

Imperial College
London

**Epigenomic profiling of oestrogen
receptor-positive breast cancer**

Darren K. Patten

CID:00329863

**Thesis submitted for the
degree of Doctor of
Philosophy**

**Supervisors: Professor Charles Coombes, Dr.
Luca Magnani & Professor Carlo Palmieri**

June 2017

Supported by
wellcometrust

Abstract

Endocrine resistance occurs in the majority of patients who have relapsing breast cancer which remains a challenge to current medical treatment. Tumour heterogeneity poses an obstacle in clinical diagnostics which can lead to suboptimal treatment and the development of therapy resistance. Although tumour heterogeneity, at an intrinsic as well as at an interpatient level, is a potential contributor to endocrine resistance in oestrogen receptor-positive breast cancer, it is not clearly understood. Epigenetically-defined regulatory regions including promoters and enhancers govern cell type specific transcription. In this study, I have processed 47 fresh-frozen primary and metastatic oestrogen receptor-positive breast cancer samples for H3K27ac (acetylation of lysine 27 on histone 3) chromatin immunoprecipitation and high throughput sequencing. The regulatory landscapes of the 47 patient samples were analysed and key regulatory drivers identified. Results demonstrate that although enhancer elements are strongly associated with epigenetic heterogeneity, dominant enhancers tend to be shared across patients. Identified regulatory drivers possess binding sites for the transcription factor YY1 which collaborates with the oestrogen receptor alpha to drive the transcription of several key genes including *SLC9A3R1*. *SLC9A3R1* has subsequently been identified as a novel breast cancer specific oncogene and is shown in this study to stratify oestrogen receptor-positive breast cancer patients. This study demonstrates for the first time that epigenetic annotations can be used to map regulatory, as well as phenotypic, changes through breast cancer progression.

Declaration of originality

This thesis has been submitted in fulfilment of the requirements for a degree of Doctor of Philosophy. I declare that the work reported in this write-up is my own, or if not, it is clearly stated and fully acknowledged.

Darren K. Patten

June 2017

Copyright Declaration

The copyright of this thesis is the responsibility of the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgements

Firstly, I would like to thank the Wellcome Trust for providing the funding towards this project (Grant number: 103034/Z/13/Z) which enabled me to carry out this work.

I would like to thank and express my appreciation to my supervisors, Professor Charles Coombes, Dr. Luca Magnani and Professor Carlo Palmieri. Their guidance, inspiration, patience and most of all, support, have been pivotal to my project and progression during this PhD.

I would like to also thank all the collaborators involved in this project; Edward Curry, Sami Shousha, Balázs Györfy, Edina Erdős, Kate Goddard, Andrea Vingiani, Lőrinc Sándor Pongor, Dimitri J Hadjiminis, Gais Schiavon, Peter Barry, Peter Scacheri and Giancarlo Pruneri.

I would also like to express my gratitude to Mr. Giacomo Corleone who was meticulously involved in the bioinformatic analysis conducted in this project. I would like to thank all my present colleagues, Nikolaos Trasanidis, Professor Sung Pil Hong, Dr. Ylenia Perone, Seham Al-Shehri and Neil Slaven for their advice support.

A big thank you goes to Miss. Asmita Thapa for all her priceless help in reading and providing useful suggestions for the thesis.

I would also like to thank the Imperial College Tissue Banks, Breast Cancer Now Tissue bank (Grant number: 2013MaySP013) and Royal Marsden NHS Foundation Trust tissue bank (Study number: CCR3995; Ethics study number: 13/LO/1535) for providing the precious tissue samples for this project.

Finally, I would like to thank my family (and in particular, Mr. Dee Mauree) and Mr. Samson Patten for their everlasting encouragement, love and support in these times.

Table of contents

Abstract	2
Declaration of originality	3
Copyright Declaration	4
Acknowledgements	5
Table of contents	6
List of figures	10
List of tables	15
Abbreviations	17
Chapter 1 Introduction	22
1.1 Cancer	22
1.2 Hallmarks of cancer	24
1.3 Breast cancer	26
1.4 The function of the oestrogen receptor in physiological human development	27
1.5 The oestrogen receptor in breast cancer and molecular subtypes	30
1.6 The treatment of breast cancer	31
1.7 Tumour heterogeneity	34
1.8 DNA Sequencing in breast cancer	40
1.9 Epigenetic modifications	41
1.10 Enhancer and super-enhancers	43
1.11 Super-enhancers in cancer	46
1.12 Study Aims	49
Chapter 2 Materials and Methods	50
2.1 Breast tumour selection criteria	50
2.2 Cell culture	51
2.3 Tumour processing	53
2.4 Preparation of cell lines for ChIP	55

2.5 Chromatin Immunoprecipitation and high throughput next generation sequencing	55
2.6 Sequence alignment and peak calling	59
2.7 Bioinformatic analysis of “training set” of tumour samples	59
2.8 Bioinformatic analysis of complete set of tumour samples; a novel approach....	61
2.9 Small interference RNA (siRNA) transfections	66
2.10 Luciferase gene reporter assay	67
2.11 Cell growth assay	68
2.12 Total RNA extraction	69
2.13 Complementary DNA (cDNA) synthesis	70
2.14 Protein lysate preparation; Radioimmune Precipitation Buffer (RIPA) buffer extraction.....	71
2.15 Protein quantification	71
2.16 Western blotting	72
2.17 Quantitative real-time PCR.....	74

Chapter 3 Collaborative work and acquired lab skills 75

3.1 “DMXL2 drives epithelial to mesenchymal transition in hormonal therapy resistant breast cancer through Notch hyper-activation”.....	75
3.2 “The pioneer factor PBX1 is a novel driver of metastatic progression in ER α -positive breast cancer”	77
3.3 Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion	79
3.4 Acquired laboratory techniques	81

Chapter 4 Results: Optimisation of chromatin immunoprecipitation and high-throughput sequencing 82

4.1 Overview of ChIP	82
4.2 The ChIP protocol revised	83
4.3 Results of the standard compared to the optimised ChIP protocol.....	88
4.4 Sequencing results obtained for YY1 ChIP-Seq in MCF7 cells	88
4.5 Sequencing results obtained for H3K27ac ChIP-Seq in murine cancer associated fibroblast (CAF) cells	91

Chapter 5 Results: Analysis of the training set	94
5.1 Aims of the analysis of the initial training set of breast tumour samples.....	94
5.2 Clinicopathological details of the tumour samples.....	94
5.3 Sequencing results.....	96
5.4 Epigenetic analysis of sequenced breast tumours	98
5.5 Super-enhancer identification and analysis.....	100

Chapter 6 Results: Epigenomic profiling of primary and metastatic breast tumour samples using H3K27ac ChIP-Seq	104
6.1 Aims of the analysis of the complete set of primary and metastatic breast tumour samples.....	104
6.2 Clinicopathological details of tumour samples.....	105
6.3 Summary statistics for ChIP-Seq analyses	111
6.4 H3K27ac ChIP-Seq coverage of enhancer and promoters in patient samples.	116
6.5 Intratumoural and intertumoural heterogeneity in ER α -positive breast cancer	117
6.6 The transcriptional landscape of ER α -positive breast cancer.....	126
6.7 YY1 enhancer as a novel driver of ER α -positive breast cancer	134
6.8 The role of YY1 in ER α binding in breast cancer.....	142

Chapter 7 Results: SLC9A3R1 and phenotypic mapping	155
7.1 The role of SLC9A3R1 in ER α -positive breast cancer	155
7.2 The application of YY1 and SLC9A3R1 enhancer activity for phenotypic mapping	166

Chapter 8 Discussion	173
8.1 Optimisation of ChIP-Seq (Chapter 4).....	173
8.2 Analysis of the training set of tumours (N=31) (Chapter 5)	174
8.3 Epigenomic profiling of primary and metastatic breast tumour tissue samples using H3K27ac ChIP-Seq and SLC9A3R1 (Chapter 6 and Chapter 7).....	178
8.3.1 ChIP-Seq results in the tissue samples.....	180
8.3.2 Intratumoural and intertumoural heterogeneity in ER α -positive breast cancer and the use of epigenetic information	182

8.3.4 The transcriptional landscape of ER α -positive breast cancer.....	184
8.3.5 YY1 a novel driver in ER α -positive breast cancer	184
8.3.6 The role of SLC9A3R1 in ER α -positive breast cancer.....	186
 8.3.7 Conclusions	 188
References	190
Appendix	232

List of figures

Figure 1.1 Cancer incidence in the U.K.	23
Figure 1.2 Cancer mortality in the U.K.	24
Figure 1.3 The hallmarks of cancer.	25
Figure 1.4 The average number of new female breast cancer cases per year with age-specific incidence rates in the U.K. between 2011- 2013.	26
Figure 1.5 Overview of ER- α function.	29
Figure 1.6 Tumour heterogeneity in diagnostics.	39
Figure 1.7 Epigenetic regulation of ER α -target genes in breast cancer cells.	42
Figure 1.8 Enhancer –gene interaction.	43
Figure 1.9 Definition and general properties of super-enhancers.	45
Figure 1.10 Comparison of gene expression coordinated by “common” enhancer gene and super-enhancer gene interaction.	46
Figure 1.11 Different types of malignancy can utilise different enhancers to target the same gene.	47
Figure 4.1 The standard ChIP-Seq protocol adapted from Schmidt <i>et al.</i>	84
Figure 4.2 The optimised ChIP-Seq protocol modified from Schmidt <i>et al.</i> [139]	87
Figure 4.3 H3K27ac peaks called by MACS 2.0 comparing standard (blue bar) protocol to the optimised (green bar) protocol.	93
Figure 4.4 YY1 peaks called by MACS 2.0 comparing standard (blue bar) protocol to the optimised (green bar) protocol.	93
Figure 5.1 Principal Component analysis on H3K27ac tumour profiles (N=31).	98
Figure 5.2 Correlation heat map representation for significantly differentially enriched sites for H3K27ac marker in poor, good and metastatic outcome samples.	99

Figure 5.3 Graphical and numerical representation of the called super-enhancers in Good, Poor and Metastatic tumours samples	101
Figure 5.4 Heat map: clustering of tumours (N=31) based on the called super-enhancers.	102
Figure 5.5 Multidimensional scaling plot of Jaccard distances, based on identified super-enhancers, of the Good, Poor, Metastatic tumour groups (N=31).....	103
Figure 6.1 MACS 2.0 generated peaks (q-value = 00.1) for H3K27ac ChIP-Seq in 48 tumour samples following exclusion of the samples which generated less than 2000 peaks.....	115
Figure 6.2 (A) Saturation plots for promoters and enhancers in N= 47 tumour samples generated by H3K27ac-enriched regions.....	116
Figure 6.3 MCF7 H3K27ac ChIP-qPCR.....	118
Figure 6.4 Spiking ER α ChIP experiment using MCF7 cells at different percentages +/- oestradiol stimulation.	119
Figure 6.5 Linear regression analysis of ranking index (RI – obtained from H3K27ac data from patient samples) vs sharing index (which indicates the number of patients sharing a particular regulatory region).....	122
Figure 6.6 Linear regression graphs using patient-derived RNA levels from the METABRIC, TCGA and Affymetrix (AFFY) datasets	125
Figure 6.7 Schema of the bioinformatic framework for analysis of transcription factor (TF) binding sites within the patient-derived H3K27ac data.	128
Figure 6.8 Overlap between imputed DHS/H3K27ac regions with in vivo-derived ER α binding sites.	129
Figure 6.9 Distribution of in vivo-derived ER α binding sites compared to the number of patients in which these binding sites were observed.....	130

Figure 6.10A Transcription factor motif analysis applied to sharing index (SI) bins of enhancers followed by unsupervised clustering.	132
Figure 6.10B Transcription factor motif analysis applied to sharing index (SI) bins of promoters followed by unsupervised clustering.....	133
Figure 6.11 Chromatin interaction analysis.	136
Figure 6.12 Graph of RI in YY1 determined within patient-derived samples (N=47).	137
Figure 6.13 The clonal YY1-enhancer A (pictured in the top right-hand corner of the figure) using Epigenomic Roadmap 155 H3K27ac data in normal tissues classified according to high (H), medium (M) and low (L) ranking index (RI).	138
Figure 6.14 YY1 RNA expression in breast cancer datasets.....	139
Figure 6.15 Immunohistochemistry staining of YY1 in breast tissue.	141
Figure 6.16 YY1 ChIP-Seq data in ER α -positive MCF7 cells stimulated with (control) vehicle (ETOH = ethanol) vs oestradiol (E2) for 45 minutes.	143
Figure 6.17 Heatmaps of YY1 overlap with H3K27ac enriched regions associated with active regulatory regions.	144
Figure 6.19 A) In vivo core ER α binding sites are strongly enriched for YY1 binding	145
Figure 6.20 Western blot analysis for YY1 depletion using two types of siRNA (a/b) in MCF7 cells stimulated with vehicle vs oestradiol for 6 hours.	146
Figure 6.21 Luciferase tagged with oestrogen responsive element (ERE) promoter reporter assay performed in MCF7 cells	147
Figure 6.22 A) Proliferation assay of MCF7 cells grown in vehicle control vs. oestradiol (E2) with transfection of silencer control (siCtrl), siRNA (a) and siRNA (b).	147

Figure 6.23 A) Proliferation assay of ZR75-1 cells transfected with silencer control (siCtrl), siRNA (a) and siRNA (b).....	148
Figure 6.24 Gene expression data in MCF7 cells transfected with siRNA for YY1 and silencer control (siCtrl).....	150
Figure 6.25 ChIP-Seq carried out in MCF7 (with oestradiol stimulation, E2) and LTED cells for YY1.....	151
Figure 6.26 ChIP-Seq conducted in LTED cells for YY1 overlapped with ER α ChIP-Seq.....	152
Figure 6.27 The relative enrichment of ER α -YY1 bound enhancers near oestrogen responsive genes that do not respond to Tamoxifen therapy.....	154
Figure 7.1 SLC9A3R1 and ESR1 expression in MCF7 cells and MCF7 variants...	155
Figure 7.2 Expression data from MCF7 cell lines after silencing of the oestrogen receptor.....	156
Figure 7.3 Meta-analytical study assessing the transcriptional response of SLC9A3R1 in response to oestrogens	157
Figure 7.4 Expression levels of SLC9A3R1 (top) and progesterone receptor (PGR) expression in 51 patients undergoing neo-adjuvant aromatase inhibitors prior to surgery for ER α -positive BC.....	158
Figure 7.5 Global Kaplan-Meier analysis summarises univariate analysis for each gene included in the Affymetrix [222] microarray platform (N=724 patients).	159
Figure 7.6 Survival data in three independent datasets.	160
Figure 7.7 SLC9A3R1 expression in various cell lines.	162
Figure 7.8 The regulatory landscape of SLC9A3R1 (chr17:72, 721, 520-72, 788, 719).....	163
Figure 7.9 Patient-derived mRNA expression levels of SLC9A3R1	164

Figure 7.10A Western blot analysis for SLC9A3R1 depletion using two types of siRNA (a/b) in MCF7 cells stimulated with vehicle vs oestradiol for 6 hours.	164
Figure 7.10B Proliferation assays performed in MCF7 (Left) and LTED (Right) when silencing SLC9A3R1.	165
Figure 7.11 The ranking index (RI) in SLC9A3R1 enhancer activity determined within patient-derived samples (N=47).	167
Figure 7.12 YY1 enhancer activity in MCF7 (M) and MCF7 derivative cell lines. ...	168
Figure 7.13 SLC9A3R1 enhancer activity in MCF7 (M) and MCF7 derivative cell lines.	169

List of tables

Table 2.1 Breast cancer cell lines with corresponding hormone receptor status	51
Table 2.2 Reagents used for chromatin immunoprecipitation.....	54
Table 2.3 Primer sequences used for ER- α and H3K27ac ChIP qPCR and revalidation of library prepared samples	58
Table 2.4 Primer sequences used for YY1 ChIP QPCR and revalidation of library prepared samples.....	58
Table 2.5 Reagents required per reaction to reverse transcribe 1 μ g of total RNA to cDNA using the iScript kit.....	70
Table 2.6 Primer sequences of targets used for RT-qPCR experiments.....	74
Table 4.1. Showing sequencing results obtained from the standard and optimised ChIP protocols.....	91
Table. 5.1 Showing the characteristics of the collected tumour samples which have been processed for ChIP-Seq.	94
Table 5.2 Percentage (%) alignment of the raw (FASTQ) files to the hg19 reference genome and number of positive peaks called.	97
Table 6.1 (A/B) Characteristics of the collected primary (Table 6A) and metastatic (Table 6B) tumour samples which have been processed for H3K27ac ChIP-Seq..	105
Table 6.1A Clinicopathological information for primary breast tumour samples collected and processed for H3K27ac ChIP-Seq.	106
Table 6.1A Showing clinicopathological information for primary breast tumour samples collected and processed for H3K27ac ChIP-Seq.....	107
Table 6.1B Clinicopathological information for metastatic breast tumour samples collected and processed for H3K27ac ChIP-Seq.	108

Table 6.1B Clinicopathological information for metastatic breast tumour samples collected and processed for H3K27ac ChIP-Seq.	109
Table 6.2A Alignment rates generated from Bowtie [135], number of raw reads before and after redundant rates are applied and peaks called by MACS 2.0 [136] (q-value = 0.01) for all the primary breast cancer samples.	112
Table 6.2A Alignment rates generated from Bowtie [135], number of raw reads before and after redundant rates are applied and peaks called by MACS 2.0 [136] (q-value = 0.01) for all the primary breast cancer samples.	113
Table 6.2B Showing alignment rates generated from Bowtie [135], number of raw reads before and after redundant rates are applied and peaks called by MACS 2.0 [136] (q-value = 0.01) for all the metastatic breast cancer samples.	114
Table 7.1 Available formalin-fixed paraffin embedded (FFPE) slides for the 23 of the 47 patient-tumour samples processed for H3K27ac ChIP-Seq.	172

Abbreviations

5-methyl cytosine = 5mC

ATCC = American Tissue Type Culture Collection

EAS = European age-standardised

AI = Aromatase inhibitor

bp = Base pair

BSA = Bovine serum albumin

BC = Breast cancer

BCa = Breast cancer

BCN = Breast Cancer Now

BCS = Breast conserving surgery

CB = Cholesterol biosynthesis

ChIP = Chromatin immunoprecipitation

CSC = Cancer stem cell

ChIA-PET = Chromatin interaction analysis by paired-end tag sequencing

CR = Contralateral recurrence

DCIS = Ductal carcinoma in situ

DCFCS = Charcoal-stripped fetal calf serum

ddH₂O = Double distilled water

DHS = DNase I hypersensitivity sites

DMEM = Dulbecco's Modified Eagle's Medium

DNMT = DNA methyl-transferase

DXT = Radiotherapy

E1 = Oestrone

E2 = Oestradiol

E3 = Oestriol

ECACC = European Collection of Cell Cultures

ECM = Extracellular matrix

enh = Enhancer

ER = Oestrogen receptor

ER+ = Oestrogen receptor positive

ER- = Oestrogen receptor negative

ER α = Oestrogen receptor alpha

ER β = Oestrogen receptor beta

ERE = Oestrogen responsive element

eRNA = Enhancer RNA

ET = Endocrine therapy

FCS = Fetal calf serum

FOXA-1 = Forkhead box A-1

GnRH = Gonadotrophin releasing hormone

GREAT = Genomic Regions Enrichment of Annotations Tool

GRN = Gene-regulatory network

H3K27ac = Acetylation of lysine 27 of histone H3

H3K4me2 = Dimethylation of lysine 4 of histone H3

H3K4me1 = Monomethylation of lysine 4 of histone H3

H3K4me3 = Trimethylation of lysine 4 of histone H3

H3K27me3 = Trimethylation of lysine 27 of histone H3

H&E = Haematoxylin and eosin

HER2 = Human epidermal growth factor receptor 2

ICHTB = Imperial College Healthcare NHS trust Tissue Bank

ICR = Institute of Cancer Research

IDC = Invasive ductal carcinoma

IHC = Immunohistochemistry

ILC = Invasive lobular carcinoma

IMC = Invasive mucinous carcinoma

ITH = Inter-tumour heterogeneity

LR = Local recurrence

LTED = Long-term oestrogen deprived MCF7 cells

M = molar

μ M = Micromolar

mM = Millimolar

Mb = Megabase

Mbase = Million bases

MCF7FulvR = MCF7 fulvestrant resistant cells

MDS = multidimensional scaling

mESC = Murine embryonic stem cell

NFR = Nucleosome-free region

O.S./OS = Overall survival

PBS = Phosphate buffered saline

PBST = Phosphate buffered saline tween

PCA = Principal component analysis

PBX1 = Pre-B cell leukaemia transcription factor 1

PF = Pioneer factor

POL II = RNA polymerase 2/II

PR/PGR = Progesterone receptor

qPCR = Quantitative polymerase chain reaction

PSG = 100 U penicillin/0.1 mg mL⁻¹ streptomycin, 2mM L-glutamine

rcf = relative centrifugal force

RT-qPCR = Real time quantitative polymerase chain reaction

RD = Regulatory driver

RN = Regulatory noise

ROSE = Ranked ordering of super-enhancers

RPKM = Reads per kilobase of transcript per million mapped reads

RT = Reverse transcriptase

SE = Super-enhancer

SERM = Selective oestrogen receptor modulator

siRNA = Small interference RNA

SNP = Single nucleotide polymorphism

SRB = Sulphorhodamine B

TAD = Topologically associated domain

T-ALL = T cell acute lymphoblastic leukaemia

TAM-R = Tamoxifen resistant

TCA = Trichloroacetic acid

TF = Transcription factor

TMA = Tissue microarray

TNBC = Triple negative breast cancer

TSS = Transcription start site

VERB = A Study to Measure In Vivo Changes in Oestrogen Receptor DNA Binding Events in Breast Cancer Treated with Endocrine Therapy for Primary or Recurrent Disease

Chapter 1 Introduction

1.1 Cancer

Cancer continues to be a major health issue and is still responsible for the leading causes of death around the globe. Worldwide cancer statistics of 2012 demonstrate 14.1 million newly diagnosed cases (excluding non-melanoma skin cancer) and 8.2 million attributable deaths. It has been predicted that 23.6 million new cancer cases will be diagnosed globally by 2030, which is 68% more cases than in 2012 [1]. The four most commonly diagnosed cancers include 1. Lung, 2. Breast (female), 3. Bowel and 4. Prostate [2].

In the U.K. alone, 352,000 people were diagnosed with cancer in 2013. This equates to a European age-standardised (EAS) incidence rate of 605 people per 100,000 of the human population [3]. In 2013, male cancer diagnoses reached 179,000 and in females, 173,000. Incidence rates for all cancers combined have increased in the U.K. by approximately 30% since the late 1970s; this is exemplified by a 37% and 17% increase in female and male cases, respectively (Figure 1.1). Lung, breast, bowel and prostate cancers together accounted for 53% of all new cases in the U.K. [3]. Lung cancer remains the most common cause of cancer death in the U.K., accounting for 22% of all cancer-related deaths in both sexes (2014), followed by bowel (10%), breast (7%) and prostate (7%). Cancer-related mortality is decreasing in the U.K. despite the aforementioned increases in incidence (Figure 1.2). Earlier diagnosis coupled with improved methods of treatment have largely contributed to the fall in cancer mortality. There was a 27% decrease in the EAS cancer mortality rates in males and 20% decrease in females reported between 1990-1992 and 2010-2012 [4-6]. These rates have diminished over the last 10 years with AS mortality

rates declining by 13% and 9% in males and females, respectively. A further 17% fall in cancer mortality rates is expected between 2011 and 2030 in the U.K [7, 8].

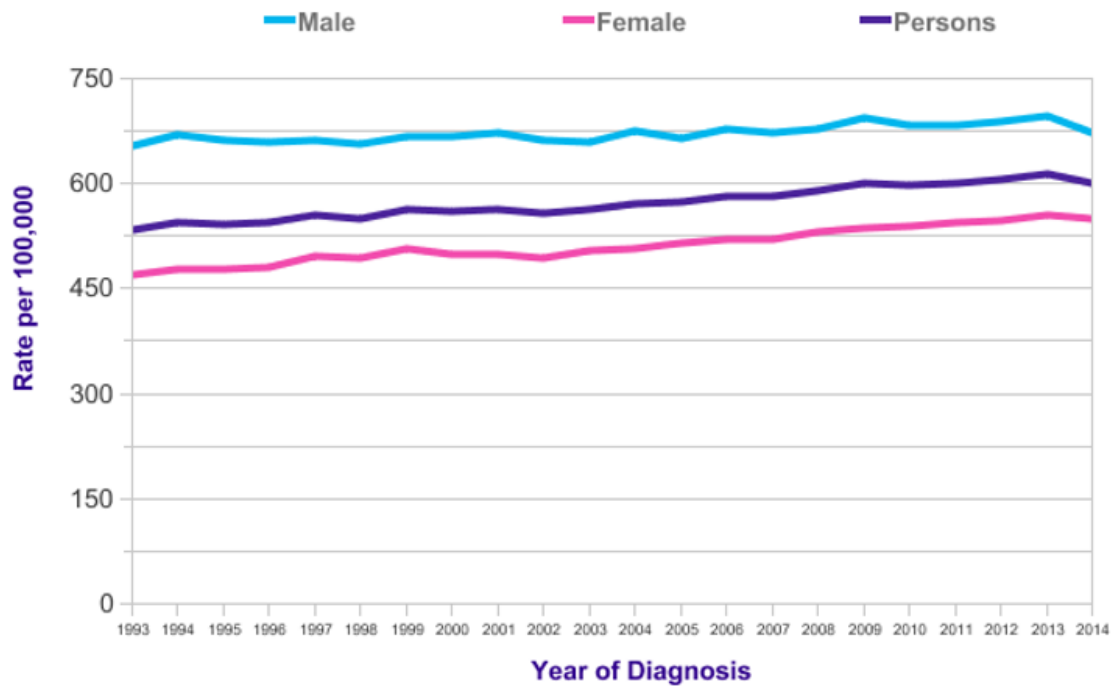


Figure 1.1 Cancer incidence in the U.K. Data collected from 1993 to 2014 showing European Age-Standardised incidence rates of all cancers (excluding non-melanoma skin cancer). [Modified from the Cancer Research UK website, 2016] ^[3]

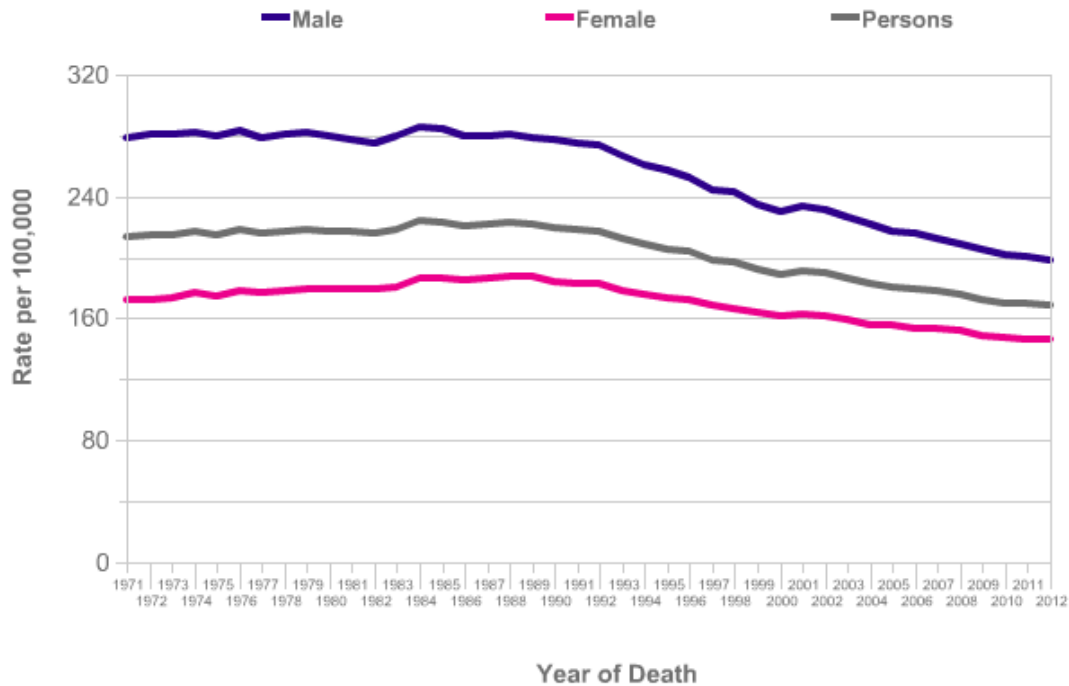


Figure 1.2 Cancer mortality in the U.K. Data collected from 1971 to 2012 showing European Age-Standardised mortality rates of all cancers. [Modified from the Cancer Research UK website, 2016] ^[3]

1.2 Hallmarks of cancer

Cancer is postulated to arise from one single cell. The oncogenic transformation of a normal cell into a malignant one, is a multistage process involving contributory genetic changes which lead to the gradual conversion of normal human cell processes to an oncogenic state [9, 10].

Advances in cancer research over the past ten years have elucidated the stages of development and progression. Cancer cells are not just thought of as a mass of uncontrolled proliferating cells; they are also composed of various cell types, such as normal cells contributing to the tumour-associated stroma which are involved in initiation, development and finally metastasis. Six common traits outlining the hallmarks of cancer were initially proposed by Hanahan and Weinberg in 2000 [10]. These include 1. Self-sufficiency in growth signals 2. Insensitivity to anti-growth

signals 3. Evading programmed cell death 4. Limitless replicative potential 5. Developing blood vessels and 6. Tissue invasion and metastasis. These aforementioned traits were then revisited in 2011 following which, Hanahan and Weinberg updated the hallmarks with 4 further characteristics; 1. Deregulated metabolism 2. Evading the immune system 3. Genome instability and 4. Inflammation. (Figure 1.3) [9, 10].



Figure 1.3 The hallmarks of cancer. Biological traits outlining the main characteristics shared in common by all human cancers. [Modified from Hanahan and Weinberg, 2011] ^[10]

1.3 Breast cancer

On a global scale, breast cancer (BC) is currently the second most common cancer after lung with 1,676,633 and 1,824,701 cases diagnosed in 2012, respectively. BC is the most common cancer in females accounting for 25% of the 6.7 million female population diagnosed in 2012. Approximately 1 in 8 women will develop BC in the U.K. at some point during their lifetime; 53,700 new cases of BC were diagnosed accounting for 15% of all new cancer cases in 2013. In males 340 (0.6%) new cases were diagnosed leaving 53,360 (99.4%) new female cases. Approximately 46% of female BC cases in the U.K. each year are diagnosed in females over the age of 65 (Figure 1.4). In comparison, 51% of male BC cases in the U.K. each year are detected in men over the age of 70 [11].

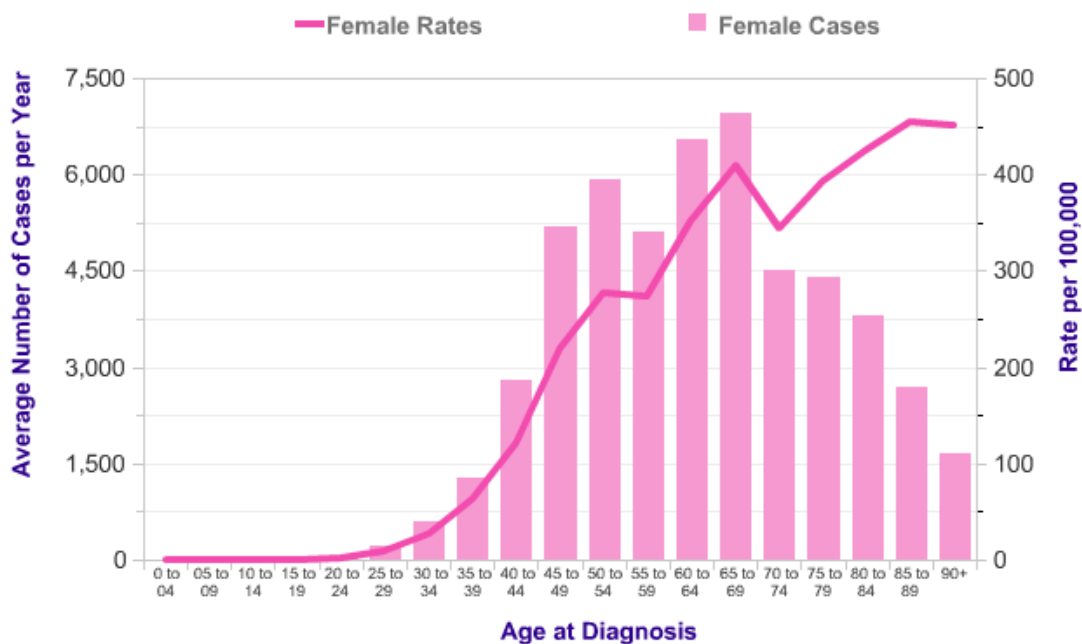


Figure 1.4 The average number of new female breast cancer cases per year with age-specific incidence rates in the U.K. between 2011- 2013. [Modified from the Cancer Research UK website, accessed 2016] ^[11]

The incidence of breast cancer has increased by approximately 54% in the U.K. since the late 1970s. However, over the last ten years BC incidence in the U.K. has

increased by approximately 4% [11]. In addition, 11,433 deaths were reported in 2013 attributable to BC. This equates to 28 deaths per 100,000 women per annum, which is amongst the highest in the world [11].

1.4 The function of the oestrogen receptor in physiological human development

The oestrogen receptor (ER) alpha (α), a nuclear receptor, is a master transcription factor involved in physiological breast development. It is central to the pathogenesis of BC, and is a key therapeutic target [12, 13]. Unoccupied ER α is predominantly situated in the nucleus, either bound or unbound to its consensus motif. It is activated by oestrogens (steroid hormones) causing it to dimerise and translocate into the nucleus [14-16]. This then allows ER α to subsequently bind to its respective target sites across the genome, ultimately controlling cell proliferation and treatment response (Figure 1.5). Oestrogens are involved in regulating the development and maintenance of female characteristics which include the reproductive organs and the mammary glands [17]. In addition, oestrogens play an important role in maintaining normal physiology in tissues from the central nervous, cardiovascular and skeletal system [18-20]. There are three types of oestrogens: 1. Oestrone (E1), 2. 17 β -oestradiol (E2) and 3. Oestriol (E3). Of the three, E2 is the most predominant as well as the most active form of oestrogen [17]. In premenopausal women, E1 and E2 are biosynthesised and secreted by the ovaries and this process is under strict negative feedback regulation through the hypophyseal (anterior pituitary gland) gonadal axis [17, 21]. In postmenopausal women, oestrogen biosynthesis occurs through

aromatisation of circulating adrenal androgens at extragonadal sites inducing the adipose tissue, the cardiovascular and central nervous system, and bone [17].

There are two types of oestrogen receptor (ER α and ER β) which are both members of the nuclear receptor superfamily pertaining to the ligand inducible transcription factors (TFs). The two forms of the oestrogen receptor are encoded from different genes and are cell-type-specific. With regard to the physiological mammary gland, ER α can be detected in the ductal epithelia whilst ER β is expressed in epithelial, stromal and endothelial cells, and fibroblasts [22]. Mammary gland development is dependent on the action of oestrogens, which is demonstrated by ER α and aromatase knockout mouse models; impaired mammary duct formation as well as breast tissue development seen beyond the pre-pubertal stage [23, 24].

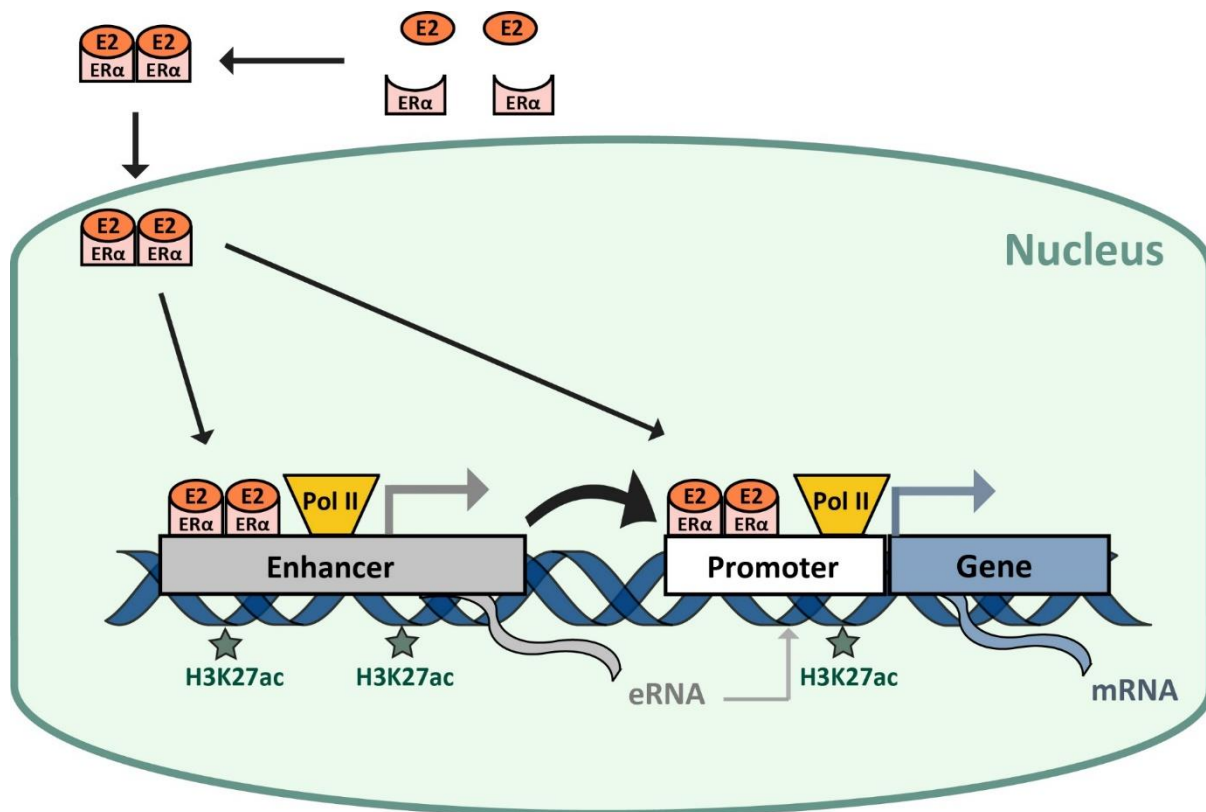


Figure 1.5 Overview of ER- α function. Upon binding with its ligand oestradiol (E2), and dimerising, ER- α binds to its respective oestrogen responsive element (ERE) at promoters and enhancers, along with other cofactors, to result in the recruitment of RNA polymerase II (Pol II). The sites of active transcription are flanked by H3K27 acetylation histone modifications and once RNA polymerase II is recruited transcription of ER target genes occurs at promoters and enhancers leading to the formation of messenger RNA (mRNA) and enhancer RNA (eRNA), respectively. [Figure created in collaboration with Nikolaos Trasanidis]

However, knockout ER β mice exhibit physiological mammary duct and breast tissue formation with normal lactation. This data strongly suggests that ER α plays a fundamental role in regulating oestrogen functions in physiological development including pubertal characteristics [25]. On assessment of mammary tissue, ER α -positive cells can be visualised throughout the luminal epithelium. The latter ER α -positive cells do not proliferate in normal breast tissue. However, the actively dividing cells, which are often ER α -negative, can be seen in close proximity to the ER α -

positive cells. It has been postulated that the ER α -positive cells, in the luminal breast epithelium, act as sensor cells for oestrogen and subsequently secrete paracrine hormones/factors which stimulate the growth of the adjacent cells [26].

1.5 The oestrogen receptor in breast cancer and molecular subtypes

ER α expression is increased in breast epithelial hyperplasia and breast tumours compared to normal breast parenchyma. Furthermore, in BC, a high number of the ER α -positive cells are actively dividing. There is an aberrant proliferation of the ER α expressing cells which would usually not undergo cell division in normal breast tissue. Collectively these findings provide pathological information about the development of BC [27]. Although the ER α is an established key driver in the development and progression of BC, the ER β has been shown to antagonise the actions of ER α [28]. When compared to ER α , ER β expression in breast epithelial hyperplasia and in breast neoplasms is significantly lower than in normal breast tissue. There is a negative correlation between ER β expression and proliferation markers in BC which would imply that ER β has a protective effect against the proliferative actions of oestrogen as well as tumour initiation [29].

Although pathological characterisations of BC have been long established, BCs can also be categorised into various molecular subtypes, which include 1. Luminal A; oestrogen receptor (ER) positive (+), progesterone receptor (PR) positive and human epidermal growth factor 2 receptor (HER2) negative (-), 2. Luminal B; ER+, PR lowly expressed or absent, HER2 occasionally overexpressed with the tumour exhibiting high proliferation rates, 3. HER2+; ER-, PR- and HER2 overexpressed, 4. Basal-like;

this includes the triple negative subtype ER- PR- HER2- and 5. Claudin-low that are also triple negative but resemble cells undergoing epithelial-mesenchymal transition (EMT) [30]. The latter molecular subtypes of BC all exhibit variable responses to treatment and also differ in incidence and survival. Luminal A and luminal B (i.e. ER positive BC) molecular subtypes represent over two thirds of all BCs and are amenable to endocrine therapy (ET) which aims to target the ER [30].

1.6 The treatment of breast cancer

Achieving long-term local disease control coupled with minimal local morbidity, are the goals of local treatment of BC [31]. Breast conserving surgery (BCS) or mastectomy, with or without axillary node clearance, are the usual surgical modalities employed in BC treatment. Surgery aims to completely excise the BC tumour, aiming for at least microscopically disease-free margins [32, 33]. In addition, the resection margin should be clear of tumour tissue (i.e. presence of normal tissue) by $\geq 2\text{mm}$ at all radial margins of the carcinoma [34]. A 2mm margin decreases the risk of ipsilateral breast tumour recurrence when compared to smaller negative margins [34]. Studies have demonstrated that wider excision margins do not decrease the rate of local BC recurrence. Incomplete excision, however, defined by presence of tumour tissue within resection margins, results in unacceptable rates of local control [32, 33]. The application of radiotherapy (DXT) following surgery in BCS remains a fundamental adjuvant treatment modality. A study by the Early Breast Cancer Trialists' Group (EBCTCG), showed that DXT post BCS reduces the 10-year risk of any locoregional or distant first recurrence as well as decreasing the patient death rate by one sixth [35]. In addition, post-operative DXT is associated with a

reduction in ipsilateral recurrence of ductal carcinoma in situ (DCIS) and the development of invasive BC [36-42].

Since Beatson's observation that, over a century ago, the induction of regression in advanced BC could be propagated by bilateral oophorectomy [43], ET has shown to be one of the most fundamental treatment modalities in cancer medicine [31]. Furthermore, it has been shown that over half of females with operable BC, who receive surgery and DXT only, will die from metastatic disease. This indicates the presence of micro-metastatic disease at the onset of clinical diagnosis [31].

Endocrine therapy forms part of the key treatment modalities in the management of ER α -positive BC. ET can be administered in the pre-operative (neoadjuvant), post-operative (adjuvant) and in the advanced (metastatic) disease setting. The aims of current ET are to control or inhibit peripheral oestrogen production or the function of ER α within the BC cells [44].

Patients diagnosed with ER α -positive BC in the pre-menopausal setting are managed with postoperative Tamoxifen (a selective oestrogen receptor modulator (SERM)) for 5 years. However combined results from randomised control trials show a statistically significant reduction in BC mortality, and improvement in overall survival (O.S), with a 10-year treatment with Tamoxifen [45, 46]. Ovarian suppression, which can be achieved surgically or through the use of gonadotrophin releasing hormone (GnRH) analogues, in addition to Tamoxifen, chemotherapy or both, is associated with reduced disease recurrence and death after recurrence in premenopausal patients with ER α -positive BC [47]. The use of aromatase inhibitors (AIs), which inhibit the peripheral production of oestradiol, in combination with ovarian suppression do not appear to add additional benefit when compared to

patients taking Tamoxifen on the background of ovarian suppression treatment [44, 48].

