derived samples from early ($n = 195$, (EBC)) and metastatic ($n = 112$, (MBC)) breast cancers, we evaluated the role of CSC activity in breast cancer metastasis. We assessed both mammosphere colony formation *in vitro* and tumour initiation *in vivo* to determine the relationship between CSC activity and disease progression. Tumour implantation *in vivo* resulted in the production of patient-derived xenograft (PDX) models, and metastasis in these models was correlated to CSC activities in the patient-derived samples. We show that both mammosphere formation *in vitro* and tumour take *in vivo* are increased in metastatic samples compared to early breast cancer samples, and mammosphere formation *in vitro* predicts for metastasis to the lung in PDX models *in vivo*. We thus conclude that cancer stem cell activity *in vitro* correlates with metastasis *in vivo*, and that the mammosphere assay should be further investigated as a tool for identifying metastasis preventing drugs.

Materials and Methods

Sample Collection

195 early breast cancer surgical specimens were collected by the Manchester Cancer Research Centre Biobank from patients undergoing surgery for primary tumour removal at University Hospital of South Manchester, Salford Royal and The Pennine Acute Hospitals NHS Trusts. 112 unrelated metastatic samples (pleural effusion or ascitic fluid) were collected from patients during standard therapeutic drainage procedures at The Christie NHS Foundation Trust. All patients underwent fully informed consent as either "basic consent" or "animal consent" in accordance with local research ethics committee guidelines (see Compliance with Ethical Standards section for consenting procedures). Clinical information for samples used in this study (grade, nodal involvement, oestrogen receptor (ER) status, Her2 receptor status, radiotherapy, chemotherapy and endocrine therapy prior to sample collection, Nottingham Prognostic Index (NPI) score) is detailed in Table 1. Cells from samples with basic consent were isolated (see below) and cancer stem cell activity was assessed *in vitro* only. Samples with animal consent were also implanted into mice (see Compliance with Ethical Standards section for animal ethics approvals) to assess tumour initiation *in vivo* (See Supplementary Fig. 1 for sample pathway).

In vivo Implantation

120 early breast cancer samples and 24 metastatic breast cancer samples were implanted into female NSG (NOD.Cg-

Table 1 Summary of clinical data for patients included in the study. 307 patients (195 early and 112 metastatic) were included. A summary of their clinical characteristics is shown. NA; data not available

		EBC (*n* = 195)	MBC (*n* = 112)
Grade	1	8 (4 %)	5 (4.5 %)
	2	83 (42.6 %)	46 (41.1 %)
	3	95 (48.7 %)	34 (30.4 %)
	NA	9 (4.6 %)	27 (24.1 %)
Nodal involvement	Yes	79 (40.5 %)	NA
	No	95 (48.7 %)	NA
	NA	21 (20.7 %)	NA
ER	Pos	133 (68.2 %)	81 (72.3 %)
	Neg	58 (29.7 %)	24 (21.4 %)
	NA	4 (2.1 %)	7 (6.3 %)
Her2	Pos	33 (16.9 %)	10 (8.9 %)
	Neg	148 (75.9 %)	91 (81.3 %)
	NA	14 (7.2 %)	11 (9.8 %)
Radiotherapy	Yes	4 (2.1 %)	NA
	No	191 (97.9 %)	NA
Chemotherapy	Yes	8 (4.1 %)	68 (60.7 %)
	No	187 (95.9 %)	17 (15.2 %)
	NA	0 (0 %)	27 (24.1 %)
Endocrine therapy	Yes	25 (12.8 %)	72 (64.3 %)
	No	170 (87.2 %)	13 (11.6 %)
	NA	0 (0 %)	27 (24.1 %)
NPI	Good (<3.4)	20 (10.3 %)	NA
	Moderate/Poor (>3.4)	152 (77.9 %)	NA
	NA	23 (11.8 %)	NA

Prkdcscid Il2rgtm1Wjl/SzJ) mice in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986. All tumours were implanted subcutaneously bilaterally into at least two mice per patient sample, either fresh or following one freeze/thaw. Early breast cancers were implanted as 2x2mm^3 fragments, and metastatic samples were injected as 1×10^6 isolated cancer cells. Metastatic cells were injected in 100 μl of a 50:50 mix of matrigel and mammosphere media (components detailed below). The addition of matrigel to early breast cancer fragments was tested but this did not improve tumour take rate, therefore early breast cancers were implanted without matrigel. The implantation procedure did not change during the duration of the study. Oestrogen supplementation was provided in drinking water for mice with ER positive tumours at a concentration of 8 μg/ml. Tumour growth was measured twice weekly using callipers. When tumours reached 1.3cm^3 mice were culled and tissue fragments were either implanted into a further generation of mice, frozen for later use, or fixed and examined histologically.

In vitro Sample Processing

Early breast cancers were disaggregated by mincing with a scalpel, prior to digestion in 4.7 ml RPMI medium plus enzymes from the Miltenyi tumour dissociation kit (130–095-929) on a rotating platform at 150 rpm for 2 h. Optimisation was carried out by splitting tumours in half and comparing the following methods: rotating platform vs. gentleMACS tubes, 2 h digestion vs. overnight digestion, Miltenyi tumour digestion enzymes vs. collagenase. In all cases the number of viable cells (assessed using Trypan Blue on a haemocytometer) were counted following digestion. Two hours digestion on a rotating platform with Milteniyi tumour digestion enzymes was found to be optimal. Following digestion, cells were strained through a 70 μM filter (BD, 352340), using a plunger from a 5 ml syringe (Terumo, SS05SE1) to very gently massage the undigested tissue over the filter, and the filter was then rinsed with 3 × 1 ml RPMI medium. Cells were then further strained through a 40 μM filter (BD, 352340), the filter was rinsed with 3 × 1 ml RPMI medium and centrifuged at 1000 g for 5 min at 4 °C to pellet cells. Supernatant was removed and the cell pellet resuspend in 100 μl of ice-cold PBS. Live cells were counted using Trypan Blue (Gibco, 15,250–061) on a haemocytometer, and cells were then cultured as mammospheres (see below). Metastatic samples were first centrifuged at 1000 g for 10 min to pellet cells. Pellets were resuspended in PBS and blood cells were removed by centrifugation of the cell suspension through 0.5 volumes of Lymphoprep solution (Axis