In the postmenopausal setting, patients with ER α -positive BC who have undergone surgery are usually started on either adjuvant Tamoxifen or aromatase inhibitors (anastrozole, letrozole). AIs have shown to be associated with an increased disease-free survival compared to Tamoxifen alone [49-51]. The combination of Tamoxifen or AI (i.e. Tamoxifen for 2-3 years followed by AI to complete a total of 5 years of treatment) or monotherapy (i.e. Tamoxifen or AI alone) is administered for 5 years [44]. More recently, the use of inhibitors to the cyclin-dependent kinase-retinoblastoma (CDK-Rb) pathway, have shown promising results. Specific CDK4/6 inhibitors such as palbociclib, ribociclib and abemaciclib have been shown to block cell cycle progression [52]. Randomised phase 2 and a phase 3 trial of palbociclib with the addition of ET compared to ET alone, have demonstrated a substantial increase in progression free survival in the CDK4/6 plus ET group in patients who have hormone receptor-positive and HER2-negative (HER2-) BC. Several phase 3 trials are still ongoing regarding the use of CDK4/6 inhibitors in hormone receptor positive and HER2- BC [52].

Neoadjuvant ET and/or chemotherapy are commonly used for locally advanced BC, inflammatory BC and downstaging of large tumours for BCS. It is also commonly used for earlier, operable BC patients. BCS improves the patient's psychosocial wellbeing, provides a better cosmetic result and shares outcome equivalence when compared to mastectomy [53-57]. In addition, neoadjuvant therapy is also used to assess treatment response in patients with early stage, aggressive BC phenotypes. Although Tamoxifen or AI can be administered in the neoadjuvant setting for ER α -

positive BC, chemotherapy remains the first-line therapy to down-stage BC tumours irrespective of ER positivity [44].

Despite treatment with ET, 30-40% of patients presenting with ER-positive primary disease relapse within 10-15 years of treatment owing to endocrine resistance [58-60].

Despite advances in adjuvant therapy for BC, metastatic BC remains a major challenge. In some cases, the hormone receptor status between the primary and metastatic disease can be completely different and hence undertaking a clinical biopsy of the recurrence would aid the physician in tailoring more focal treatment [44]. Depending on the metastatic BC tumour hormone status and the patient's wellbeing, proposed treatments include, ET with Tamoxifen, third generation AIs (which have shown to be more beneficial than Tamoxifen in recurrent BC), fulvestrant (a 7 α -alkylsulphonyl analogue of 17 β -oestradiol) which is a selective ER downregulator (SERD) and trastuzumab (a monoclonal antibody to the HER2); usually administered to HER2-positive as well as hormone receptor positive metastatic BC patients.

1.7 Tumour heterogeneity

Through the assessment of cell morphology, tumour heterogeneity has been documented for many years [61-63]. It is now widely accepted that all types of carcinomas contain a high number of different populations of cells [62]. These cells exhibit inter-tumour heterogeneity (ITH) at a genetic, epigenetic and phenotypic level. Studies using next-generation sequencing and micro-array technologies have provided insights in the complexity surrounding ITH [64]. Furthermore, ITH has been

deemed to be a major cause of targeted therapy failure as well as disease resistance [64]. Tumour heterogeneity can be divided into genetic and non-genetic causes [62].

Cancer formation and development are dependent on the acquisition of multiple driver mutations that are able to trigger pro-oncogenic pathways, which can then subsequently disable tumour suppressors [65]. Compared to experimental tumours and cell lines which can be induced by a set of defined genetic events, spontaneous tumours tend to form through Darwinian-like somatic evolution [62, 66]. During this latter process, key driver mutations are acquired to provide the tumour with a survival advantage [67]. Due to the fact that tumours change over time, which are secondary to changes in the mutational landscape, classifying mutations into driver and passenger types can be confusing (i.e. what is considered as a passenger mutation could essentially become a driver mutation over time or the value of the passenger-mutation changes over time) [62]. Although genetic heterogeneity is an unlikely major cause for phenotypic heterogeneity, it substantiates the process through which tumour cells acquire heritable differences and therefore provides a conceptual framework to explain tumour evolution in the context of tumour growth and resistance to targeted therapy [62].

The causes of non-genetic tumour heterogeneity can be divided into 1. Cancer stem cell (CSC) theory, 2. The effects of the heterogeneous tumour environments and 3. Changes in the epigenetic landscape and expression noise [62].

The rise in CSC research and subsequent findings have given the CSC perspective as one of the dominant factors in non-genetic sources of phenotypic heterogeneity [68]. Therefore, phenotypic heterogeneity in carcinomas potentially occurs as a mirror effect from the differentiation hierarchies that exist in normal tissue. In

addition, phenotypic heterogeneity that is regulated by differentiation states has been associated to pivotal outcome parameters such as clinical prognosis, treatment-resistance and the potential for metastatic spread. It is presumed that in tumours, there are a higher number of cells carrying stem cell markers which are associated with poor clinical outcome. Therefore, it is these cells which are more resistant to therapy and possess metastatic potential compared to the non-CSC containing cancer cells [62]. However, applying the CSC theory to cancer cell phenotypes based on differentiation states alone may not be feasible due to the fact that the altered differentiation states exist in cancer [62].

Changes in the tissue microenvironment, including alterations in its components and properties of stromal cells as well as remodelling of the extra-cellular matrix (ECM), plays an important role in tumour progression [69-71]. It has been proposed that tumour microenvironment changes occur as participants in tumourigenesis and are not just passive consequences of genetic evolution [70, 72]. Tumour microenvironments can subsequently mould tumour cell phenotypes, with mathematical modelling demonstrating that heterogeneity within the tumour microenvironment can possibly result in the selection of more aggressive/invasive phenotypes [73, 74].

In addition, normal tissues and organs consist of highly organised structural units. These are not seen in the case of tumour tissue, suggesting a loss of anatomical organisation. Therefore, in comparison to normal cells, which are subjected to a limited number of microenvironmental changes, tumour cells demonstrate loss of anatomical organisation in tandem with many changes in the microenvironment [62]. For example, neovascularisation, which is required for tumour growth, leads to poorly organised vasculature surrounding the tumour leading to normoxic and

hypoxic regions [75]. The inter-tumour differences seen within the stroma have been associated with changes in the biology of the tumour as well as clinical outcome [70, 76, 77]. However, changes in the microenvironment are not only seen between tumours or between matched primary and metastatic cancers. This change in the microenvironment has also been noted at an intra-tumour level and affects the phenotypic heterogeneity of tumour cells which serves as an obstacle to cancer treatment [62].

The valleys within the epigenetic landscape, a paradigm introduced by Waddington, of cells represent stable/semi-stable cellular phenotypes [78]. In addition, the valleys can correspond to the gene-regulatory network (GRN) of higher probability. Applying this model, there are two ways in which cellular phenotypes can change; 1. Noise-driven transitioning between particular attractor states or, 2. Changing the topology of the landscape [62]. Hence, fluctuations in cellular processes, including changes in the gene expression profile, can cause shifts in cellular position with regard to their epigenetic landscape. If these alterations surpass a particular threshold, the cell can shift to another attractor state which, in turn, leads to transitioning of one phenotypic state to the next. To apply this to Waddington's model, the cell would have to "jump" from one valley to the next to change phenotype [62]. The epigenetic landscape is built by the culmination of intrinsic factors, including the genetic makeup and pre-existing differentiation state, and extrinsic factors, which are the signals that cells receive from their microenvironment. Both of these aforementioned factors, which remain stable within normal tissues, can be disrupted/modified in tumours [62]. The GRN of a tumour can be changed through the genotypic differences present at the individual-cell level within that tumour. This translates into corresponding alterations in the topology of the epigenetic landscape, even in a homogenous environment. In

addition, the inactivation of non-essential genes can result in the change of the GRN architecture [62, 79]. This implies that the multitude of accumulating passenger mutations accrued during tumour development might also participate in forming the tumour landscape topology and contribute to noise-driven transition potential [62]. Furthermore, with regard to changes in the micro-environment (mentioned above), this can result in the creation of distinct landscape topologies that would not be accessible for cells pertaining to normal tissue [62]. The abnormal gene expression profiles exhibited by carcinomas are not just limited to transcriptional alterations. Saturated heat shock protein responses in tumour cells, (secondary to cellular stresses, allelic imbalance and oncogene expression) can lead to an increase in the molecular noise at the level of protein expression which, in turn, can result in change in phenotypic manifestation of silent mutations [80]. For example, it has been demonstrated that HRAS-driven transformation leads to elevated spontaneous dedifferentiation in the human mammary epithelial cell, suggesting that following oncogenic transformation, noise-driven phenotypic transitions might be improved [81].

Amalgamating all factors which potentially contribute to tumour heterogeneity, it is clearly evident why this phenomenon poses as a challenge in clinical diagnostics and subsequent therapy. Firstly, ITH exists amongst patients. Secondly, intra-tumour heterogeneity exists within the same tumour. Therefore, conducting a clinical biopsy of a presumed tumour may only capture one dominant clone, amongst several, as opposed to the whole picture of mixed clonal dominance within the same tumour (Figure 1.6).

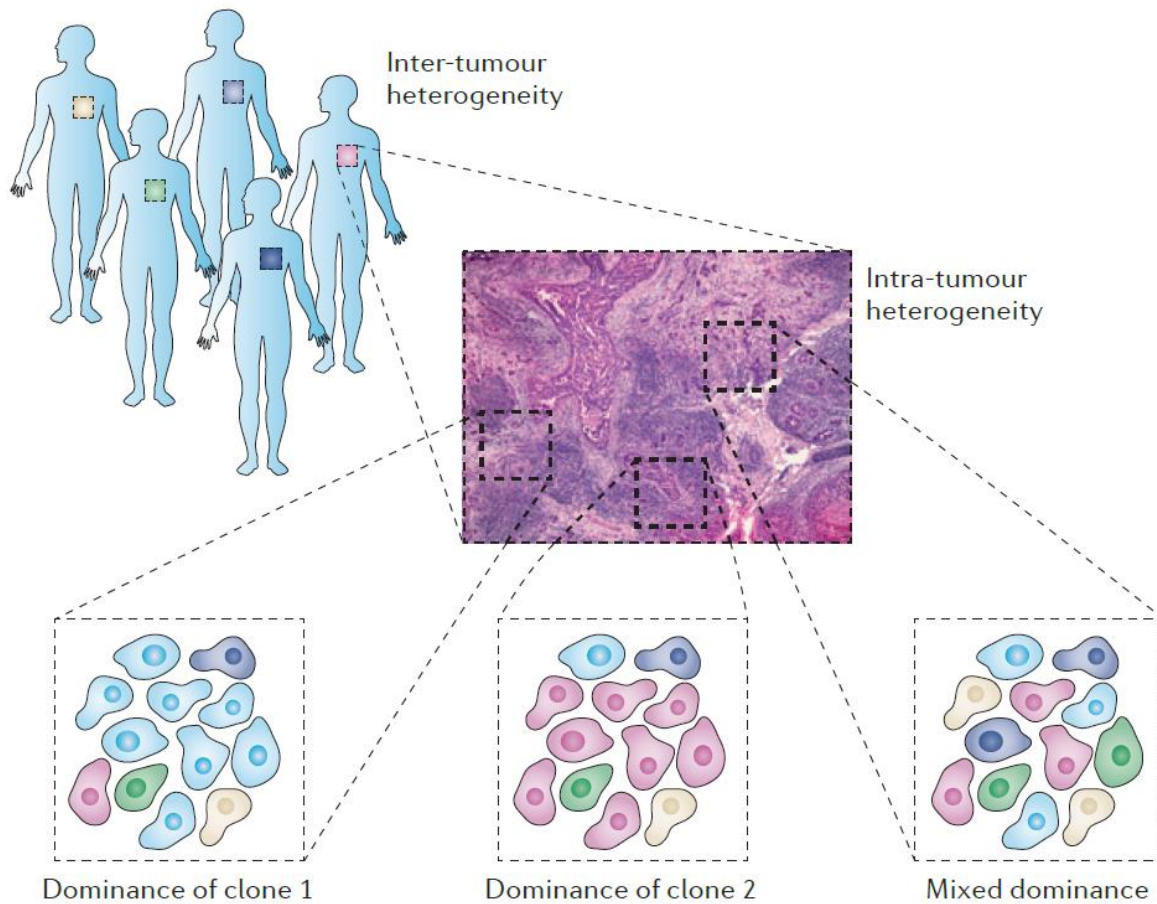


Figure 1.6 Tumour heterogeneity in diagnostics. ITH resulting from genetic and non-genetic factors (also seen in inter-tumour heterogeneity) inevitably complicates definitive clinical diagnostics which subsequently acts as an obstacle in therapy planning. Spatial phenotypic heterogeneity (i.e. only capturing one of many dominant clones within the biopsy specimen) whereby the biopsy specimen does not provide a true reflection of the cellular diversity present within the tumour. Furthermore, decisions made based on scoring the dominant phenotype within a biopsy sample may mislead the diagnostician especially if the cells do not account for minor subpopulations which possess clinically significant information which could alter clinical outcome if taken into account. [Modified from Marusyk *et al.*] ^[62]

1.8 DNA Sequencing in breast cancer

A recent study was conducted by Nik-Zainal *et al.* [82], looking at the landscape of somatic mutations in 560 (i.e. 366 (65%) ER+ and 194 (35%) ER-) breast cancers with matched normal tissue DNA (e.g., peripheral blood lymphocytes, adjacent normal breast tissue or skin) undergoing whole-genome sequencing at a coverage of 40.4x and 30.2x, respectively. Probable driver mutations were found in 93 protein-coding cancer genes [82]. 5 of the 93 cancer genes, where evidence implicating them in BC previously unknown or equivocal, were found. These include *MED23*, *FOXP1*, *MLLT4*, *XBP1*, *ZFP36L1*. Non-coding somatic mutations were fewer in number but exhibited high mutation frequencies with most having distinctive structural features. The authors did not identify driver mutations in the non-coding regions albeit evidence demonstrates that most of the genome is accounted by non-coding regulatory regions [82].

Chromatin immunoprecipitation followed by high throughput sequencing (ChIP-Seq) has been used to map genome-wide ER-binding events in primary breast cancers [83, 84]. ER binding has been found to be heterogeneous and used to retrospectively classify ER positive BC [85]. ER regulatory regions associated with poor clinical outcome observed in primary tumours reveal gene signatures that predict clinical outcome in ER-positive disease exclusively. The findings of this study demonstrated that oestrogen responsive element (ERE) motifs were enriched for good outcome tumours whilst EREs and forkhead box A-1 (FOXA-1) motifs were both enriched in the poor outcome tumour groups. *In vitro* experiments revealed differences in ER-binding events in Tamoxifen resistant cell-lines (TAM-R; MCF7 derivatives and BT-474) compared to Tamoxifen responsive cell-lines (MCF7, T47-D and ZR75-1) [85]. Furthermore, it is known that ER α , upon activation, also binds to

distant regulatory elements, and this process is facilitated by 'pioneer factors' that include FOXA-1, pre-B cell leukaemia transcription factor 1 (PBX1) and the GATA family [84, 86, 87]. Pioneer factors facilitate binding of transcription factors to DNA, either in *cis* or *trans*, by relaxing the chromatin and keeping DNA accessible [87, 88].

1.9 Epigenetic modifications

Epigenetics can be defined as heritable states of gene expression that are not attributed to the DNA sequence. Epigenetic factors such as DNA methylation, various types of histone marks and noncoding RNAs, are responsible for regulating the structural organisation of DNA-histone complexes (also known as chromatin) [89]. DNA (approximately 147 base pairs) wrapped around an octamer of histone proteins, where two of four core histones (H2A, H2B, H3 and H4) are present, is termed a nucleosome [89]. Histone proteins possess a N-terminal tail which protrudes out through the nucleosome structure allowing them to be subjected to post-translational modifications such as phosphorylation, ubiquitination, isomerization, deamination, ADP ribosylation, acetylation and methylation [90-92]. This ultimately leads to a multitude of possible combinations of post-translational modifications on a single nucleosome. Nucleosomes which are trimethylated on lysine 4 and 27 of histone H3 (H3K4me3 and H3K27me3) are associated with transcriptionally inactive promoters whereas the acetylation of lysine 27 (H3K27ac) along with lysine 4 monomethylation (H3K4me1) and Lysine 4 dimethylation (H3K4me2), on histone H3, can be located at transcriptionally active promoters [93, 94]. Furthermore, the histone modification, H3K27ac, has been shown to mark

nucleosome-depleted regions highlighting transcriptionally active promoter and distal regulatory elements (e.g., enhancers) [95-97].

During the process of transcriptional activation, pictured in Figure 1.7, histone marks H3K4me1/H3K4me2 and H3K27ac are placed on nucleosomes; this, in addition to the action of pioneer factors (not shown in the diagram), causes increased chromatin accessibility and thus allows ER α to bind to its target gene and subsequent activation of RNA polymerase II [87]. Transcriptional repression is represented by compacted chromatin, absence of pioneer factors along with transcriptionally inactive histone mark representatives and DNA hypermethylation of promoter regions where a cytosine is preceded by a guanine nucleotide (CpG islands) [87].

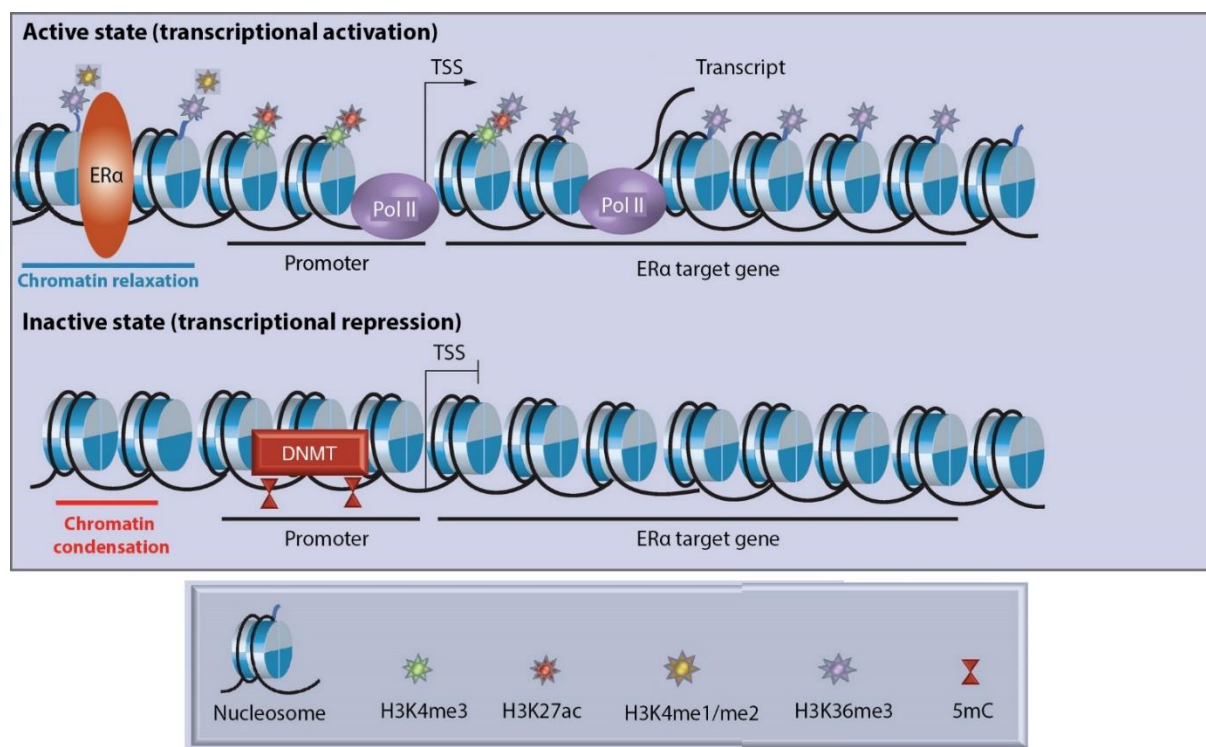


Figure 1.7 Epigenetic regulation of ER α -target genes in breast cancer cells. Schematic illustration of epigenetic modifications and chromatin states demarcating expression and repression of ER α -target genes in breast cancer cells. (ER α : Oestrogen receptor α , DNMT: DNA methyl-transferase; 5mC: 5-methyl cytosine; TSS: Transcription start site; Pol II: RNA polymerase 2) [Modified from Magnani *et al.*] [87]

1.10 Enhancer and super-enhancers

Enhancers belong to the group of *cis*-regulatory modules (regions of non-coding DNA which regulate the transcription of genes) and are classified as DNA elements as they activate the transcription of a gene from a distance, independent of their orientation on DNA [98, 99]. Enhancers consist of multiple motifs for binding of multiple transcription factors which recruit co-activators and co-repressors such that the resultant effect of the enhancer activity is determined by the combined regulatory cues of all bound factors [100, 101]. Figure 1.8 shows that enhancers, bound by transcription factors, may act distally or proximally to their target genes. The position of putative enhancers can be inferred through deoxyribonuclease I (DNase I) hypersensitivity regions (DHS) and also with specific post-translational modifications found on nucleosomes [102, 103]. Enhancers may interact directly or through a looping mechanism with promoters of their respective target genes. These mechanisms of interaction are required for enhancers to maintain cell-type specific gene expression which are fundamental to physiological development [101, 103, 104].

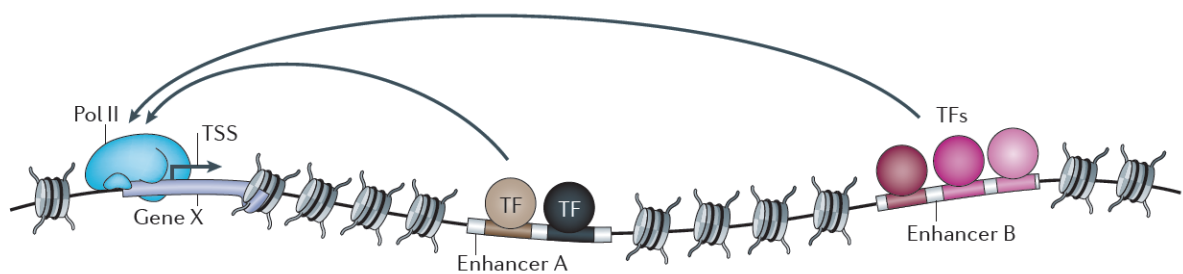


Figure 1.8 Enhancer –gene interaction. Enhancers, contain sequence motifs to which transcription factors (TFs) bind regulating the activity of the enhancer on its target gene. Enhancers may act proximally (Enhancer A) or distally (Enhancer B) from their target gene (Gene X) to regulate the transcription and facilitating the recruitment of RNA polymerase II (Pol II). (TSS; transcription start site) **[Modified from Shlyueva *et al.*]^[101]**

Identification of mammalian enhancers has been assisted by applying modern DNA sequencing techniques such as DNase I hypersensitivity site profiling and ChIP-Seq [105-109]. Studies using ChIP-Seq have shown that certain factors (e.g., p300) and histone marks (e.g., H3K4me1 or H3K27ac) are associated with active enhancer sites [97, 110-114].

Although enhancers exhibit activator binding along with chromatin modifiers, such as histone marks, there are subsets of enhancers which demonstrate a relatively stronger affinity for activator binding and also possess stronger chromatin modification signals. This subset of enhancers has been termed super-enhancers. These represent a small percentage of the entire enhancer population within a given cell type or tissue and have been implicated to be involved in cell type-specific gene expression patterns [98, 115-117]. Whyte *et al.* demonstrated that super-enhancers are found proximal to genes encoding proteins (e.g., Oct4, SOX2 and Nanog) to maintain pluripotency in murine embryonic stem cells (mESCs) [116]. Furthermore, it was also shown that super-enhancers may be defined in any cell type (e.g., T-helper cells, pro-B cells, and myotubes) and not just in stem cells [116]. Hnisz *et al.*, using the signal intensity of the H3K27ac epigenetic mark, identified putative enhancers and super-enhancers in 86 different tissues. Most of the identified super-enhancers were cell type specific and were found to be enriched for sequence motifs corresponding to cell-type specific master transcription factors [96].

Compared to enhancers, super-enhancers 1. display a significantly higher (by orders of magnitude) Med1 signal (Figure 1.9 a) 2. are fewer in number within a given cell type (Figure 1.9 b) 3. cover larger genomic regions (Figure 1.9 b) and 4. exhibit higher enrichment for other epigenetic marks such as Med1, H3K27ac,

H3K4me1 and DNaseI (Figure 1.9 b, bottom table; Total signal and density at constituents) [115-117].

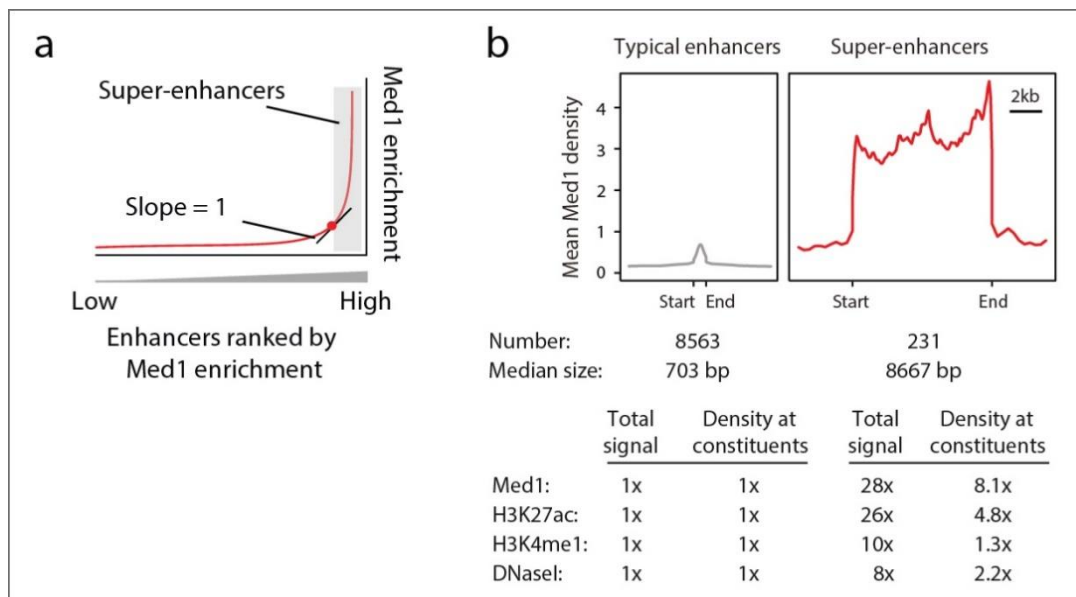


Figure 1.9 Definition and general properties of super-enhancers. (a) Distribution of Mediator 1 (Med1) ChIP-Seq signal across enhancer regions identified in embryonic stem cells. All enhancers (total 8,794) are ranked on the x axis based on the Med1 signal intensity plotted on the y axis. The grey area corresponds to the subset of enhancers displaying exceptionally high Med1 enrichment and defined as super-enhancers. (b) General attributes of super-enhancers in comparison with normal enhancers. The differences rely on (i) their number (231 versus 8563), (ii) their length (8667 bp versus 703 bp), (iii) Med1 occupancy (28-times higher in super-enhancers), (iv) enrichment for other epigenetic marks (e.g. H3K27ac) and (v) constituent densities (bottom table), [Modified from Whyte *et al.*, 2013 and Pott and Lieb, 2015] ^[100, 116]

In comparison to cell-type specific genes, housekeeping genes are regulated by simple promoters which either do not require enhancer activity or contain an integrated enhancer. Cell-type specific genes are controlled by complex regulatory mechanisms and their expression is regulated by a minimal active promoter in combination with multiple distal enhancer elements [118]. The activation of individual enhancers necessitates the binding of multiple transcription factors. In comparison, multiple enhancers tend to function independently of each other so that the activation

of any one enhancer element leads to the expression of the target gene. In the context of super-enhancers, whereby clusters of enhancers are working together, the latter are able to drive relatively higher levels of expression of their corresponding shared target gene or genes (Figure 1.10) compared to enhancers [118].

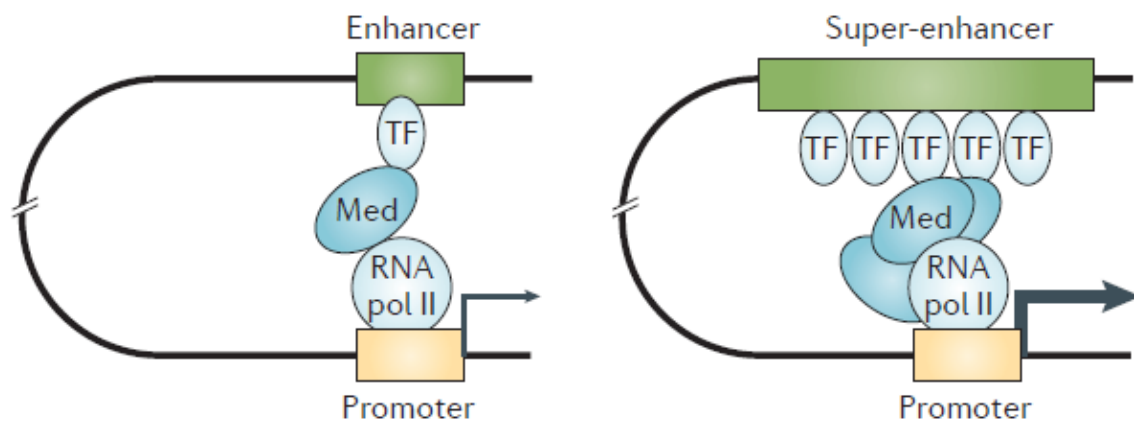


Figure 1.10 Comparison of gene expression coordinated by “common” enhancer gene and super-enhancer gene interaction. Compared to enhancers super-enhancers are, 1. Bound by several transcription factors (TFs) owing to their relatively larger spanning genomic size, 2. Are enriched for the Mediator (Med) protein complex which mediates the interaction between TFs and RNA polymerase II (RNA pol II). Therefore, higher levels of gene expression occur at super-enhancer gene interactions to bring about cell-type specific phenotypes. [Modified from Taipale and Sur, 2016] ^[118]

1.11 Super-enhancers in cancer

Cancer cells differ from their non-neoplastic counterparts due to the fact that they are driven through a modified cell cycle pattern which renders them to proliferate uncontrollably and disseminate. This would imply that super-enhancers would be situated proximal to, and control the expression of, genes with oncogenic potential. To support this idea, studies have shown that, for example, super-enhancers have

been located near the genomic loci governed by *MYC* in a broad range of cancers (Figure 1.11) [116, 117, 119].

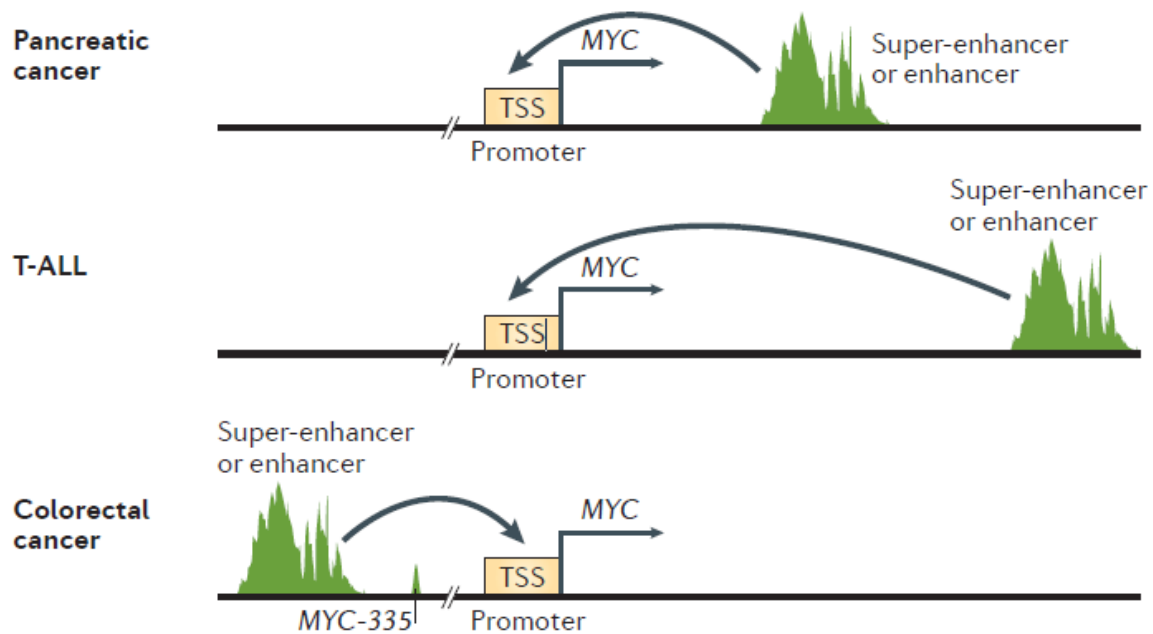


Figure 1.11 Different types of malignancy can utilise different enhancers to target the same gene. Taking the *MYC* gene for example, the super-enhancer formed in pancreatic carcinoma is located near the transcription termination site of *MYC*. In comparison, the super-enhancer activating *MYC* in T cell acute lymphoblastic leukaemia (T-ALL) is situated more than 1 megabase downstream of *MYC*. In the colorectal cancer cell line HCT116, *MYC* makes use of a super-enhancer formed 500 kilobases (kb) upstream its transcription start site (TSS). In addition, the enhancer element MYC-335 which possesses the single nucleotide polymorphism rs6983267 is located 335 kb upstream of the *MYC* TSS. **[Modified from Taipale and Sur, 2016]** ^[118]

Furthermore, in other cancers, such as multiple myeloma, super-enhancers have been shown to have high enrichment for the binding of the co-activator BRD4, a member of the bromodomain and extra terminal domain (BET) family [119-121]. BRD4 has been shown to bind acetylated histones and directly interact with the Mediator complex and elongation factors to promote transcription [121, 122]. Although BRD4 is found in many cell types, the above study provided the explanation as to why cancers cells are specifically sensitive to BRD4 inhibition

[117]. Hnisz *et al.*, in their study of colorectal cancer, revealed that, compared to normal colon tissue, colorectal cancer cells (HCT-116) acquired a super-enhancer at the *c-MYC* locus using the H3K27ac signal [117]. In addition, using H3K27ac ChIP-Seq, this also showed that compared to normal breast epithelium cells, ER-positive MCF7 breast cancer cell lines acquired a super-enhancer at the *ESR1* locus [117]. Several studies have demonstrated that the progression of carcinogenesis involves re-programming of genomic regions with many cancer types developing super-enhancers at oncogenes which warrant further scientific investigation in BC [116, 119, 123-129].

It has been shown that enhancers can be enriched for disease-associated variants which tend to be more prominent in the tissue involved in the pathogenesis of the disease in question. Subsequently, studies have shown that this enrichment for disease-associated variants and single nucleotide polymorphisms (SNPs) has been also identified in the defined super-enhancers in the tissue or cell type where the disease originates (e.g., 1. Alzheimer's disease; super-enhancers have been identified to possess disease associated SNPs within brain tissue 2. Rheumatoid arthritis; disease associated SNPs located within identified super-enhancers in B and T cells [116, 130-132]. Interestingly, a study conducted by Mansour *et al.* investigating the effects of *TAL1* oncogene in T-cell acute lymphoblastic leukaemia (T-ALL) showed that in a subset of T-ALL heterozygous somatic mutations are acquired at non-coding sites. This resulted in the creation and introduction of binding motifs for the MYB transcription factor thus establishing a super-enhancer upstream of the *TAL1* oncogene and driving disease progression. This study was essential in highlighting the fact that oncogenic super-enhancers can also be formed through non-coding somatic mutation using T-ALL as an example [133].

1.12 Study Aims

Non-coding regions of the human genome represent 96% of the total genetic code in each cell and cis-regulatory elements such as enhancers have shown to be controllers of differential gene expression leading to cell-type specificity. Several studies using the epigenetic histone mark H3K27ac for ChIP-Seq have demonstrated that regulatory regions including enhancers and promoters can be mapped using the H3K27ac signal [108, 134]. To further understand phenotypic heterogeneity that exists in BC, the H3K27ac ChIP-Seq data that I have generated, through processing of patient-derived samples, will undergo bioinformatic analysis to:

1. Identify enhancers and promoters within my BC patient-derived sample-set and subsequently build a database representing the regulatory landscape of ER α -positive BC.
2. Identify key drivers within the regulatory landscape database built from the patient-derived samples.
3. Ascertain whether epigenetic annotations can be used to infer phenotypic heterogeneity in primary and metastatic BC.

Collectively, these study objectives aim to firstly, further understand phenotypic heterogeneity and its implication in endocrine therapy resistance in BC and secondly, obtain a more comprehensive picture of dominant/non-dominant clones present within the tumour which can assist in tailoring more focally efficient treatment plans and improve prognosis.

Chapter 2 Materials and Methods

2.1 Breast tumour selection criteria

ER-positive primary BC samples and metastatic samples (were acquired from four sources (1) Imperial College Healthcare NHS Trust Tissue Bank (ICHTB), (2) 'A Study to Measure In Vivo Changes in Oestrogen Receptor DNA Binding Events in Breast Cancer Treated with Endocrine Therapy for Primary or Recurrent Disease' (VERB study) (REC reference: 11/LO/1966), (3) Breast Cancer Now Tissue Bank (Grant number: 2013MaySP013) and Institute of cancer research (ICR) tissue bank. The metastatic samples were selected irrespective of ER status provided that the primary disease was deemed ER-positive. Corresponding clinic-pathological and treatment information were collected.

Imperial College statisticians were consulted to ascertain whether a power calculation could be performed (i.e. indicating how many BC samples would be required for this study). The power calculation could not be calculated in this type of study due to the lack of effect size. Ross-Innes et al., in their study of differential oestrogen receptor binding and clinical outcome in breast cancer, used 18 (7 Good outcome, 8 Poor outcome and 3 Metastatic) breast tumour samples to conduct ER- α ChIP-Seq [85]. Therefore, the aim was to collect up to 60 patient BC samples for this study.

A total of 55 breast (including metastatic) tumour samples were collected and for the purpose of the study, were divided into 2 main groups: 1. Primary breast cancer (N=39) (Inclusion criteria: 1. ER α -positive primary breast cancer) and 2. Metastatic breast cancer group (N=16) (Inclusion criteria: 1. Background of previous primary ER α -positive BC, 2. Adjuvant, with or without neo-adjuvant, therapy for primary

disease was received and 3. ER α -positive or ER α -negative metastatic samples may be selected providing the primary diagnosis was found to be ER α -positive).

Within the first 18 months from the start of the project, 31 tumour samples (Primary breast cancer, N = 19 and Metastatic breast cancer, N = 12) were processed for H3K27ac ChIP-Seq and were categorized as the initial “training set”. The remaining 24 tumour samples, obtained from BCN and ICR 6 months later, underwent H3K27ac ChIP-Seq and the sequencing data combined according to the predefined “primary” and metastatic tissue groups with the 31 “training set” tumour samples with a new analysis strategy; described in section 2.8.

2.2 Cell culture

MCF7 [135] cells lines and variants were obtained from Shaw *et al.* [136] and T47D [137] and ZR75-1 [138] cell lines were obtained from the American Tissue Type Culture Collection (ATCC) and the European Collection of Cell Cultures (ECACC) (Table 2.1).

Table 2.1 Breast cancer cell lines with corresponding hormone receptor status

Cell Line	Tissue	Cell type	ER status	PR status	HER2 status
MCF7	Mammary gland, breast; derived from metastatic pleural effusion	Epithelial	+	+	-
ZR75-1	Mammary gland, breast/duct; derived from metastatic fluid: ascites	Epithelial	+	+	-
T47D	Mammary gland, breast; derived from metastatic pleural effusion	Epithelial	+	+	-

Cell lines were maintained in a Galaxy CO₂ incubator (Eppendorf; #170R) at 37°C in humidified air with 5% supplemented CO₂ and passaged when a 70-85% confluency was reached. Cell culture and related experiments were conducted in a NUAIRE

class II, ducted laminar flow safety cabinet (NUAIRE; #DH Autoflow). Cell lines were routinely maintained at a logarithmic phase of growth in T75 tissue culture flasks (Corning; #430725U) and were tested every 3 to 6 months to rule out mycoplasma infection. Furthermore, the cell lines were genotyped for authenticity on an annual basis by sending genomic DNA to LGC Standards (Cologne, Germany).

Cell culture media was suctioned from the culture flask after which, cells were washed twice with pre-warmed (to 37°C) 0.02% of EDTA solution (Sigma; #E8008). Cells were then trypsinised by adding 3mL of 1X trypsin solution (Sigma), and incubated for 3 to 5 minutes at 37°C to allow the cells to detach. 3mL of culture media (at 37°C) was then added to the cells to quench the action of the trypsin. The cell suspension was collected and re-suspended in fresh medium and transferred to new flasks. Cells were serially passaged at a ratio ranging from 1:3 and 1:6.

MCF7 cells were cultured using Dulbecco's modified Eagle's medium (DMEM) (Sigma) containing 10% fetal calf serum (FCS) (First Link Ltd, U.K.) and 100 U penicillin/0.1 mg mL⁻¹ streptomycin, 2mM L-glutamine (PSG) (Sigma) plus 10⁻⁸ 17- β -oestradiol (Sigma; #E8875). MCF7 fulvestrant resistant (MCF7FulvR) cells were cultured using Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS) and PSG plus 10⁻⁷ fulvestrant (Sigma; #I4409). MCF7 long term oestrogen deprived (LTED) cells were grown in phenol-free DMEM (Invitrogen) with 10% charcoal-stripped FCS (DCFCS) (First Link Ltd, U.K.) and PSG. T47D and T47D-LTED cells were passaged using DMEM containing 10% FCS and PSG and phenol-free DMEM with 10% DCFCS and PSG, respectively. ZR75-1 cells were grown in DMEM containing 10% FCS and PSG.

In general, aqueous ligand stocks were added into the culture media at a dilution of 1 in 1000. An equal volume of molecular grade ethanol or DMSO was added to the “no ligand” (vehicle) controls.

2.3 Tumour processing

Breast tumour samples were collected by the tissue bank technicians immediately following surgery. The excision specimens were taken immediately to the department of Histopathology where the tumours were identified and sections (approximately 1cm x 1cm; this would vary depending on the size of the tumour) would be cut and handed over to the tissue bank team who would snap freeze the tumour sample using dry ice (-78.5°C) and the latter finally stored in -80°C conditions. Tumour samples were then received in dry ice (-78.5°C) and transferred for storage in the laboratory -80°C freezer until required. Upon selection, the tumour sample was placed on a petri dish, over dry ice, and macroscopic adipose tissue was carefully dissected from the breast tumour tissue with a disposable scalpel (Swann-Morton; #0565). This process was performed under aseptic conditions and once complete, the tumour was then cut into small fragments (approximately 2mm x 2mm) and fixed using 1% formaldehyde in “solution A” (Table 2.2) for 20 minutes. Cold glycine (1M) was added (1/10 volume) to the formaldehyde-fixed tissue and incubated at 4°C for 10 minutes. The samples were then centrifuged (Eppendorf; #5418R) at 4°C at a speed of 2000 relative centrifugal force (rcf) for 5 minutes and the supernatant discarded. Once fixed, the tumour fragments were placed into a ceramic mortar which was placed over dry ice. 30-50 mL of liquid nitrogen (-195.79°C) is carefully poured over the tumour fragments and the tissue was homogenised using a pestle until a consistent fine powder was obtained.

Table 2.2 Reagents used for chromatin immunoprecipitation

Reagents	Composition/Concentrations	Storage
Solution A + Formaldehyde	50 mM Hepes-KOH, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 1% formaldehyde	Room temperature
Glycine	1M	4 °C
Lysis Buffer 1 (LB1)	50 mM Hepes-KOH, pH 7.5; 140 mM NaCl; 1mM EDTA; 10% Glycerol; 0.5% NP-40 or Igepal CA-630; 0.25% Triton X-100	4 °C
Lysis Buffer 2 (LB2)	10 mM Tris-HCL, pH8.0; 200 mM NaCl; 1 mM EDTA; 0.5 mM EGTA	4 °C
Lysis Buffer 3 (LB3)	10 mM Tris-HCl, pH 8; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% Na-Deoxycholate; 0.5% N-lauroylsarcosine	4 °C
RIPA buffer	50 mM Hepes-KOH, pH 7.5; 500 mM LiCl; 1 mM EDTA; 1% NP-40 or Igepal CA-630; 0.7% Na-Deoxycholate	Room temperature

2.4 Preparation of cell lines for ChIP

Cell lines were seeded at a density of 4×10^6 cells in a 15cm² dish (Corning; #CLS430597). The cells were maintained for 48 hours to allow a 75-85% confluency to be reached. The cells were either treated or not depending on the experimental design prior to fixation.