Shield, Dundee, UK) at 800 g for 20 min. Epithelial cells were removed from the interface and diluted with PBS before centrifugation at 800 g for 2 min at 4 °C to pellet cells. The supernatant was removed and a cell count performed using Trypan Blue on a haemocytometer. Cells were then cultured as mammospheres (see below).

Mammosphere Assay

Mammosphere culture was performed as previously described [12]. A single cell suspension was prepared by manual disaggregation (25 gauge needle) and a total of 500 cells/cm^2 were plated in appropriate polyHEMA (Poly (2-hydroxyethylmethacrylate)) coated tissue culture plates in mammosphere medium (phenol red-free DMEM/F12 (Gibco, 21,041)) containing B27 supplement (no vitamin A; Invitrogen, 12,587), rEGF (20 ng/ml; Sigma, E-9644) and Pen-Strep). Methylcellulose was not added to mammosphere media. Cells were cultured for seven days before mammospheres with diameter greater than 50 μm were counted by visual inspection at ×40 magnification using a microscope fitted with a graticule. Percentage primary mammosphere-forming efficiency (MFE) was calculated by dividing the number of mammospheres formed by the number of cells plated and expressed as a percentage. To assess self-renewal, mammospheres were counted, centrifuged (115×g), and dissociated into a single cell suspension by incubation for 2 min at 37 °C in trypsin EDTA 0.125 % (Sigma), followed by mechanical dissociation (25 gauge needle). Single cells were re-plated at 500 cells/cm^2 and the number of secondary mammospheres counted after 7 days. Mammosphere self-renewal was calculated by dividing the number of secondary mammospheres formed by the number of primary mammospheres formed.

Immunohistochemistry

All tumours/cell pellets were fixed in formalin and paraffin embedded. Tumour staining for ER, PR, Ki67 and Her2 was performed in The Christie Hospital Pathology Department. Antibodies used were anti-ERα (Thermo, SP1), anti-PgR (Dako, M3569), anti-Ki67 (Dako M7240) and anti-Her2 (Vector Laboratories, VP-C380). Antigen retrieval was performed either using Target Retrieval Solution pH 9 (Dako S2367, for ER, PR and Ki67) or in 10 mM Citrate buffer (Her2). Antibodies were detected using Dako EnVision Detection System Peroxidase/DAB, Rabbit/Mouse (Dako, K5007) and sections were counterstained with haematoxylin. Hormone receptor positivity was defined as follows; oestrogen receptor (ER) > 5 % positive cells, progesterone receptor (PR) > 5 % positive cells, Her2 receptor 3+ or 2+ confirmed by positive FISH analysis. Human cells were detected in PDX models using a human specific anti-mitochondrial antibody (Abcam, ab92824). Lungs, livers, femurs and lymph nodes were examined for metastases post-mortem in each mouse where subcutaneous tumour growth (P1) was seen. 5 x step sections (40 μM) of each organ were stained with human specific mitochondrial antibody. Sections were stained on a Leica Bond system using standard protocol F and primary antibody conditions of 1:1000 for 15 min. Metastases were detected by visual examination of stained sections. Organs where at ≥1 human cell was observed were classed as metastases positive.

Statistical Analysis

Statistical analysis was performed using SPSS. Data are represented as mean ± SEM. Statistical significance was measured using parametric testing, assuming equal variance, in the majority of experiments with standard t-tests for two-paired samples used to assess difference between test and control samples. Chi squared analysis was used for categorical variables, The Mann-Whitney U test was used for data which was not normally distributed. Differences were considered statistically significant if the probability value (p) was ≤0.05.

Results

Breast Cancer Stem Cell Activity is Increased in Metastatic Breast Cancer Samples

In total, 307 patient-derived samples were collected during this study. These comprised 195 early breast cancers and 112 unrelated metastatic breast cancers. For samples with animal consent, fragments were first removed for implantation before the remainder of the sample was digested for *in vitro* experiments. For samples with basic consent, the whole sample was digested for *in vitro* experiments. Following this sample pathway, 144 samples in total were implanted into mice (120 early breast cancers and 24 metastatic breast cancers). 131 early breast cancer samples and 67 metastatic breast cancer samples were digested for *in vitro* experiments. Of these, 117 early breast cancer samples and 63 metastatic samples yielded viable cells for experiments. The median number of cells isolated from early breast cancer samples was 35,000 and the median number of cells isolated from metastatic breast cancer samples was 5,250,000.

Cancer stem cell activity was assessed using mammosphere colony formation and self-renewal assays in 117 early breast cancers and 63 metastatic breast cancer cases (Representative images of mammospheres in culture Fig. 1a). Patient-derived samples collected represented all molecular subtypes of breast cancer defined by their oestrogen, progesterone and Her2 receptor expression (Clinical information of all samples used in this study is summarised in Table 1).

Metastatic tumour cells formed mammospheres (i.e. mammosphere forming efficiency > 0) more frequently than early breast cancers (61/63 (97 %) MBC vs. 90/117 (77 %) EBC; $p = 0.0003$) (Fig. 1b), and had higher mammosphere formation efficiency (0.9 % MBC vs. 0.6 % EBC; $p < 0.0001$) (Fig. 1c). Mammosphere self-renewal rates (measured via secondary mammosphere formation) were similar for early and metastatic breast cancers (0.41 % EBC vs. 0.29 % MBC, $p = 0.074$) (Fig. 1d). Tumour initiation *in vivo* was more frequent in metastatic than early breast cancer samples, with 15/24 (63 %) metastatic samples forming tumours in mice compared to 46/120 (38 %) early samples ($p = 0.04$) (Fig. 1e).

Next, we assessed the relationship between mammosphere colony formation / tumour initiation and clinical parameters in both early and metastatic samples. A summary of all samples which formed mammospheres *in vitro* or tumours *in vivo* is given in Table 2. Mammosphere formation was higher in metastatic samples from patients with ER negative disease than ER positive disease ($p = 0.02$), but not in early breast cancer samples (Supplementary Fig. 2a, b). No association was seen between mammosphere colony formation and tumour grade or molecular subtype for early or metastatic breast cancers (Supplementary Fig. 2c–f), or between mammosphere colony formation and Nottingham Prognostic Index (derived from tumour size, number of involved lymph nodes and grade. <3.4 = good prognosis, ≥3.4 = poor prognosis) in early breast cancers (Supplementary Fig. 2g). Neither mammosphere self-renewal nor *in vivo* tumour initiation ability correlated to the above clinical parameters, although the available numbers of cases where these assays were performed were smaller than those examined for primary mammosphere colony formation (self-renewal: $n = 36$ EBC and $n = 21$ MBC, *in vivo* tumour initiation: $n = 120$ EBC and $n = 24$ MBC).

90 breast cancer samples were both implanted *in vivo* and grown as primary mammospheres *in vitro* to assess correlations between *in vivo* and *in vitro* stem cell activity (70 EBC,

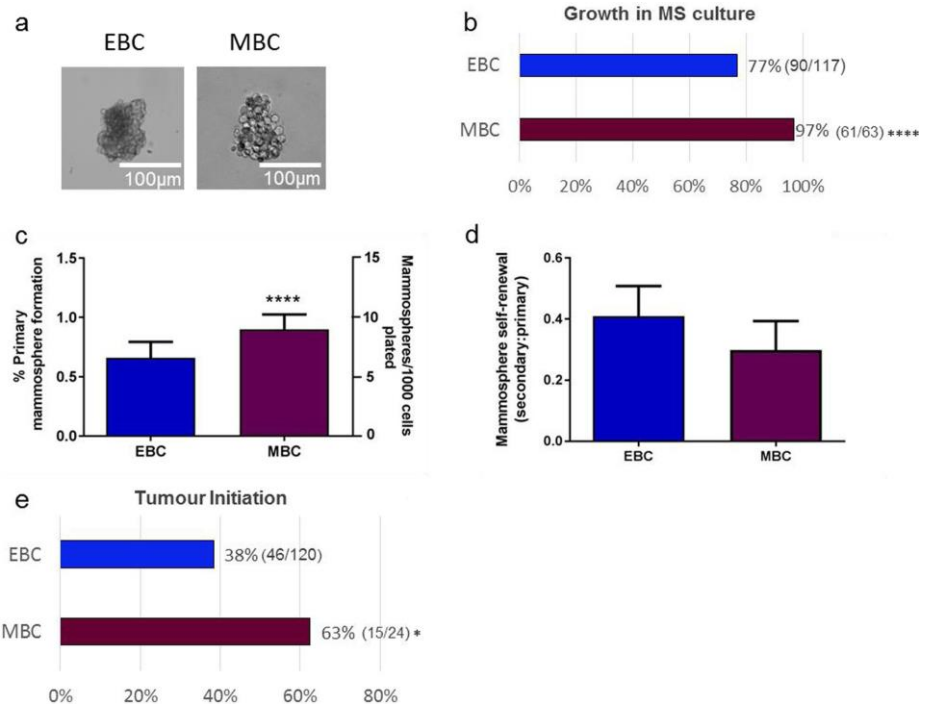


Fig. 1 Breast cancer mammosphere colony formation and *in vivo* tumour-initiating activity is increased in metastatic compared to early breast cancers. Cells were isolated from breast cancer samples and grown in suspension culture as mammospheres (representative mammospheres from early (EBC) and metastatic (MBC) samples shown in **a**). Metastatic breast cancer samples were more likely to form mammospheres (MS formation >0) than early breast cancers cultured under the same conditions **b** and had higher primary mammospheres forming efficiencies **c** (Data presented both as % mammosphere formation, and mammospheres formed/1000 cells plated). No difference in secondary mammosphere formation (self-renewal), defined as a ratio of secondary mammospheres: primary mammospheres, was observed between early and metastatic samples **d**. Metastatic samples had significantly higher *in vivo* xenotransplantation potential than early breast cancer samples, over an average period of 200 days, irrespective of tumour phenotype ($p = 0.04$), where *in vivo* growth is defined as tumour formed to size limit ($1.3cm^3$) **e**. Data are presented as mean ± SEM. *$p < 0.05$ ****$p < 0.0005$. Statistical analyses: Chi Squared tests (b and e) and two tailed t-test (c and d)

20 MBC). Of these, 17 formed tumours when implanted *in vivo* and 53 formed mammospheres *in vitro*. Mammosphere colony formation *in vitro* did not predict tumour initiation *in vivo* (Supplementary Fig. 2h), and there was no difference in the percentage mammosphere formation *in vitro* in samples which were able to initiate tumours *in vivo* (data not shown).