10mL of 1% formaldehyde in “Solution A” (Table 2.2) was added to each dish and incubated at 37°C for 10 minutes. Cold glycine (1M) was then added at a ratio of 1/10 the volume of the media in the 15cm² dish and incubated at 4°C for 8 minutes. The media was then removed and the cells were washed three times with cold autoclaved phosphate buffered saline (PBS). The cells were harvested using a cell scraper (Corning; #CLS3008) using 500µL of cold PBS supplemented with 1X EDTA-free protease inhibitor (PI) (Sigma; #S8830) to each dish. The cell suspension was collected in a 14mL and centrifuged at 3000rpm for 3 minutes (Eppendorf; #5810). The fixed cell pellets were either stored at -80°C, until required, or subsequently used for ChIP.

2.5 Chromatin Immunoprecipitation and high throughput next generation sequencing

Once the cells (originating from cell lines or tumour tissue) have been lysed, and the DNA extracted and sonicated (Diagenode; #B01060001), they undergo chromatin immunoprecipitation as per protocol, described by Schmidt *et al.* [139] Immunoprecipitation was carried out using 4ug of antibody per ChIP sample: H3K27ac (Abcam; #ab4729), ERα (Santa Cruz; #sc-543), YY1 (Santa Cruz; #sc-

281). The DNA samples ('input' samples – no antibody treatment and in antibody treated (ChIP) samples) are then quantified using either Picogreen double stranded DNA quantification assay (ThermoFisher Scientific; #P7581) or the Qubit (ThermoFisher Scientific; Qubit 3.0 Fluorometer; #Q33216) high sensitivity assay (ThermoFisher Scientific; #33216). Quantitative polymerase chain reaction (qPCR) was then carried out (Applied Biosystems; #7900HT Real time PCR, #StePOnePlus) to ensure enrichment of 'pulled-down' DNA fragments from the ChIP experiment are present using amplicons which are known to be ER α DNA (Table 2.3) and YY1 targets (Table 2.4). If sufficient enrichment is seen, (following normalisation to the quantified DNA concentrations), in the antibody treatment samples over the 'input' samples, and compared to internal negative controls, these undergo DNA library assembly and multiplexing (NEB; NEBNext Ultra DNA library prep kit for illumina, #E7370L, NEBNext Multiplex Oligos for Illumina, #E7335L) and size selection (SPRIselect magnetic beads, Beckman Coulter; B23318) which aims to retain DNA fragments between 200-300 base pairs (bp), recognisable for the Illumina sequencer (Illumina; #HiSeq2500, #NextSeq500) . qPCR and high sensitivity DNA quantification and size selection measurement (Agilent Bioanalyser 2100 system + High sensitivity DNA measurement assay; #5067-4626) are carried out to, firstly, ensure that the captured DNA selected fragments are between 200-300 bp and, secondly, exhibit enrichment at the same DNA targets which were used for the initial qPCR after the ChIP is performed, respectively.

Briefly, reactions were carried out in 10 μ L volume containing 5 μ L of SYBR Green mix (ABI; 4472918), 0.5 μ L of primer (5 μ M final concentration), 2.5 μ L of genomic DNA and 2 μ L of DNASE/RNASE-free water. A three-step cycle programme and a melting analysis were applied. The cycling steps were as follows: 10s at 95°C, 30s at

60°C and 30s at 72°C, repeated 40 times. Following this, the samples are then sent for sequencing (Imperial College London -MRC Genomics Centre, NIHR Imperial BRC Genomics Facility).

Table 2.3 Primer sequences used for ER- α and H3K27ac ChIP qPCR and revalidation of library prepared samples

Annotation	Targeted region	Forward Primer (5'→3')	Reverse Primer (5'→3')
Hetero-chromatinic region	<i>Rho</i> : <i>chr3</i> :129532090 - 129532110	TGGGTGGTGTCATCTGGTAA	GGATGGAATGGATCAGATGG
Hetero-chromatinic region; non-functioning ERE	<i>chr10</i> :102,411,493-102,414,495	CTTCTTCCTTCCGGCTTTCT	AGCTGGGAGAGGACACACAC
Gene	<i>TFF1</i> : <i>chr21</i> :42808765-42810311	GAATCTGCCCCACTTATCCA	AAGACCTCTGGAAGCACGAA
Enhancer region	<i>EGR3(e)</i> : <i>chr8</i> :22648025-22648951	ACCTCCAAGAGGGAGAGGAG	CTGTCCAGCCGGAGTTAGAG
Validation primer in previous H3K4me2 ChIP	<i>chr2</i> :110829130-10829790	GGTCCAGGGTGAGAGAAACA	CGCTGCTGTCTGGTGTGTAT
Validation primer in previous H3K4me2 ChIP	<i>chr17</i> :37606730-37607272	ATTCAGGCGTCTCTTCTGGA	GCTCGACTAGGGCACTTTTG
Validation primer in previous H3K4me2 ChIP	<i>chr18</i> :55297615-55297883	GCTCCTCATAACAGGCTTCG	GTA CTGGGACAGGGAAACGA

Table 2.4 Primer sequences used for YY1 ChIP QPCR and revalidation of library prepared samples

Annotation	Targeted region	Forward Primer (5'→3')	Reverse Primer (5'→3')
Hetero-chromatinic region	<i>Rho</i> : <i>chr3</i> :129532090 - 129532110	TGGGTGGTGTCATCTGGTAA	GGATGGAATGGATCAGATGG
Hetero-chromatinic region; non-functioning ERE	<i>chr10</i> :102,411,493-102,414,495	CTTCTTCCTTCCGGCTTTCT	AGCTGGGAGAGGACACACAC
Promoter	<i>TLE3</i> : <i>chr15</i> :70094916-70103615	GCGTCCCCGGATCTATATTT	AAGATGCAAAGCGAAC
Promoter	<i>FOXA1</i> : <i>chr14</i> :37593766-37597365	AGGCTAGGGAAAGAGGGAAA	GTTCCGATCCAGAGCCCTA
Validation primer in previous H3K4me2 ChIP	<i>chr17</i> :37606730-37607272	ATTCAGGCGTCTCTTCTGGA	GCTCGACTAGGGCACTTTTG
Validation primer in previous H3K4me2 ChIP	<i>chr18</i> :55297615-55297883	GCTCCTCATAACAGGCTTCG	GTA CTGGGACAGGGAAACGA

2.6 Sequence alignment and peak calling

Raw sequencing files processed by the Illumina HiSeq2500 sequencer were obtained in "FASTQ" format. The raw sequencing files were then aligned to the genome using Bowtie 1.11 short reads sequence aligner using the human reference genome 19 (Hg19) as the reference genome [140].

The output of Bowtie 1.1.1 [141] is the "SAM" file extension format, for both input (control) and ChIP samples, which were then used by Model-based analysis for ChIP-Seq (MACS) version 1.42 for peak calling; with all peaks called at a p-value cut-off of 10^{-5} and default settings applied [142]. MACS outputs results in "BED" file (showing sites of peaks with genomic location and intensity) format and "WIG" files (displays the peaks called, graphically).

2.7 Bioinformatic analysis of “training set” of tumour samples

The integrated genomics viewer (IGV) [143] was utilised to display the peak calling data (BED and WIG files) processed by MACS [142]. In addition, the online bioinformatics tool Galaxy/Cistrome Analysis Pipeline [144] was used to conduct further analysis of the "BED" peak calling files obtained from MACS [142].

Super-enhancer calling was performed using computational software called Ranked Ordering of Super-enhancers (ROSE) as described by Whyte *et al.* and Loven *et al.* [116, 119]. The purpose of this program is to create stitched enhancers and to separate super-enhancers from typical enhancers using sequencing data. The aligned "SAM" files obtained from Bowtie were then converted to "BAM" files (for

control and ChIP samples) which act as input data for ROSE along with a predefined file (".gff" format) of identified constituent enhancers.

Differential analysis of H3K27ac enrichment amongst all the tumour samples was performed using the Bioconductor package "DiffBind" [85, 145]. Epigenomic profiles defined by "BED" files obtained by MACS [142] significance peak calling analysis were comparatively assessed and visualised via a correlation heat map plot and a principal component analysis (PCA) plot.

Multidimensional scaling (MDS) plots were used to visualise the level of similarity between bivalent signatures of the processed tumours (N= 31). MDS plots were used to display the information contained in a distance matrix on a 2-dimensional scale, in this case the Jaccard distance. The distance matrix in this study matched each tumour sample against all other samples defining the Jaccard distance between them. The Jaccard distance was used to measure comparisons between genomic feature data sets. The Jaccard test measures overlaps between two interval sets by calculating the amount of intersection between two interval sets, divided by the length of their union, so two datasets that overlap entirely have a union that is equal to their intersection, hence the ratio is 1. In simpler terms, it is the proportion of ranges or intervals in either sets that overlap with at least one range or interval in the other set. The Jaccard distance is obtained by subtracting the Jaccard coefficient from 1. It then places each tumour in N-dimensional space so that the distance between tumours are preserved as well as possible.

2.8 Bioinformatic analysis of complete set of tumour samples; a novel approach

Processing of raw sequencing data

Raw sequencing reads were quality controlled with FastQC v0.11.5 [146] and aligned to the human hg38 reference genome using bowtie v1.1.2 [141] with the default parameters. The generated sequence alignments were converted into its binary format (BAM), then sorted and indexed using the SAMtools v1.3 utilities [147]. H3K27ac peaks were called with MACS2 v2.1.1 -callpeak with parameters --format AUTO -B --SPMR --call-summits -q 0.01 using input DNA as a control [142].

H3K27ac positive peaks were quality controlled and samples with less than 2000 peaks or more than 200,000 peaks were removed.

Functional characterisation of peaks

The identification of promoter and enhancer peaks was performed using an in-house pipeline based on BEDTOOLS suite v2.25.0 and BASH scripting [148]. A promoter annotation which classifies the promoter as the region 1kb upstream proximal to the transcription-start site (TSS) was generated using UCSC table browser (assembly: hg38; groups: Genes and Gene predictions; track: GENCODE v24) [149].

Peaks were intersected using BEDTOOLS intersect (standard parameters) to identify the promoter specific peaks. Annotated promoters which were not overlapping with the patient signal were considered not-active promoters. A multiple intersection between the promoter peaks was performed using BEDTOOLS multiinter to identify

the core signal. The book-ended regions from the core signal file were merged using BEDTOOLS merge, then intersected against the original peak calls and sorted in order to produce an output which identify the active promoter set [148].

All the peaks which were not overlapping with the promoter annotation table were considered enhancers and the signal was processed as described above with a final identification of the active enhancer set.

Assessment of the level of heterogeneity

Active promoter and enhancer sets were processed in order to reveal the genomic coverage achieved with the available dataset. The saturation analysis was performed with ACT SaturationPlotCreator with the default parameters [150]. The frequency distribution and the average peak size distribution of each regulatory region was calculated intersecting all the patient peaks with the active promoter and enhancer set and then plotted using BASH and R in-house scripts. The size of each peak was obtained from the MACS2 output (_peaks.xls) and the peaks are binned by sharing index [142].

The ranking approach applied in breast cancer tumour sample data

Active enhancers and promoters were associated with the respective share index (SI) and rank index (RI) which typify the number of patients in the dataset sharing the regulatory region (SI) and the average level of enrichment of each regulatory region (RI).

SI was calculated as the number of times a regulatory region overlaps the patient peak dataset with at least 40% of its size. RIs were assigned to each peak call. Duplicate reads from the ChIP-Seq treatment files were filtered out using PICARD v2.1.1 MarkDuplicates REMOVE_DUPLICATES=true and only the uniquely mapped reads were retained and counted for the further analysis [151]. Peak read count was obtained using BedTools Multicov function [148] and this value was normalized with the following equation:

$$Nscore = ((peak\ read\ count / peak\ size) \cdot 10^6) * 10^3 / Sample\ total\ mapped\ reads$$

Peak calls in each sample were categorized as promoter or enhancer, then sorted by their Nscore and assigned to their respective intra-sample category percentile score where 1 is highest enrichment and 100 is the lowest. The peak calls were then intersected to the active promoter and enhancer set and the average RI for each promoter and enhancer was calculated.

Ranking approach in cancer cell lines and normal tissue epigenomes

ChIP-Seq data of H3K27ac profiles across 33 different Encode cell lines and 37 Epigenomic Roadmap tissues with a total of 337 epigenomes, were analysed [108, 134]. Matching “.bam” and “.bed” profiles from Encode and matching raw reads of input and antibody from Epigenomic Roadmap were downloaded. The epigenomic profiles of Encode cell lines from the human hg19 reference genome were lifted to the human hg38 assembly using CrossMap v0.2.3 [152] and peaks from Epigenomic Roadmap were called following the procedure above in 6.2.1. The BC active promoter and enhancer set were intersected with all the epigenomic profiles and the RI calculation of each peak was repeated as described in 6.2.4.

Transcription factor profiling

The profile of the BC cistrome was imputed by taking all the potential accessible regions encoded in the active promoter and enhancer set. H3K27ac ChIP-Seq provides the location of the enriched histones while the transcription factors bind the accessible regions in the nucleosome-free region (NFR). NFRs were putatively characterized by the analysis of DNaseI-hypersensitivity site (DHS) from 220 different Encode cell lines at [108]:

<http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeUwDnase/>

<http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeOpenChromDnase/>

DHS profiles were generated using MACS2 with the following parameters: `--format AUTO --nomodel --shift -100 --extsize 200 -B --SPMR --call-summits -q 0.01` and lifted to the human hg38 assembly [142, 153]. All the DHS peaks were then concatenated into one sorted bed file. NFRs were identified as the regions between 2 sub-peaks at a distance of ± 71 bps from the subpeak summit and the region between 2 broad-peaks distant at the most 500bps. DHS signals overlapping the NFRs were retained for the analysis. The retained DHS sites were sorted and elongated using Bedtools merge to have a unique DHS signal for all the NFRs. Promoter and enhancer specific DHS signals in all BC dataset were used, separately, to carry out the motif enrichment analysis using HOMER [154] function `findMotifsGenome.pl` with parameters: `-size given --preparse`. And thus, the promoter and enhancer specific TFs were determined. Furthermore, the highest 50 ranked TFs in the 2 groups were selected and graphed in polar histograms with a custom R script.

Promoters and enhancers were then binned according to SI after which, the NFRs identified above were overlapped and subsequent motif enrichment analysis was run (as described above) for each promoter and enhancer bin. The motif enrichment results were filtered for statistical significance ($q\text{-value} < 0.05$) and integrated with the observed/expected ratio (OEr) of each putative TF. Thus, a matrix containing the OEr of each TF within each bin was built with a custom R script. Finally, 2 heatmaps (promoter and enhancer datasets) of TF OEr amongst the bins was obtained by using `heatmap.2` from the `gplot` library [155].

2.9 Small interference RNA (siRNA) transfections

Small interfering RNA (siRNA) against SLC9A3R1 (Gene ID; 9368: Ambion; s17919, s17920), YY1 (Gene ID; 7528: Ambion; s14958, s14959, s14960) and Silencer negative control (Ambion; AM4611) were used for transfection experiments. The siRNAs were re-suspended in nuclease-free water at a concentration of 50 μ M, equivalent to 50pmol/ μ L. 1.5×10^5 cells were seeded, per well, using a 6-well plate (Costar; #3516).

MCF7 cells were seeded in phenol-free DMEM with 10% DCFCS and PSG in a 6-well plate; at a density of 3×10^3 cells. Following 24 hours, the cells were then transfected with siRNA using Lipofectamine 3000 (Invitrogen; L3000015). Using the manufacturer protocol for a 6-well plate transfection, per well, 7.5 μ L of Lipofectamine 3000 was added to 125 μ L of Opti-MEM in one eppendorf whilst the 50 μ M stock siRNA (5nM to 20nM final concentration per well) was added to 125 μ L of Opti-MEM. After five minutes, contents of both eppendorfs were mixed thoroughly and incubated for 15 minutes. The complex was then added to the wells containing phenol-free DMEM with 10% DCFCS and PSG. Compared to Lipofectamine 2000 (Invitrogen; #11668019) protocol the “siRNA-Lipofectamine 3000-Opti-MEM” can be added to culture media with PSG without a decrease in transfection efficiency. T47D and ZR75-1 cells were seeded in DMEM containing 10% FCS and PSG. Following 24 hours, the cells were then transfected as previously described above. Cells were harvested for protein/survival/gene expression analysis following at least 48 hours of transfection.

2.10 Luciferase gene reporter assay

MCF7 cells were seeded, using phenol-free DMEM with 10% DCFCs and PSG, in 24-well plates at a density of 5×10^4 cells per well and incubated for 24 hours prior to transfection. Transfection of the plasmid DNA and siRNA were carried out using Lipofectamine 3000 (Invitrogen; #11668019). The amount of expression plasmid used was as follows: 100ng of the firefly luciferase oestrogen responsive element (ERE-Luc) reporter (obtained from Professor Simak Ali's Group, Imperial College London, U.K.), 10ng of renilla luciferase control plasmid (pRL-CMV), 10ng of ER α (pSG5-ER) and 280ng of bluescribe DNA (BSM). BSM was added to adjust the total amount of DNA to 400ng per well. siRNA (on-target and off-target) was also added to the mix (unless otherwise stated in the figure legends) to reach a final concentration of 25nm per well. The DNA/siRNA mix was diluted with Opti-MEM to obtain a final concentration of 0.02 μ g/ μ L. Lipofectamine 3000 and P3000 reagent (Invitrogen; #11668019) were added in quantities of 1.5 μ L and 2 μ L per well, respectively. The Lipofectamine 3000 was allowed to complex with the DNA/siRNA for 15 minutes at room temperature. 45 μ L of the DNA/siRNA-Lipofectamine 3000 mixture is then added to each well and incubated overnight. The following day, the medium was replaced. To determine the effect of the ligands, these were added in the fresh medium according to the study design. After 24 hours of incubation, the medium was aspirated fully after which, 50 μ L of passive lysis buffer (Promega; #E1941), added per well and left on a shaking-device, at room temperature, for 30 minutes. The firefly and renilla luciferase activity was determined using the DualGlo luciferase assay kit (Promega; # E2920) according to the manufacturer's protocol. In brief, the cell lysates were transferred into 96-well opti-plate (PerkinElmer; #6005290). 50 μ L of luciferase reagent was added to each well. Following incubation

for 10 minutes, firefly luciferase activity was determined by the TopCount NXT microplate scintillation and luminescence counter (PerkinElmer). This was followed by the addition of 50µl of Stop and Glo reagent and incubation for 10 minutes before measurement of renilla luciferase activity. The renilla luciferase activity served as control for transfection efficiency, so the firefly luciferase reading was normalized to the reading for renilla luciferase.

2.11 Cell growth assay

The Sulphorhodamine B (SRB) growth assay was utilised to assay cell growth based on the measurement of cellular protein composition. Cells were seeded, in 96-well plates, at a density of 3×10^3 cells per well. Cells were allowed to attach overnight after which, the first plate (Day 0) was assayed after the cells have become adherent. Prospective plates were assayed sequentially after 3 days, 5 days and 7 days. The cells were fixed by adding 100uL of cold 40% (weight/volume) of trichloroacetic acid (TCA) to each well and incubated at 4°C for at least 60 minutes. On the day of measurement, the plates were washed 5 times in double-distilled water (ddH₂O) after fixation, and 100µL of 0.4% SRB (Sigma) in 1% acetic acid was added to each well and incubated for 30 minutes, and left on a shaking-device, at room temperature. Following SRB staining, the plates were washed 5 times with 1% acetic acid to remove unbound dye and allowed to air-dry overnight. The bound dye was then solubilised by adding 100µL of 10mM Tris (HCl pH 8.0) and placed on a shaker for 30 minutes at room temperature. The absorbance was read at 492nm using a Tecan Infinite M200 plate reader.

2.12 Total RNA extraction

In preparation for RNA extraction, cells were washed twice with PBS and harvested by scraping with 350µL of RLT buffer (QIAGEN), containing 1% (vol/vol) β-Mercaptoethanol (Sigma). Cells were homogenised by centrifugation through a QIASHredder spin column (QIAGEN) at 13,000 rpm for 2 minutes. RNA extraction was carried out using the RNeasy Mini Kit (Qiagen; #74106), according to the manufacturer's protocol. The RNA was then eluted in 30µL of RNase-free water. RNA concentration and purity were determined, using the Nanodrop device (Thermofisher; #ND-3300), by measuring absorbance at 260 and 280 nm.

2.13 Complementary DNA (cDNA) synthesis

cDNA was prepared from total RNA using (thin-walled) PCR tubes and reagents from reverse transcriptase (RT) iScript cDNA synthesis kit (Bio-Rad; #1708891). Reactions were prepared from 1µg of RNA, as shown in Table 2.5. The reaction mix was then incubated in a thermal cycler (Applied Biosystems; Veriti; # 4375786) using the following thermal profile settings: 1. Priming; 5 minutes at 25°C, 2. Reverse transcription; 20 minutes at 46°C, RT inactivation; 1 minute at 95°C and Optional step; 4°C holding temperature. cDNA samples are diluted 1 in 10 with nuclease-free water and subsequently, 2.5µL of diluted cDNA was used for each qPCR reaction.

Table 2.5 Reagents required per reaction to reverse transcribe 1µg of total RNA to cDNA using the iScript kit.

Component	Volume per reaction (µL)
5X iScript Reaction Mix	4
iScript Reverse Transcriptase	1
Nuclease-free water	Variable
RNA template (100fg-1µg)	Variable
Total volume	20

2.14 Protein lysate preparation; Radioimmune

Precipitation Buffer (RIPA) buffer extraction

Cells were washed twice with ice-cold PBS, trypsinised and cell pellet collected and stored in a 1.5mL eppendorf in dry ice. Ice-cold RIPA buffer (50mM Tris- HCl at pH 8.0, with 150mM sodium chloride, 1.0% Igepal CA-630 (NP-40), 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulfate) (Sigma; #R02780), supplemented with 1X protease (Roche; #11697498001) and 1X phosphatase (Sigma; #93482) inhibitor cocktail was added to the cell pellet (50µL per 1.5mL eppendorf). The cell pellet and RIPA were mixed by pipetting up and down, incubated at 4°C for 30 minutes and vortexed every 5 minutes. Cell lysates were then centrifuged at 13,000 rpm for 30 minutes at 4°C. The supernatants were transferred to a new 1.5mL eppendorf tube and the pellets were discarded.

2.15 Protein quantification

Protein concentrations were ascertained using the Pierce BCA Protein Assay Kit (ThermoFisher Scientific; #23227), according to the manufacturer's product instructions. In brief, bovine serum albumin (BSA) standards were generated (1 to 30µg) in a 50µL volume. 5µl protein lysates were diluted 10x in ddH₂O to obtain a 50µL volume. BCA reagents A and B were mixed with a ratio of 50 to 1, and 1mL of mixed BCA reagent were added to each sample and standard. The samples were mixed and incubated at 37°C for 30 minutes. The samples were then transferred into a micro-cuvette and the optical density was recorded at absorbance reading of 562 nm. The protein concentration of the samples was determined by comparing the measured protein values to the BSA standard curve.

With regard to western blotting, 20µg of protein, per sample, were mixed with 4X Bolt sample buffer (Life Technologies; #B0007), 10X Bolt sample reducing agent (Life Technologies; #B0009), ddH₂O and heated at 95°C prior to loading.

2.16 Western blotting

Cell lysates were heated to 95°C for 5 minutes and ice-cooled prior to loading. 20µg of RIPA protein lysate were loaded onto BOLT 4-12% Bis-Tris Plus Gel (Life Technologies; NW04120BOX). The pre-made gel was placed into a mini gel tank (Life Technologies; #A25977) containing 1X Bolt running buffer (Life Technologies). Electrophoresis was carried out at 90V for 35 minutes to allow proteins to adequately run through and also until the bromophenol blue dye reached the bottom of the gels. The gels were transferred onto a Biotrace nitrocellulose membrane (VWR; #PN66485) using a TE-22 transfer unit (Hoefer GE Healthcare) at 100V for 90 minutes. The membrane was incubated in blocking buffer for 45 minutes at room temperature to reduce non-specific binding of primary antibody. The membrane was then incubated with the diluted primary antibody (YY1 antibody: Santa Cruz; #sc-281 (1:500 dilution), SLC9A3R1 antibody: Atlas antibodies; #HPA027247 (1:1000 dilution)) in blocking buffer at 4°C and allowed to shake overnight. After primary antibody incubation, the membrane was washed four times in PBST (5 minutes per wash on a rocking platform) and then incubated for 1 hour with the HRP-GAPDH (Abcam; #ab9385 (1:5000 dilution)) conjugated antibody (for the loading control membrane) which was diluted in blocking buffer. The membranes (including the loading control membrane) were washed three times in PBST. Proteins were then visualised using goat anti-mouse (ThermoFisher Scientific; #31446) or anti-rabbit

(ThermoFisher Scientific; #31462) HRP conjugated secondary antibodies. Amersham ECL start Western Blotting Detection reagent (GE Healthcare Life Sciences; #RPN3243) was used for chemiluminescent imaging using the Fusion solo (Vilber; Germany) imager.

2.17 Quantitative real-time PCR

Real time-qPCR (RT-qPCR) reactions were carried out in 10uL volume containing 5µL of SYBR Green mix (ABI; 4472918), 0.5µL of primer (2.5uM final concentration), 2.5µL of genomic DNA and 2µl of DNASE/RNASE-free water. A three-step cycle programme and a melting analysis were applied. The cycling steps were as follows: 10s at 95°C, 30s at 60°C and 30s at 72°C, repeated 40 times. For the case of assessing gene expression, the expression of the test gene was normalised to the 28S housekeeping gene and relative to the appropriate control using the $2^{-\Delta\Delta CT}$ method. Primer sequences used for gene expression profiling can be viewed in Table 2.6)

Table 2.6 Primer sequences of targets used for RT-qPCR experiments

Targeted region	Forward Primer (5'→3')	Reverse Primer (5'→3')
<i>28s</i>	CGATCCATCATCCGCAATG	AGCCAAGCTCAGCGCAAC
<i>TFF1</i>	CACCATGGAGAACAAGGTGA	TGACACCAGGAAAACCACAA
<i>GREB1</i>	ATCAGCTGCTCGGACTTGCTG	TGAGCTCCGGTCCTGACAGATG
<i>EGR3</i>	CAATCTGTACCCCGAGGAGA	GGAAGGAGCCGGAGTAAGAG
<i>CA12</i>	GTGGTGTCCATTTGGCTTTT	CAAGGTCCTTCCTGGATGTG
<i>XBP1</i>	CTGAGCCCCGAGGAGAAG	TGTTCCAGCTCACTCATTCTG
<i>EGR1</i>	GCCATAGGAGAGGAGGGTTC	GTTGGCCAATAGACCTTCCA
<i>YY1</i>	ATAAAGGCTGCACAAAGATG	TTCTCTCCAGTATGAACCAG
<i>SLC9A3R1</i>	CAGGAGCACCTGAATGGTCC	CAGGGCTTCACGACTGTTCT

Chapter 3 Collaborative work and acquired lab skills

3.1 “DMXL2 drives epithelial to mesenchymal transition in hormonal therapy resistant breast cancer through Notch hyper-activation”

Article citation: DMXL2 drives epithelial to mesenchymal transition in hormonal therapy resistant breast cancer through Notch hyper-activation. Faronato M, Nguyen VT, **Patten DK**, Lombardo Y, Steel JH, Patel N, Woodley L, Shousah S, Pruneri G, Coombes RC, Magnani L. *Oncotarget*. 2015 Sep; 6(26):22467-79 [156].

The Notch signalling pathway has been shown to be dysregulated in the context of ER α -positive BC which poses as a great challenge. The latter's dysregulation allows for further study with regard to therapeutics [157]. The four Notch receptors that have been identified have five corresponding Notch ligands [158]. Although the exact mechanisms remain to be defined, the activity of the Notch receptor and its ligands are thought to be regulated by endocytic trafficking [158]. Literature suggests that dysfunctional endosome maturation could subsequently lead to abnormal Notch signalling through interference with γ -secretase mediated Notch cleavage [159]. Although Notch can be targeted therapeutically with γ -secretase inhibitors and immunotherapy, the side effect profile of these developed drugs remains high [160, 161]. As well as Notch being involved in pathological process such as carcinogenesis, cancer stem cell maintenance, angiogenesis and epithelial to mesenchymal transition, it also plays an imperative physiological role during embryonic development [162-168]. Abnormal Notch signalling has been associated with BC initiation and progression [169]. Although an exact mechanism of endocrine therapy resistance in BC has not been ascertained, it is postulated that the increased

Notch signalling, arising from epigenetic programming, may serve as contributor [170, 171].

DMXL2 has been proposed to modulate the Notch signalling pathway, through regulation of acidification of intracellular compartments through the vacuolar proton pump (V-ATPase) in *Drosophila* and mammals [172, 173]. The V-ATPase pump is involved in vesicular trafficking including endocytic and exocytic traffic [174]. The purpose of this study was to investigate the role of DMXL2 in the context of endocrine therapy resistance in BC using long-term oestrogen deprived (LTED) and native MCF7 breast cancer cell lines and patient derived clinical samples [135, 136, 175-177]. DMXL2 is a novel driver of the Notch pathway which is implicated in endocrine resistant patients with ER α -positive BC. The results of the project demonstrate that DMXL2 is overexpressed in a subset of endocrine resistant BC cell lines where it promotes epithelial to mesenchymal transition through hyper-activation of Notch signalling via V-ATPase dependent acidification. It has also been demonstrated that DMXL2 protein levels are significantly increased in ER α -positive BC patients that progress after endocrine therapy. Furthermore, DMXL2 has been shown to be a transmembrane protein with a potential extra-cellular domain. This study identifies DMXL2 as a novel, functional biomarker for ER α -positive BC [156].

My contribution to this project entailed performing sulforhodamine B (SRB) assays; culturing MCF7, MCF7-Tamoxifen resistant, MCF7-Fulvestrant resistant, LTED, LTED-Tamoxifen resistant and LTED-Fulvestrant resistant cell lines in differing concentrations of the antibiotic Bafilomycin A1 which, inhibits the V-ATPase pump driven by DMXL2. The experiments were performed at least three times and results revealed that a concentration of 125pM of Bafilomycin was required to inhibit the function of the V-ATPase pump but not cell proliferation.

3.2 “The pioneer factor PBX1 is a novel driver of metastatic progression in ER α -positive breast cancer”

Article citation: The pioneer factor PBX1 is a novel driver of metastatic progression in ER α -positive breast cancer. Magnani L, **Patten DK**, Nguyen VT, Hong SP, Steel JH, Patel N, Lombardo Y, Faronato M, Gomes AR, Woodley L, Page K, Guttery D, Primrose L, Fernandez Garcia D, Shaw J, Viola P, Green A, Nolan C, Ellis IO, Rakha EA, Shousha S, Lam EW, Györfy B, Lupien M, Coombes RC. The pioneer factor PBX1 is a novel driver of metastatic progression in ER α -positive breast cancer. *Oncotarget*. 2015 Sep 8; 6(26):21878-91 [178].

PBX1 has been described as novel pioneer factor in ER α -positive BC. PBX1 has been shown to be fundamental in maintaining chromatin accessibility, allowing ligand-activated ER α recruitment to many regulatory elements [178]. In addition, PBX1 has been shown to be required for transcriptional regulation of a subset of oestrogen target genes associated with poor prognosis. The upregulation of PBX1, *in vitro*, has been associated with the development of endocrine resistance. Furthermore, mRNA levels of genes controlled via Notch-PBX1 crosstalk are associated with poor clinical outcome in patients with ER α -positive BC [170].

With regard to oestradiol independent activation and recruitment of ER α , epidermal growth factor (EGF) driven activation of ER α was not yet known [179]. The aims of this study were to investigate the role of PBX1 in mediating EGF signaling in ER α -positive BC. Results revealed that PBX1 regulates a subset of EGF-ER α genes highly expressed in aggressive tumours. Retrospective stratification of luminal patients using PBX1 protein levels in primary cancer further demonstrates that elevated PBX1 protein levels correlate with earlier metastatic progression. PBX1

protein levels are significantly upregulated during metastatic progression in ER α -positive breast cancer patients. It has also been revealed that PBX1 upregulation in aggressive tumours is partly mediated by genomic amplification of the PBX1 locus which clinically corresponds to poor survival. PBX1 amplification can also be identified in tumour-derived circulating free DNA of ER α metastatic patients who also have a shorter relapse-free survival. Finally, the data suggests PBX1 amplification as a functional hallmark of aggressive ER α -positive BCs [179].

My contribution to this project entailed, 1. Culturing MCF7 and T47D breast cancer cell lines and transfecting them with siRNA for PBX1 along with off-target siRNA (control). Proliferation assays were performed using the aforementioned transfected cells, in various conditions (E.g., In the presence of 1. Vehicle, 2. Epidermal growth factor, Oestradiol) to ascertain the effects of silencing PBX1 vs control in the two cell lines and 2. 2000 human tissue micro arrays (TMAs) of normal and primary BC cases underwent immunohistochemistry (IHC) staining for PBX1 protein. I along with a histopathologist undertook the scoring of these TMAs; >90% inter-rater concordance was achieved and 3. I produced a database of clinical corroboration of the cases along with survival data for the patients with BC where PBX1 amplification was assessed.

3.3 Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion

Article citation: Nguyen VT, Barozzi I, Faronato M, Lombardo Y, Steel JH, Patel N, Darbre P, Castellano L, Györfy B, Woodley L, Meira A, **Patten DK**, Vircillo V, Periyasamy M, Ali S, Frige G, Minucci S, Coombes RC, Magnani L. Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion. *Nat Commun.* 2015 Nov 27; 6: 10044 [180].

Cell identity is established during development through epigenetic activation of cell-type specific distal regulatory elements which are regions found at chromatin accessible sites containing motifs for cell-type-specific transcription factors (TFs)[108, 181]. The degree of plasticity which exists in the epigenome allows for cell identity to be epigenetically reprogrammed subsequent to exposure from external stimuli such as aberrant expression of TFs [182]. In the same way, cancer cells can undergo epigenetic reprogramming in response to chronic exposure to therapeutics agents [87, 183, 184]. This is exemplified when ER α -positive cell lines undergo oestrogen deprivation. During this they have been shown to undergo epigenetic reprogramming at distal regulatory elements such as enhancers [87]. Regulatory elements associated with sites of active transcription, and more complex regulatory regions, like super-enhancers (SEs) can be located measuring the signal for histone modification H3- acetylated lysine 27 (H3K27ac) [97, 99, 116, 185, 186]. The purpose of this study was to investigate the nature of drug adaptation to a

diverse range of endocrine therapies in isogenic ER α -positive BC cells (MCF7, MCF7-Tamoxifen resistant, MCF7-Fulvestrant resistant, Long-term oestrogen-deprived MCF7-LTED, MCF7-LTED-Tamoxifen resistant and MCF7-LTED-Fulvestrant resistant, T47D and T47D-LTED) using integrative and orthogonal epigenomic approaches [180].

Results identify individual drug-induced epigenetic states involving large topologically associating domains (TADS) and the activation of super-enhancers. Aromatase inhibitor (AI) resistant BC cells were found to activate endogenous cholesterol biosynthesis (CB) through stable epigenetic activation seen *in vitro* and *in vivo*. CB enables the activation of ER α in AI-resistant cells through the biosynthesis of 27-hydroxycholesterol. Using HMG-CoA reductase inhibitors (“statins”), ER α binding is decreased and cell invasion is prevented. This study also showed that epigenomic-driven stratification can predict resistance to AIs in a subset of ER α -positive patients [180].

My contribution to this project entailed performing ChIP for H3K27ac in primary and metastatic clinical BC samples. I subsequently performed ChIP-qPCR to investigate the level of H3K27ac at distal regulatory elements associated with CB genes between primary and metastatic clinical BC samples. *In vitro* data from H3K27ac ChIP-qPCR demonstrated that a strong increase in H3K27ac, at enhancers associated with CB genes, were seen in the endocrine therapy resistant MCF7 and the T47D-LTED cells. This pattern was reflected in the clinical cohort whereby a significant increase in H3K27ac signal was observed from a metastatic AI-resistant clinical sample compared to a primary, endocrine therapy sensitive primary sample.

3.4 Acquired laboratory techniques

Over the course of the PhD project from October 2013 to March 2017 I have learnt various laboratory techniques to a satisfactory proficiency. These include:

1. “Wet” laboratory techniques
 - a. Tissue culture
 - b. Cell line transfection
 - c. Cell proliferation assay (SRB)
 - d. Luciferase reporter assay
 - e. Total RNA extraction
 - f. cDNA preparation
 - g. DNA extraction from cell lines/primary tumour samples/metastatic tumour core biopsy samples.
 - h. Gel electrophoresis
 - i. Western blotting
 - j. RT-qPCR
 - k. ChIP and library preparation of ChIP DNA samples for sequencing
2. “Dry” laboratory techniques
 - a. Basic/Intermediate Linux/Unix
 - b. FastQ file alignment (Bowtie 1.11/2.0) [141, 187]
 - c. Peak calling (MACS 1.42/2.0) [142]
 - d. Peak file analysis (Peakanalyser) [188]
 - e. Cistrome advanced pipeline [144]
 - f. Use of integrative genomics viewers (IGV/UCSC/ENSEMBLE)[143, 149, 189]
 - g. qPCR primer design (Primer 3 plus) [190]

Chapter 4 Results: Optimisation of chromatin immunoprecipitation and high-throughput sequencing

4.1 Overview of ChIP

In order to understand transcriptional regulation, genome-wide mapping of protein-DNA interactions is essential. Mapping of binding sites for TFs and core transcriptional machinery and other DNA-binding proteins is vital for deciphering the gene regulatory networks that underpin and control several biological processes [191].

The combination of nucleosome positioning and dynamic modification of DNA and histones plays a fundamental role in development and differentiation [110, 192-194]. Chromatin states can influence transcription by either changing the packaging of DNA to permit or prevent access to DNA-binding proteins, or modify the nucleosome surface to facilitate or impinge on the recruitment of effector protein complexes.

Published literature suggests that the crosstalk between chromatin and transcription is a dynamic one, and more complex than once postulated [195]. Hence profiling the epigenomes in multiple cell types, at different stages, may be required for further comprehending and distinguishing between developmental processes and pathological states [196].

ChIP has become a very important tool in the artillery for understanding the dynamic mechanisms that are involved in transcriptional regulation and involves assaying protein-DNA binding *in vitro* and *in vivo* [197].

Laboratory techniques mapping TF occupancy genome-wide by using ChIP were developed over ten years ago [197-201]. ChIP assays involved taking a TF, cofactor,

or other chromatin protein of interest and enriched by immunoprecipitation from cross-linked cells, along with its corresponding DNA [202]. The genomic DNA sites which were enriched were then identified by DNA hybridisation to microarray (ChIP-chip) [197-201] which then progressed to ChIP coupled with next generation sequencing (NGS), commonly known as ChIP-Seq [109, 110, 203]. ChIP-Seq enables the biologist to sequence millions of small DNA fragments in a single sequencing run, which allows for large-scale experiments to be carried out [191]. NGS has its applications in sequencing whole genomes (whole genome sequencing (WGS)), sequencing of mRNA to ascertain gene expression profiles (RNA-seq) [204-206], characterising structural variation [207], profiling DNase I hypersensitivity sites [142], detecting fusion genes from mRNA transcripts [203], and in the discovery of new classes of small RNAs [208]. The goal of ChIP is to completely map the binding site of the target protein with maximal signal-to noise ratio across the genome [202].

4.2 The ChIP protocol revised

The protocol for ChIP-Seq has been detailed by *Schmidt et al.* [139] and can also be found in Chapter 2 (Sections 2.3 to 2.5). To obtain ChIP and genome-wide (“Input”) DNA, including ChIP RT-qPCR validation and library preparation, the process takes 6 days as shown in Figure 4.1. which details the fundamental steps required for processing 6 samples, each containing 1,000,000 cells or more, for ChIP-Seq.

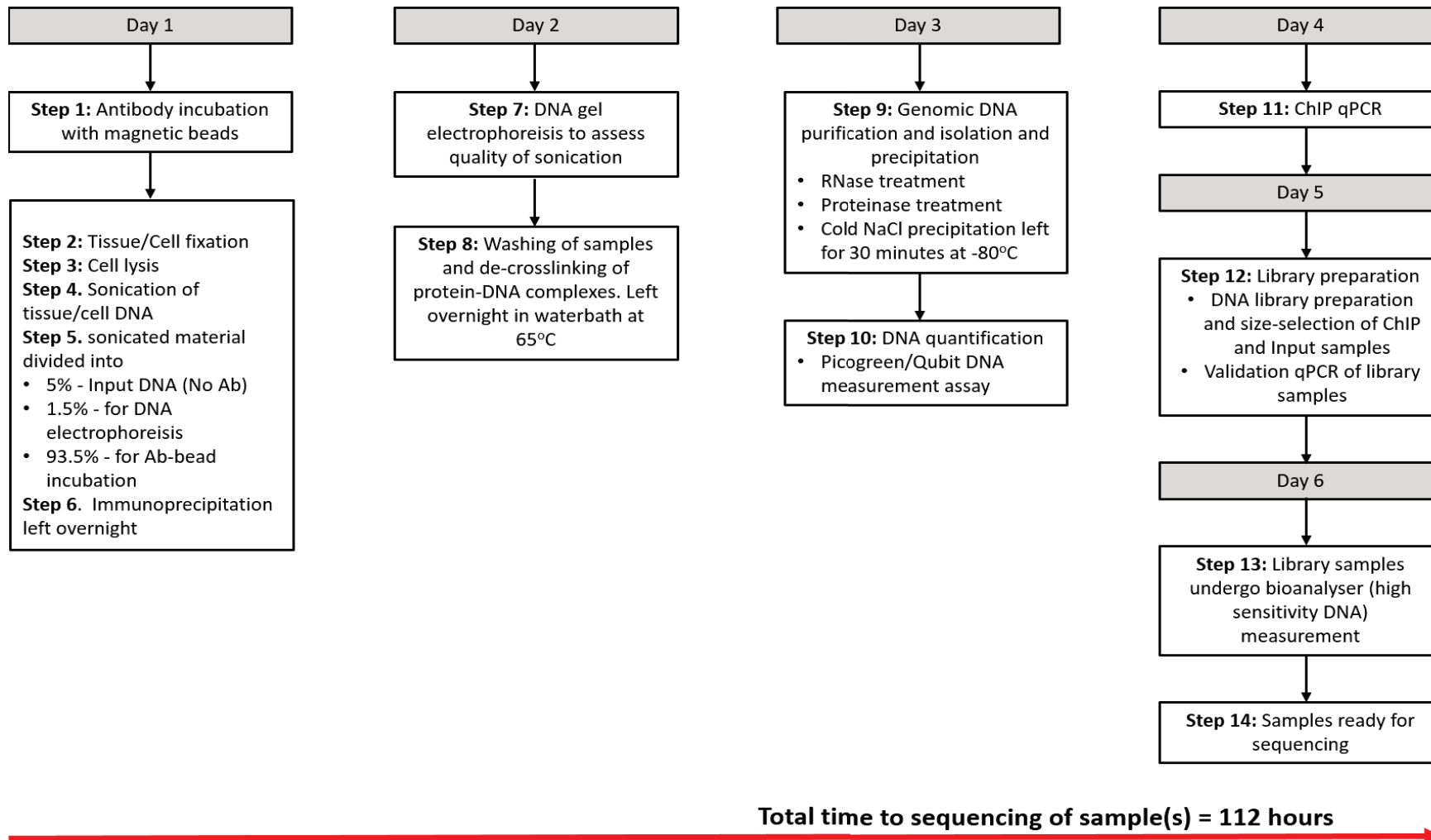


Figure 4.1 The standard ChIP-Seq protocol adapted from Schmidt et al. [139] This figure illustrates the pivotal steps involved in ChIP-Seq and the time taken to process 6 cell lines or tissue samples ($\geq 1,000,000$ cells/sample) through to obtaining sequence-ready DNA (Input DNA: Genome-wide DNA sample, ChIP sample – immunoprecipitated DNA, Ab – antibody)

Although this protocol is established and proven to work efficiently using cell lines or tumour tissue [139], my objective was to ascertain whether the procedure could be shortened without a substantial decrease in quality of the sequencing data.

I began with optimisation of steps from Day 2 of the protocol. I performed DNA extraction of the gel samples followed by gel electrophoresis to assess for sonication efficiency. In addition, de-crosslinking of the samples was initiated at the beginning of the day and left on a thermomixer (Eppendorf; #5384000020) for 6 hours at 65°C (between 800 to 1000 rpm).

After 6 hours, the Input samples were left on ice, whilst the ChIP samples were placed on a magnetic rack (Invitrogen; #CS15000) and the supernatant collected. Input samples usually have an approximate volume of 120uL and the ChIP samples vary between 105 to 110uL/sample. During this process of volume measurement, I ensured that the pipettes used were fully calibrated. Equal volumes of T.E buffer were then added and the samples vortexed for at least 30 seconds followed by microcentrifugation.