Patient-Derived Xenografts Recapitulate Clinical Tumours

In total, 144 samples were subcutaneously implanted *in vivo* into NSG mice (120 EBC, 24 MBC). Table 2 provides a summary of all samples implanted, and those which formed a tumour in P1, or a stable PDX model. Of the 144 samples implanted, 61 formed tumours in the first passage (P1) in mice (46 EBC, 15 MBC), and 20 stable PDX tumour models at generation 2 or greater have been established (13 EBC, 7 MBC). These are derived from 7 ER + Her2-, 4 ER + Her2+, 1 ER-Her2+ and 8 ER-Her2- patient samples (Table 3). Of the 13 early breast cancers from which stable PDX lines were derived, 10 patients were treatment naïve, 2 had received endocrine therapy (BB6RC37 and BB6RC160, both treated with Arimidex) and treatment information was not available for 1 patient (BB2RC08). All 7 of the MBC PDX models were derived from patients who had received prior therapy.

PDX tumours grown in mice resembled patient-derived tumours histologically (Fig. 2a) and retained hormone receptor expression of the patient-derived tumour (Fig. 2b). There was no association between the ability of an early breast cancer sample to form a stable PDX and clinical characteristics (grade, NPI group, ER/PR status, Her2 status) (data not shown).

Spontaneous Lung Metastasis of PDX Tumour models Correlates with Mammosphere colony Formation

When subcutaneous tumour growth (P1) was seen, the lymph nodes, bones, lungs and livers were examined for the presence of metastases post-mortem. This totalled 34 models, comprising 25 derived from early breast cancers and 9 derived from metastatic breast cancers. Metastatic cells were not observed in the lymph nodes, bones or livers of any of the mice, however lung metastases were detected in mice from 14/34 (41 %) models (representative H&E of lung metastases are shown in Fig. 3a, details of mice with lung metastasis are given in Supplementary Table 1). These included 11 of our 20 stable PDX models (detailed in Table 3). Tumours which metastasised to the lung comprised 7 ER + Her2- models, 3 ER + Her2+ models, 1 ER-Her2+ models and 3 ER-Her2-, from 8 early breast cancer and 6 metastatic samples. Human origin of the cells was confirmed using a human-specific mitochondrial antibody (Fig. 3b). Lung metastases retained the hormone receptor status of the primary tumour (Fig. 3c–e shows ER (Fig. 3c), PR (Fig. 3d) and Her2 (Fig. 3e) from an ER+, PR+, Her2- tumour) and contained Ki67 positive dividing cells (Fig. 3f). Mammosphere colony formation (primary mammosphere generation) by patient-derived tumour cells before implantation into mice predicted the ability of a PDX tumour to metastasise *in vivo*, with samples with a higher percentage mammosphere forming efficiency being significantly more likely to metastasis to the lungs in PDX models ($p = 0.05$) (Fig. 3g).

Table 2 Summary of samples exhibiting cancer stem cell properties. Clinical data for all samples which formed mammospheres *in vitro* (n = 90 EBC, n = 61 MBC), a tumour in passage 1 *in vivo* ($n = 46$ EBC, $n = 15$ MBC) or a stable PDX model ($n = 13$ EBC, $n = 7$ MBC) is presented. Data is presented as successful growth / number of samples tested for each clinical characteristic. ER; oestrogen receptor, PR; progesterone receptor, Her2; Her2 receptor, NPI; Nottingham Prognostic Index. NA; data not available

		Formed Mammospheres		Formed tumour *in vivo* (P1)		Formed stable PDX model	
		EBC ($n = 90$)	MBC ($n = 61$)	EBC ($n = 46$)	MBC ($n = 15$)	EBC (n = 13)	MBC (n = 7)
Grade	1	3/4 (75 %)	3/3 (100 %)	1/6 (17 %)	NA	NA	NA
	2	34/45 (76 %)	27/28 (96 %)	22/50 (44 %)	8/14 (57 %)	4/22 (18 %)	3/8 (38 %)
	3	48/63 (76 %)	18/18 (100 %)	30/59 (34 %)	4/6 (67 %)	8/20 (40 %)	2/4 (50 %)
	Unknown	5/5 (100 %)	13/14 (93 %)	3/5 (60 %)	3/4 (75 %)	1/4 (25 %)	2/3 (67 %)
Hormone receptor status	ER+Her2-	50/65 (77 %)	46/48 (96 %)	26/71 (36 %)	10/19 (53 %)	2/26 (8 %)	5/8 (63 %)
	ER+Her2+	12/13 (92 %)	3/3 (100 %)	4/11 (36 %)	1/1 (100 %)	3/4 (75 %)	1/1 (100 %)
	ER-Her2+	6/10 (60 %)	1/1 (100 %)	4/9 (44 %)	1/1 (100 %)	1/4 (25 %)	0/1 (0 %)
	ER-Her2-	20/25 (80 %)	8/8 (100 %)	11/26 (42 %)	3/3 (100 %)	7/11 (64 %)	1/3 (33 %)
	Unknown	2/4 (50 %)	3/3 (100 %)	1/4 (25 %)	0/0 (0 %)	0/1 (0 %)	0/2 (0 %)
NPI	Good (<3.4)	10/12 (83 %)	NA	3/13 (23 %)	NA	0/2 (0 %)	NA
	Moderate/Poor (>3.4)	72/95 (76 %)	NA	33/91 (36 %)	NA	8/22 (36 %)	NA
	NA	8/10 (80 %)	NA	10/16 (63 %)	NA	5/22 (23)	NA