Furthermore, I ensured that the samples did not contain bubbles prior to the addition of the SPRI beads (Beckman Coulter; #B23318). This can alter the required volume of SPRI magnetic beads (added in the ratio 1:1) resulting in a decreased efficiency of DNA capture. The SPRI beads were, 1. Vortexed thoroughly to ensure bead homogenisation within solution and 2. Allowed to equilibrate to room temperature. Equal volumes of SPRI beads were then added to each sample and mixed by pipetting up and down at least 20 times to allow the beads to mix with the DNA samples. I observed that, by applying this technique, the Input samples required more mixing with the beads compared to the ChIP samples owing to their higher

DNA concentration. After 10 minutes, the samples were microcentrifuged and placed on a magnetic rack for 10 minutes to allow the SPRI beads to collect. I noticed that the Input samples, when incubated with SPRI beads, tend to stick to the microcentrifuge tubes even when placed on a magnetic rack. This issue is usually solved by gently pipetting up and down whilst the tubes are on the magnetic rack. This will slowly free the stuck beads and allow them to congregate more efficiently.

The supernatants were discarded and the beads washed with freshly prepared 80% ethanol by adding 200uL for 30 seconds. This step was repeated once and the beads were allowed to air dry for 10 minutes and then eluted with 45uL of 10mM Tris (HCl pH 8.0). The beads and Tris mixture were pipetted up and down to ensure efficient mixing. The samples were left to incubate at room temperature for 5 to 10 minutes after which, they were placed onto the magnetic rack for 5 minutes; this provides enough time for the beads to separate and stick to the magnetic rack. 42uL of supernatant for each sample was carefully collected and put into new microcentrifuge tubes.

The DNA samples (Input and ChIP) can then be quantified using either Picogreen (Invitrogen; #P7589) or Qubit high sensitivity (quantitation range: 0.2 – 100 ng) DNA assays (Invitrogen; #Q32853). If using the Qubit DNA assay, the Input sample will require the broad range Qubit assay (quantitation range: 2 – 1, 000 ng) due to the high DNA concentration. Furthermore, in the case of performing a ChIP for TFs, where the ChIP sample DNA concentration may be less than 0.2 ng, using the Picogreen (quantitation range: 0.05 – 2, 000 ng) assay is advised. Once DNA quantifications were performed, the samples underwent ChIP qPCR on Day 2 of the protocol and library preparation was then commenced on Day 3.

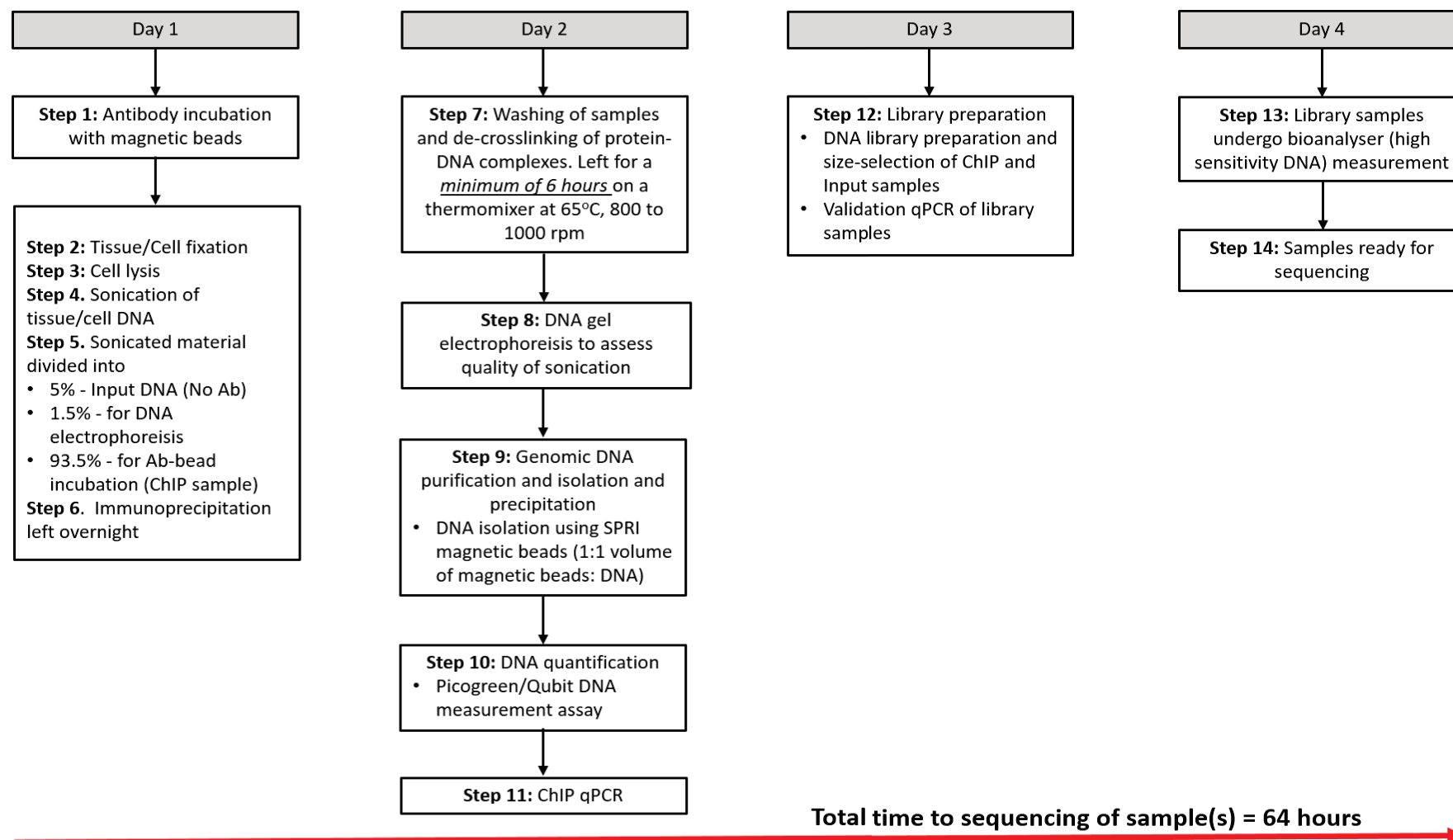


Figure 4.2 The optimised ChIP-Seq protocol modified from Schmidt et al. [139] This figure illustrates the pivotal steps involved in the optimised ChIP-Seq protocol and the time taken to process 6 cell lines or tissue samples ($\geq 1,000,000$ cells/sample) through to obtaining sequence-ready DNA (Input DNA: Genome-wide DNA sample, ChIP sample – immunoprecipitated DNA, Ab – antibody)

4.3 Results of the standard compared to the optimised

ChIP protocol

In order to assess the quality of the results obtained from the optimised ChIP protocol, ChIP experiments were set up in one biological replicate for H3K27ac in murine cancer associated fibroblasts (cell pellets were obtained already fixed with 1% formaldehyde and snap frozen from Fernando Calvo group; ICR, London, U.K.) using the standard and optimised protocols after which, the results were compared.

Another ChIP experiment for TF YY1 was set up with the same parameters. MCF7 seeded at a density of 5×10^6 cells in 15cm² dishes in the presence of phenol-free DMEM with 10% DCFCS and PSG, for 48 hours. Two of the four 15cm² dishes of MCF7 cells were treated with 1/1000 volume molecular grade ethanol and the remaining two 10^{-8} 17- β -oestradiol (Sigma; #E8875) for 45 minutes. The aims and results of this experiment will be explained in more detail in Chapter 6 but in summary, this was to investigate the binding of TF YY1 in ER α -positive BC cell line under the effect of ligand stimulation versus vehicle.

4.4 Sequencing results obtained for YY1 ChIP-Seq in

MCF7 cells

The MCF7 YY1 ChIP was performed in biological triplicates, of which only two replicates were processed for sequencing due to cost implications. The third replicate, having undergone the optimised protocol, also passed ChIP qPCR validation but was not taken to the library stage. With reference to Table 4.1, the first 4 samples underwent ChIP for YY1 under the standard protocol. 8×10^6 cells were

used per ChIP. The yield denotes the sum of all bases, per sample, in reads that passed the filtering step, of which the output is given in million bases (Mbase) per lane. For the standard protocol in the YY1 ChIP experiment, this ranges between 3415 – 5978 Mbases (mean = 4927 Mbases) compared to 3326 – 4305 Mbases (mean = 3436 Mbases).

The number of reads generated (not shown in Table 4.1), per sample, in the standard protocol ranged from 29587896 – 68308594 (mean = 51649440) compared to 18817221 – 66527141 (mean = 45239778) in the optimised protocol. The percentage of raw reads per sample ranged from 7.3% – 27.2 % and 4.7% - 26.5%, comparing the standard to the optimised protocol, respectively.

Base calling accuracy is measured by the Phred quality score (Q-score) and indicates the probability that a given base is called incorrectly by the sequence. The Phred quality score (Q-score), is defined as a property that is logarithmically related to the base calling error probabilities (P)². A Q-score of 30 (Q30) is equivalent to the probability of an incorrect base call 1 in 1000 times. Hence, this implies that the probability of a correct base call is 99.9%. For example, a lower base call accuracy of 20 (i.e., Q20) will have an incorrect base call probability of 1 in 100 times which implies that for every 100 base pairs (bp) sequencing read, there will likely be an error. Although not shown in the Table 4.1, between 93.3% - 96.5% of reads per sample reached a Q-score of 30 or greater in the standard ChIP protocol compared to 94.4% – 96.7% in the samples used for the optimised protocol.

Enrichment (in which a minimum of 1.5-fold enrichment of the ChIP sample compared to the Input sample was used as minimal cut off acceptance) was

observed for both standard and optimised protocols for ChIP and library sample qPCR.

Alignment of the raw reads to the GRCh38 Human genome [153] using Bowtie [141], per sample, ranged from 96.25% – 98.65% (mean = 97.93%) and 96.71% - 98.65% (mean = 98.1%) comparing the standard versus the optimised ChIP protocols, respectively.

Peak calling refers to a computational method whereby regions in the genome that have enriched aligned reads are identified. MACS2 is an established peak-calling software which identifies fold-enrichment of genomic loci by comparing Input sample aligned reads to the ChIP sample aligned reads [142]. The Q-value of a test measures the proportion of false positives incurred (also known as the false discovery rate; FDR) when that particular test is deemed significant. The Q-value are adjusted p-values found using an optimised FDR approach by utilising characteristics of the p-value distribution which then produces a list of q-values. A Q-value of 0.01 indicates that 1% of all peaks called by MACS2 with a Q-value less than this are expected to be false positives. The results from the standard protocol show that (Table 4.1 and Figure 4.4) 2551 YY1 peaks were called in the MCF7 ethanol stimulated group compared to 6745 YY1 peaks in the same group which underwent the optimised protocol. The MCF7 oestradiol stimulated group generated 65032 compared to 59936 YY1 peaks. Both groups underwent MACS2 peak-calling at a conservative Q-value of 0.03.

4.5 Sequencing results obtained for H3K27ac ChIP-Seq in murine cancer associated fibroblast (CAF) cells

Referring to Table 4.1 and Figure 4.3, 10×10^6 cells were used per ChIP experiment and the mean number Yield was 1501 Mbases compared to 4012 Mbases comparing standard and optimised protocols, respectively. The mean number of raw reads generated in the standard ChIP protocol was 30019025 compared to 27105125 in the optimised protocol.

The mean percentage of raw reads in the standard ChIP versus the optimised ChIP protocol groups was 21% and 7%, respectively while a Q-Score of greater than 30 was 91.7% and 95.4%, respectively.

Mean percentage alignment rate to the Mm10 murine genome was 91.5% compared to 65.1% in the standard versus the optimised ChIP protocols. The Input sample of the CAF cells undergoing the optimised protocol produced a 33.61% alignment rate to the Mm10 genome. Further investigation for this revealed that 66.39% of the remaining reads were not able to be aligned using Bowtie. This resulted from a technical issue from the sequencing process. However, the number of significant peaks called by MACS2 were 26980 (q-value 0.01) in the standard ChIP group versus 47459 (q-value 0.02) in the optimised ChIP group.

Table 4.1. Showing sequencing results obtained from the standard and optimised ChIP protocols. All samples were sequenced using Illumina technology and 5ng of each library sample was provided. For library preparation 20ng of Input and ChIP DNA was used and the samples all underwent 10 cycles of PCR amplification.

Sample name	Standard/Optimised ChIP protocol?	Species	Number of cells used	Yield (Mbases)	Number of Reads	ChIP qPCR enrichment	Library ChIP qPCR enrichment	Genome version aligned to	Alignment rate (Bowtie)	Number of peaks obtained with MACS2	MACS2 Q-Value
MCF7-Etoh-Input-Rep1	Standard	Human	8 million	3415	68308594	✓	✓	GrCH38	98.65	2551	0.03
MCF7-Etoh-YY1-Rep1		Human		4379	29587896			GrCH38	98.15		
MCF7-E2-Input-Rep1	Standard	Human	8 million	3415	68308594	✓	✓	GrCH38	98.65	65032	0.03
MCF7-E2-YY1-Rep1		Human		5978	40392674			GrCH38	96.25		
MCF7-Etoh-Input-Rep2	Optimised	Human	8 million	3326	66527141	✓	✓	GrCH38	98.56	6745	0.03
MCF7-Etoh-YY1-Rep2		Human		4305	29087610			GrCH38	98.26		
MCF7-E2-Input-Rep2	Optimised	Human	8 million	3326	66527141	✓	✓	GrCH38	98.65	59936	0.03
MCF7-E2-YY1-Rep2		Human		2785	18817221			GrCH38	96.71		
CAF-Input Rep1	Standard	Murine	10 million	747	14931973	✓	✓	Mm10	87.27	26980	0.01
CAF-H3K27ac-Rep1		Murine		2255	45106077			Mm10	95.72		
CAF-Input Rep2	Optimised	Murine	10 million	1702	11500862	✓	✓	Mm10	96.62	47459	0.02
CAF-H3K27ac-Rep2		Murine		6321	42709388			Mm10	33.61		

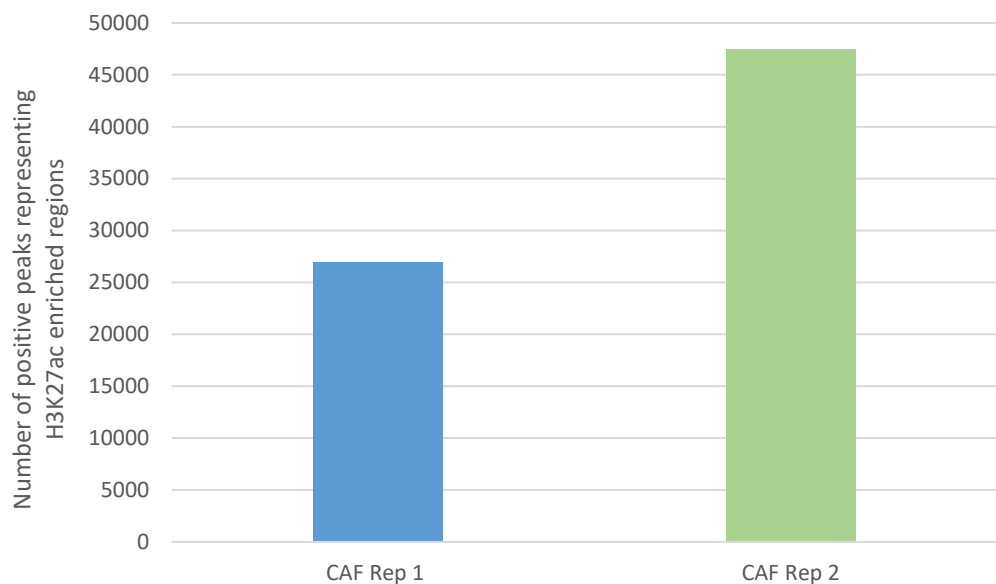


Figure 4.3 H3K27ac peaks called by MACS 2.0 comparing standard (blue bar) protocol to the optimised (green bar) protocol. Peaks were called with a q-value of 0.01 in the standard protocol and 0.02 in the optimised protocol.

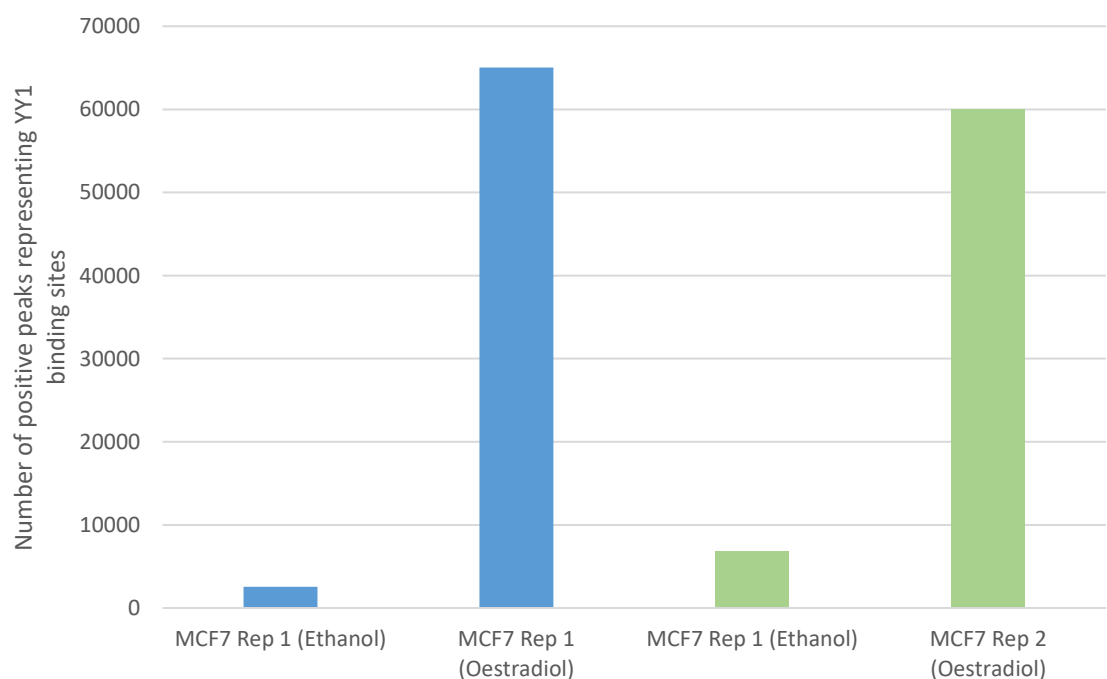


Figure 4.4 YY1 peaks called by MACS 2.0 comparing standard (blue bar) protocol to the optimised (green bar) protocol. Peaks were called with a q-value of 0.03 in both standard protocol and optimised protocols.

Chapter 5 Results: Analysis of the training set

5.1 Aims of the analysis of the initial training set of breast tumour samples

Sourcing and collection of BC tumour samples began within three months of the project commencing and tumour samples were processed for H3K27ac ChIP-Seq without delay.

The aims of the initial analysis were to gain further understanding of the cis-tromic regulatory network underlying ER α -positive BC. Firstly, epigenetic profiling was conducted on primary and metastatic breast tumours, using H3K27ac ChIP-Seq, and underwent bioinformatic analysis. Secondly, sequencing data from the epigenetic profiling was further analysed for the identification of super-enhancers in order to ascertain whether stratification of tumours could be carried out based on the called super-enhancers.

5.2 Clinicopathological details of the tumour samples

Table. 5.1 Showing the characteristics of the collected tumour samples which have been processed for ChIP-Seq. The tumours, are untreated postmenopausal, have been grouped according to "Good", "Poor" and "Metastatic" outcome arms. (IDC = invasive ductal carcinoma; ILC = invasive lobular carcinoma; IMC = invasive mucinous carcinoma; ER = oestrogen receptor; PR = progesterone receptor; HER2 = human epidermal growth factor receptor 2; "+" = positive; "-" = negative)

Tumour Sample	Good Outcome	Poor Outcome	Metastatic	Body site	Histology details	Hormonal Status
Good-1	✓			Breast	IDC	ER+, PR+, HER2-
Good-2	✓			Breast	IDC	ER+, PR-, HER2+
Good-3	✓			Breast	IDC	ER+, PR+, HER2-
Good-4 + (Good-9)	✓			Breast	IDC	ER+, PR-, HER2+
Good-5	✓			Breast	IDC	ER+, PR+, HER2-
Good-6	✓			Breast	IDC	ER+, PR+, HER2+
Good-7	✓			Breast	IDC	ER+, PR+, HER2-
Good-8	✓			Breast	IDC	ER+, PR+, HER2-
Good-10	✓			Breast	ILC	ER+, PR+, HER2-
Good-11	✓			Breast	IDC	ER+, PR+, HER2-
Good-12	✓			Breast	ILC	ER+, PR-, HER2-
Poor-1		✓		Breast	IDC	ER+, PR+, HER+
Poor-2 + (Poor-6)		✓		Breast	IDC	ER+, PR+, HER2-
Poor-3 + (Poor7)		✓		Breast	IDC	ER+, PR-, HER2-
Poor-4		✓		Breast	IDC	ER+, PR+, HER2-
Poor-5		✓		Breast	IDC	ER+, PR+, HER2-
Met-1			✓	Lymph node	IDC	ER+, PR-, HER2+
Met-2			✓	Liver Biopsy	IDC	ER-, PR-, HER2-
Met-3			✓	Chest wall mass	IMC	ER+, PR-, HER2-
Met-4			✓	Breast recurrence	IDC	ER+, PR-, HER2+
Met-5			✓	Breast recurrence	IDC	ER+, PR-, HER2+
Met-6			✓	Internal thoracic mass	IDC	ER-, PR-, HER2+
Met-7			✓	Groin lymph node	IDC	ER-, PR+, HER2-
Met-8			✓	Axillary Lymph Node	IDC	ER+, PR+, HER2-
Met-9			✓	Liver Biopsy	IDC	ER+,PR+, HER2-
Met-10			✓	Liver Biopsy	IDC	ER+, PR+, HER2-
Met-11			✓	Sternal Mass	IDC	ER-, PR-, HER2-
Met-12			✓	Liver Biopsy	IDC	ER+, PR-, HER2-

A total of 12, 7 and 12 tumour samples (N=31) were retrieved from the ICHTB and the VERB study, and categorised into "Good", "Poor" and "Metastatic" outcome groups, respectively. All the Good and Poor outcome tumours were deemed ER-positive upon IHC staining prior to selection for ChIP-Seq. The metastatic samples (ER positive or negative) were only selected for if the patient had a previous background of ER positive breast cancer (Table 5.1). In addition, 3 tumour samples were processed as technical replicates. This includes Good-4 and Good-9, Poor-2 and Poor-6, Poor-3 and Poor-7).

5.3 Sequencing results

Raw sequencing files were obtained with number of sequencing reads ranging between 11,600,000 – 34,000,000 amongst the 31 tumours processed for H3K27ac ChIP-Seq. Alignment of the raw sequencing files to the hg19 reference genome demonstrated percentage alignment ranging from 89.1% - 97.3% (Table 5.2) [140]. Peak calling was successfully performed for all the 31 aligned tumours using MACS (Range: 71- 49262) (Table 5.2) [142]. Tumours Good-2 and Good-4 show significant less peaks than the rest of the 29 sequenced tumours.

Table 5.2 Percentage (%) alignment of the raw (FASTQ) files to the hg19 reference genome and number of positive peaks called.

Tumour Sample	Percentage read alignment (%) to Reference Genome hg19 called by Bowtie [141]	Number of positive peaks called by MACS [142]
Good-1	95.6	15423
Good-2	89.1	71
Good-3	93.7	58147
Good-4	89.6	603
Good-5	89.5	29896
Good-6	89.1	7215
Good-7	94.6	4684
Good-8	88.2	30574
Good-9	89.7	40874
Good-10	92.7	25403
Good-11	91.2	22642
Good-12	91.7	28722
Poor-1	94.8	49262
Poor-2	90.6	28675
Poor-3	89.1	39407
Poor-4	91.7	14025
Poor-5	89.4	10423
Poor-6	90.3	28675
Poor-7	89.2	37723
Met-1	89.5	1937
Met-2	91.3	27369
Met-3	89.4	10847
Met-4	97.3	36276
Met-5	96.4	37132
Met-6	92.1	32476
Met-7	90.8	31511
Met-8	94.8	37141
Met-9	87.6	13562
Met-10	92.5	17802
Met-11	90.8	2940
Met-12	90.1	13132

5.4 Epigenetic analysis of sequenced breast tumours

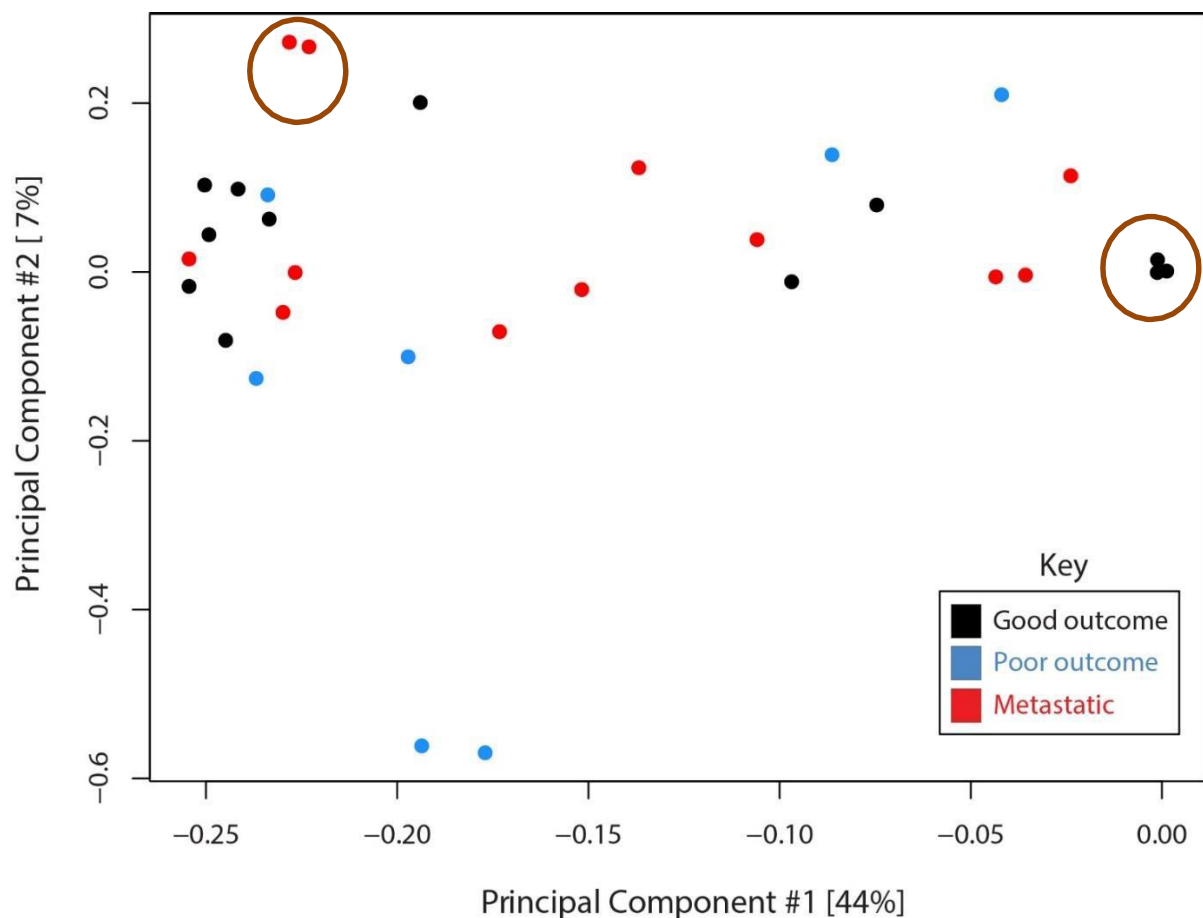


Figure 5.1 Principal Component analysis on H3K27ac tumour profiles (N=31). PCA displays how the groups cluster based on differentially enriched sites. The closer the dots are, the greater the similarity in characteristics.

[Analysis conducted by Nikolaos Trasanidis and Darren K. Patten]

Principal component analysis (PCA) pictured in Figure 5.1 demonstrates the similarities observed in each (represented by the colour coded dot) of the 31 sequenced tumours. The distance between each dot represents a measure of difference of each tumour sample from its counterpart(s). With reference to Figure 5.1, no grouping of the tumours within the Good, Poor and Metastatic groups can be observed. However, there is one area (displayed within the brown circle) where three Good outcome tumours are grouped close together signifying high similarity. There

are also two metastatic samples (displayed within the brown circle), to the left top corner of the PCA, which exhibit similarity.

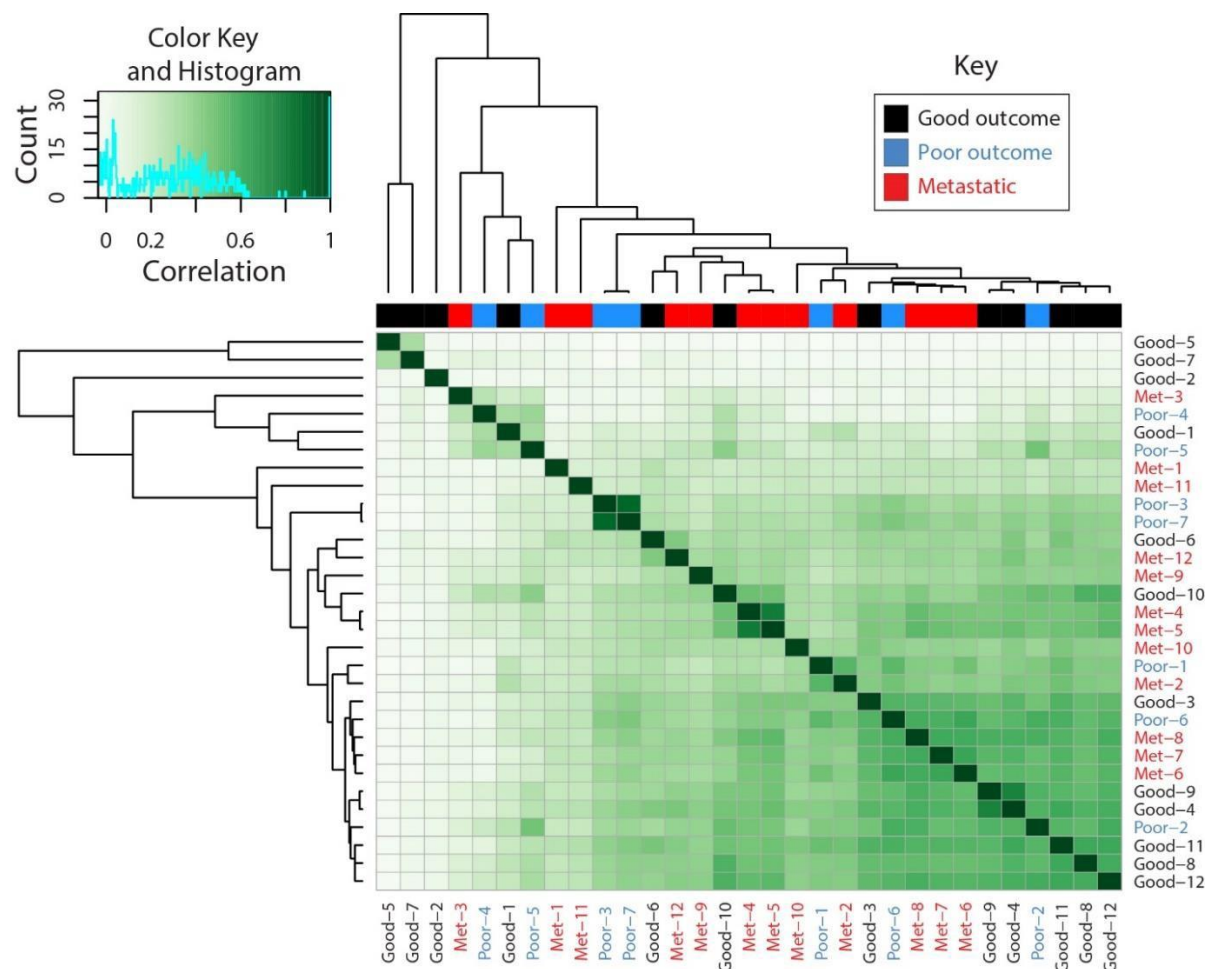


Figure 5.2 Correlation heat map representation for significantly differentially enriched sites for H3K27ac marker in poor, good and metastatic outcome samples [Analysis performed by Nikolaos Trasanidis and Darren K. Patten]

The peak calling data was further analysed to assess for similarity amongst the individual samples (N=31). The correlation heat map, generated by DiffBind [85, 145] pictured in Figure 5.2, demonstrates a cluster correlation plot based on the enriched genomic regions (called peaks) for the 31 (H3K27ac ChIP-Seq) sequenced breast tumours. There is no distinct overall grouping/clustering amongst the Good, Poor

and Metastatic groups. However, there are 9 main areas which show differentially stronger correlation than other areas within the heat map.

Reading from the vertical axis, these include: a) Good-5 and Good-7, b) Met-3 Poor-4, Good-1 and Poor-5, c) Poor-3 and Poor-7, from d) Poor-3 to Good-12 shows a correlation of approximately 0.3, e) Met-4 and Met-5, f) Good-10 to Good 12 (correlation of 0.4), g) Good-3 to Good-12 (correlation of 0.5), h) Good-9 and Good-4, which are one of the three technical replicates (correlation of 0.8) and finally i) Good-11 to Good-12 (correlation = 0.5). The Good outcome tumours tend to cluster relatively stronger than the Poor or Metastatic outcome groups. Further analysis reveals no trend based on hormonal phenotype of the tumours and the areas of clustering observed.

5.5 Super-enhancer identification and analysis

Super-enhancers were successfully called in all the tumour samples (N=31) applying the ROSE computational software (Figure 5.3) with the aligned (BAM) files from the H3K27ac ChIP-Seq data. The average number of super-enhancers called in Good (N=12), Poor (N=7) and Metastatic (N=12) tumour samples was 742 (Range: 12-1174), 826 (Range: 162-1045) and 774 (Range: 228-1099), respectively [116, 119].

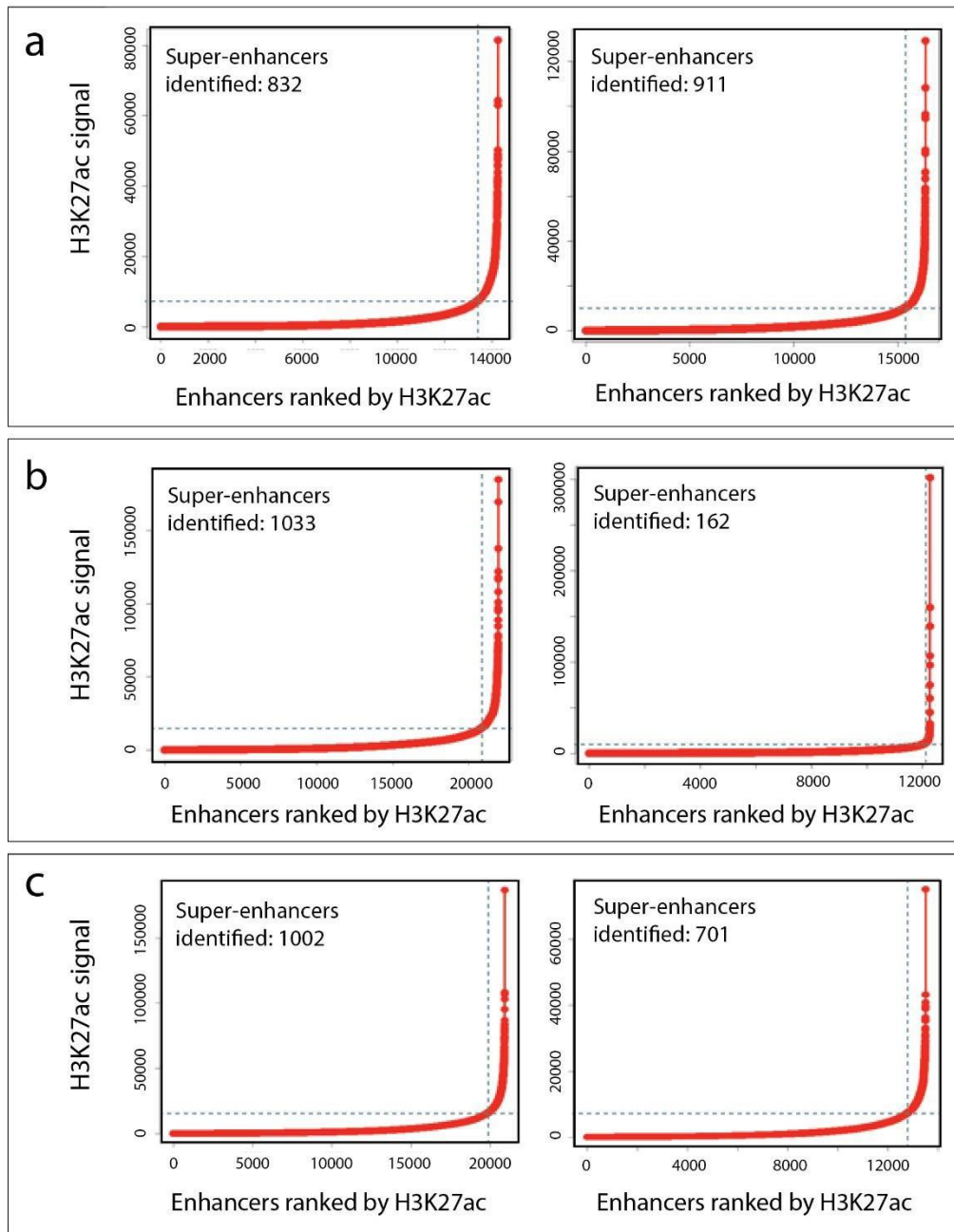


Figure 5.3 Graphical and numerical representation of the called super-enhancers in Good, Poor and Metastatic tumours samples (The figure demonstrates 2 randomly picked samples from each outcome group; (a) Good, (b) Poor, (c) Metastatic). The X-axis represents the number of enhancers identified and the Y-axis demonstrates the corresponding H3K27ac signal. All the graphs (N=31) exhibit the same tangential pattern whereby the signal intensity rises exponentially at the point where the gradient of the curve approaches 1. At this “cutoff” point, super-enhancers are called; the H3K27ac signal increases by orders of magnitude. The method applied for super-enhancer calling was applied as described by Whyte *et al.* and Loven *et al.* [116, 119] **[Analysis performed by Luca Magnani and Darren K. Patten]**

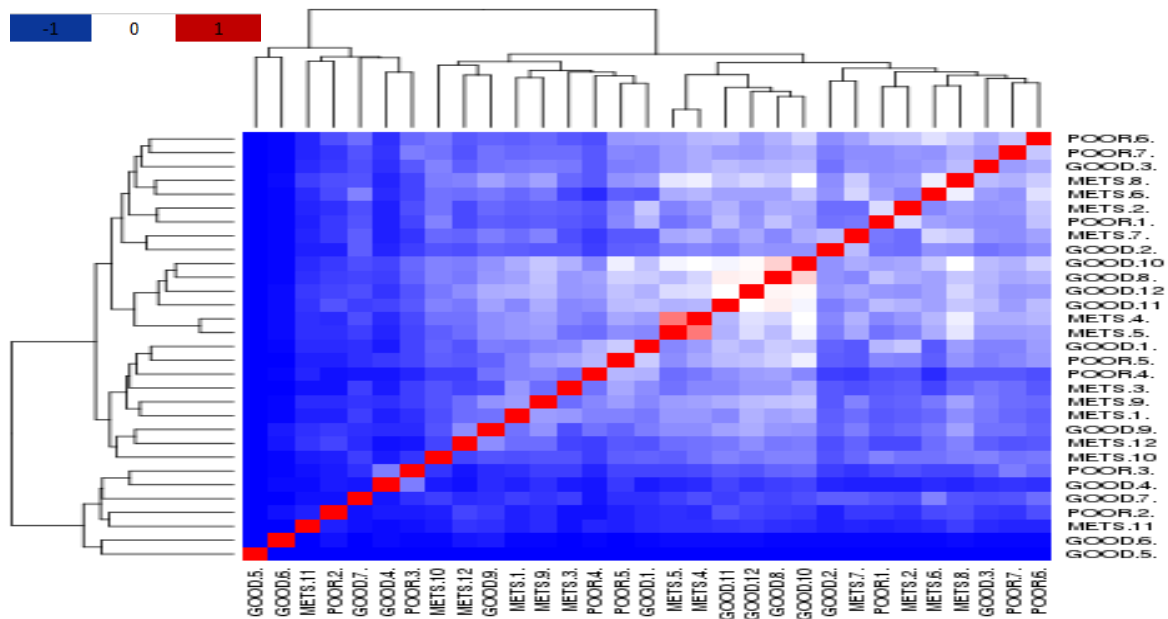


Figure 5.4 Heat map: clustering of tumours (N=31) based on the called super-enhancers. [Analysis conducted by Edward Curry]

To ascertain whether the Good, Poor and Metastatic groups share similarity in super- enhancer positions, a heat map cluster plot across all 31 tumours was applied using R which is displayed in Figure 5.4. There is no definitive pattern to show clustering patterns amongst all Good, Poor and Metastatic samples. Good-8, Good-10 and Good-12 as well as Metastatic samples -4 and -5 correlate relatively higher with each other than the rest of the samples on the heat map. Subsequently, the super-enhancer data was further analysed using a multidimensional scaling (MDS) plot of Jaccard distances (Figure 5.5). Dissimilarity amongst samples is depicted by the distance between one or more sample with others in the same plot. Although there is no distinct grouping amongst the samples within the same outcome group, there are specific samples which share similarity.

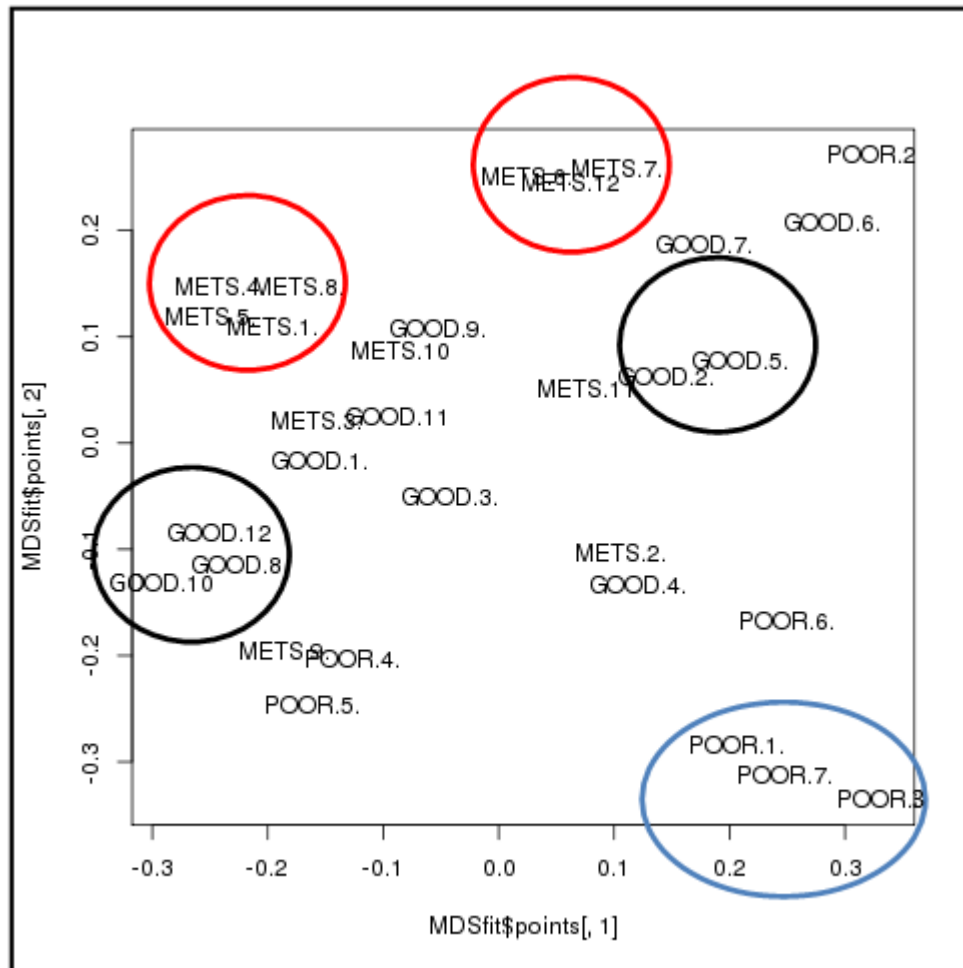


Figure 5.5 Multidimensional scaling plot of Jaccard distances, based on identified super-enhancers, of the Good, Poor, Metastatic tumour groups (N=31). The red, black and blue circles are highlighting short Jaccard distances amongst Metastatic, Good and Poor groups. **[Analysis conducted by Edward Curry]**

There are 2 groups within the Metastatic samples (a) METS -1, -4, -5 and -8 and b) METS -6, -7, -12) and Good samples (a) Good -8, -10, -12 and b) Good -2 and -5) and 1 group within the Poor samples (a) Poor -1, -3, -7) which demonstrate similarity, depicted by short Jaccard distances. The results obtained from the heat map cluster plot are mirrored in the Jaccard distance analysis which, in addition, has displayed additional samples sharing similar characteristics.