Table 3 Summary of Patient-Derived Xenografts created in this study. 20 stable PDX models were created during this study. Receptor status information is presented for clinical tumours and PDX tumours. The highest passage number for each model is presented, as is if PDX models spontaneously metastasise to the lung. EBC; early breast cancer. MBC; metastatic breast cancer. IDC; Invasive Ductal Carcinoma, DCIS; Ductal Carcinoma in situ, NA; data not available, NE; not examined

		Clinical information				PDX information				
Model	Type	Pathology	ER	PR	Her2	ER	PR	Her2	Passage	Lung Mets?
BB2RC08	EBC	NA	Pos	NA	Neg	Pos	Neg	Neg	P4	Yes
BB6RC80	EBC	IDC	Pos	Pos	Neg	Pos	Pos	Neg	P3	NE
BB6RC39	EBC	IDC	Pos	Pos	Pos	Pos	Pos	Pos	P3	Yes
BB6RC87	EBC	IDC/DCIS	Pos	Neg	Pos	NA	NA	NA	P1	NE
BB6RC160	EBC	IDC/DCIS	Pos	Neg	Pos	NA	NA	NA	P2	Yes
BB6RC148	EBC	IDC	Neg	Neg	Pos	NA	NA	NA	P3	Yes
BB6RC37	EBC	IDC	Neg	Neg	Neg	Neg	Neg	Neg	P3	No
BB6RC52	EBC	IDC	Neg	Neg	Neg	NA	NA	NA	P4	Yes
BB6RC69	EBC	IDC/DCIS	Neg	Neg	Neg	Neg	Neg	Neg	P3	No
BB6RC88	EBC	IDC	Neg	Pos	Neg	NA	NA	NA	P2	Yes
BB6RC153	EBC	IDC	Neg	Neg	Neg	Neg	Neg	Neg	P2	Yes
BB6RC191	EBC	NA	Neg	Neg	Neg	Neg	Neg	Neg	P2	NE
BB6RC193	EBC	IDC	Neg	Neg	Neg	NA	NA	NA	P1	No
BB3RC29	MBC	NA	Pos	Neg	Neg	Neg	Neg	Neg	P9	Yes
BB3RC31	MBC	IDC	Pos	Pos	Neg	Pos	Pos	Neg	P8	Yes
BB3RC32	MBC	IDC	Pos	Pos	Neg	Pos	Pos	Neg	P4	Yes
BB3RC50	MBC	IDC	Pos	Neg	Neg	Pos	Neg	Neg	P3	Yes
BB3RC72	MBC	NA	Pos	Pos	Neg	NA	NA	NA	P1	NE
BB3RC71	MBC	IDC	Pos	Pos	Pos	NA	NA	NA	P2	NE
BB3RC84	MBC	NA	Neg	Neg	Neg	Neg	Neg	Neg	P2	No

Finally we investigated whether any clinical parameters correlated to the ability to metastasise *in vivo*. PDX tumours established from high grade, early breast cancer samples were more likely to metastasise to the lungs, with 6/9 high grade (Grade 3) tumours metastasising compared to 1/10 low grade tumours (Grade 1 or 2) ($p = 0.02$) (Fig. 3h).

Discussion

There is evidence to suggest that breast cancer metastases are initiated by stem-like cells [11]. In the current study, we investigated breast CSC activity and metastasis using prospectively collected early and metastatic patient-derived samples *in vitro* and *in vivo*. We established that cells derived from metastatic fluids possessed an increased capacity to form both mammospheres *in vitro* and tumours *in vivo* compared to early breast cancer samples. Furthermore, we demonstrated that samples which metastasised to the lung in mice were those which possessed significantly higher mammosphere-forming efficiency *in vitro*.

To our knowledge, this is the largest prospective study of patient-derived tumour mammosphere formation. We compared mammosphere formation to clinical parameters across the whole patient cohort and found no correlation of mammosphere formation to clinical parameters such as tumour grade, molecular subtype or Nottingham Prognostic Index (NPI, derived from tumour size, number of involved lymph nodes and grade). However, higher mammosphere forming efficiency was observed in patients with ER-negative metastatic disease. This supports previous work where expression of the cancer stem cell marker ALDH1 has been shown to be more common in ER- tumours [13, 14]. *In vivo* tumour initiation capability did not correspond to the tumour characteristics of tumour grade, molecular subtype, hormone receptor status or NPI. The relationship between tumour characteristics and *in vivo* engraftment is controversial, with some studies showing hormone receptor status of the primary tumour to predict engraftment [15–18], and others showing no relationship [19, 20]. Our data adds to those studies suggesting that hormone receptor status does not predict engraftment *in vivo*.

Throughout this study we have utilised mammosphere culture to assess CSC activity *in vitro*. It should be noted that there are limitations to this technique. Aggregation of cells can occur, leading to a misinterpretation of results. To minimise cellular aggregation, we have fully optimised our protocols with regard to seeding density, culture time and culture

conditions [12]. Further, we have previously established that mammospheres can be generated from a single cell [21], and that when single mammospheres are disaggregated and re-plated at one cell per well, only one mammosphere will form [7, 8]. These results demonstrate that the mammospheres reported in our study are not a result of cellular aggregation. Despite limitations, the mammosphere assay provides the advantage that it can be performed on a small number of isolated cells. Given the low yields generated from some of our early breast cancer samples (median cell number isolated; 35,000), this assay allowed us to gain an *in vitro* measure of CSC activity in a far larger number of samples than would have been achievable using FACS-based assays.

We did not observe a correlation between mammosphere formation *in vitro* and tumour initiation *in vivo*, our two measures of CSC activity. These have been previously shown to correlate in samples taken from metastatic breast cancer patients [22]. However, although tumour-propagating ability can reflect sphere-forming capacity, they do not always correlate. This is likely to be related to differences in environmental factors *in vitro* and *in vivo*. In fact, we previously reported a strong positive correlation between mammosphere formation and tumour initiation in limiting dilution assays once PDX models are established [23].

Patient samples derived from metastases had increased CSC activity compared to early tumours both as mammospheres *in vitro* and by initiating tumours *in vivo*. These results contrast to one previous study, which reported no difference between tumour initiation ability between early and metastatic samples [19], but support recent work demonstrating that recurrent tumours have a higher engraftment rate *in vivo* [18]. The increased CSC activity observed in metastatic samples suggests that CSCs are those which are able to survive conventional treatment and become metastatic in breast cancer patients. This supports results of two previous studies. Firstly it has been demonstrated that high expression of CD44+/CD24- CSCs in tumours favours distant metastases [24], and secondly patient treatment with chemotherapy has been shown to increase CSC activity (as assessed by CD44+/CD24- expression and mammosphere forming efficiency) in patient-derived tumours [25]. As all of the patients in our metastatic cohort had been exposed to chemotherapy, this may provide an explanation for the increase in CSC activity in this group. It is unknown if chemotherapy promotes the expansion of CSCs, or if CSCs are selected by chemotherapy based on their intrinsic drug resistance. A further hypothesis is that metastatic niche selectively supports the growth of cancer stem cells resulting in an expansion of their numbers in the metastatic setting. Although the mechanism remains unclear, our findings emphasise the importance of developing anti-CSC treatments to prevent and treat metastatic disease.

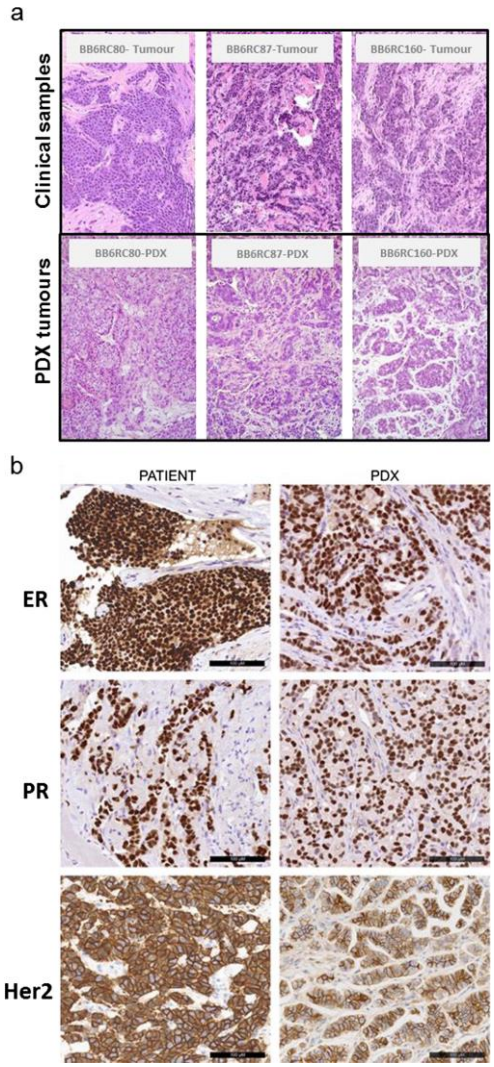
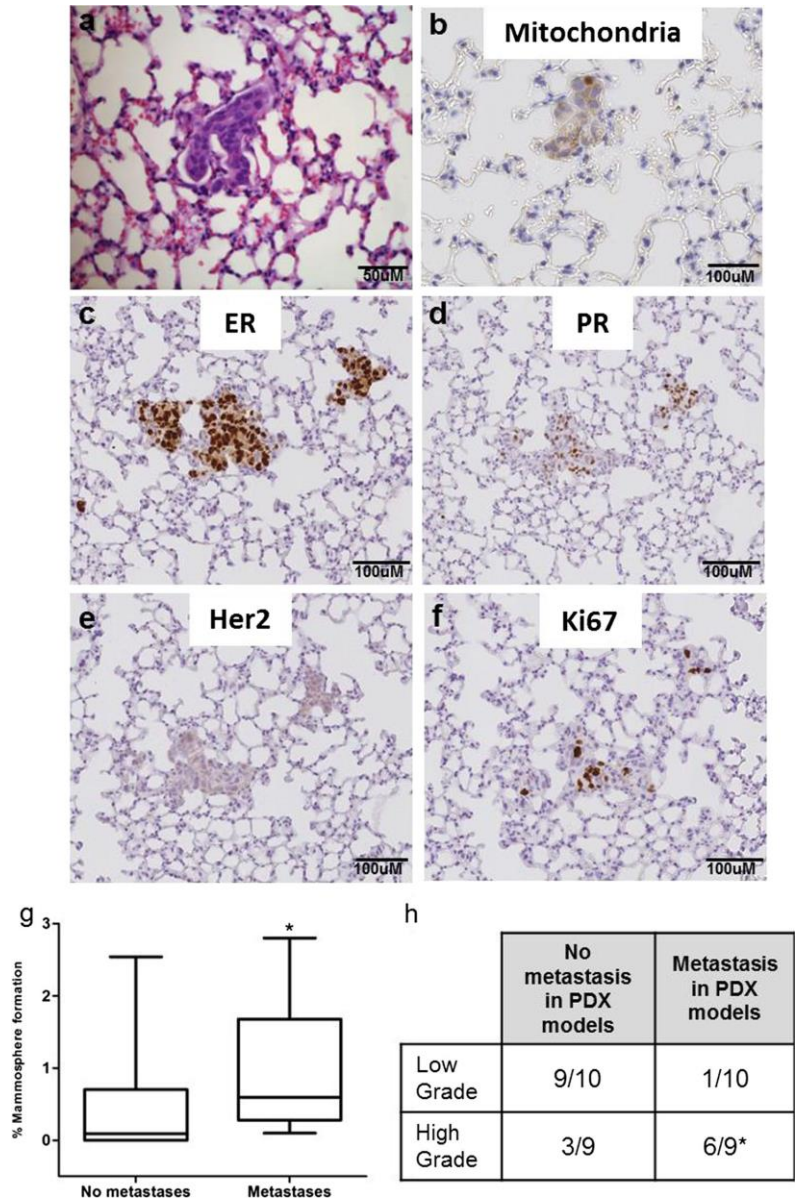