Chapter 6 Results: Epigenomic profiling of primary and metastatic breast tumour samples using H3K27ac ChIP-Seq

6.1 Aims of the analysis of the complete set of primary and metastatic breast tumour samples

An alternative analytical approach was adapted for investigation of the H3K27ac ChIP-Seq data from the breast tumour samples due to the fact that, stratifying the tumour samples based on outcome or called super-enhancers did not reveal any distinguishing factors from the previous analysis reported in Chapter 5.0. H3K27ac ChIP-Seq data for all primary and metastatic breast cancer tumour samples were initially analysed as one group without stratification founded on clinical outcome. Original bioinformatics approaches, aimed to deconvolute the complex regulatory landscape of ER α -positive BC, were applied.

6.2 Clinicopathological details of tumour samples

Table 6.1 (A/B) Characteristics of the collected primary (Table 6A) and metastatic (Table 6B) tumour samples which have been processed for H3K27ac ChIP-Seq. (ER = oestrogen receptor; PR = progesterone receptor; HER2 = human epidermal growth factor receptor 2; "+" = positive; "-" = negative; IDC = invasive ductal carcinoma; ILC = invasive lobular carcinoma; IMC = invasive mucinous carcinoma; A = Adjuvant hormonal therapy; N = Neoadjuvant hormonal therapy CR = contralateral recurrence; LR = Local recurrence; D = Distant recurrence; L = Lost to follow-up; N/S = Not specified)

Table 6.1A Clinicopathological information for primary breast tumour samples collected and processed for H3K27ac ChIP-Seq.

Tumour Sample	Tumour grade	Tumour size (mm)	Receptor status	Histology	Lymph node positivity	Adjuvant(A) and/or Neoadjuvant (N) therapy	Date of diagnosis	Recurrence date	Recurrence type	Alive (+ censor date) or Deceased (+ Date)	Time to relapse (months)
G_DKP_001	2	35	ER +, PR +, HER2 -	IDC	15 of 34	A	19/03/2003	None		Alive	
G_DKP_002	2	15	ER +, PR +, HER2 -	IDC	5 of 22	A	24/11/2004	None		Alive	
G_DKP_003	2	15	ER +, PR +, HER2 -	IDC	5 of 22	A	24/11/2004	None		Alive (12/05/2016)	
G_DKP_004	2	30	ER +, PR +, HER2 -	IDC	6 of 15	A	24/05/2005	None		Alive (12/05/2016)	
G_DKP_005	2	50 + 10	ER +, PR -, HER2 +	IDC	1 of 8	A	27/09/2005	None		Alive (12/05/2016)	
G_DKP_006	3	38	ER +, PR -, HER2 -	IDC	1 of 23	A	10/01/2007	None		Alive (12/05/2016)	
G_DKP_007	2	15	ER +, PR +, HER2 -	IDC	5 of 22	A	24/11/2007	None		Alive (12/05/2016)	
G_DKP_008	2	50 + 10	ER +, PR -, HER2 +	IDC	1 of 8	A	24/11/2004	None		Alive (12/05/2016)	
G_DKP_009	2	71	ER +, PR -, HER2 -	ILC	0 of 10	N	27/09/2005	None		Alive (12/05/2016)	
G_DKP_010	2	16	ER +, PR +, HER2 -	IDC	0 of 18	A	04/12/2007	None		Alive (12/05/2016)	
G_DKP_011	2	20	ER +, PR +, HER2 -	ILC	0 of 6	A	13/08/2008	None		Alive (12/05/2016)	
G_DKP_012	2	19	ER +, PR +, HER2 -	IDC	1 of 6	A	22/05/1991	None		Alive (12/05/2016)	
G_DKP_013	2	4	ER +	IDC	0 of 1	No therapy	14/05/1993	16/03/2005	CR	Alive (12/05/2016)	142
G_DKP_014	2	20	ER +, PR +, HER2 -	IDC	3 of 7	Unknown	15/03/1994	None		Alive (12/05/2016)	
G_DKP_015	2	25	ER +, PR +, HER2 -	IDC	4 of 6	A	22/05/1991	01/06/2006	LR	Alive (12/05/2016)	181
G_DKP_016	2	12	ER +, PR +, HER +	IDC	1 of 7	No therapy	13/09/1991	None		L	
G_DKP_017	2	8	ER +	IDC	0 of 2	A+N	30/09/1991	02/03/2005	CR	Alive (12/05/2016)	162
G_DKP_018	2	20	ER +, HER2 -	IDC	0 of 8	No therapy	17/01/1992	None		L	
G_DKP_019	2	15	ER +, PR +, HER2 -	IDC	2 of 5	A	12/01/1994	27/10/2009	LR	Alive (12/05/2016)	189
G_DKP_020	2	20	ER +, PR +, HER2 -	IDC	2 of 6	A	15.08/1994	None		Alive (12/05/2016)	
G_DKP_021	2	16	ER +, PR +, HER2 -	IDC	0 of 18	A	15/10/2005	None		Alive (12/05/2016)	

Table 6.1A Showing clinicopathological information for primary breast tumour samples collected and processed for H3K27ac ChIP-Seq.

Tumour Sample	Tumour grade	Tumour size (mm)	Receptor status	Histology	Lymph node positivity	Adjuvant(A) and/or Neoadjuvant (N) therapy	Date of diagnosis	Recurrence date	Recurrence type	Alive (+ censor date) or Deceased (+ Date)	Time to relapse (months)
P_DKP_001	1	43 + 8	ER +, PR +, HER2 -	IDC	0	A	11/01/2009	None		Deceased (04/12/2010)	
P_DKP_002	3	30	ER +, PR -, HER2 -	IDC	1 of 12	A	16/11/2005	None		Deceased (23/03/2008)	
P_DKP_003	3	64	ER +, PR -, HER2 -	IDC	7 of 20	A	10/12/2007	20/11/2010	LR	Deceased (2012)	35
P_DKP_004	2	20 + 16	ER +, PR +, HER2 -	IDC	0 of 20	No therapy	17/01/2007	26/10/2010	D	Deceased (12/07/2012)	45
P_DKP_005	2	50	ER +, PR +, HER2 -	IDC	10 of 22	A	20/04/2010	20/04/2010	D	Deceased 25/01/2013	0
P_DKP_006	3	30	ER +, PR -, HER2 -	IDC	1 of 12	A	16/11/2005	None		Deceased (23/03/2008)	
P_DKP_007	3	64	ER +, PR -, HER2 -	IDC	7 of 20	A	10/12/2007	20/11/2010	LR	Deceased (2012)	35
P_DKP_008	2	31	ER +, PR +, HER2 -	IDC	0 of 4	A	25/09/2012	None		Alive (12/05/2016)	
P_DKP_009	2	32	ER +, PR +, HER2 -	IDC	2 of 13	A	15/11/2010	None		Deceased (25/05/2011)	
P_DKP_010	2	17	ER +, PR +, HER2 -	IDC	0 of 8	A	15/06/2011	None		Alive (12/05/2016)	
P_DKP_011	2	25	ER +, PR +, HER2 -	IDC	0 of 1	A	15/11/2012	None		Alive (12/05/2016)	
P_DKP_012	2	25	ER +, PR +, HER2 -	IDC	1 of 6	No therapy	22/12/1993	22/02/1996	D	Deceased (28/07/1997)	26
P_DKP_013	2	15	ER +, PR -, HER2 -	IDC	0 of 1	A	07/07/2011	None		Alive (12/05/2016)	
P_DKP_014	2	28	ER +, PR +, HER2 -	IDC	0 of 2	A	22/08/2012	None		Alive (12/05/2016)	
P_DKP_015	2	9	ER +, PR +, HER2 -	IDC	5 of 13	A	10/09/2012	None		Alive (12/05/2016)	
P_DKP_016	2	39	ER +, PR -, HER2 -	IDC	6 of 21	A	10/09/2012	None		Alive (12/05/2016)	
P_DKP_017	2	18	ER +, PR -, HER2 +	IDC	0 of 4	A	17/05/2013	None		Alive (12/05/2016)	
P_DKP_018	2	29	ER +, PR +, HER2 -	IDC	0 of 2	A	19/06/2013	None		Alive (12/05/2016)	

Table 6.1B Clinicopathological information for metastatic breast tumour samples collected and processed for H3K27ac ChIP-Seq.

Tumour sample	Recurrence site	Size of metastasis (mm)	Grade	Histology details	Lymph nodes positive	Receptor status	Primary disease diagnosis date	Date of metastatic diagnosis	Alive (+ censor date) or Deceased (+ Date)	Time to relapse (months)
M_DKP_001	Liver	N/S	N/S	IDC	N/S	ER +, PR -, HER2 -	05/02/1989	04/02/2005	Deceased (04/03/2008)	192
M_DKP_002	Liver	42 x 35	N/S	IDC	N/S	ER -, PR -, HER2 -	04/11/2010	03/08/2012	Deceased (17/08/2014)	21
M_DKP_003	Axillary Lymph node	N/S	N/S	IDC	N/S	ER +, PR +, HER2 -	20/05/2005	03/06/2014	Alive (12/05/2016)	109
M_DKP_004	Chest wall mass	100 x 83 x 102	N/S	IMC	N/S	ER +, PR -, HER2 -	03/04/1989	19/08/2014	Alive (12/05/2016)	304
M_DKP_005	Breast recurrence	38	3	IDC	N/S	ER +, PR -, HER2 +	28/02/2014	28/02/2014	Alive (12/05/2016)	0
M_DKP_006	Breast recurrence	38	3	IDC	N/S	ER +, PR -, HER2 +	28/02/2014	28/02/2014	Alive (12/05/2016)	0
M_DKP_007	Internal thoracic mass	N/S	2	IDC	N/S	ER -, PR -, HER2 +	30/01/2013	25/03/2014	Deceased (01/05/2014)	14
M_DKP_008	Liver	24	N/S	IDC	N/S	ER +, PR +, HER2 -	19/12/2001	27/03/2014	Alive (12/05/2016)	147
M_DKP_009	Liver	20	N/S	IDC	N/S	ER +, PR +, HER2 -	20/02/2007	11/04/2014	Deceased (26/10/2015)	86

Table 6.1B Clinicopathological information for metastatic breast tumour samples collected and processed for H3K27ac ChIP-Seq.

Tumour sample	Recurrence site	Size of metastasis (mm)	Grade	Histology details	Lymph nodes positive	Receptor status	Primary disease diagnosis date	Date of metastatic diagnosis	Alive (+ censor date) or Deceased (+ Date)	Time to relapse (months)
M_DKP_010	Sternal mass	56	N/S	IDC	N/S	ER -, PR -, HER2 -	19/03/2010	11/04/2014	Deceased (21/03/2016)	49
M_DKP_011	Groin lymph node	34.9	N/S	IDC	N/S	ER -, PR +, HER2 -	14/06/1996	15/05/2011	Alive (12/05/2016)	179
M_DKP_012	Axillary lymph Node	12	N/S	IDC	1 of 27	ER +, PR +, HER2 -	25/07/2014	25/07/2014	Alive (12/05/2016)	0
M_DKP_013	Liver	40	N/S	IDC	N/S	ER +, PR +, HER2 -	21/11/2013	21/11/2013	Deceased (13/08/2015)	0
M_DKP_014	Liver	20	N/S	IDC	N/S	ER +, PR +, HER2 -	22/11/2013	22/11/2013	Deceased (13/08/2015)	0
M_DKP_015	Liver	34	N/S	IDC	N/S	ER +, PR +, HER2 -	23/11/2013	23/11/2013	Deceased (13/08/2015)	0
M_DKP_016	Liver	29	N/S	IDC	N/S	ER +, PR +, HER2 -	24/11/2013	24/11/2013	Deceased (13/08/2015)	0

A total of 39 primary tumours samples were retrieved (Table 6.1A), 19 samples from the ICHTB and the VERB study and 20 samples from Breast Cancer Now tissue bank. With regard to the metastatic sample collection, 16 cases were collected (Table 6.1B); the ER α -negative metastatic samples were collected on the proviso that the primary disease was deemed ER α -positive. All 55 samples underwent H3K27ac ChIP-Seq as described in Chapter 2; Section 2.3 to 2.5.

All of the tumours were reported to have undergone haematoxylin and eosin (H&E) staining and tumour burdens of each sample was given as a percentage of the total cell number within that particular sample. A 70% tumour burden cut-off was applied before inclusion into this study for ChIP-Seq. Tumour burdens for the ICHTB samples were all guaranteed above 70-75% by the tissue bank pathologist, which were checked at the time the samples were stored for research purposes. The H&E slides for the ICHTB samples were not available upon request. Samples obtained from the Breast Cancer Now tissue bank underwent H&E staining and subsequent tumour burden measurement. The mean tumour burden of the samples was 81.9% (Range: 71% - 91%).

With reference to Table 6.1A, a number of primary tumour samples, from the ICHTB, were retrospectively discovered as biological replicates for the same tumour despite having different tissue bank identifiers. These included, 1. G_DKP_002 and G_DKP_003, 2. G_DKP_005 and G_DKP_008, 3. G_DKP_010 and G_DKP_021, 4. P_DKP_003 and P_DKP_007 and 5. P_DKP_002 and P_DKP_006.

Within the metastatic samples (Table 6.1B), M_DKP_005 and M_DKP_006 were biological replicates of the same tumour. M_DKP_013 to M_DKP_016 were 4 separate liver metastatic BC deposits from the same patient.

6.3 Summary statistics for ChIP-Seq analyses

Raw sequencing files were aligned to the human genome (GrCh38) [153] using Bowtie [141] (Table 6.2A). The average alignment rate from the primary (N=39) BC tissue ranged from 87.3% - 97.9% (average alignment rate = 91.8%) and 88.1% - 97.3% (average alignment rate = 92.0%) in the Input (control) and ChIP samples, respectively. For the metastatic group (Table 6.2B) the alignment rate ranged between 87.4% - 97.0% (average alignment rate = 91.9%) and 89.5% - 97.3% (average alignment rate = 92.0%) for Input and ChIP samples, respectively.

The number of initial reads, redundant rates (of reads) and reads after filtering are displayed, in tables 6.2A and 6.2B, for both Input and ChIP samples. The corresponding number of MACS 2.0 [142] (q-value = 0.01) peaks called are also shown. For the primary samples the number of H3K27ac peaks ranged between 0 – 364,228 (average = 64,042 peaks) and for the metastatic samples the range of H3K27ac peaks was between 0 – 172,092 (average = 63,691 peaks).

Samples generating less than 2000 peaks (N=7) which included G_DKP_003 (204 peaks), G_DKP_006 (0 peaks), P_DKP_010 (0 peaks), P_DKP_013 (1467 peaks), M_DKP_003 (1039 peaks), M_DKP_010 (404peaks), M_DKP_016 (825 peaks), were excluded from further downstream analysis. Figure 6.1 graphically demonstrates the numbers of peaks generated in the tumour samples (N=48) following exclusion of the above 7 samples. The sample G_DKP_013 generated 364, 228 peaks and was deemed an outlier based on the distribution of peaks across the other samples. This sample was therefore excluded leaving 47 tumour sample ChIP-Seq tracks for downstream analysis.

Table 6.2A Alignment rates generated from Bowtie [141], number of raw reads before and after redundant rates are applied and peaks called by MACS 2.0 [142] (q-value = 0.01) for all the primary breast cancer samples.

Sample Number	Bowtie hg38 alignment rate (%) in control + ChIP samples	Number of reads in control sample	Redundant rate	Number of reads in control sample after filtering	Number of reads in treatment sample	Redundant rate	Number of reads in treatment sample after filtering	MACS2 Peaks
G_DKP_001	95.6/94.9	29416384	0.15	25123727	48617322	0.41	28822320	15957
G_DKP_002	88.1/88.2	45513399	0.05	43348259	26915513	0.09	24404518	4175
G_DKP_003	92.7/92.6	39116635	0.05	37014317	2091525	0.01	2075954	204
G_DKP_004	90.6/91.2	51250051	0.3	36085942	44081910	0.66	14957984	3807
G_DKP_005	90.4/90.3	46998364	0.14	40483700	47732047	0.12	42051904	11441
G_DKP_006	89.1/88.9	54050875	0.88	6709108	82756265	0.45	45725207	0
G_DKP_007	94.6/94.3	38557798	0.05	36643309	37704908	0.04	36170771	114806
G_DKP_008	88.2/88.1	46998364	0.14	40483700	7604832	0.03	7392475	69859
G_DKP_009	89.8/89.9	29647976	0.12	26156591	24409362	0.11	24409362	52073
G_DKP_010	92.7/93.0	35527333	0.1	31848417	16132966	0.08	14801314	16207
G_DKP_011	91.7/91.9	8482217	0.11	7585968	10730688	0.07	9969157	18098
G_DKP_012	94.8/94.3	59423819	0.15	51155594	59423819	0.14	51155594	85829
G_DKP_013	87.3/88.1	43873611	0.15	37140133	38059931	0.07	35237112	364228
G_DKP_014	91.3/91.4	21693843	0.15	18482495	37052893	0.13	32246197	83447
G_DKP_015	95.6/95.4	28242679	0.09	25822782	37386385	0.1	33811074	61201
G_DKP_016	96.4/96.3	23202730	0.16	17582561	23202730	0.24	17582561	96973
G_DKP_017	90.6/91.7	36681717	0.29	24717941	49067405	0.33	34973407	97466
G_DKP_018	91.5/92.6	40816138	0.41	23896190	33080341	0.4	19807861	32818
G_DKP_019	95.6/95.7	32033225	0.09	29127585	33331400	0.11	29824074	123661
G_DKP_020	97.9/97.3	28670863	0.09	26157793	23914487	0.1	21598647	79846
G_DKP_021	95.4/96.1	41584558	0.17	34432211	41584558	0.22	34432211	19134

Table 6.2A Alignment rates generated from Bowtie [141], number of raw reads before and after redundant rates are applied and peaks called by MACS 2.0 [142] (q-value = 0.01) for all the primary breast cancer samples.

Sample Number	Bowtie hg38 alignment rate (%) in control + ChIP samples	Number of reads in control sample	Redundant rate	Number of reads in control sample after filtering	Number of reads in treatment sample	Redundant rate	Number of reads in treatment sample after filtering	MACS2 Peaks
P_DKP_001	94.8/94.7	33868872	0.11	30260525	33868872	0.11	30260525	130031
P_DKP_002	90.6/90.2	26270448	0.15	21408480	23002643	0.07	21408480	52672
P_DKP_003	89.1/89.7	18990914	0.05	18747394	19954126	0.06	18747394	100068
P_DKP_004	91.7/91.9	24755410	0.13	21589384	31318859	0.27	43116870	10202
P_DKP_005	89.4/90.1	3033924	0.1	2745484	3033924	0.08	2745484	2383
P_DKP_006	90.3/90.7	26270448	0.15	22222658	31013027	0.17	25605492	55272
P_DKP_007	89.2/90.1	10325143	0.58	4303907	25407101	0.12	22421625	42429
P_DKP_008	88.5/88.9	6954287	0.03	6742122	61660871	0.14	52984880	17986
P_DKP_009	94.2/93.9	7304342	0.09	6653058	75611766	0.17	63052925	2521
P_DKP_010	90.5/90.3	6857077	0.1	6153624	63757687	0.26	47271136	0
P_DKP_011	89.7/89.4	49675611	0.12	44726242	35312910	0.1	31111580	172092
P_DKP_012	95.7/95.8	24787260	0.13	21673823	46059302	0.18	37686881	18340
P_DKP_013	92.8/92.8	24074930	0.06	22528925	46123154	0.08	42345493	1468
P_DKP_014	91.7/91.8	57577050	0.19	46357312	73139872	0.17	60360449	155186
P_DKP_015	93.5/93.4	25984041	0.05	24777058	36541319	0.21	28962803	101444
P_DKP_016	92.1/95.1	51860134	0.12	45390661	41575023	0.22	32249718	171338
P_DKP_017	88.4/88.1	17994528	0.06	16960718	40731205	0.21	32111033	82394
P_DKP_018	89.8/90.1	38222773	0.12	33794658	62871556	0.3	44085078	30615

Table 6.2B Showing alignment rates generated from Bowtie [141], number of raw reads before and after redundant rates are applied and peaks called by MACS 2.0 [142] (q-value = 0.01) for all the metastatic breast cancer samples.

Sample Number	Bowtie hg38 alignment rate (%) in control + ChIP samples	Number of reads in control sample	Redundant rate	Number of reads in control sample after filtering	Number of reads in treatment sample	Redundant rate	Number of reads in treatment sample after filtering	MACS2 Peaks
M_DKP_001	89.5/89.3	23851177	0.05	22756985	22526970	0.04	21681987	8431
M_DKP_002	91.3/91.4	14117528	0.15	11969698	12706557	0.15	10792912	30011
M_DKP_003	89.4/89.3	9463740	0.68	3016791	30311922	0.74	7944040	1040
M_DKP_004	97.3/97.0	34440294	0.58	14625360	48687811	0.58	20663350	5133
M_DKP_005	96.4/95.2	24582698	0.03	23809766	61786583	0.12	54499692	99738
M_DKP_006	92.1/92.7	26186664	0.04	25169712	25958127	0.06	24314944	89530
M_DKP_007	90.8/90.9	28787681	0.05	25709307	28647301	0.1	25709307	60791
M_DKP_008	94.8/95.2	29260547	0.42	16831215	26329506	0.47	13990704	7416
M_DKP_009	87.6/87.4	30119569	0.3	21228854	26207315	0.36	16746266	9864
M_DKP_010	92.5/92.3	23859790	0.18	19624489	27669040	0.14	23867566	404
M_DKP_011	90.8/91.1	46279779	0.04	44568755	31207234	0.03	30243502	53772
M_DKP_012	90.1/91.5	73251397	0.09	66598280	36195802	0.12	31696346	82105
M_DKP_013	95.6/97.0	38574636	0.12	34118135	24677684	0.17	20574946	3559
M_DKP_014	91.5/94.1	27466355	0.09	24858027	47227405	0.1	42434157	16226
M_DKP_015	92.8/92.6	15713576	0.19	13262804	15713576	0.16	13262804	62707
M_DKP_016	89.4/91.1	18633770	0.14	15998892	21114290	0.16	17636173	825

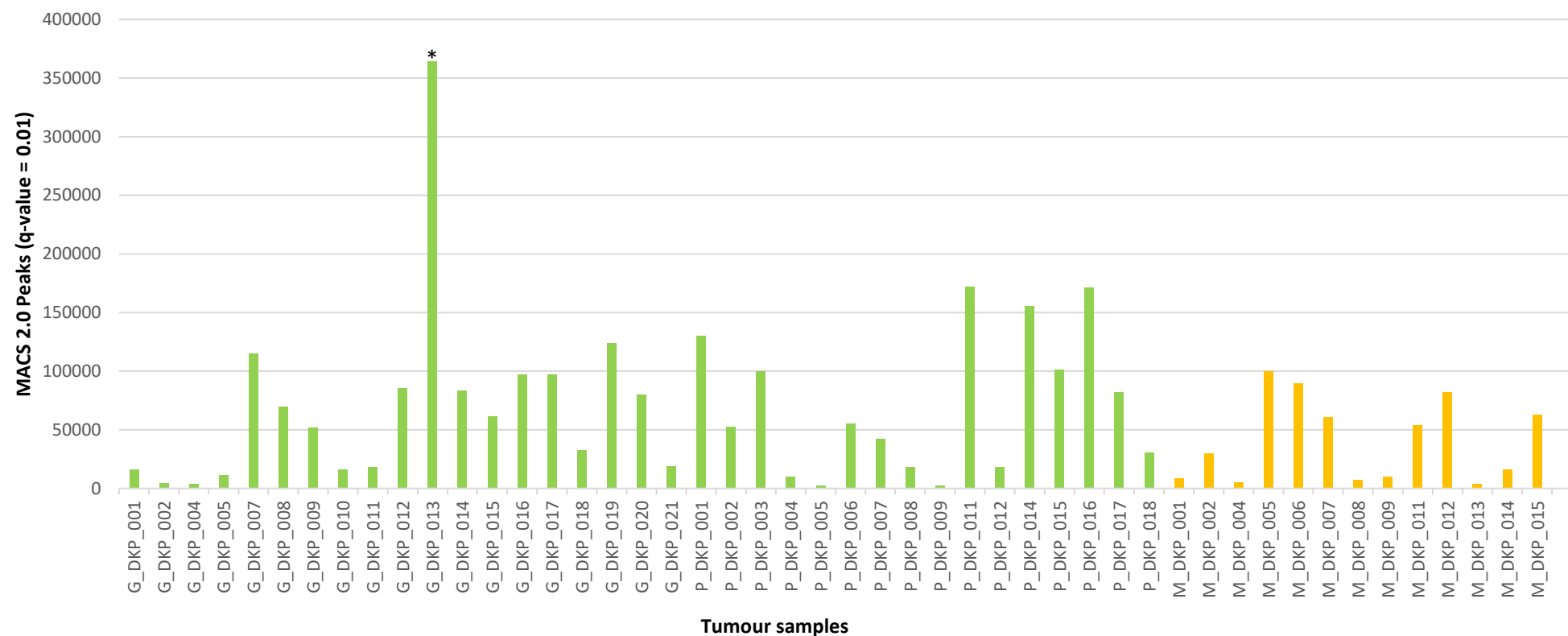


Figure 6.1 MACS 2.0 generated peaks (q-value = 0.01) for H3K27ac ChIP-Seq in 48 tumour samples following exclusion of the samples which generated less than **2000 peaks**. The sample G_DKP_013 which has been marked by an asterisk is an outlier and has also been excluded from further downstream analysis. (Primary samples and metastatic breast cancer samples are depicted by the green and orange bars, respectively).

6.4 H3K27ac ChIP-Seq coverage of enhancer and promoters in patient samples

In order to ascertain whether the number of patient tumour samples were adequate, in terms of genomic coverage, H3K27ac-enriched regions were classified into 23,976 gene-proximal promoters (1Kb upstream of the transcription start site (TSS)) and 326, 719 gene-distal enhancers (obtained by subtracting the calculated gene-proximal promoters from the H3K27ac-enriched regions). Figure 6.2A shows a permutation saturation plot (ACT [150]) of H3K27ac-enriched promoters and enhancers (in megabase (Mb) coverage) with patient sample number. With regard to promoters a maximum of 4 patient samples are required to reach saturation. The enhancers require 47 patient samples to capture 80% of the active regulatory

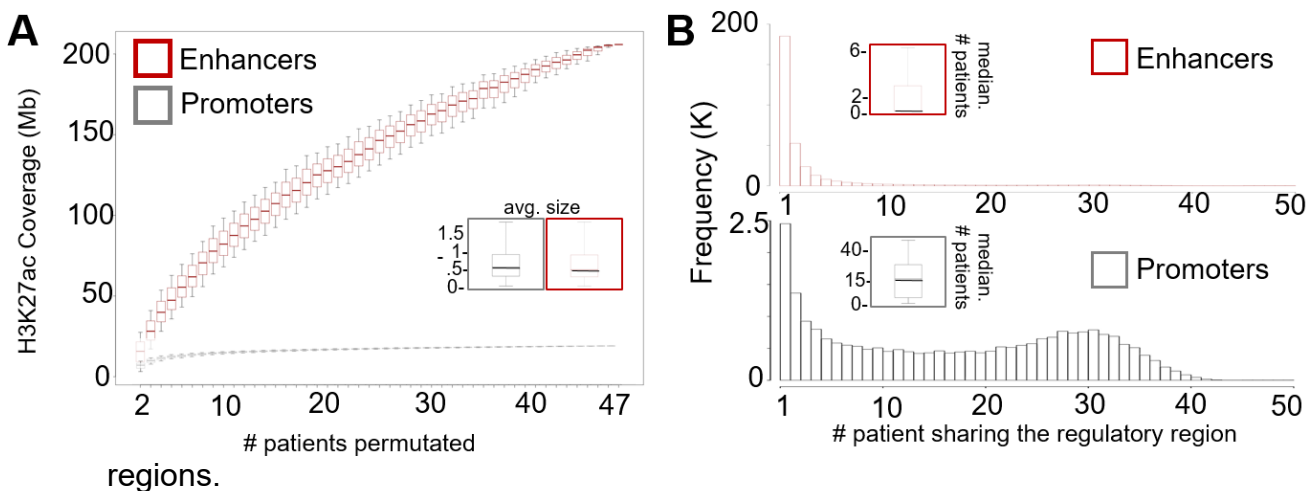


Figure 6.2 (A) Saturation plots for promoters and enhancers in N= 47 tumour samples generated by H3K27ac-enriched regions. Promoters are depicted by the red line and enhancers the grey line. The average peak size in promoters and enhancers is 517 and 463bp, respectively. **(B) Sharing of H3K27ac-enriched enhancers (red plot, top graph) and promoters (grey plot, bottom graph) across 47 patient samples.** The median (represented by the black lines in the small inset diagrams) number of patient samples sharing the same enhancer peak = 1 and the median number of patient samples sharing the same promoter peak = 15. **[Analysis performed by Giacomo Corleone and Darren K. Patten]** ^[209]

The graphs in figure 6.2B show the number of patients sharing regulatory regions (in terms of H3K27ac-enriched regions) in promoters (in grey) and enhancers (red). The first bar of each graph displays the unique peaks in any one sample. Promoters show less variability compared to enhancers with 3500 unique peaks vs 260,000 unique peaks in the first bar, respectively. The median number of patient samples sharing the same peak in promoters is 15 while in enhancers, this is 1. This data strongly suggests that the enhancers account for the majority of heterogeneity in ER α -positive BC.

6.5 Intratumoural and intertumoural heterogeneity in ER α -positive breast cancer

It has been widely recognised that intratumoural (i.e. cancer cells residing within the same tumour) heterogeneity arising from clonal tumour evolution exists in solid tumours [210-213]. Breast tumours are known to exhibit genetic and phenotypic heterogeneity at an intertumoural as well as intratumoural level [214, 215]. Due to the fact that a large number of mutations are passengers [216], it remains unclear how genetic heterogeneity translates to phenotypic heterogeneity. It has been shown that differing cell phenotypes within solid brain tumours can override genetic hierarchies [213]. The assessment of phenotypic heterogeneity can be inferred by using IHC to measure protein intensity within a particular group of cells or tissue. Studying global phenotypic heterogeneity using transcriptional data is unfeasible due to the fact that each cell within a tissue can potentially contribute random amounts of RNA [209].

Therefore, using the patient derived ChIP-Seq data, it was hypothesised that the latter may be used to approximate phenotypical heterogeneity. Hence, the H3K27ac ChIP-Seq can indicate whether a nucleosome is activated or not by using the K27ac signal presence as a proxy for the nucleosome to be on [209].

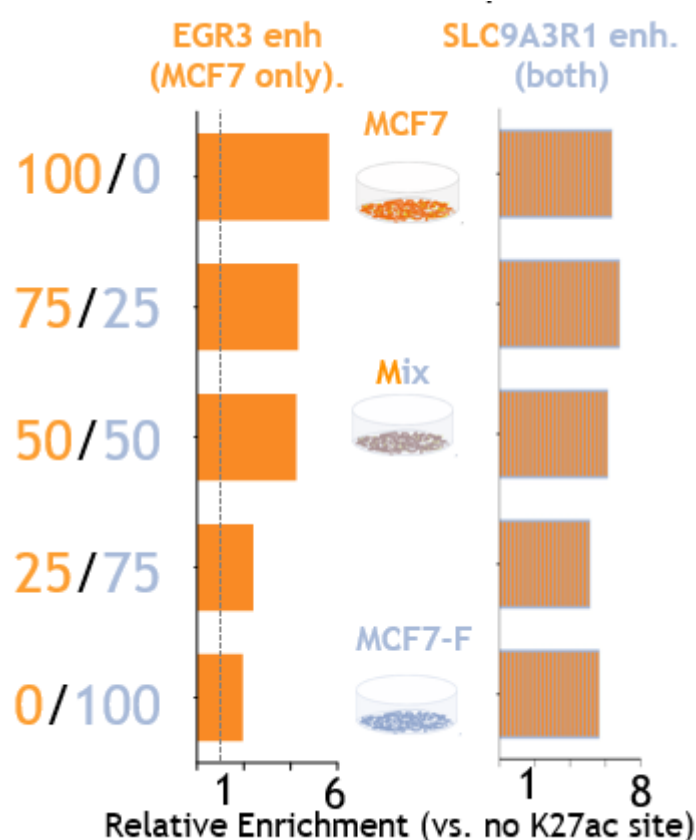


Figure 6.3 MCF7 H3K27ac ChIP-qPCR. A spiking experiment involving MCF7 and MCF7 fulvestrant resistant cells mixed at various known percentages in terms of cell number. The EGR3 enhancer (enh) is present in the MCF7 but not the MCF7 fulvestrant cells. The SLC9A3R1 enhancer is present in both cell lines. H3K27ac ChIP-qPCR demonstrates the varying enrichment seen at the EGR3 enhancer with increasing the number of MCF7 fulvestrant cells. [Experiment performed at least three times by Darren K. Patten; Figure produced in collaboration with Luca Magnani] [209]

To support the statement that the H3K27ac ChIP-Seq signal presence is a proxy for nucleosome activity, a spiking experiment using a known number of (cells with

similar genetic background [180]) MCF7 and MCF7 fulvestrant resistant cells was carried out. The H3K27ac ChIP-qPCR results shown in Figure 6.3 demonstrate clearly that the relative enrichment seen at the *EGR3* enhancer is dependent on the number of cells carrying that particular signal. The *SLC9A3R1* enhancer is present in both cells and the relative enrichment does not differ greatly with changes in the mixture of the MCF7 and MCF7 fulvestrant cell lines.

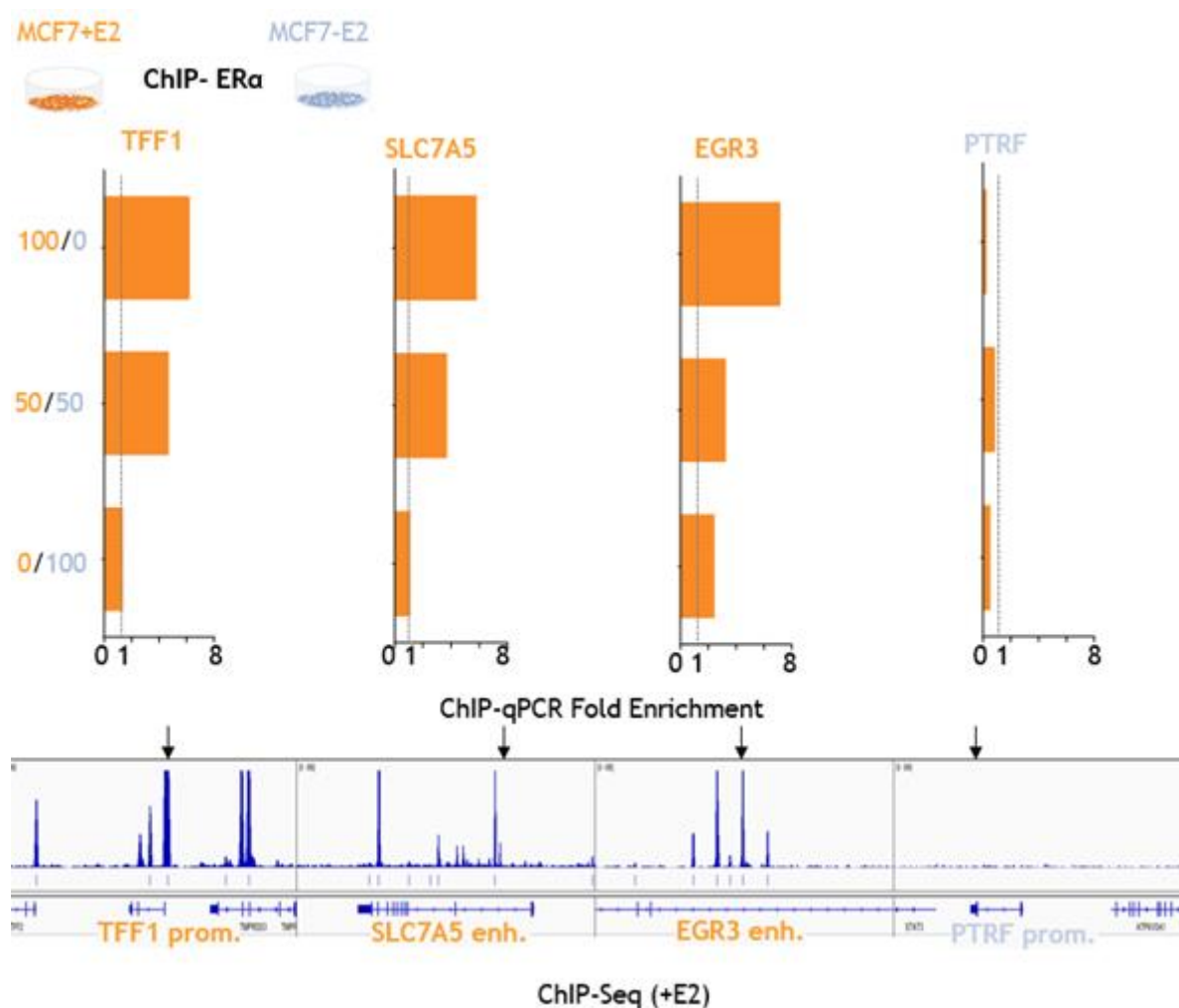


Figure 6.4 Spiking ERα ChIP experiment using MCF7 cells at different percentages +/- oestradiol stimulation. ChIP-qPCR was performed at known ERα target sites which can be seen by the corresponding MCF7 oestradiol stimulated ERα ChIP-Seq IGV tracks pictured at the bottom of the figure **[Experiment performed at least three times by Darren K. Patten; Figure produced in collaboration with Luca Magnani]**

[209]

The spiking experiment was repeated using ER α as the pulldown antibody and performing ChIP-qPCR in MCF7 cells stimulated with oestradiol for 45 minutes vs vehicle. ChIP-qPCR was performed at 3 known ER α target sites (*TFF1* promoter, *SLC75A* enhancer, *EGR3* enhancer) and 1 control region where ER α does not bind (*PTRF* promoter) despite stimulation with oestradiol. The ER α binding sites for *TFF1* enhancer, *SLC75A* enhancer, *EGR3* enhancer and the *PTRF* promoter can be visualised by the ER α ChIP-Seq IGV tracks in Figure 6.4. As the mixture of MCF7 cells (with or without stimulation of oestradiol) increases towards the cells not stimulated with oestradiol, the relative enrichment seen at these positive sites decreases. The relative enrichment at the *PTRF* enhancer remains less than 1, as expected, despite any change in the oestradiol- stimulated/non-stimulated MCF7 cells. The second spiking (ER α ChIP-qPCR; Figure 6.4) experiment corroborates with the first (H3K27ac ChIP-qPCR; Figure 6.3) such that relative enrichment for ER α binding correlates with the number of cells carrying that particular binding event.

Using a ranking strategy (performed in collaboration with Giacomo Corleone and Luca Magnani), each H3K27ac peak from the patient-derived dataset was assigned a ranking index based on the relative strength of the H3K27ac peak; 1 = strongest peak and 100 = weakest peak. This approach was adapted in order to normalise the varying H3K27ac signals across patient samples so that comparisons could be drawn from the peak strength of each regulatory region. Statistical binning was applied for each patient based on the ranking index (RI) of each H3K27ac peak. A sharing index (SI) was also created which indicated the number of patients sharing a particular regulatory region [209].

Figure 6.5 shows a linear regression analysis of RI vs SI from the patient-derived H3K27ac ChIP-Seq data. Signals obtained from low RIs (equating to the stronger

H3K27ac peaks) mark regulatory regions active across the majority of the cell population whilst the high RIs (corresponding to weaker H3K27ac peaks) are associated with heterogenous enhancer activity. The relationship between RI and SI was further investigated and correlation between the latter two indices was found to be high which indicates that clonal regulatory regions are relatively common between patients while subclonal regulatory elements are not. Therefore, the low RIs associated with high SIs were termed regulatory drivers (RD) with an $SI > 21$ and high RIs associated with low SIs were termed regulatory noise (RN) (with an $SI < 21$) possibly originating from technical noise or weak enhancers [209]. Furthermore, a small proportion of the promoters and enhancers did not follow this trend having extremely low RI despite possessing a high RI whilst being shared, as depicted by the dot-boxes.

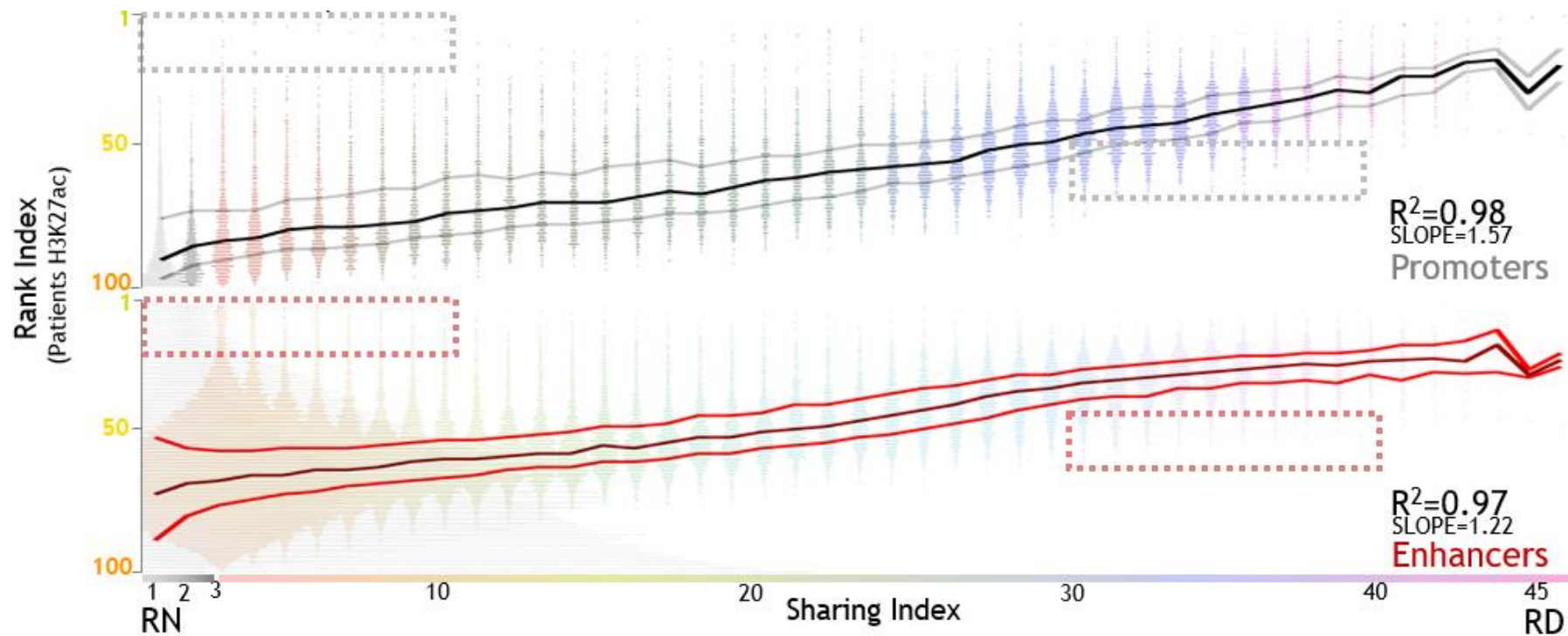


Figure 6.5 Linear regression analysis of ranking index (RI – obtained from H3K27ac data from patient samples) vs sharing index (which indicates the number of patients sharing a particular regulatory region). Each dot corresponds to the median RI for the regulatory regions amongst all patients. The interpolating lines represent the median RI value and interquartile ranges calculated for each SI bin. (RN = regulatory noise; RD = regulatory drivers; R^2 = the statistical measure of how close the data are to the fitted regression line). [Analysis performed by Giacomo Corleone. Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

The identified enhancers from the patient-derived samples were further mined by testing their relationship with DNA risk variants specifically associated with BC [217, 218]. The results showed that most of the BC risk variants were significantly contained (i.e. high degree of overlap) within the BC patient-derived enhancer database implying that a large amount of BC predisposition occurs through enhancer variants. When applying the same patient derived enhancer database to colorectal cancer risk variants, this association is not replicated implying that the enhancers identified above are specific to ER α positive BC.