Fig. 2 Characterisation of Patient-Derived Xenograft tumours. Of the 144 tumours implanted in this study, 61 grew in the first passage in mice and 20 stable PDX models were created. Comparisons of representative patient tumour samples and their corresponding patient-derived xenograft in passage 1 **a**. Patient-derived xenografts retained ER, PR and Her2 status of the patient tumour. Representative images in **b**. Scale bar =100uM

In our study, we prospectively derived 20 stable breast cancer patient-derived xenograft (PDX) models from 144 implanted tumours. Importantly, 11 of 20 were derived from ER positive tumours, a subtype which has historically been more



	No metastasis in PDX models	Metastasis in PDX models
Low Grade	9/10	1/10
High Grade	3/9	6/9*

Fig. 3 Mammosphere colony formation predicts metastasis in PDX models. Lung metastases were detected in 14/34 models where a tumour formed in passage 1 by H&E **a** and confirmed by staining with a human specific mitochondrial antibody **b**. Lung metastases retained the ER **c**, PR **d** and Her2 **e** status of the primary tumour and contained ki67 positive dividing cells **f**. Primary mammosphere formation predicted PDX metastases, with samples which metastasised to the lungs *in vivo* having a significantly higher % mammosphere forming efficiency *in vitro* **g**. Metastases are more likely to form in PDX models from high grade EBC tumours (grade 3) than low grade (grade 1 and 2) **h**. Data are presented as mean ± SEM. *$p < 0.05$ Statistical analysis: Mann -Whitney U Test

difficult to successfully engraft [26]. Where we assessed spontaneous metastases in the models that formed tumours in the first generation, 14/34 (41 %) developed lung metastases, similar to previously published studies where 38/70 (54 %) PDX models were reported to develop lung metastases [15, 16, 19, 20, 27].

We are the first to assess metastatic capability in PDX models in relation to CSC frequency, and show that tumours which metastasise to the lung are those which have significantly higher CSC activity *in vitro*. Supporting this finding, a recent study demonstrated that single cells which metastasise in PDX models possess a gene signature similar to stem cells

[11]. A limitation of studying metastasis in PDX models is that the frequency and sites of metastases in PDX models may differ from the patient [28]. Follow up data at 5 or more years post diagnosis will be important to help determine if metastases in our models correspond to that in the patients from which they were generated. Independent of this, these models will provide a useful biological resource for the breast cancer community, and will now be used as a platform to test new breast cancer anti-metastasis therapies.

In summary, both *in vitro* and *in vivo* CSC activities are increased in metastatic samples, and CSC activity *in vitro* predicts metastasis *in vivo*. These results suggest that breast CSC activity may predict for poor outcome tumours and that anti-CSC treatment should be utilised in the prevention and treatment of breast metastasis. Further validation of these results is now required, both in larger patient/PDX cohorts and by assessing patient follow-up data.

Acknowledgments We would like to thank all patients who donated tissue to this study, and the Manchester Cancer Research Centre Biobank for consenting patients and collecting tissue.

Compliance with Ethical Standards

Funding This study was funded by Cancer Research UK and Breast Cancer Now. MK was funded by S cience and Technology Development Fund fellowship No. 6123.

Research involving Human Participants All patients underwent fully informed consent in accordance with local research ethics committee guidelines. Ethical approval for metastatic samples was granted by the Central Office for Research Ethics Committee study number 05/Q1402/25. Early breast cancer samples were collected via the Manchester Cancer Research Centre Biobank which is licensed by the Human Tissue Authority (licence number: 30,004) and has been ethically approved as a research tissue bank by the South Manchester Research Ethics Committee (Ref: 07/H1003/161 + 5).