Comparing the H3K27ac from the 47 patient samples to H3K27ac data generated from a panel of treatment-naïve and drug-resistant BC cell lines, it was observed that the H3K27ac loci from the cell lines forms a subset of the patient-derived enhancer panel which would imply that the cell lines are not activating new regulatory regions [209]. However, two distinct patterns were noticed, 1. For the RN regions, the patient-derived SI could not reliably predict the cell line RI; for example, clusters of enhancers characterised by a low SI and high RI in patients might have a low median RI in cell lines which would imply that a low-frequency clone may have expanded in culture and 2. For the RD regions, the patient-derived SI could reliably predict the RI. This data indicates that cell lines possess dominant phenotypic clones from the tumour of origin whereby certain epigenetic characteristics are further selected which can occur during changes in the microenvironment or alterations in cell culture constituents.

The results presented so far indicate that RD regions are activated in the majority of tumour cells driving disease progression but RN regions are frequently inactivated. This should imply that gross RNA output, from genes related with the identified regulatory regions from the patient-derived samples, should increase in parallel to

the RI/SI indices; this prediction would still apply for random transcriptional output from a single-cell perspective [219].

Using three METABRIC [220], TCGA [221] and Affymetrix [222] BC expression datasets, the relationship between SI and RNA expression was hence further studied (Figure 6.6). The regulatory regions were paired to potential targets taking into account CTCF boundaries [223]. The results of the analysis demonstrate that increases in mRNA levels are commensurate with increases in associated SI which further supports the fact that RDs drive RNA expression as the cell number increases. Furthermore, when mining transcriptional data from normal breast tissue (Figure 6.6; small graphs) and comparing this to our BC patient-derived dataset, the results did not follow a strong relationship insinuating that the patient-derived data has identified a subset of regulatory regions possessing signatures associated with malignant cell proliferation.

Summarising the results from this subsection, the data suggest that clonal enhancers mould the transcriptional landscape of ER α -positive BC with phenotypic heterogeneity being present in patients. Furthermore, the information obtained from our patient-derived dataset suggests that BC cell lines can be utilised to model the epigenome through contribution of dominant epigenomic signatures but develop differing levels of heterogeneity which is not related to the clinical material analysed in this project [209].

6.6 The transcriptional landscape of ER α -positive breast cancer

It is known that TFs interact with regulatory regions via protein-DNA interactions which is dictated by TF affinity for its respective DNA-binding motif [107, 224]. In order for TFs to bind DNA, the chromatin landscape must be accessible. In this subsection, the patient-derived H3K27ac was further studied, through computational imputation, to ascertain which TFs were binding to the potential chromatin-accessible regulatory regions.

The first step of this analysis was to calculate, from the patient-derived data (n=47), all potentially accessible regulatory regions through imputation (Figure 6.7). This involved comparing imputed chromatin-accessible regions with publicly available ER α binding data obtained from BC patients [85]. It was observed that RD enhancers tend to overlap with a high number of ER α sites (9% vs. 70%: SI 1 vs. SI 39) which was not seen in the case of promoters (15% vs. 6%: SI 1 vs. SI 39) (Figure 6.8). This data strongly suggests that ER α preferentially binds enhancer elements which has been reported [83, 225]. Although this data implies that the ER α is more likely to bind enhancers which are highly shared across patients, on closer inspection, the vast majority of ER α binding events tend to be exclusive to each patient (Figure 6.9). This would suggest that a high level of phenotypical heterogeneity exists in ER α -positive BC. This is supported by the fact that highly shared ER α binding sites have a higher number of individual ER α motifs in addition to DNA motifs for its associated pioneer factors [226] situated near and adjacent to the central binding site. This data suggests that the activity of enhancers is influenced by the regulatory information stored in the non-coding regions of DNA.

Moreover, the findings from this analysis demonstrates that only a small subset of solitary ER α binding may be functional whilst additional TFs may be required in order to contribute to BC disease progression [209].

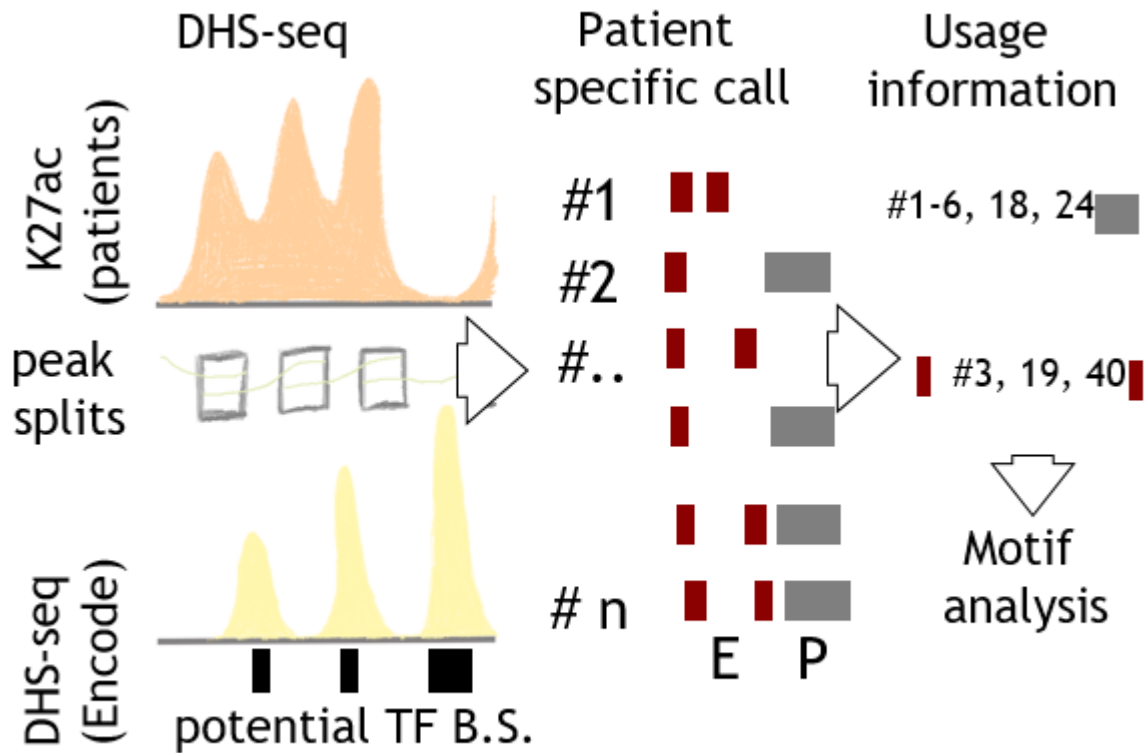


Figure 6.7 Schema of the bioinformatic framework for analysis of transcription factor (TF) binding sites within the patient-derived H3K27ac data. The H3K27ac MACS [142] peaks were split to identify nucleosome-free regions (subpeaks) which were then integrated with ENCODE-derived DNASE 1 hypersensitivity sites (DHS) for identification of valleys where TF binding occurs [108]. The imputed DHS were then given sharing index (SI) values based on the number of patients sharing this region. (B.S. = binding site; E = enhancer element; P = promoter element) [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

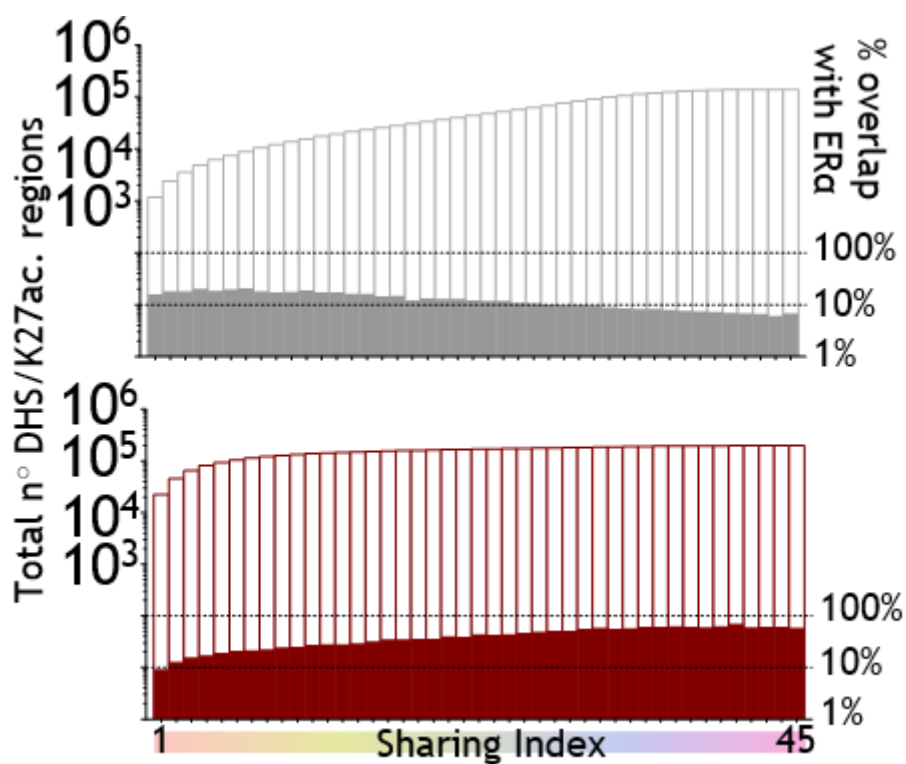


Figure 6.8 Overlap between imputed DHS/H3K27ac regions with in vivo-derived ER α binding sites. The Y-axis on the left corresponds to the imputed DHS regions and the Y-axis on the right indicated the percentage of overlap based on the total number of DHS in each sharing index (SI) bin. Promoters are represented by the grey bars and enhancers by the red bars [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

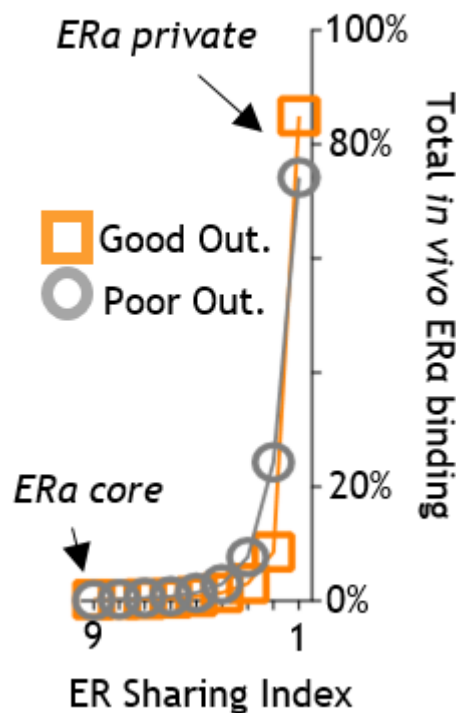


Figure 6.9 Distribution of in vivo-derived ER α binding sites compared to the number of patients in which these binding sites were observed. Approximately 70% -75% of ER α is exclusive to each patient where approximately 0% to 25% of ER α core binding is shared [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

TF motif analysis was conducted by applying stratification based on regulatory regions (i.e. promoters and enhancers) according to their corresponding SI. Unsupervised clustering was then performed identifying two biological groups (Figure 6.10A/B). The regulatory noise regions and regulatory driver regions showed clear separation confirming that the regions contain different regulatory information.

It was also observed that ER α -specific oestrogen responsive element (ERE) motifs were most residing in the RN enhancer regions, which is in agreement with ER α being involved in a vast majority of heterogenous low-affinity binding loci displayed by Figure 6.10A. The RN enhancer regions (Figure 6.10A) showed enrichment for a

larger set of TF motifs with a few exceptions to this pattern. Amongst all of the TF motif enrichments seen, the TF motif for YY1 was enriched exclusively in the RD enhancers (Figure 6.10A) but latter was also enriched in the RD promoters (Figure 6.10B). In summary, the data from this subsection shows that clonal enhancers contain distinct regulatory information, pointing to the TF YY1 as a novel, and potential, driver of ER α -positive BC [209].

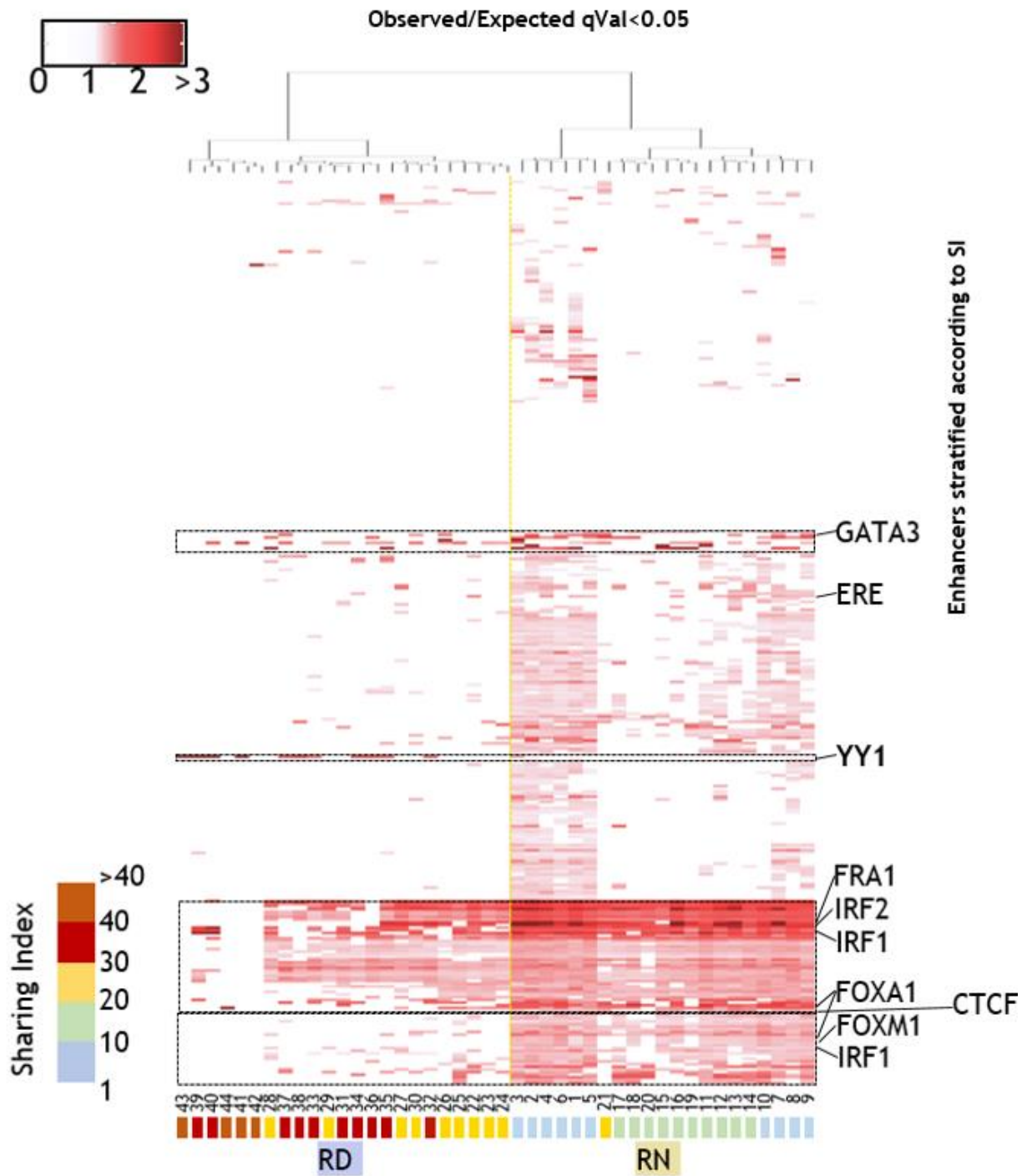


Figure 6.10A Transcription factor motif analysis applied to sharing index (SI) bins of enhancers followed by unsupervised clustering. Regulatory noise (RN) regions and regulatory driver (RD) regions cluster separately in two distinct biological groups. (The observed/expected ratio was set at a q-value < 0.05) **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

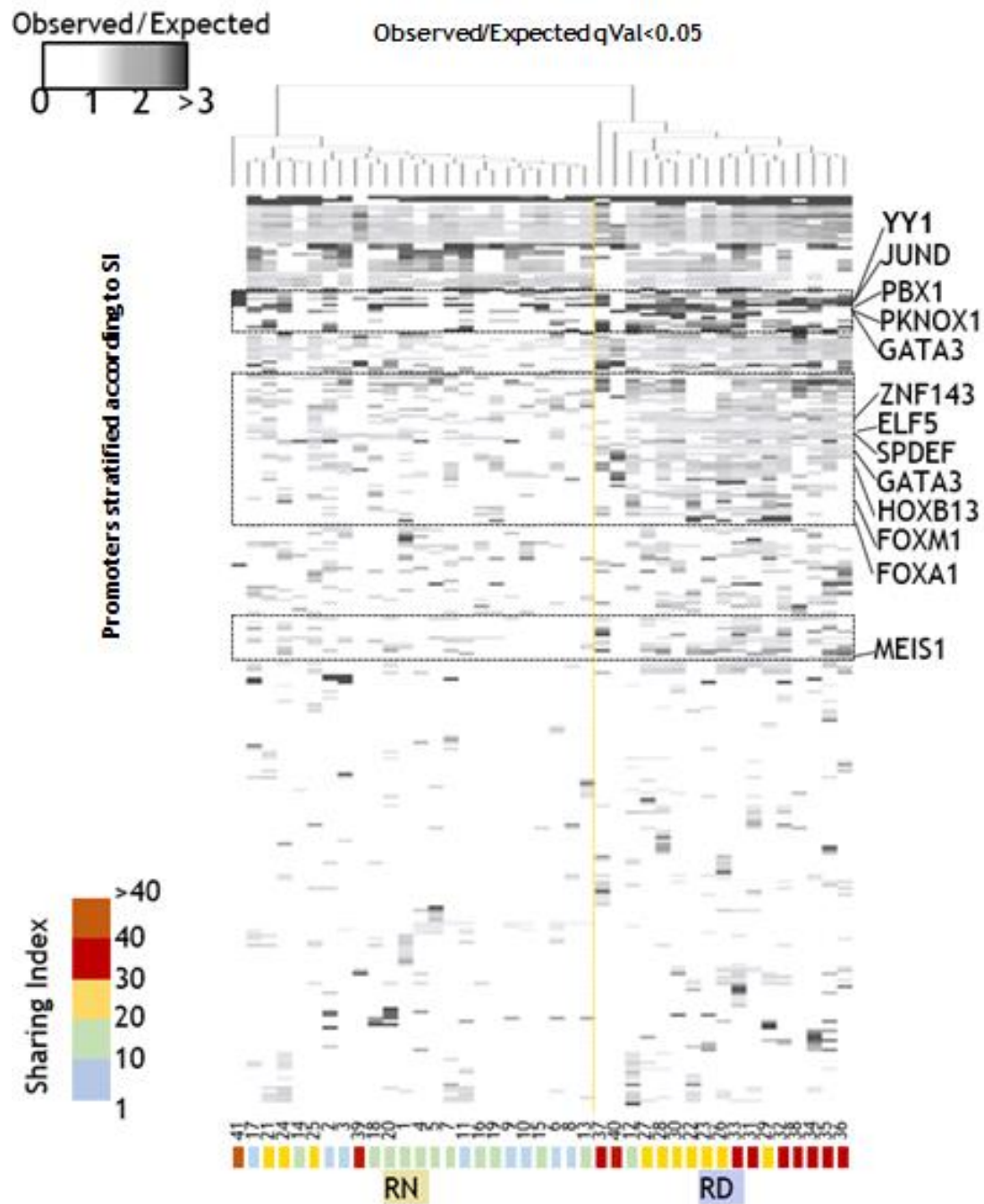


Figure 6.10B Transcription factor motif analysis applied to sharing index (SI) bins of promoters followed by unsupervised clustering. Regulatory noise (RN) regions and regulatory driver (RD) regions cluster separately in two distinct biological groups. (The observed/expected ratio was set at a q-value < 0.05) (Figure produced in collaboration with Giacomo Corleone and Luca Magnani) ^[209]

6.7 YY1 enhancer as a novel driver of ER α -positive breast cancer

YY1 is known to be a ubiquitously expressed TF which has been shown to act as an activator or repressor through binding with DNA, RNA and chromatin modifiers [227, 228]. The analysis described in subsection 6.6 implies that YY1 might act as a regulatory driver for clonal enhancers associated with a majority of ER α -positive BC patients. Hence, most patient tumour samples used for this analysis should contain a proportion of cells exhibiting YY1 activity. Therefore, applying the epigenomic mapping analysis used in this subsection, YY1 enhancers should possess very low-ranking index and corresponding high sharing index.

The first step was to identify genuine enhancers looping at the YY1 promoter with the aid of 3-dimensional (3D) chromatin maps [229] (Figure 6.11). Three potential enhancers were identified with very high SI (enhancer A, SI =41; enhancer B, SI = 33 and enhancer C, SI = 26). The enhancer A was then studied further due to its high SI and was found to be ranked as a clonal enhancer in most of the patient-derived BC tumour samples (Figure 6.12), implying that YY1 transcription would therefore be occurring in a high number of BC tumour cells. When assessing the activity of the YY1 enhancer A within normal tissues from the H3K27ac Epigenomic Roadmap ¹⁵⁵ the data is variable predicting the presence of varying degrees of YY1-positive subpopulations within normal tissues (Figure 6.13). It was also observed that tissues with low RI possess a higher number of YY1-positive cells whilst tissues with a higher RI contain a lower number of YY1-positive cells. (e.g., High RI tissues: Ovary, oesophagus, large intestine, heart muscle and low RI tissues: stomach and adipose tissues) ¹⁵⁹. This supports the use of epigenomic mapping to assess for

intra-sample heterogeneity and, moreover, the data presented in this subsection suggest that the activation of the YY1 enhancer represents a dominant phenotypic clone in ER α -positive BC.

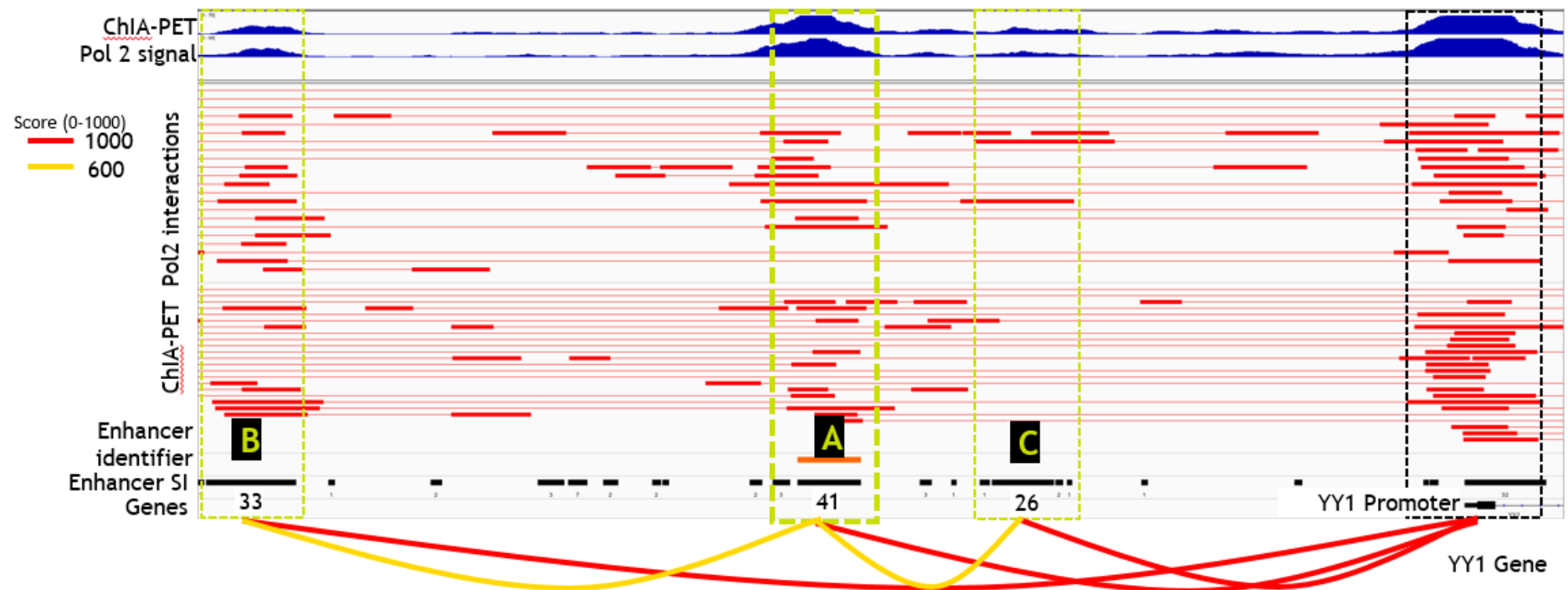


Figure 6.11 Chromatin interaction analysis. Chromatin interaction analysis by paired-end tag sequencing (CHIA-PET) and RNA polymerase 2 (Pol 2) signals (top) and with CHIA-PET Pol interactions at identified enhancers (A, B and C) for YY1 and their corresponding sharing index below. The YY1 promoter is pictured and the bottom left hand corner of the figure showing promoter-enhancer interactions with all three YY1 enhancers identified within the 47 breast cancer patient samples. Out of three YY1 enhancers which are all interacting with the YY1 promoter, enhancer A possesses the highest sharing index (41) and is shown to interact with enhancers B and C as well as the YY1 promoter. Visualisation using IGV. [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[143, 209]

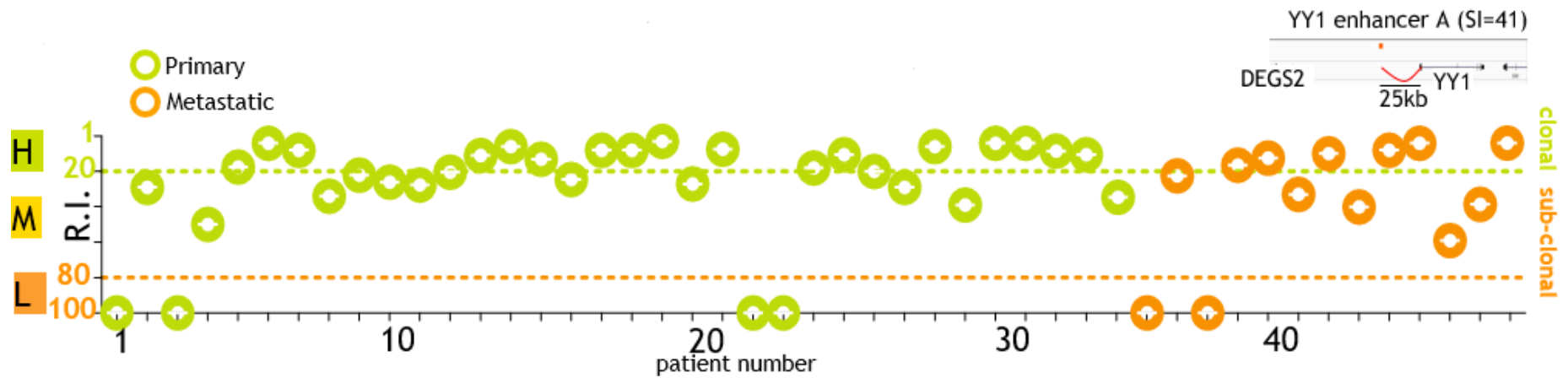


Figure 6.12 Graph of RI in YY1 determined within patient-derived samples (N=47). Each green circle represents a primary breast cancer sample whereas each orange circle represents a metastatic sample. The location of the YY1 enhancer A is pictured at the top right-hand corner of the figure. Most of the patient samples are found to possess low RI for the YY1 enhancer implying that this enhancer is clonal. [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

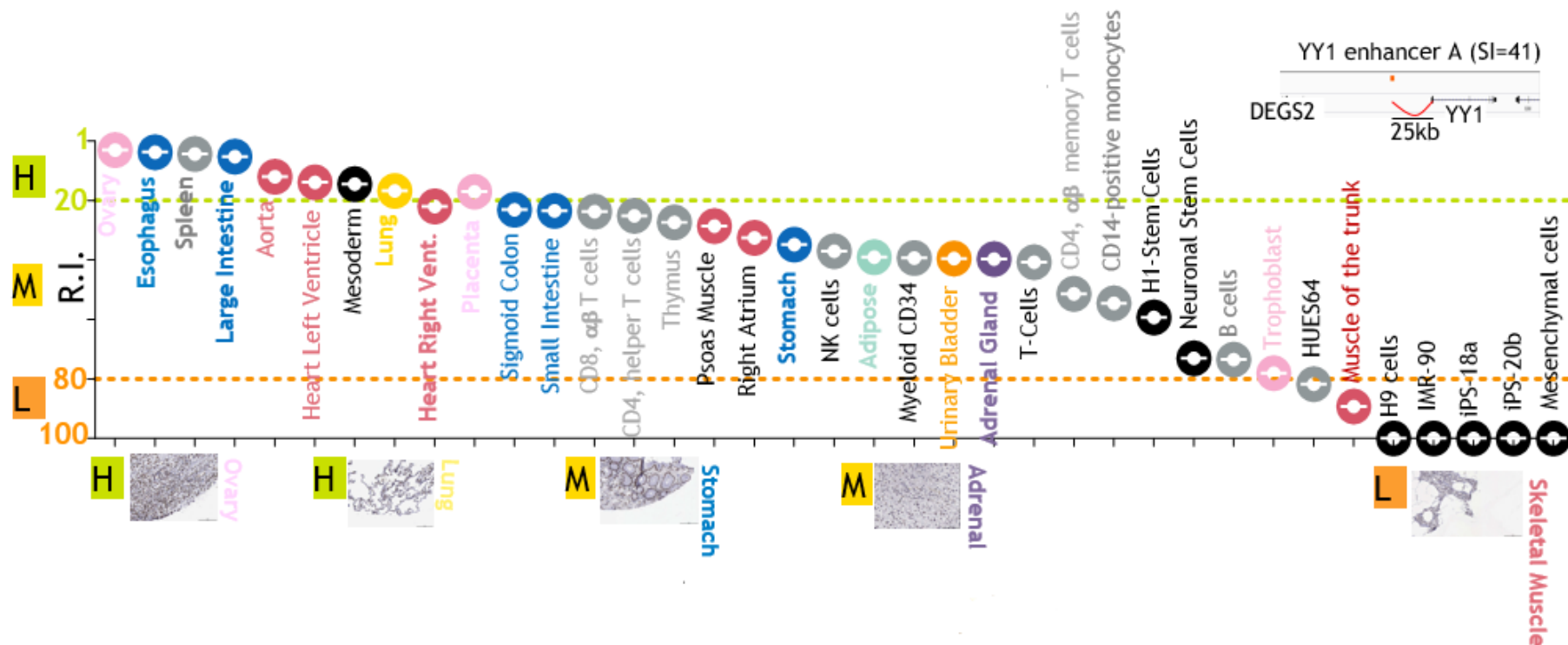


Figure 6.13 The clonal YY1-enhancer A (pictured in the top right-hand corner of the figure) using Epigenomic Roadmap 155 H3K27ac data in normal tissues classified according to high (H), medium (M) and low (L) ranking index (RI). Patients were classified as clonal YY1 (RI<20) or subclonal YY1 (20<RI<80) and YY1 negative (RI>80) Tissues with a low RI for the YY1-enhancer A show a higher number of YY1 positive cells which diminishes as the RI of the tissue increases. Tissues that have undergone immunohistochemistry staining for YY1 protein (represented by the smaller insets under the X-axis) include ovary (high RI), lung (high RI), stomach (medium RI), adrenal glands (medium RI) and skeletal muscle (low RI). **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

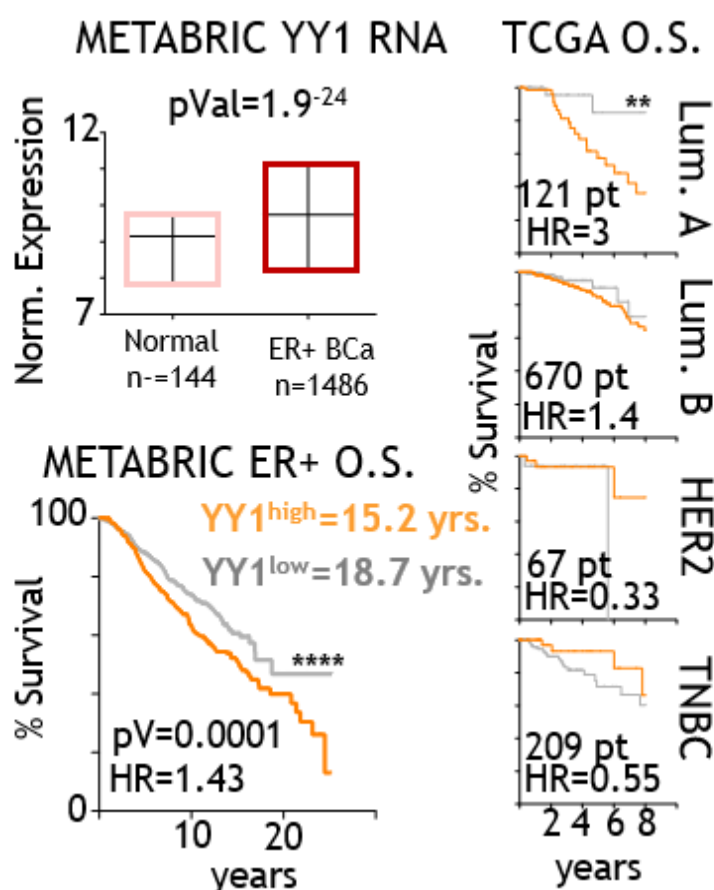


Figure 6.14 YY1 RNA expression in breast cancer datasets. The METABRIC [220] dataset of 1486 ER α -positive BC cases (top left-hand corner of figure) shows significant (p-value = 1.9×10^{-24}) increase in YY1 gene expression when compared to normal breast tissue. In addition, when investigating overall survival (O.S.) in the same dataset, Kaplan-Meier analysis demonstrates that cases expressing high levels of YY1 have a significant (p-value = 0.0001) reduction in O.S. compared to the ones with low expression of YY1 within the same dataset (High YY1: Low YY1: 15.2 years: 18.7 years). YY1 was also used to stratify patients within the TCGA [221] dataset (pictured on the right of the figure). Of the four molecular subtypes of BC, the Luminal A patients (N=121) who possess low expression of YY1 have a significant increase in O.S. compared to the high expressing patients who are 3 (hazard ratio (HR) = 3) times more likely to have a reduced O.S. Stratification did not occur for the other molecular subtypes of BC based on YY1 expression (i.e., Luminal B, HER2 positive and triple negative BC (TNBC)). (**** = pvalue of 0.0001, ** = p-value of 0.01) **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** [209]

The YY1 mRNA output is found to be high in many cancers, including all invasive subtypes of BC, when compared to their corresponding tissue of origin. When patients from the METABRIC dataset cohort are stratified based on ER α positivity, YY1 emerges as a significant prognostic biomarker associated with poor outcome (Figure 6.14). These findings were reproducible using the TCGA RNA-Seq dataset [221] but in Luminal A stratified BC patients which is a cohort associated with good prognosis (Figure 6.14). The data above supports the idea that a BC cell possessing the YY1 phenotypic clone increases during the development/progression of ER α -positive BC. In order to validate this, 300 diagnostic histological cases of Luminal A and Luminal B BC, where there was evidence of early relapse, were identified and subsequently underwent IHC staining for YY1 protein. This was performed (in collaboration with the Department of Histopathology, IEO, Milan, Italy) in order to ascertain the number of positive YY1 cells at diagnosis [209]. Although the IHC staining is still on-going, preliminary results demonstrate that the normal breast ducts are relatively deficient of YY1 positive cells with only a few cells exhibiting positive YY1 staining. When assessing adjacent ER α -positive BC cells, YY1 positive cells are highly prevalent (Figure 6.15).

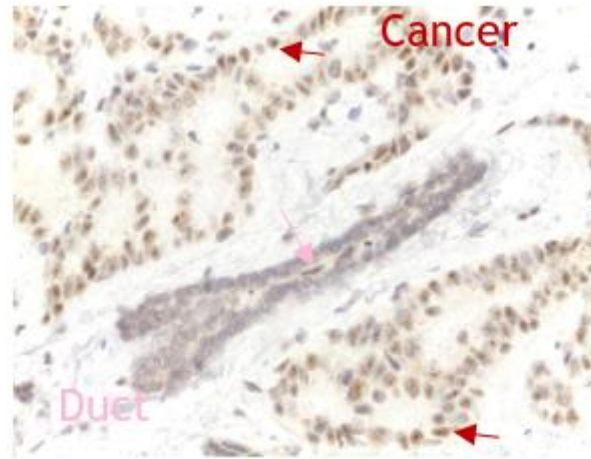


Figure 6.15 Immunohistochemistry staining of YY1 in breast tissue. Islands of invasive ER α -positive BC (highlighted by the red arrows) and a central normal breast duct (highlighted by the pink arrow) (magnification x10). YY1 shows intensified staining within the BC islands but is weakly positive at the normal ductal epithelium. (Preliminary IHC analysis performed by Dr. Giancarlo Pruneri in collaboration with the Department of Histopathology, IEO, Milan, Italy) ^[209]

The data in this subsection demonstrates that YY1 positive cells exist in high number in ER α -positive BC and might be used as a prognosticator in a subset of ER α -positive patients (i.e. Luminal A subtype of BC). In summary, the data shows that YY1 is a novel driver of ER α -positive BC.

6.8 The role of YY1 in ER α binding in breast cancer

The enrichment observed, for the motif of the TF YY1, was significantly high within the patient-derived H3K27ac enhancer data. Furthermore, the YY1 motif was found to be enriched within the most highly shared and strong enhancers. In order to further understand the role of YY1 in ER α -positive BC, YY1 ChIP-Seq was performed in MCF7 cells; MCF7 cells were grown in phenol-free DMEM with 10% DCFCS for 48 hours in four 15cm² dishes. Oestradiol was added to two of the dishes whilst the remaining two dishes were treated with (control) vehicle (molecular grade ethanol) for 45 minutes. It has been shown that maximum genome-wide chromatin binding occurs at 45 minutes [83]. The cells were then processed for YY1 ChIP-Seq as described in Section 2.4 and 2.5.

YY1 occupies a small set of enhancers and promoters situated near housekeeping genes within the MCF7 cells stimulated with vehicle of which the interactions of YY1 with its respective motif exhibit high affinity. The oestradiol treated group, however, showed a 23 -fold increase in the number of binding sites to gene characterising ER α -BC and genes involved in epigenetic modifications (Figure 6.16).

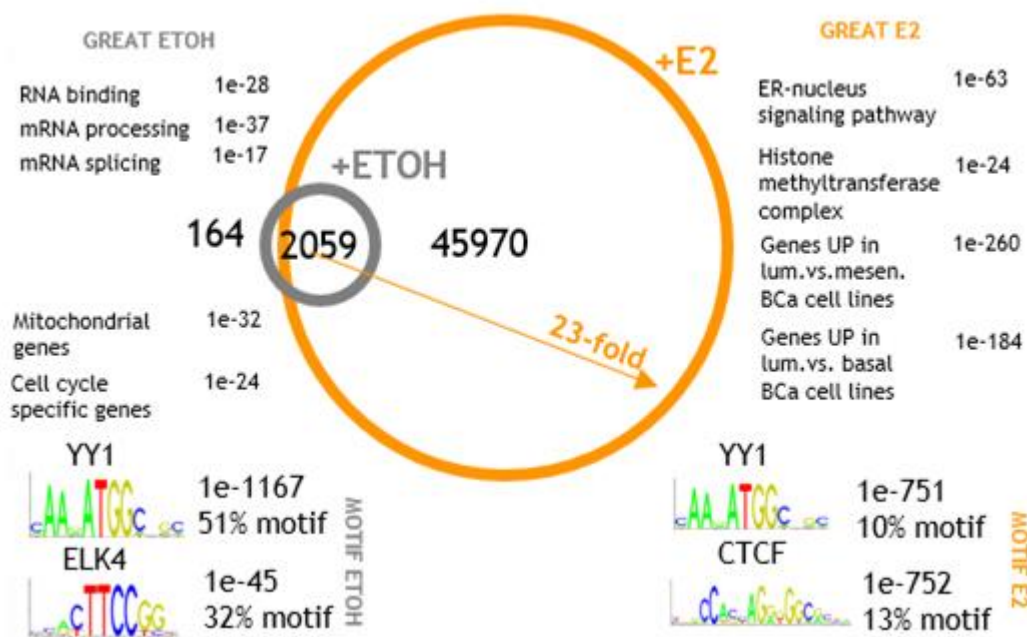


Figure 6.16 YY1 ChIP-Seq data in ER α -positive MCF7 cells stimulated with (control) vehicle (ETOH = ethanol) vs oestradiol (E2) for 45 minutes. The top left-hand side of the figure shows the Genomic Regions Enrichment of Annotations Tool (GREAT) function predictions with their corresponding p-values for the control group. The bottom left hand corner shows motif analysis with p-values for the control group. The central Venn diagram shows the total number of binding sites identified within both control and oestradiol groups and the number of shared binding sites. On the right top hand corner are pictured the GREAT function predictions for the oestradiol group with relevant p-values. The bottom right hand corner displays the motif enrichment analysis in the E2 group with corresponding significant p-values. (Experiment performed three times but only two replicates were taken for ChIP-Seq) [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

Only 10% of all the binding events of YY1 show high affinity for the YY1 motif which suggests that YY1 may be interacting with other TFs or binding directly to the modified nucleosomes (Figure 6.16). Further comparative analyses demonstrate that YY1 binding involves almost all active genes and it is strongly associated with H3K27ac marks and a high degree of chromatin accessibility (Figure 6.17).

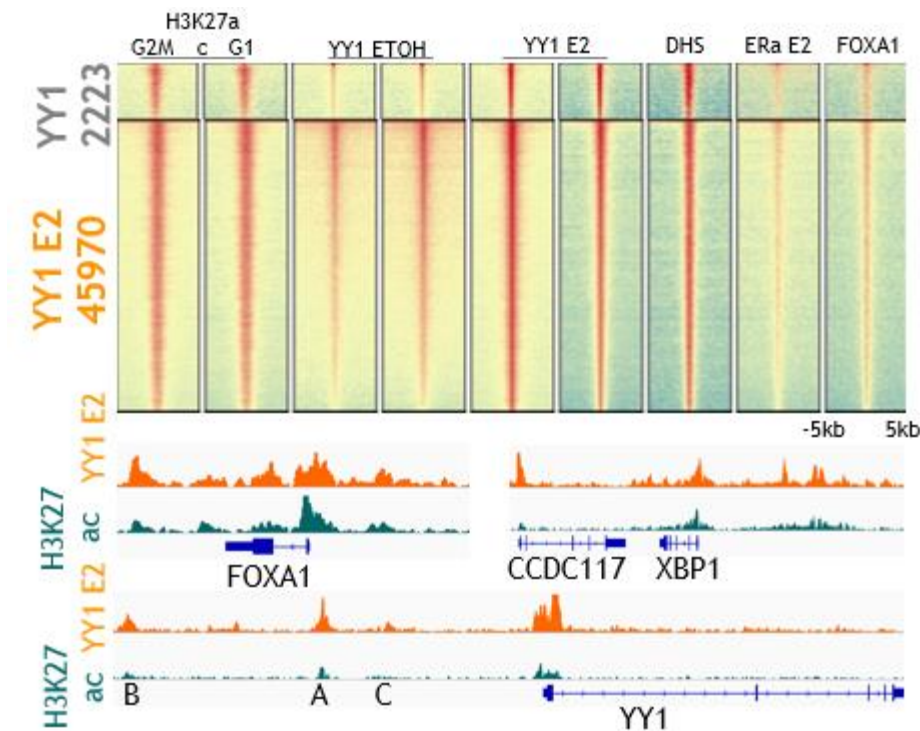


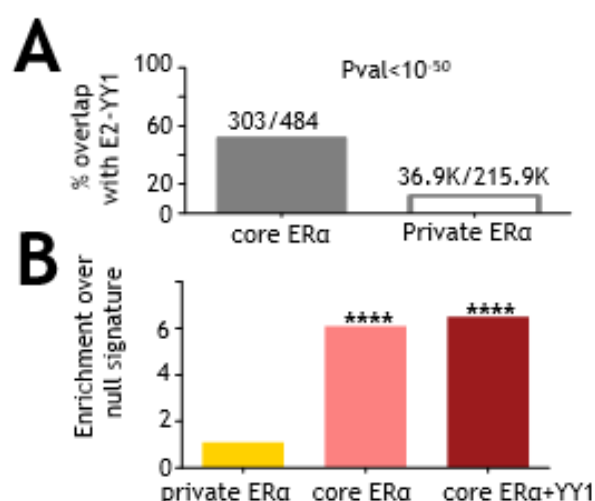
Figure 6.17 Heatmaps of YY1 overlap with H3K27ac enriched regions associated with active regulatory regions. YY1 binding marginally overlaps with ERα and its pioneer factor FOXA1. YY1 binding is also seen at ERα target genes enhancer of *XBP1* which is shown to be an active regulatory region by the enriched H3K27ac track at the *XBP1* enhancer below. The bottom last two IGV tracks display YY1 (E2 stimulated) binding with the most bottom track showing H3K27ac enriched regions at the three identified enhancers for YY1 (enhancers A/B/C – please refer to section 6.7) and at the YY1 promoter. **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

The number of YY1 binding events, in the MCF7 oestradiol stimulated cells, only marginally overlaps with ERα and its pioneer factor (FOXA1) (Figures 6.17) which would imply that the majority of YY1 recruitment is independent of ERα (Figure 6.18). However, the ERα-YY1 increases when taking into account the atypical ERα core binding (Figure 6.19A) sites, which are common to 75% of patients and BC cell lines [85]. Collectively, the data from this subsection demonstrates that YY1 may act as a stabiliser for ERα binding at a small subset of transcriptionally functional enhancers

common to most BC patients [228]. To support this statement, on inspection of the TCGA [221] data for the cohort representing the Luminal subtype of BC, the genes defining the Luminal A subtype are strongly enriched for ER α core binding with YY1 and not the private ER α binding with YY1 (Figure 6.19B).



Figure 6.18 Showing the overlap between ER α (Era), oestradiol stimulated (E2), ChIP-Seq binding events vs. YY1 (E2 stimulated) ChIP-Seq binding events in MCF7 cells. The overlap between YY1 binding and ER α



is marginal. [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

Figure 6.19 A) In vivo core ER α binding sites are strongly enriched for YY1 binding whereas this is not the case for the private ER α binding sites (p -value $< 10^{-50}$). **B)** Genes that classify luminal BC from the TCGA dataset [221]. The results show that within the core ER α genes ER α -YY1 binding sites are strongly enriched. [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

The next step was to further investigate the role of YY1 in ER α -positive BC cell line models. siRNA for YY1 was transfected into MCF7 cells as described in Section 2.9. Western blot analysis shows significant reduction in YY1 protein with both siRNAs compared to control with or without oestradiol stimulation (6 hours) (Figure 6.20). To further understand whether YY1 shares co-factor attributes with ER α , MCF7 cells were co transfected with siRNAs for YY1 along with an ERE tagged luciferase reporter (obtained from Professor Simak Ali's group, Department of Surgery and Cancer, Imperial College London, U.K.) as described in Section 2.10. The results demonstrate that YY1 depletion using both siRNAs is sufficient to reduce the activity of the ERE-luciferase reporter in MCF7 cells stimulated with vehicle (ethanol) and oestradiol stimulation (results were normalised with the control renilla-luciferase (Figure 6.21).

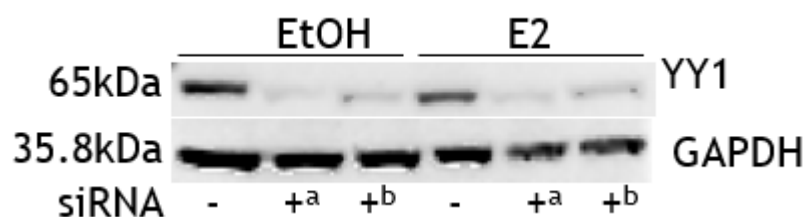


Figure 6.20 Western blot analysis for YY1 depletion using two types of siRNA (a/b) in MCF7 cells stimulated with vehicle vs oestradiol for 6 hours. Compared to controls (-) YY1 protein is significantly reduced > 80% in both control and oestradiol treated groups. siRNA “-” represents the MCF7 cells transfected with silencer negative control.

MCF7 cells were then grown in DMEM supplemented with 10%DCFCS for 48 hours after which, cells were transfected with either siControlRNA or one of the two types of siRNA (a/b) for YY1. Cells were divided into two groups; ethanol (control vehicle) vs. oestradiol. Proliferation assays were set up as described in Section 2.11 and measurements were taken after seven days (Figure 6.22A).

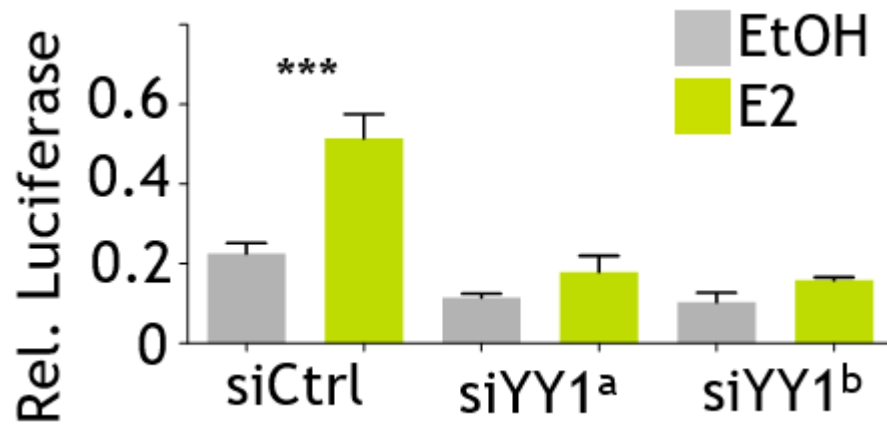


Figure 6.21 Luciferase tagged with oestrogen responsive element (ERE) promoter reporter assay performed in MCF7 cells stimulated with control vehicle ethanol (EtOH, in grey) vs. oestradiol (E2, in green) for 24 and transfected with silencer negative control (siCtrl) or siRNA “a” or siRNA “b”. Relative luciferase activity was normalised with the control renilla-luciferase activity. (Experiment performed 3 times, *** = p-value = 0.001, Statistical analysis performed = one-way ANOVA). [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

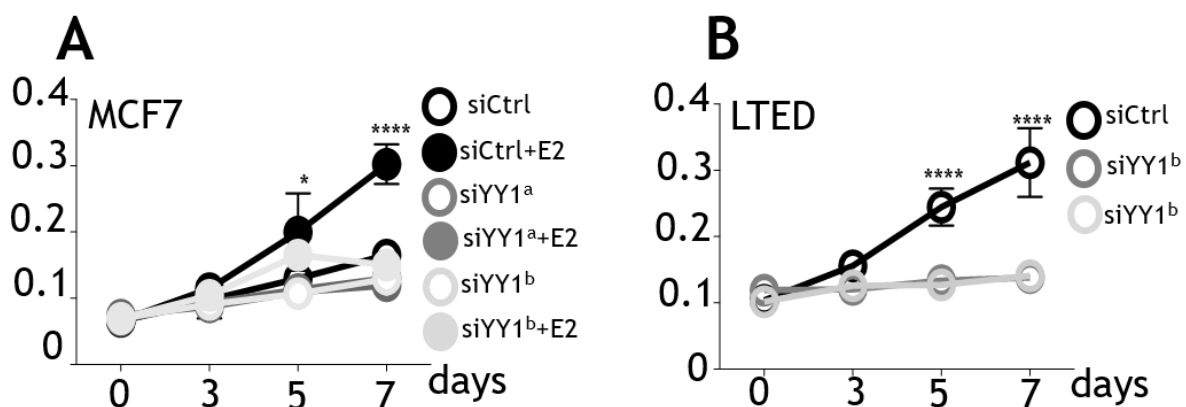


Figure 6.22 A) Proliferation assay of MCF7 cells grown in vehicle control vs. oestradiol (E2) with transfection of silencer control (siCtrl), siRNA (a) and siRNA (b). Compared to the control groups, both siRNAs abrogate growth in MCF7 cells. B) Proliferation assay in LTED cells. Abrogation of growth is exhibited in the siRNA group compared to control (SiCtrl). (Experiment performed 3 times, Statistical analysis performed = ne-way ANOVA, * = p-value = 0.05, **** = p-value = 0.0001). [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

LTED cells, (MCF7-derived BC cell line model which grow in the absence of oestradiol stimulation via constitutive activation of ER α signalling [180, 219]), were seeded in DMEM supplemented with 10% DCFCS and transfected with one of the two working siRNAs for YY1 (siRNA “b”). A proliferation assay was also set up as described in Section 2.11 and measurements of growth taken after 7 days (Figure 6.22B).

The proliferation assay involving MCF7 cells (Figure 6.22A) clearly shows abrogation of growth (p-value = 0.0001) using both siRNAs to YY1 compared to siCtrl in both vehicle control (ethanol) and oestradiol stimulated groups. The same results were obtained with the LTED cells, whereby growth of the cells was significantly halted (p-value = 0.0001) in the siRNA group compared to control (siCtrl) (Figure 6.22B).

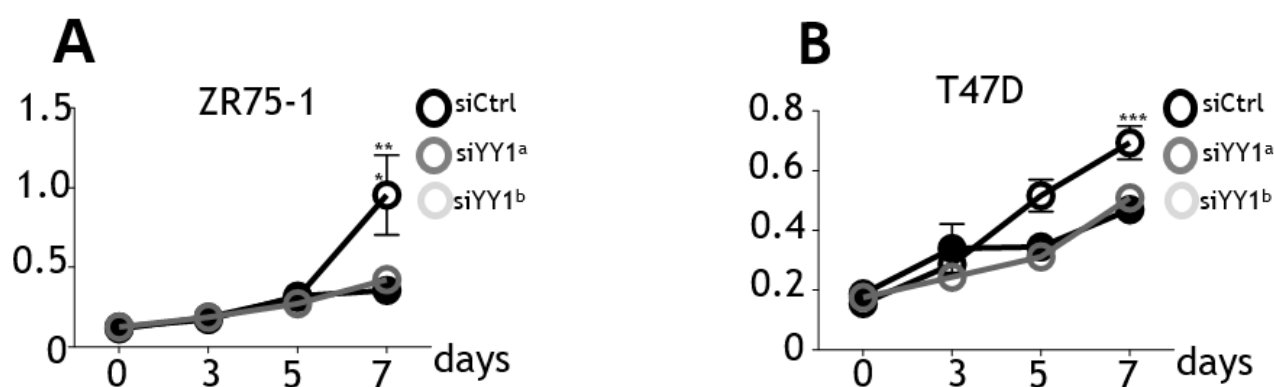


Figure 6.23 A) Proliferation assay of ZR75-1 cells transfected with silencer control (siCtrl), siRNA (a) and siRNA (b). Compared to the control groups, both siRNAs abrogate growth ZR75-1 cells **B) Proliferation assay in T47D cells.** T47D cell line exhibits abrogation of growth in both siRNA groups compared to control (siCtrl). (Experiment performed three times; Statistical analysis performed = one-way ANOVA; ** = p-value = 0.01, *** = p-value = 0.001). **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

To further test the effects of YY1 depletion in ER α -positive BC, proliferation assays were setup in using two different ER α -positive BC cell lines (i.e., ZR75-1 and T47D) and growth measured over the course of 7 days comparing silencer control to siRNAs for YY1. ZR75-1 cells exhibit abrogation of growth (p-value = 0.01) at 7 days in both siRNA groups compared to the control group (siCtrl). For the T47D cells, there is still growth of the cells in both siRNA groups compared to control but the cells in both siRNA groups are still significantly (p-value = 0.001) reduced in number compared to the control group after 7 days (Figure 6.23A/B).

MCF7 cells were then transfected with either of the two siRNAs for YY1 or the silencer control RNA (siCtrl) and cultured in the presence of either oestradiol or ethanol for 6 hours. After 6 hours, total RNA was extracted per treatment group and cDNA was synthesised, as described in Sections 2.12 and 2.13. Real time-qPCR (RT-qPCR) reactions were carried out at known ER α target genes. If YY1 acts as a co-factor for ER α at a subset of binding sites, depletion of YY1 should therefore disrupt the transcriptional output at ER α target BC genes.

Figure 6.24 shows the mRNA expression at three ER α BC target genes (*TFF1*, *CA12* and *GREB1*). Compared to the siCtrl groups, mRNA levels of all three genes are significantly reduced, using both siRNAs, in both ethanol and oestradiol treated groups.

Collectively, the data shows that 1. YY1 is involved with ER α at a subset of core ER α binding sites 2. YY1 collaborates with ER α to establish the luminal BC transcriptional profile, and 3. YY1 depletion stunts growth in ER α -positive BC cell lines.

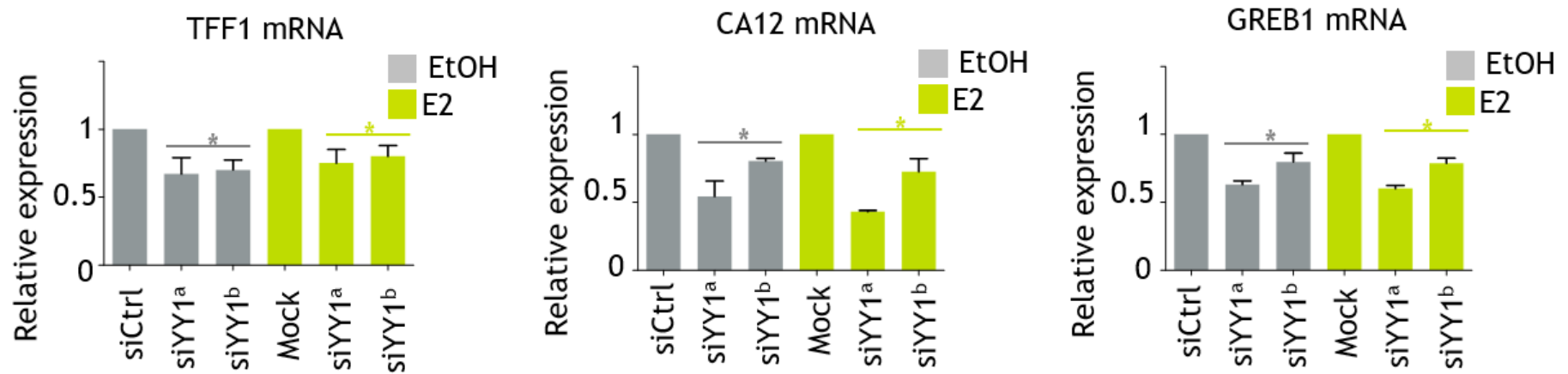


Figure 6.24 Gene expression data in MCF7 cells transfected with siRNA for YY1 and silencer control (siCtrl). Cells were then divided into two groups, and one group-arm was treated with ethanol (EtOH, in grey) and the other arm, oestradiol (E2, in green) for 6 hours. mRNA relative expression is shown for *TFF1*, *CA12* and *GREB1* ER α target genes. (Experiments performed three times by Darren K. Patten and Edina Erdos; Statistical analysis = one-way ANOVA; * = p-value = 0.05). **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

With reference to Figure 6.22B, the LTED cells exhibited growth arrest upon YY1 depletion. This would imply that YY1 serves a vital function in the progression of oestrogen independent BC progression. To further investigate this, ChIP-Seq for YY1 was conducted in MCF7 (stimulated with oestradiol) and LTED cells revealing 59, 676 binding sites for YY1 in LTED and 49, 645 binding sites in the MCF7 cells (Figure 6.25). Of note, 38, 981 YY1 binding sites are shared between the MCF7 and LTED cell lines implying that there is a large subset of YY1 binding sites which may be very important in allowing these BC cell lines to proliferate. This would suggest that it is these 38, 981 binding sites which might be affected upon YY1 depletion.



Figure 6.25 ChIP-Seq carried out in MCF7 (with oestradiol stimulation, E2) and LTED cells for YY1. The results show that 38, 981 binding sites are shared between MCF7 and LTED cells. **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

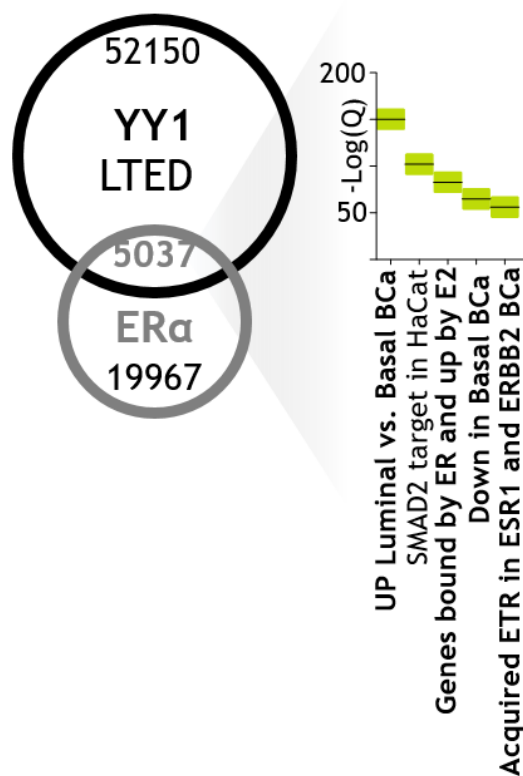


Figure 6.26 ChIP-Seq conducted in LTED cells for YY1 overlapped with ERα ChIP-Seq. Genes upregulated by the ERα-YY1 bound enhancers (5037 binding sites) are displayed to the right of the figure in Log fold. Genes involved in the Luminal phenotype of BC are strongly increased as those involved in endocrine resistance in BC. (Experiment performed by Darren k. Patten) **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

The LTED cells are the model of endocrine resistant breast cancer because they proliferate despite the absence of peripheral effects of oestradiol. Looking at the potential role of YY1 in contributing to BC progression and endocrine therapy resistance, it was observed that the ERα-YY1 bind a different set of enhancers compared to MCF7 cells (3598 out of 5037 binding sites). The enhancers involved in ERα-YY1 binding in LTED cells are associated with genes involved with acquired endocrine therapy resistance (Figure 6.26).

To further investigate the effects of YY1 involvement in the context of endocrine resistance in ER α BC, a set of oestrogen-induced genes that are not antagonized by Tamoxifen treatment were examined in relation to YY1 [230]. Consequently, a very strong enrichment for ER α -YY1 binding near these genes was observed. ER α -YY1 was found close *CXXC5* and *SLC9A3R1* genes which are also ranked the first and second strongest oestrogen-inducible genes that cannot be antagonised by Tamoxifen, respectively (Figure 6.27). On assessment of the *in vivo* sharing indices generated from the patient-derived samples, the *SLC9A3R1* enhancer possesses a SI of 34 making the latter a highly-shared region [209].

The data presented in this section demonstrates that 1. YY1 has a key role in orchestrating disease progression and also endocrine therapy resistance in ER α BC and, 2. YY1- ER α binding shows high an *in vivo* SI for the *SLC9A3R1* region which is ranked as the second most strongly oestrogen-induced gene that cannot be antagonised by Tamoxifen, hence warranting further investigation.

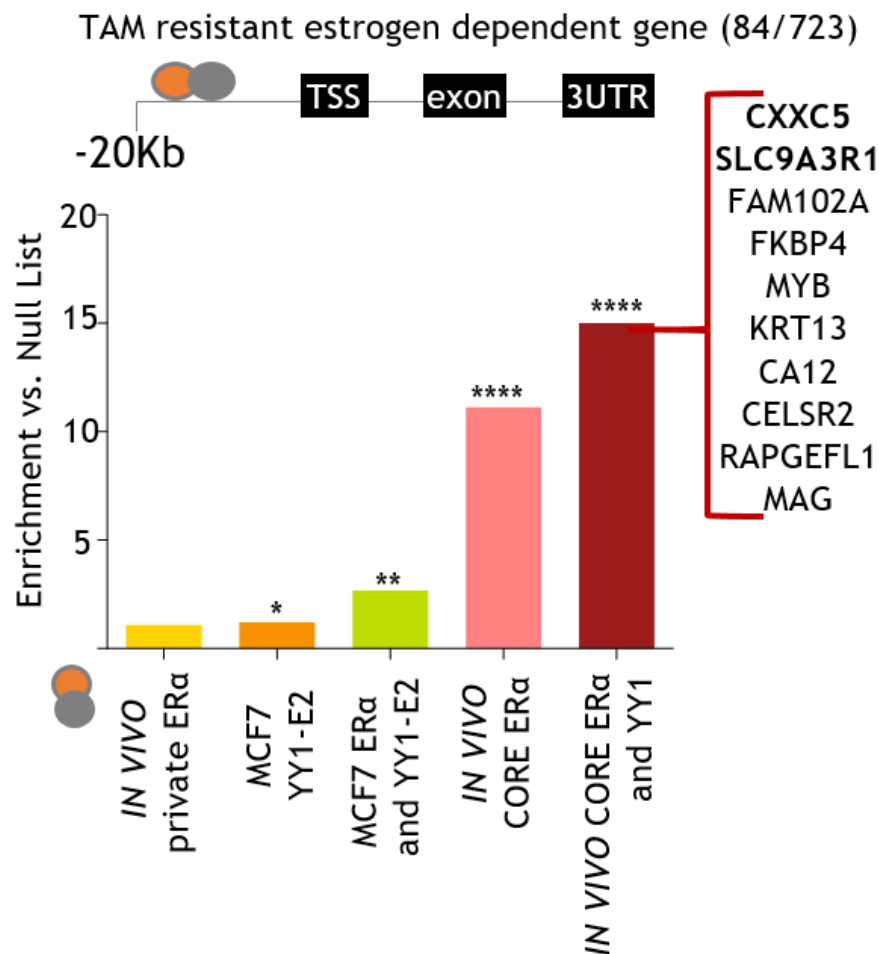


Figure 6.27 The relative enrichment of ERα-YY1 bound enhancers near oestrogen responsive genes that do not respond to Tamoxifen therapy. The set of enhancers that are bound by ERα -YY1 are enriched 15-fold near genes which are not antagonised by Tamoxifen treatment (red). *In vivo* private enhancers where private ERα binding occurs shows the least enrichment (yellow), The set of enhancers governed by YY1-oestradiol-induced binding in MCF7 cells shows a significant 1-fold enrichment (orange). ERα-YY1 binding in MCF7 (green) under oestradiol stimulation demonstrates a significant 2.5-fold enrichment. *In vivo* core binding (with ERα) set of enhancers are strongly enriched at 11-fold which is highly significant (pink). (Analysis performed by Luca Magnani; Statistical analysis = one-way ANOVA; * = p-value = 0.05, ** = p-value = 0.01, **** = p-value = 0.0001)[Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

Chapter 7 Results: *SLC9A3R1* and phenotypic mapping

7.1 The role of *SLC9A3R1* in ER α -positive breast cancer

On closer inspection of the role of *SLC9A3R1*, it has been documented that this gene cannot be suppressed with Tamoxifen [230] which would suggest its corresponding enhancer(s), which is/are bound by ER α remain(s) functional. Examining BC cell lines with their endocrine resistant counterparts, it is observed that the *SLC9A3R1* remains transcriptionally active in cell lines (Figure 7.1) where ER α is expressed whilst depleting ER α renders *SLC9A3R1* transcription inactive (Figure 7.2). In addition, *SLC9A3R1* is not antagonised by either oestrogen deprivation or Raloxifene (a SERM) (Figure 7.3) and is insensitive to neo-adjuvant aromatase inhibitor therapy in ER α -positive BC patients (Figure 7.4).

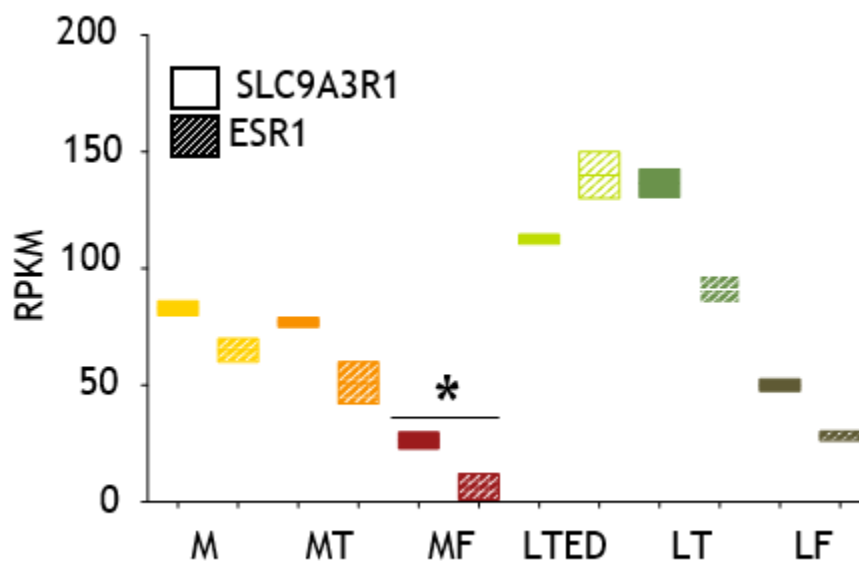


Figure 7.1 *SLC9A3R1* and ESR1 expression in MCF7 cells and MCF7 variants. MCF7(M; yellow), MCF7 Tamoxifen resistant (MT; orange), MCF7 fulvestrant resistant (MF; red), LTED (light green), LTED Tamoxifen resistant (LT; dark green) and LTED fulvestrant resistant (LF; grey) BC cell lines. *SLC9A3R1* expression remains present in the presence of ESR1 expression. In the fulvestrant resistant cell lines, where ESR1 is depleted by the actions of fulvestrant, *SLC9A3R1* expression remains at its lowest compared to the other variant cell lines displayed. (Analysis performed by Luca Magnani; Statistical test = one-way ANOVA; * = p-value = 0.05). **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

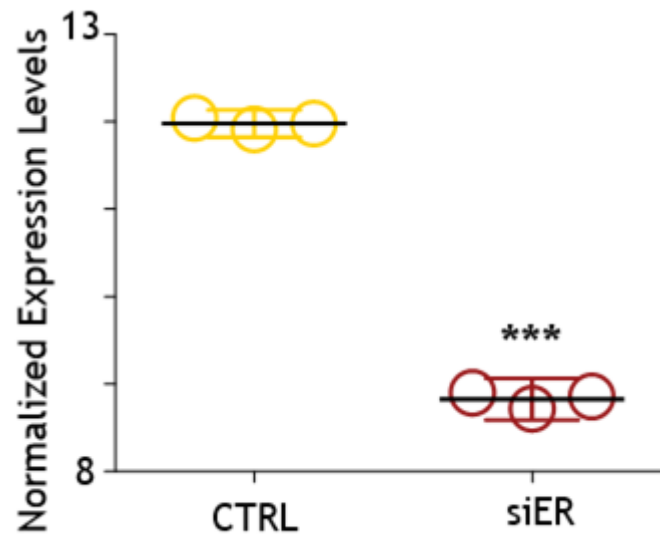


Figure 7.2 Expression data from MCF7 cell lines after silencing of the oestrogen receptor. The graph shows normalised expression levels of *SLC9A3R1* in control (CTRL) MCF7 vs. MCF7 with ESR1 depletion (siER) using siRNA. Compared to the control group, there is a significant 3-fold reduction in *SLC9A3R1* expression (Analysis performed by Luca Magnani; Statistical analysis = Paired T-test; *** = p-value=0.001.

[Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

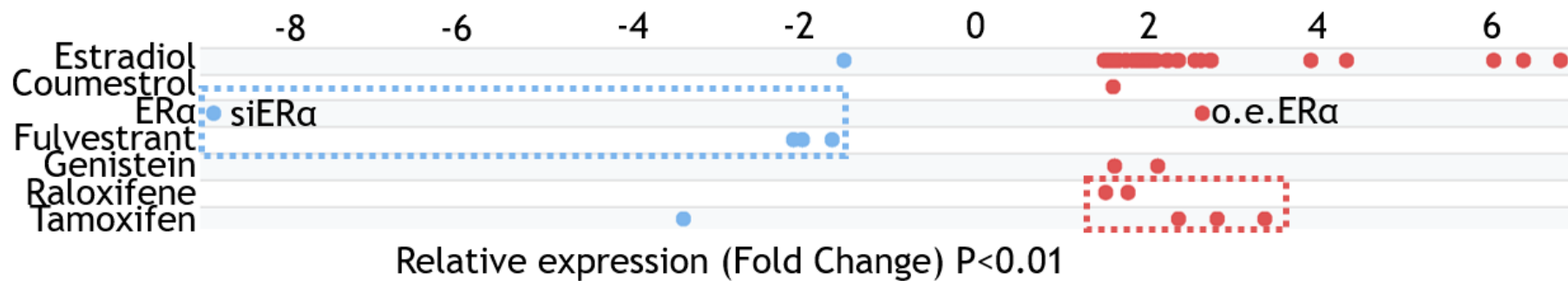


Figure 7.3 Meta-analytical study assessing the transcriptional response of *SLC9A3R1* in response to oestrogens (oestradiol), Coumestrol (phyto-oestrogen), ERα (and its depletion, siERα), Fulvestrant (oestrogen receptor downregulator), Genistein (phyto-oestrogen), Raloxifene and Tamoxifen (SERMs). When ERα is silenced (siERα) there is a > 10-fold reduction in the expression of *SLC9A3R1* compared to native ERα. Fulvestrant is also capable of reducing the expression of *SLC9A3R1* by 4-fold. Phyto-oestrogens such as coumestrol and Genistein increase *SLC9A3R1* expression by 2-fold but less than native ERα. Raloxifene and Tamoxifen both cannot decrease the transcriptional response of *SLC9A3R1* with expression levels nearing those of ERα. Blue dots represent expression levels below 0 and red dot above 0. (Relative expression in fold change; p-value = <0.01) **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

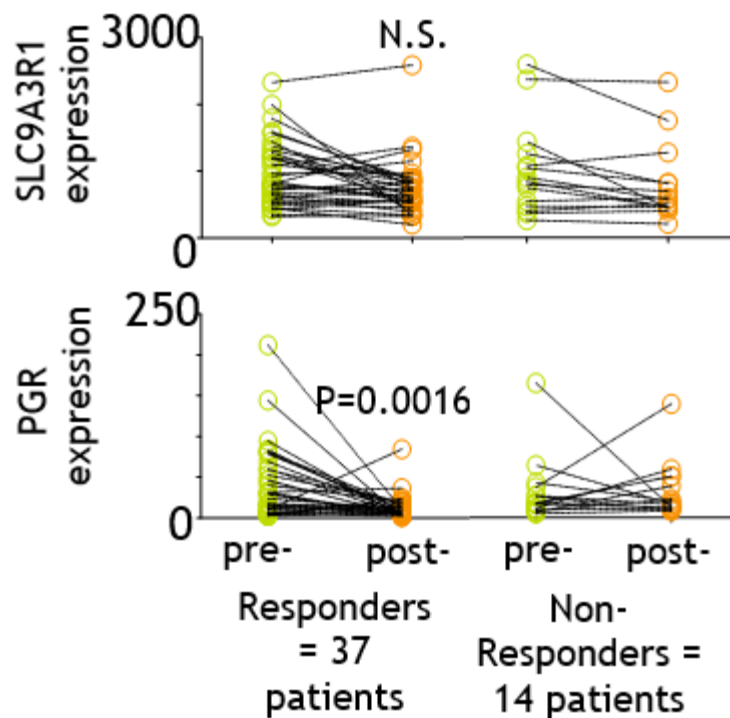


Figure 7.4 Expression levels of *SLC9A3R1* (top) and progesterone receptor (PGR) expression in 51 patients undergoing neo-adjuvant aromatase inhibitors prior to surgery for ER α -positive BC. Of the 37 responders and 14 non-responders to treatment, *SLC9A3R1* expression did not decrease to a statistically significant level. Oestrogen-dependent PGR expression was significantly reduced in the responder group (p-value=0.0016) but was variable in the non-responder group (N.S. = Not statistically significant). (Statistical test = one-way ANOVA). [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

A meta-analytical study involving patient-derived data using all available genes (N=22,777 genes) demonstrates that *SLC9A3R1* expression is amongst the top 1% of genes bearing the most determining prognostic relationship with BC relapse in a cohort of 724 endocrine treated ER α -positive patients (Figure 7.5) [231]. In addition to the latter dataset, other BC datasets show that the high expression of *SLC9A3R1* also significantly correlates with poor survival (Figure 7.6).

In summary, the data presented strongly supports that firstly *SLC9A3R1* is a direct target of ER α . Secondly, the expression of *SLC9A3R1* cannot be antagonised by first-line endocrine therapy and thirdly, the transcription of *SLC9A3R1* may play a pivotal role in endocrine resistant BC progression.

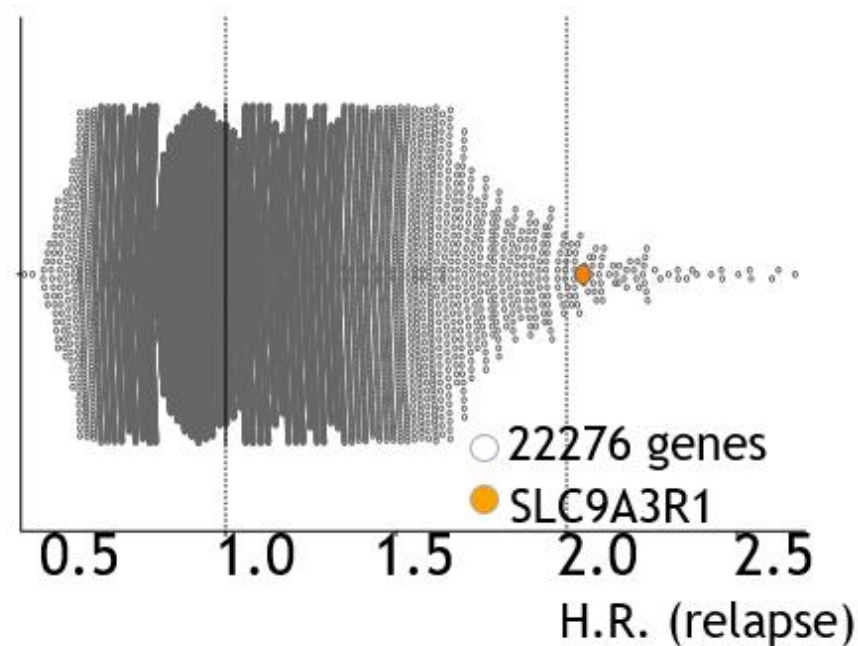


Figure 7.5 Global Kaplan-Meier analysis summarises univariate analysis for each gene included in the Affymetrix [222] microarray platform (N=724 patients). The hazard ratios analysis for disease relapse are displayed on the X-axis. *SLC9A3R1* is plotted as part of the top 1% of genes which is associated with a 2-fold more likely chance for a patient to develop relapse. **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** [209]

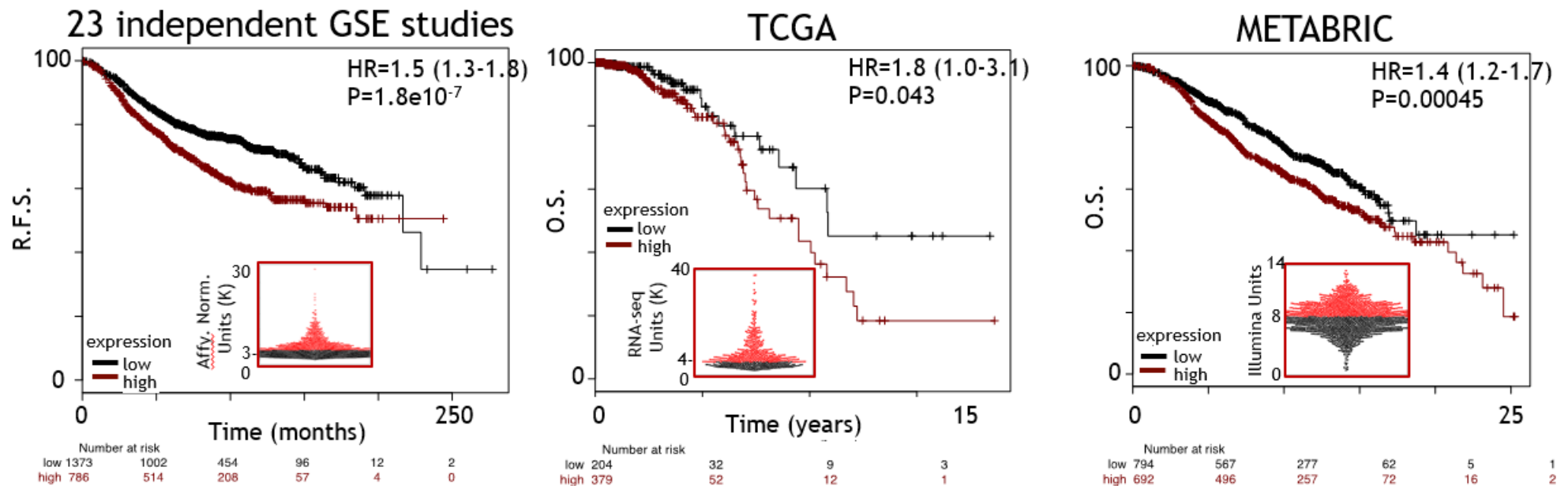


Figure 7.6 Survival data in three independent datasets. The first panel on the left represents **23 independent GSE studies (N= 2159 patients)**, **TCGA (N=583 patients)** [221] and **METABRIC (N=1486)** [220]. High expression of *SLC9A3R1* is associated with a reduced recurrence free survival (R.F.S.) (Hazard ratio = 1.5, P-value=1.8x10⁻⁷) in the GSE studies, decreased overall survival (O.S.) (Hazard ratio = 1.8, p-value=0.043) from the TCGA dataset and a reduced O.S. (Hazard ratio = 1.4, p-value=0.00045) in the METABRIC dataset. [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] [209]

In vitro analysis of *SLC9A3R1* demonstrates that ER α -positive BC cell lines exhibit high expression levels of *SLC9A3R1* (Figure 7.7). Investigation of the chromatin landscape of MCF7 and LTED cell reveals ER α and YY1 binding at three independent loci within the *SLC9A3R1* enhancer (named: E1, E2 and E3). This region is shown to directly loop to the YY1-bound *SLC9A3R1* promoter (Figure 7.8). The data is in agreement with *in vivo* transcriptional analysis that *SLC9A3R1* is the only gene near the enhancers E1-E3 that shows a significant increase in RNA levels when compared to normal breast tissue with ER α -positive BC (Figure 7.9).

Silencing YY1 using transfected siRNAs in MCF7 cells as described in Section 2.9 and measuring *SLC9A3R1* expression reveals decreased mRNA levels of *SLC9A3R1* in the siRNA compared to siCtrl groups in both ethanol (EtOH) and oestradiol (E2) treated arms (p-value=0.05).

The next step was to understand the effects of silencing *SLC9A3R1* in MCF7 (Figure 7.10A) and LTED cells on ER α -positive BC growth. MCF7 cells were transfected with siRNAs for *SLC9A3R1* as described in Section 2.11 and proliferation was assessed over 7 days. The cells were divided into ethanol (EtOH) and oestradiol treated (E2) groups. In both EtOH and E2 groups, the siRNAs transfected cells were able to abrogate proliferation. This experiment was repeated in the LTED cell lines and compared to the silencer control group (siCtrl), growth was halted in both of the siRNA groups (Figure 7.10B).

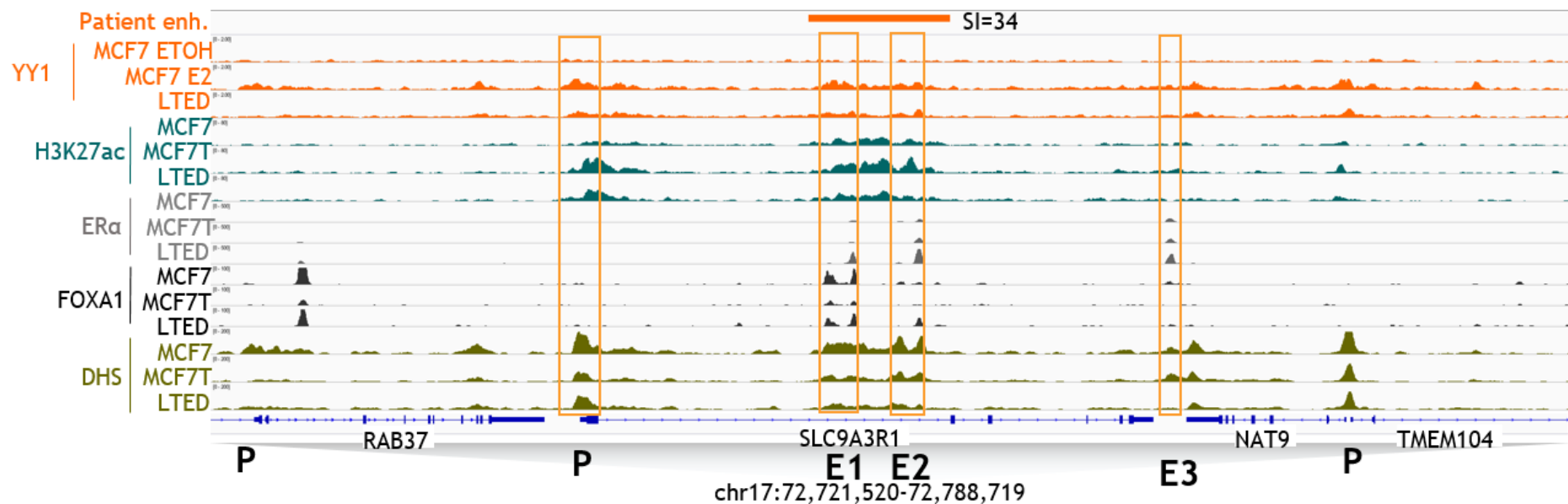


Figure 7.8 The regulatory landscape of *SLC9A3R1* (chr17:72, 721, 520-72, 788, 719). The promoters of *SLC9A3R1*/*RAB37*/*NAT9* (P), along with the identified three enhancers (E1, E2, E3) associated with *SLC9A3R1* are shown in the figure. At the top of the figure, the sharing index (SI) of E1 and E2 from the patient-derived samples (SI=34) is displayed. MCF7 ChIP-Seq without oestradiol (E2) stimulation shows no binding at the E1/E2 and E3 enhancers as well as the promoter of *SLC9A3R1*. But in the presence of E2 stimulation, YY1 binding is observed in these regions which is also supported by the YY1 ChIP-Seq track for LTED cells. H3K27ac ChIP-Seq *in vitro* data demonstrates that E1/E2 *SLC9A3R1* enhancers are active regulatory regions. ERα and FOXA1 ChIP-Seq tracks in MCF7/MCF7 Tamoxifen resistant (MCF7T)/LTED cells shows increased binding occurring preferentially at the E1 enhancer compared to E2 and E3 enhancers. Transcription factor binding at the E1-E3 enhancer is supported by the active DHS to further substantiate binding of ERα and pioneer factor FOXA1. Collectively this data clearly implicates enhancers E1 to E3 as an ERα-YY1 targets. **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

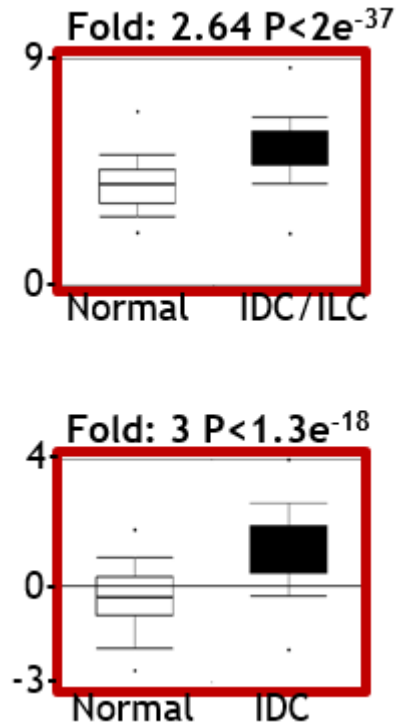


Figure 7.9 Patient-derived mRNA expression levels of SLC9A3R1 from the METABRIC¹⁶⁷ (top) (normal, n = 144; invasive ductal/lobular carcinoma = 1556) and TCGA (bottom) (normal, n = 61; invasive ductal carcinoma = 389). Both datasets reveal that compared to normal breast tissue, SLC9A3R1 expression is 2.64-fold ($p\text{-value} < 2 \times 10^{-37}$) and 3-fold ($p\text{-value} = 1.3 \times 10^{-18}$) higher. [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

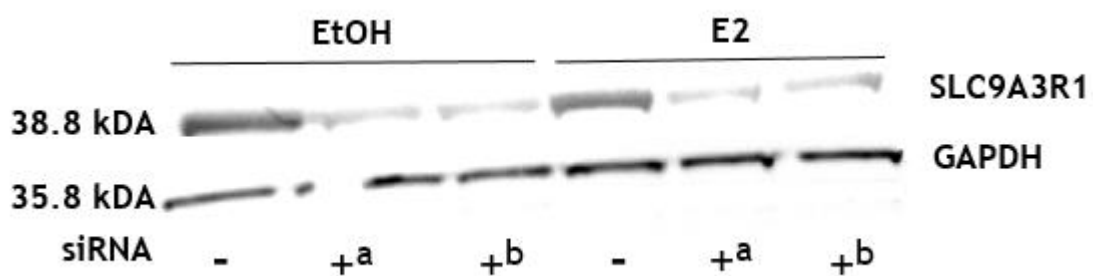


Figure 7.10A Western blot analysis for SLC9A3R1 depletion using two types of siRNA (a/b) in MCF7 cells stimulated with vehicle vs oestradiol for 6 hours. Compared to controls (-) SLC9A3R1 protein is significantly reduced > 80% in both control and oestradiol treated groups. siRNA “-” represents the MCF7 cells transfected with silencer negative control.