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. O'Shaughnessy J. Extending survival with chemotherapy in metastatic breast cancer. Oncologist. 2005;10(Suppl 3):20–9. doi:10.1634/theoncologist.10-90003-20.
2. Chambers AF, Groom AC, MacDonald IC. Dissemination and growth of cancer cells in metastatic sites. Nat Rev Cancer. 2002;2(8):563–72. doi:10.1038/nrc865.
3. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003;100(7):3983–8. doi:10.1073/pnas.0530291100.
4. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med. 1997;3(7):730–7.
5. Reynolds BA, Weiss S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. Science. 1992;255(5052):1707–10.
6. Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, Wicha MS. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. 2003;17(10):1253–70.
7. Farnie G, Clarke RB, Spence K, Pinnock N, Brennan K, Anderson NG, et al. Novel cell culture technique for primary ductal carcinoma in situ: role of Notch and epidermal growth factor receptor signaling pathways. J Natl Cancer Inst. 2007;99(8):616–27. doi:10.1093/jnci/djk133.
8. Harrison H, Farnie G, Howell SJ, Rock RE, Stylianou S, Brennan KR, et al. Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. Cancer Res. 2010;70(2):709–18. doi:10.1158/0008-5472.CAN-09-1681.
9. Charafe-Jauffret E, Ginestier C, Iovino F, Wicinski J, Cervera N, Finetti P, et al. Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. Cancer Res. 2009;69(4):1302–13. doi:10.1158/0008-5472.CAN-08-2741.
10. Croker AK, Goodale D, Chu J, Postenka C, Hedley BD, Hess DA, et al. High aldehyde dehydrogenase and expression of cancer stem cell markers selects for breast cancer cells with enhanced malignant and metastatic ability. J Cell Mol Med. 2009;13(8B):2236–52. doi:10.1111/j.1582-4934.2008.00455.x.
11. Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature. 2015;526(7571):131–5. doi:10.1038/nature15260.
12. Shaw FL, Harrison H, Spence K, Ablett MP, Simoes BM, Farnie G, et al. A detailed mammosphere assay protocol for the quantification of breast stem cell activity. J Mammary Gland Biol Neoplasia. 2012;17(2):111–7. doi:10.1007/s10911-012-9255-3.
13. Park SY, Lee HE, Li H, Shipitsin M, Gelman R, Polyak K. Heterogeneity for stem cell-related markers according to tumor subtype and histologic stage in breast cancer. Clin Cancer Res: Off J American Assoc Cancer Res. 2010;16(3):876–87. doi:10.1158/1078-0432.CCR-09-1532.
14. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell. 2007;1(5):555–67. doi:10.1016/j.stem.2007.08.014.
15. Marangoni E, Vincent-Salomon A, Auger N, Degeorges A, Assayag F, de Cremoux P, et al. A new model of patient tumor-derived breast cancer xenografts for preclinical assays. Clin Cancer Res: Off J American Assoc Cancer Res. 2007;13(13):3989–98. doi:10.1158/1078-0432.CCR-07-0078.
16. Zhang X, Claerhout S, Prat A, Dobrolecki LE, Petrovic I, Lai Q, et al. A renewable tissue resource of phenotypically stable, biologically and ethnically diverse, patient-derived human breast cancer xenograft models. Cancer Res. 2013;73(15):4885–97. doi:10.1158/0008-5472.CAN-12-4081.
17. Petrillo LA, Wolf DM, Kapoun AM, Wang NJ, Barczak A, Xiao Y, et al. Xenografts faithfully recapitulate breast cancer-specific gene expression patterns of parent primary breast tumors. Breast Cancer Res Treat. 2012;135(3):913–22. doi:10.1007/s10549-012-2226-y.
18. Moon HG, Oh K, Lee J, Lee M, Kim JY, Yoo TK, et al. Prognostic and functional importance of the engraftment-associated genes in

the patient-derived xenograft models of triple-negative breast cancers. Breast Cancer Res Treat. 2015;154(1):13–22. doi:10.1007/s10549-015-3585-y.

19. DeRose YS, Wang G, Lin YC, Bernard PS, Buys SS, Ebbert MT, et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. Nat Med. 2011;17(11):1514–20. doi:10.1038/nm.2454.
20. Visonneau S, Cesano A, Torosian MH, Miller EJ, Santoli D. Growth characteristics and metastatic properties of human breast cancer xenografts in immunodeficient mice. Am J Pathol. 1998;152(5):1299–311.
21. Singh JK, Farnie G, Bundred NJ, Simoes BM, Shergill A, Landberg G, et al. Targeting CXCR1/2 significantly reduces breast cancer stem cell activity and increases the efficacy of inhibiting HER2 via HER2-dependent and -independent mechanisms. Clin Cancer Res: Off J American Assoc Cancer Res. 2013;19(3):643–56. doi:10.1158/1078-0432.CCR-12-1063.
22. Grimshaw MJ, Cooper L, Papazisis K, Coleman JA, Bohnenkamp HR, Chiapero-Stanke L, et al. Mammosphere culture of metastatic breast cancer cells enriches for tumorigenic breast cancer cells. Breast Cancer Res: BCR. 2008;10(3):R52. doi:10.1186/bcr2106.
23. Simoes BM, O'Brien CS, Eyre R, Silva A, Yu L, Sarmiento-Castro A, et al. Anti-estrogen Resistance in Human Breast Tumors Is Driven by JAG1-NOTCH4-Dependent Cancer Stem Cell Activity. Cell Rep. 2015;12(12):1968–77. doi:10.1016/j.celrep.2015.08.050.
24. Abraham BK, Fritz P, McClellan M, Hauptvogel P, Athelogou M, Brauch H. Prevalence of CD44+/CD24−/low cells in breast cancer may not be associated with clinical outcome but may favor distant metastasis. Clin Cancer Res: Off J American Assoc Cancer Res. 2005;11(3):1154–9.
25. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst. 2008;100(9):672–9. doi:10.1093/jnci/djn123.
26. Cottu P, Marangoni E, Assayag F, de Cremoux P, Vincent-Salomon A, Guyader C, et al. Modeling of response to endocrine therapy in a panel of human luminal breast cancer xenografts. Breast Cancer Res Treat. 2012;133(2):595–606. doi:10.1007/s10549-011-1815-5.
27. Liu H, Patel MR, Prescher JA, Patsialou A, Qian D, Lin J, et al. Cancer stem cells from human breast tumors are involved in spontaneous metastases in orthotopic mouse models. Proc Natl Acad Sci U S A. 2010;107(42):18115–20. doi:10.1073/pnas.1006732107.
28. Whittle JR, Lewis MT, Lindeman GJ, Visvader JE. Patient-derived xenograft models of breast cancer and their predictive power. Breast Cancer Res: BCR. 2015;17:17. doi:10.1186/s13058-015-0523-1.