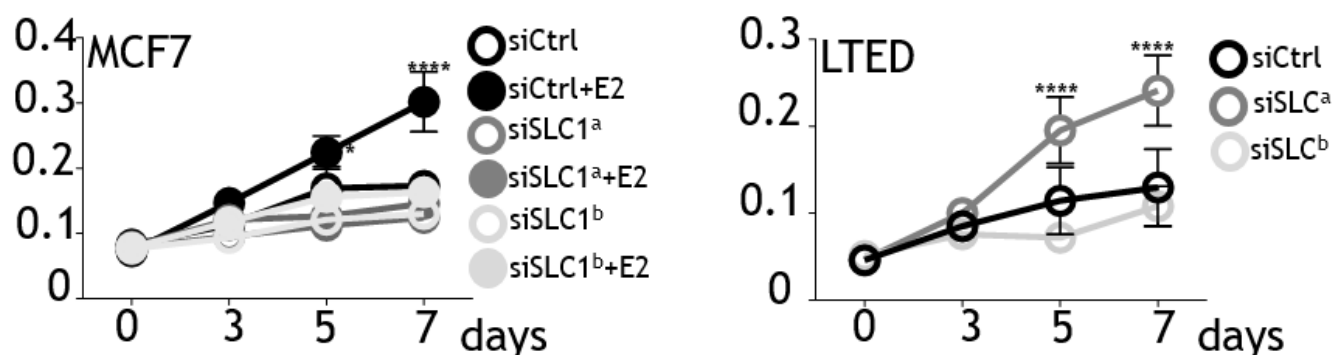


Figure 7.10B Proliferation assays performed in MCF7 (Left) and LTED (Right) when silencing *SLC9A3R1*.

In the MCF7 cells both siRNAs are able to halt growth in both ethanol and oestradiol (E2) treated groups compared to the silencer control (siCtrl) groups. In the LTED cells, growth is also abrogated in the siRNA groups compared to the siCtrl group. (Experiment performed three times; Statistical analysis = one-way ANOVA; **** = p-value = 0.0001) [Figure produced in collaboration with Giacomo Corleone and Luca Magnani] ^[209]

The data presented in this section demonstrate that firstly *SLC9A3R1* expression is driven by a breast cancer specific enhancer within the expanding ER α -YY1 clone during tumour formation. Secondly, silencing of *SLC9A3R1* protein results in halted growth in MCF7 cell lines. Furthermore, ER α -YY1 regulates *SLC9A3R1* via enhancer binding and the data presented also strongly supports the involvement of *SLC9A3R1* as a novel factor in contributing to endocrine therapy resistance in BC.

7.2 The application of *YY1* and *SLC9A3R1* enhancer activity for phenotypic mapping

YY1 and *SLC9A3R1* enhancers possess sharing indices of 41 and 34, respectively which would suggest their widespread usage amongst many BC patients. Investigating the activity of the *SLC9A3R1* enhancer within the patient-derived sample dataset reveals the latter marks itself as a dominant clone only in a subset of ER α -positive BC patients. In other patient samples, *SLC9A3R1* is more sub-clonal (Figure 7.11)

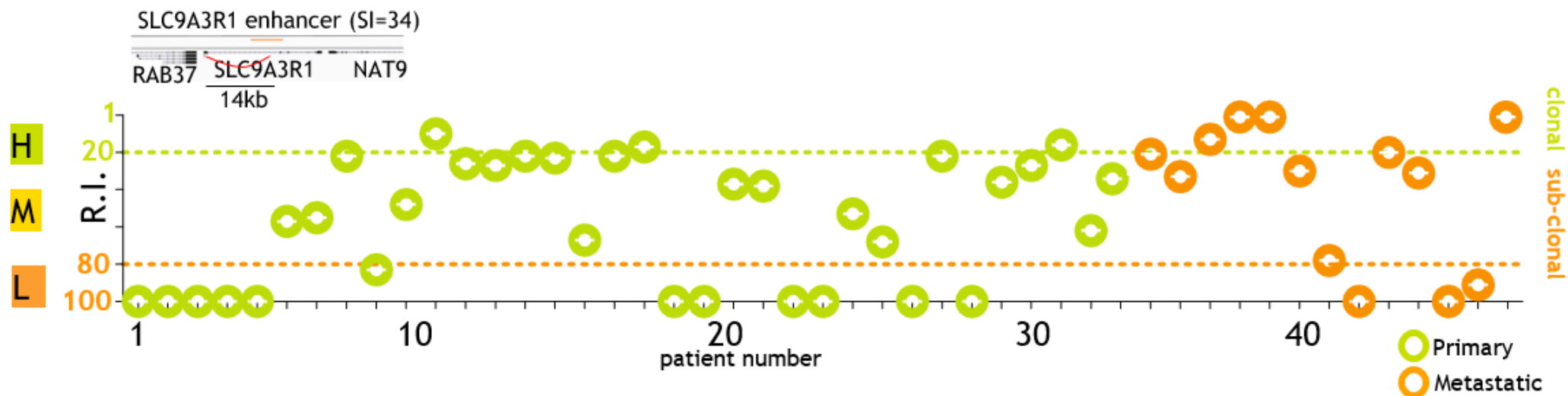


Figure 7.11 The ranking index (RI) in *SLC9A3R1* enhancer activity determined within patient-derived samples (N=47). Each green circle represents a primary breast cancer sample whereas each orange circle represents a metastatic sample. The location of the *SLC9A3R1* enhancer is pictured at the top left-hand corner of the figure. Most of the patient samples are found to possess medium to low RIs for the *SLC9A3R1* enhancer implying that this enhancer is predominantly subclonal. **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

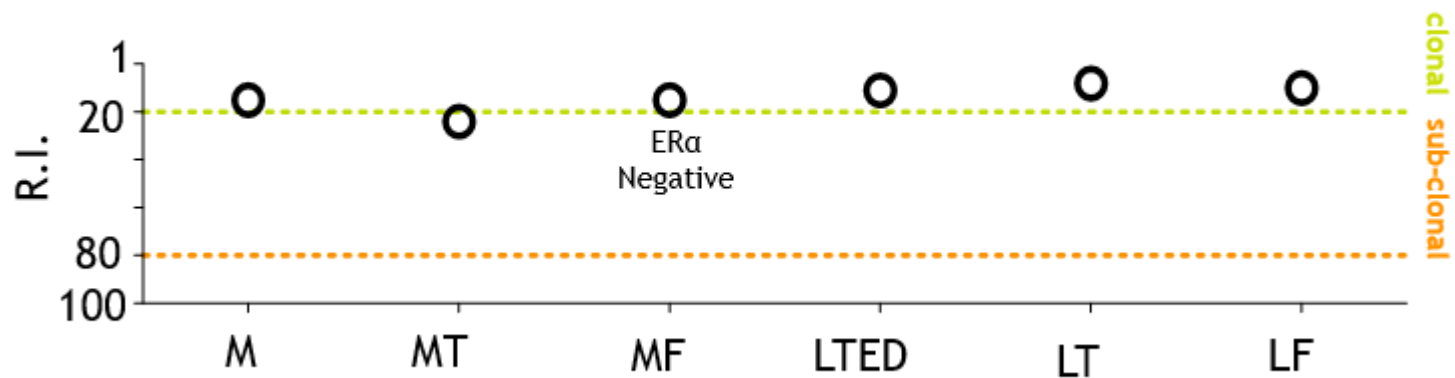


Figure 7.12 YY1 enhancer activity in MCF7 (M) and MCF7 derivative cell lines. MCF7 (M), MCF7 Tamoxifen resistant (MT), MCF7 fulvestrant resistant; ER α -negative (MF), LTED, LTED Tamoxifen resistant (LT), LTED fulvestrant resistant (LF). The YY1 enhancer activity is clonal in all of the cell lines. The MT cell line borders between a subclonal/clonal feature. (R.I. = Ranking index). [Figure produced in collaboration with Giacomo Corleone and Luca Magnani]^[209]

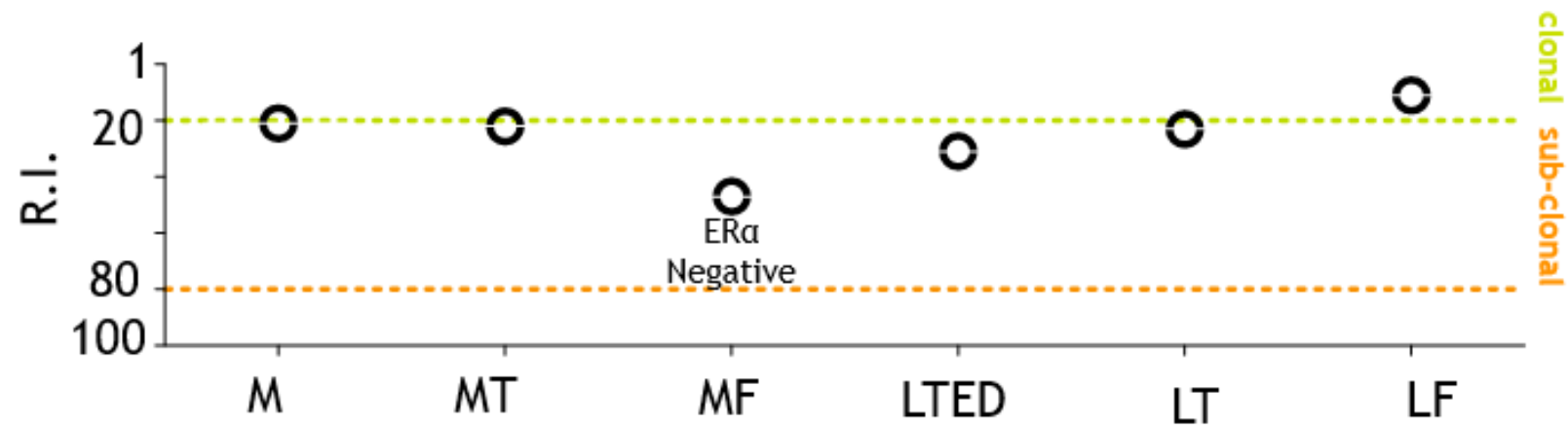


Figure 7.13 *SLC9A3R1* enhancer activity in MCF7 (M) and MCF7 derivative cell lines. MCF7 (M), MCF7 Tamoxifen resistant (MT), MCF7 fulvestrant resistant; ER α -negative (MF), LTED, LTED Tamoxifen resistant (LT), LTED fulvestrant resistant (LF). *SLC9A3R1* enhancer activity is clonal for most of the cell line derivatives except for MF (ER α -negative; with loss of ER α , the enhancer activity is reduced) and LTED which both possess subclonal *SLC9A3R1* in all of the cell lines. (R.I. = Ranking index). **[Figure produced in collaboration with Giacomo Corleone and Luca Magnani]** ^[209]

Mining of the Epigenomic roadmap database revealed that most tissues have heterogenous subpopulations highlighted by *SLC9A3R1* enhancer activity [134]. IHC meta-analysis data shows that 1. tissues with a low RI (<20) (for example, duodenum and small intestine) exhibit stronger IHC staining for *SLC9A3R1*, 2. Tissues possessing a medium RI ($50 < \text{RI} < 80$) (for example, lymph node, spleen, placenta, breast) show moderate staining for *SLC9A3R1* and 3. Tissues which have a high RI (>80) show weak to absence of staining for *SLC9A3R1* (for example, cardiac muscle, lung ovary and skeletal muscle) [232, 233]. The IHC meta-analysis identified increasing heterogeneity staining for *SLC9A3R1* protein with increasing RI values, paralleling the IHC meta-analysis staining for YY1 [232, 233].

To further validate the relationship between RI and heterogeneity, YY1 and *SLC9A3R1* activity was examined in H3K27ac MCF7 and MCF7 derivative BC cell lines. The results revealed that the YY1 enhancer activity points to the existence of a YY1-positive clonal population in all of the BC cell lines (Figure 7.12). When performing the same analysis using the *SLC9A3R1* enhancer similar results were obtained with the exception of MCF7 fulvestrant resistant cells which are ER α -negative; there is no ER α binding suggesting that the enhancer serves as a proxy for TF binding (Figure 7.13). Analysis of the ENCODE H3K27ac cancer cell line datasets [108], revealed that YY1-positive clones dominate most cell lines whereas the *SLC9A3R1* enhancer activity demonstrates that MCF7 cells contain a clonal *SLC9A3R1* population agreeing with the fact that the latter's expression is restricted to mostly ER α -positive BC cells.

There is therefore strong evidence to suggest that epigenomic mapping can be utilised to deconvolute transcriptional data to identify heterogeneous populations. This would imply that rather than increase in mRNA within single cells, transcriptional profiles can indicate changes in cell populations activating transcription [134].

Although *YY1* and *SLC9A3R1* would conventionally be considered as over-expressed genes in ER α -positive BC, with reference to the evidence presented in this Section, it is therefore proposed that these differences occur due to clonal expansion of cells marked by *YY1/SLC9A3R1* activity which can be measured through epigenetic mapping [209]. To further test this hypothesis, formalin-fixed paraffin embedded (FFPE) slides were requested, for each patient tumour sample for which ChIP-Seq data was analysed in this project (N = 47). This study was performed to assess whether the calculated RIs of *YY1/SLC9A3R1* per patient tumour sample correlated with the expected staining intensity. A total of 23 (N= 23/47) cases were obtained (Table 7.1) with the corresponding percentage tumour burden for the FFPE slides obtained. IHC staining for *YY1* and *SLC9A3R1* (i.e. IHC staining is being carried out in duplicates for *YY1* and *SLC9A3R1*) is currently underway and is being carried out by two independent centres (1. Department of histopathology, Immunohistochemistry department, Hammersmith Hospital, London, U.K. and 2. European institute of oncology, Biobank and histopathology unit, Milan, Italy).

Table 7.1 Available formalin-fixed paraffin embedded (FFPE) slides for the 23 of the 47 patient-tumour samples processed for H3K27ac ChIP-Seq. (ICTB= Imperial College tissue bank; BCN = Breast Cancer Now tissue bank; ILC = invasive lobular carcinoma; IDC = invasive ductal carcinoma; IMC = invasive mucinous carcinoma). (Please refer to Table 6.1A and 6.1B for further clinicopathological information regarding the patient samples)

Tumour sample	Tissue Bank	Histology	Tumour Burden in FFPE slide
G_DKP_011	ICTB	ILC	>50%
G_DKP_012	BCN	IDC	25%
G_DKP_013	BCN	IDC	80%
G_DKP_014	BCN	IDC	50%
G_DKP_015	BCN	IDC	60%
G_DKP_016	BCN	IDC	30%
G_DKP_017	BCN	IDC	5%
G_DKP_018	BCN	IDC	60%
G_DKP_019	BCN	IDC	45%
G_DKP_020	BCN	IDC	45%
P_DKP_008	BCN	IDC	70%
P_DKP_009	BCN	IDC	70%
P_DKP_010	BCN	IDC	45%
P_DKP_011	BCN	IDC	40%
P_DKP_012	BCN	IDC	30%
P_DKP_013	BCN	IDC	30%
P_DKP_014	BCN	IDC	80%
P_DKP_015	BCN	IDC	60%
P_DKP_016	BCN	IDC	75%
P_DKP_017	BCN	IDC	65%
P_DKP_018	BCN	IDC	35%
M_DKP_004	ICTB	IMC	>50%
M_DKP_005	ICTB	IDC	>50%

Chapter 8 Discussion

8.1 Optimisation of ChIP-Seq (Chapter 4)

The results of the ChIP-Seq data being compared using the standard (Figure 4.1) and optimised protocols (Chapter 4) demonstrate that the optimised protocol (Figure 4.2), can produce reliable results (Figure 4.3 and 4.4).

Parameters which were kept the same in both versions of the protocol include, 1. Primer sequence targets used for both ChIP and library qPCR, 2. 20 ng of DNA used from both Input and ChIP samples for initial library preparation and 3. PCR DNA amplification kept at 11 cycles during the library preparation phase. The two steps which have been introduced in the optimised protocol are 1. De-crosslinking the samples at the beginning of Day 2 for 6 hours and 2. The use of magnetic beads for DNA purification instead of using proteinase and RNase, phenol purification followed by cold DNA precipitation.

In order to further assess the optimisation of this ChIP-Seq technique and to demonstrate congruent results, this technique should be repeated in various cell lines and applied to tissue samples again comparing standard versus optimised ChIP techniques. The current tissue ChIP protocol, detailed in Chapter 4, involves pre-processing of the sample prior to fixation (Chapter 2; Section 2.3), and therefore should not confound the optimised protocol steps which are all performed on Day 2 (Figure 4.2). Cost implications concerning sequencing and time to conduct further experiments were limiting factors.

Figure 4.2 demonstrates the revised protocol which has been reduced from an average time of 112 (Figure 4.1) hours to take 6 samples for sequencing to 64 hours, resulting in a 42% reduction in the time for the samples to be ready for sequencing

submission. The optimised protocol, although produces reliable sequencing results, demands concise experimental planning owing to the numerous steps that are involved on Day 2 of the protocol (Figure 4.2).

These steps outlined by the optimised ChIP-Seq protocol are easily reproducible and do not require the user to learn new experimental procedures. A reduction in time to sequencing, demonstrated by the optimised ChIP protocol, will allow for more samples to be processed for ChIP-Seq at a faster, but reliable rate, which increases productivity by performing large-scale ChIP experiments with no loss of sequencing data quality.

8.2 Analysis of the training set of tumours (N=31) (Chapter 5)

Epigenetic mechanisms involved in breast cancer can provide insights into the complex regulatory networks involved in promoting the latter disease. The fact that ER α is involved in 70% of the total number of breast cancer warrants further understanding of the cistrome regulating ER α BC and its progression and implication in endocrine resistance.

This leads to the question whether epigenetics can be applied to decipher the mechanisms involved in disease progression and metastasis in BC. The histone mark H3K27ac has been shown to be a reliable epigenetic mark involved in highlighting regions bound by nucleosomes which show activity in promoter as well as non-coding enhancer regions of the genome [95, 98, 115-117, 119]. Studies reporting the use of histone mark H3K27ac have shown its use in reliably highlighting non-coding

enhancer/super- enhancer regions [115-117, 119]. Based on the concept that enhancers are cell type specific and can be highlighted with H3K27ac ChIP-Seq, it leads to the application of this method in the breast tumours used in this study.

The results of the computational based clustering, of the sequencing files for the 31-processed breast tumours undergoing H3K27ac ChIP-Seq, revealed that the Good, Poor and Metastatic groups did not cluster together, in terms of enriched genomic loci (peaks) called by MACS [142] (Figures 5.1 and 5.2). This implies that there are not only differences, in terms of called peaks, amongst the Good, Poor and Metastatic groups, but also significant differences amongst the samples within each group. The main factor contributing to this phenomenon is cellular heterogeneity [234].

The breast parenchyma is not only composed of one but several cell types which contribute to cellular heterogeneity [234]. In addition to stem cell plasticity, the secondary inflammatory effect caused by ongoing carcinogenesis results in the recruitment of immune cells which furthermore contributes to the increasing diversity of cells present within a tumour [234].

Using the super-enhancers called by ROSE [96, 116] computational software to cluster the tumours also did not reveal a high correlation amongst the samples within each outcome group represented by the heat map shown in Figure 5.4. The MDS plot of Jaccard distances displayed a few areas of similarity amongst two sets of Good samples and one set of Poor and Metastatic tumour groups (Figure 5.5). However, no general pattern of clustering or similarity was demonstrated within the samples of each outcome group. Using the concept of super-enhancers leading to cell type specificity, this would imply that the clustering of the tumour samples based on their called super- enhancers would lead to specific clustering amongst the samples within each group. However, this was not the case which again points to tumour

heterogeneity being a limitation of this project.

Intratumoural (occurs in the cancer cells residing within the same tumour) heterogeneity, which is thought to arise as a result of clonal tumour evolution, is acknowledged as an important obstacle in the efficacy of cancer therapy [210-212]. This is exemplified by a study conducted by *Gerlinger et al.*, who demonstrated that, sequencing different regions taken from the same renal cell carcinoma sample, revealed that greater than 60% of all the somatic mutations were not detected in every tumour biopsy [212]. Breast tumours are known to display significant genetic and phenotypic diversity including differences in rates of proliferation and invasion, metastatic potential, presence or absence of oncogenic driver mutations and response to chemotherapy (hormonal and cytotoxic). Furthermore, it is recognised that heterogeneity in BC exists at an inter-tumour (difference between tumours from different patients) as well as at an intra-tumour level [214]. The H3K27ac ChIP-Seq results obtained from the breast tumours revealed many overlapping enriched regions (peaks) amongst the samples within the Good, Poor and Metastatic groups. Grouping the samples according to similar hormonal phenotypes and analysing the peaks called did not reveal distinguishing results amongst the tumour groups.

Upon processing of the tumour specimens, the adipose tissue was thoroughly dissected but the tumours did not undergo micro-dissection prior to H3K27ac ChIP-seq. Therefore, sequenced DNA for each of the tumour samples exhibits H3K27ac signal as a representation of a mixed cellularity as opposed to the desired tumour cell specific ChIP-Seq signals. This has presented as one of the main limitations of this project and several computational methods have been used to standardise the overlapping enriched peaks amongst the tumour groups in order to obtain outcome

specific peaks. Outcome specific peak analysis using, IGV [143] Cistrome [144] advanced pipeline “Region subtraction” utility did not generate differentiating results.

Epigenetic profiling of the breast tumours is still in progress and other computational pipelines using R such as “MMDIFF” [235] (an R-based ChIP-Seq computational program algorithm which aims to statistically differentiate ChIP-Seq samples based on enriched genomic loci using BED file output from MACS [142]) will be considered for further epigenetic profiling stratification of the outcome groups.

With the breast tumour collection still underway, the aim will be to process more samples for H3K27ac ChIP-Seq in order to add to the current sample size in the primary and metastatic breast tumour groups. Increasing the sample size will allow for a more comprehensive and comparative analysis of the regulatory elements within the epigenomes of primary and metastatic breast cancer. Applications have been approved by Breast Cancer Now and ICR tissue banks for obtaining further primary ER α -positive as well as metastatic breast cancer samples, respectively.

Another proposed method of analysis would be to commence the epigenomic analysis, once the maximum number of tumour samples have been processed for H3K27ac ChIP-Seq within the next 18 months, with all primary and metastatic epigenomes together. This will involve the removal of the ‘Good’, ‘Poor’ and ‘Metastatic’ categories and analysing all H3K27ac signals to obtain and ensure maximum coverage of the epigenetic landscape of BC.

Taken together, the aim is to firstly obtain another 29 breast tumour samples (11 primary and 18 metastatic) which will then be processed for H3K27ac ChIP-Seq and secondly, supplement the signal data to the 31 H3K27ac tumour data already mentioned in this chapter to conduct the epigenomic analysis; this will bring the

sample size to N=60. Coverage of the BC genome, regulatory elements, T.F. binding and signal strength will all be studied and compared downstream of the analysis looking at the differences between primary and metastatic breast cancer.

8.3 Epigenomic profiling of primary and metastatic breast tumour tissue samples using H3K27ac ChIP-Seq and SLC9A3R1 (Chapter 6 and Chapter 7)

During the initial stages of the project, the tissue bank application was made to ICHTB to obtain fresh-frozen ER α -positive breast tumour samples based on the following criteria: A. Primary samples: 1. ER α -positive, 2. Treatment naïve, 3. Patients underwent adjuvant endocrine therapy and B. Metastatic samples: 1. ER α -positive/negative (providing the primary diagnosis was deemed ER α -positive), 2. Patients underwent adjuvant endocrine therapy for the primary ER α -positive breast tumour, 3. Samples can be distant metastasis to any site of the body.

The above inclusion criteria, with regard to the primary samples (Table 6.1A), required that a historical collection of banked samples should have been already present at the tissue bank. The ICHTB only possessed 19 primary samples which were suitable for the study. Metastatic tumour samples (Table 6.1B) were initially scarce due to the fact that current clinical National Health Service guidelines do not indicate re-biopsy for patient with BC relapse; so as not to cause the patient discomfort from an additional invasive procedure. Following discussions at the Breast Cancer Translational meeting, measures were put into place whereby the tissue collection technicians would liaise with the Charing Cross Breast Unit and

Oncology departments (Imperial College Healthcare NHS Trust, London, U.K.) to identify patients with BC relapse who had previously been diagnosed with ER α -positive BC and had undergone adjuvant endocrine following surgery. Patients were then informed regarding the research benefits of undertaking a biopsy of the relapsed disease and were subsequently consented for the latter procedure; 2 biopsy samples were therefore taken (1. Clinical; for histology and hormone receptor profiling to plan treatment, 2. Research; fresh frozen). Prospective metastatic sample collection, although, understandably a slow process was successful and has now become part of the routine investigative work-up patients with BC relapse undergo at Charing Cross hospital. Over the course of 2 years, from the start of this project, a total of 20 metastatic samples were collected from ICHTB. Only 12 of the 20 samples collected progressed to ChIP-Seq due to the fact 8 samples generated very low DNA quantities which subsequently failed the ChIP-qPCR stage.

The ICHTB primary samples which were retrospectively collected, although a record of tumour burden was made (i.e. >70%, per sample), did not have the corresponding H&E slides. Furthermore, some of the samples obtained were biological replicates (Section 6.2) Additional applications for fresh frozen primary ER α -positive BC tissue were hence made to the Breast Cancer Now Tissue bank. Despite requests were made as per the above inclusion criteria for primary tissue (from Breast Cancer Now), it was found that some patients did not undergo adjuvant endocrine therapy whilst one patient underwent adjuvant and neoadjuvant endocrine therapy and one other patient's treatment plan following surgery was unknown due to the fact that the patient was lost to follow-up (Table 6.1A). Even though these above-mentioned factors served as limitations to the study based on the inclusion criteria not being strictly met, the fact that we obtained various clinicopathological details based on

each sample implied that we were essentially capturing a wide heterogeneous population of BC tumour samples. However, without group stratification of the patient samples, which was applied in the analysis of the training set in Chapter 5 generating inconclusive results, this would make the downstream analysis and deconvolution of the BC epigenome a challenging task.

8.3.1 ChIP-Seq results in the tissue samples

A variable number of H3K27ac peaks (enriched regions) were called per sample in this study (Table 6.2A/B and Figure 6.1). 7 of the 55 tumour samples generated less than 2000 peaks although exhibiting sufficient enrichment on ChIP-qPCR. It was found that the background enrichments seen in the Input (control) samples were too high compared to the H3K27ac ChIP samples implying that when MACS [142] would apply its peak calling algorithm, a minimal number of peaks would thus be called. This arose due to technical aspects from the processing of the samples for ChIP. Although the tumour samples were macro-dissected from adipose tissue, the presence of the latter during the initial phases of ChIP may decrease the efficiency of the antibody not allowing it to efficiently pulldown the enriched regions.

Although the average number of peaks called in primary and metastatic samples were 64,042 and 63,691, respectively, this showed high variability per sample. For example, some samples generated 100,000 peaks (H3K27ac enriched regions) whereas some 25,000 and others 0 peaks. All tumour samples provided were of different sizes and did not undergo micro-dissection before they were stored for research purposes. The ischaemic time of the tissue could not be controlled for; i.e. from the time the whole tumour was excised from the patient to the time it was taken

to pathology and finally to the time the research sample was cut and stored (snap-frozen). This process should not take long if correct standard operating procedures are in place. Realistically, there are many factors which could delay this process including 1. the patient being physiologically unwell over the course of the surgery leading to a shift of priorities in the peri-operative setting 2. The lack of staff members to orchestrate the tissue collection; the sample is then left in theatre for 45 minutes to an hour before it is taken to pathology and 3. The synchronised availability of the pathologist and tissue bank technician at the time the specimen is brought to histopathology for sampling. These factors (which are a few of many) will undoubtedly have negative effects on the viability of the tissue and subsequent stability of the DNA which in turn could potentially produce variable ChIP-Seq results.

To confirm whether the number of tumour samples processed for H3K27ac ChIP-Seq was sufficient for downstream epigenomic analysis, saturation plots for active regulatory regions in promoters (calculated as 1Kb upstream the TSS; annotation obtained from UCSC Genes and Gene predictions; track: Genecode v24. [149]) and enhancers (defined as all the peaks which were not overlapping with the list of promoters obtained from the annotation table) were carried out using ACT saturation plotter (Figure 6.2A). With regard to the promoters, any 4 patient samples permutated was sufficient to reach saturation suggesting that the promoter BC epigenome was sufficiently captured. Using 47 patient samples, >80% of the enhancer landscape of the BC epigenome was obtained. The enhancer saturation plot begins to plateau at 44 samples, suggesting that processing another 5 to 10 more primary BC tumour samples would have resulted in complete enhancer saturation. The average peak size in enhancers and promoters were 517bp and

463bp, respectively. Therefore, the differences seen between promoter and enhancer saturations are due to actual differences in these active regulatory regions and not due to peaks sizes being larger/smaller in one group compared to the other. Furthermore, the results of the saturation plots would indicate that enhancers are highly heterogenous compared to enhancers within our patient sample population. To further substantiate this, the sharing of these active regulatory regions (Figure 6.2B) demonstrates that promoters (median number of patient samples sharing the same promoter peak = 15) are more widely shared amongst the patient samples compared to enhancers (median number of patient samples sharing the same enhancer peak = 1). To summarise, the data obtained from the above analysis are in agreement with enhancers being the main driver of cell-type specific transcription [111, 116, 117, 195, 209]. In addition, the data strongly suggests that enhancers account for the majority of heterogeneity in ER α -positive BC.

8.3.2 Intratumoural and intertumoural heterogeneity in ER α -positive breast cancer and the use of epigenetic information

The application of systematic epigenetic profiling of luminal BC reveals numerous fundamental principles underlying phenotypic-functional heterogeneity and the progression of BC. Profiling H3K27ac enriched regions, using ChIP-Seq, in primary and metastatic fresh-frozen BC tumour samples has led to the identification of new key players contributing to BC progression [209]. Genomic profiling of BC has unveiled extensive clonal heterogeneity which has proven to be very useful in understanding the concept of tumour evolution, especially when mutations directly

contribute to fitness [213, 236] and with the vast majority of the mutational burden composed of passenger mutations [215]. Transcriptional profiling of cancer cell populations tends to reflect the cumulative phenotypic status of the latter and is generally obtained from bulk tissue and therefore cannot be informative on the existence of low frequency distinct clonal/subclonal subpopulations [209]. Furthermore, molecular pathology provides information on the relative amount of protein present at the single-cell level but can be laborious and unfeasible when testing multiple targets at the same time.

The results from this analysis demonstrate that epigenomic analyses can be applied to extrapolate phenotypic heterogeneity in solid tumours samples. The ChIP-Seq signal can be used as a proxy for correlating the number of cells in the population carrying specific epigenetic information. *In vitro* experiments (Figure 6.3 and Figure 6.4) substantiate this statement whereby relative ChIP-qPCR enrichments, observed for a particular region, is dependent on the number of cells carrying that particular signal [209].

Application of the ranking approach strategy (Figure 6.5) enabled to further distil which of the identified regulatory regions were deemed as epigenomic “noise” (RN) vs. those that were functional drivers (RD). The linear regression analysis of ranking index (RI) vs. sharing index (SI) applied to the H3K27ac ChIP-Seq patient-derived samples, demonstrated that peaks which were highly shared subsequently possessed a low RI and vice versa. The identified enhancers within the patient-derived H3K27ac dataset were tested against known BC risk variants revealing that most of the latter were contained within the N=47 patient-sample enhancers; this would suggest that a vast amount of BC predisposition occurs through enhancer variants which are BC specific. In addition, when comparing the patient-derived

H3K27ac ChIP-Seq data with H3K27ac ChIP-Seq data in BC cell lines, it was found that 1. cell lines form a subset of the patient-derived enhancer panel and hence possess dominant phenotypic clones and 2. BC cell lines were not activating new regulatory regions and hence consist of dominant phenotypic clones from the tumour of origin. The data thus implies that solid tumours consist of dominant (clonal) and sub-clonal cell populations which are under represented by BC cell lines [209].

8.3.4 The transcriptional landscape of ER α -positive breast cancer

Unsupervised clustering of RN and RD regions in promoters (Figure 6.10B) and enhancers (Figure 6.10A), based on TF motif analysis, resulted in clear separation into two distinct biological groups, suggesting that RN and RD regions contain different regulatory information. Focusing on the enhancer regions, it was observed that mostly EREs were residing in the RN regions which is in agreement with heterogenous low-affinity ER α binding. Although the RN enhancer regions displayed enrichment for a larger set of TF motifs, strikingly the TF YY1 was exclusively enriched at the RD enhancers as well as the RD promoters.

8.3.5 YY1 a novel driver in ER α -positive breast cancer

The Ying Yang 1 (YY1), multifunctional protein, has been documented to be an important regulator of differential epigenetic regulation in gene expression and protein modifications [237]. YY1 is ubiquitous in nature and is highly conserved protein from *Xenopus* to human. YY1 functions as a transcription factor which either

activates or represses its target gene, dependent on its recruited co-factors [238-241]. YY1 is involved in the regulation of various genes enabling cells to differentiate and proliferate [239-243]. YY1 gene expression is regulated by numerous growth stimuli [240] and in addition, the latter has been found to be one of the Polycomb group proteins that contribute to epigenetic aberrations in cancers [244].

YY1 has been proposed to possess a regulatory role in cancer development orchestrating the expression of many cancer-related genes such as *MYC* and *c-FOS* [245, 246]. There is evidence to suggest that YY1 has been found to be highly expressed in BC, cooperating with activator protein 2 to stimulate expression of *ERBB2* (*Her2/neu*) which is a proto-oncogene overexpressed in approximately 30% of BCs and is correlated with poor prognosis [247].

Having identified the YY1 enhancer (Figure 6.11), within the N=47 patient derived samples, the latter was found to be highly shared (SI=41) (Figure 6.11) and also a clonal enhancer. Preliminary IHC staining in ER α -positive breast cancer FFPE slides demonstrates that YY1 exhibits intense staining with invasive BC but weakly positive at normal breast ductal epithelium (Figure 6.15).

ChIP-Seq analysis of YY1 in the BC cell line MCF7 and its oestrogen independent derivative LTED, demonstrates that although the majority of YY1 recruitment is independent of ER α , there is a subset of ER α -YY1 occurring as part of ER α core binding which is common to 75% of patients and BC cell lines. Exploration of ER α -YY1 binding reveals genes defining the Luminal A subtype of ER α -positive BC. Interestingly, when analysing the ER α -YY1 bound enhancers in the LTED cells, the genes associated with these enhancers are strongly enriched for acquired endocrine resistance in ER α -positive BC (Figure 6.27) [209].

Furthermore, silencing of *YY1* (Figure 6.20) reveals reduced relative luciferase-ERE activity implying that it shares co-factor attributes with ER α (Figure 6.21) in oestrogen-deplete or oestrogen-induced conditions. In addition, silencing *YY1* abrogates growth in MCF7, ZR75-1 and T47D (lesser extent) BC cell lines which further supports *YY1* positive involvement in BC progression (Figure 6.22 and 6.23). Inspection of the mRNA expression of ER α target genes upon *YY1* silencing, reveals a reduction of *TFF1*, *CA12* and *GREB1* in both ethanol (control) and oestradiol (E2) treated groups (Figure 6.24) [209].

8.3.6 The role of SLC9A3R1 in ER α -positive breast cancer

SLC9A3R1 has been identified as scaffolding protein Na⁺/H⁺ exchanger regulatory factor 1 (NHERF1) and ezrin-radixin-moesin binding phosphoprotein (EBP50). It is a ubiquitous protein which was first identified as a co-factor essential for protein kinase A-mediated inhibition of Na⁺/H⁺ exchanger isoform 3 (NHERF3) [248, 249]. SLC9A3R1 is a crucial component for the sorting and recycling of several receptors, ion channels, transporters and chemokine receptor internalisation and signal transduction [250-253]. In human pathology, altered *SLC9A3R1* expression has been associated with hypertension, acute kidney failure, and breast cancer [254]. SLC9A3R1 regulates cell signalling pathways related to cancer progression, including PTEN, β -catenin, PDGFR, EGFR and HER2/neu. IHC staining of SLC9A3R1 reveals that its subcellular localisation is altered in various cancers with a predominant cytoplasmic expression [248].

In the context of endocrine resistant BC, ER α -YY1 bound enhancers were tested against a set of oestrogen responsive genes that cannot be antagonised by

Tamoxifen therapy (Figure 6.27). *SLC9A3R1* was found not only to be the 2nd most Tamoxifen-resistant gene but also possessed an SI of 34 suggesting that it is widely shared amongst the patient-derived samples. In the absence of ER α the expression of *SLC9A3R1* in MCF7 cell lines is significantly reduced. Meta-analytical studies demonstrated that *SLC9A3R1* cannot be antagonised by Tamoxifen or Raloxifene (Figure 7.3) and it is also insensitive to neo-adjuvant aromatase inhibitor therapy (Figure 7.4). In addition, 3 independent clinical studies confirm that *SLC9A3R1* is associated with poor outcome (Figure 7.6) [209].

Three enhancers (E1-E3) were identified from the patient-derived H3K27ac data for *SLC9A3R1*. ER α -YY1 binding is present in these enhancers (Figure 7.8) further demonstrating that *SLC9A3R1* is an ER α target and is regulated by the ER α -YY1 collaborative binding. Silencing of *SLC9A3R1* in MCF7 cells (Figure 7.10A) and LTED demonstrates abrogation of proliferation (Figure 7.10B) in both BC cell lines. The data strongly suggest that *SLC9A3R1* expression is driven by a BC specific enhancer within an expanding clone during tumour formation and potentially contributes to endocrine therapy resistance in BC [209].

8.3.7 Conclusions

The results of the analysis have identified two novel drivers of luminal BC. Of note, YY1 has been classed as a fundamental TF associated with dominant enhancers and promoter in BC patients and acts as a global co-activator in cancer cells involved in orchestrating the cancer epigenome in BC. In addition, YY1 collaborates with ER α , as a co-activator by stabilising ER α binding and increasing transcription at a subset of enhancers specific to BC. The net effect of YY1 influence on ER α might overcome the benefit of activating the progesterone receptor (PgR) to induce a good-outcome ER α binding signatures [209]. This will allow for stratification of Luminal A BC patients who may be identified and subsequently treated with alternative chemotherapy/endocrine therapy regimens.

With regard to the ER α and YY1-bound *SLC9A3R1* enhancer, which cannot be antagonised by Tamoxifen or aromatase inhibitors, this enhancer remains active possibly due to the fact that ER α and YY1 binding provides stability to this enhancer rendering it active and insensitive to external stimuli.

More importantly, using the concept of epigenetic signals, such as nucleosome activation (i.e. H3K27ac ChIP-Seq) as a digital signal, the work I have performed coupled with the collaborative analyses demonstrate that the epigenetic signal (i.e. “on or off”) can be used as a potential proxy for inferring phenotypic heterogeneity. To our laboratory group’s knowledge and with reference to published literature, this study is novel in showing the latter. Future studies will entail applying this concept of using epigenetic marks to assess for phenotypic heterogeneity in other cancers to ascertain whether this can be applied in pan cancer types. Furthermore, 1. Analysis of IHC staining for YY1/SLC9A3R1 of the patients samples used in this study

(N=23), which is still underway, to ascertain whether the RI determined by our epigenomic analysis corroborates with staining intensity (i.e. fortifying the proposed postulation that by using epigenomic mapping, phenotypic heterogeneity can be inferred , 2. Performing IHC staining on other clinical BC samples comparing normal tissue to BC; this will further substantiate whether YY1 and SLC9A3R1 are prognostic biomarkers in the context of ER α -positive BC and 3. Further characterisation of YY1 and SLC9A3R1 as biomarkers in ER α -positive BC leading to subsequent drug discovery research in potentially targeting these two novel players in ER α -positive BC.

References

1. UK, C.R. *Worldwide cancers statistics*. [cited 2016 July 2016]; Available from: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer#heading-Zero>.
2. Ferlay J, S.I., Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11* International Agency for Research on Cancer 2013 [cited 2016 July 2016]; Available from: <http://globocan.iarc.fr>
3. *Cancer Incidence*. Cancer Research UK [cited 2016 July 2016]; Available from: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence#heading-Zero>.
4. *Office for National Statistics*. Data were provided by the Office for National Statistics on request 2013 [cited 2016 July 2016]; Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2015-07-10>.
5. *Cancer statistics*. ISD Scotland 2014 [cited 2016 July 2016]; Available from: <http://gro-scotland.gov.uk/statistics/theme/vital-events/general/ref-tables/index.html>.
6. *Northern Ireland Cancer Registry* Cancer statistics December 2013 [cited 2016 July 2016]; Available from: <http://www.nisra.gov.uk/demography/default.asp22.htm>
7. Mistry, M., D. Parkin, A.S. Ahmad, and P. Sasieni, *Cancer incidence in the United Kingdom: projections to the year 2030*. British journal of cancer, 2011. **105**(11): p. 1795-1803.

8. Sasieni P, e.a. *Cancer mortality projections in the UK to 2030 (unpublished)*. .
Analyses undertaken and data supplied on request; September 2013 2013
[cited 2016 July 2016]; Available from:
<http://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality/all-cancers-combined#r4iWTAhYkfVY0H2J.99>.
9. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. cell, 2000. **100**(1): p. 57-70.
10. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. cell, 2011. **144**(5): p. 646-674.
11. *Cancer statistics by cancer type: Breast cancer*. Cancer Research UK 2014
[cited 2016 July 2016]; Available from:
<http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero>.
12. Ali, S. and R.C. Coombes, *Endocrine-responsive breast cancer and strategies for combating resistance*. Nature Reviews Cancer, 2002. **2**(2): p. 101-112.
13. *The Mammary Gland : Development, Regulation, and Function*. Neville, M.C., ed. C.W. Daniel. 2014: Springer; Softcover reprint of the original 1st ed. 1987 edition (12 Sept. 2014).
14. Htun, H., L.T. Holth, D. Walker, J.R. Davie, and G.L. Hager, *Direct Visualization of the Human Estrogen Receptor α Reveals a Role for Ligand in the Nuclear Distribution of the Receptor*. Molecular Biology of the Cell, 1999. **10**(2): p. 471-486.
15. Monje, P., S. Zanello, M. Holick, and R. Boland, *Differential cellular localization of estrogen receptor alpha in uterine and mammary cells*. Mol Cell Endocrinol, 2001. **181**(1-2): p. 117-29.

16. King, W.J. and G.L. Greene, *Monoclonal antibodies localize oestrogen receptor in the nuclei of target cells*. Nature, 1984. **307**(5953): p. 745-7.
17. Gruber, C.J., W. Tschugguel, C. Schneeberger, and J.C. Huber, *Production and actions of estrogens*. New England Journal of Medicine, 2002. **346**(5): p. 340-352.
18. Riggs, B.L., S. Khosla, and L.J. Melton III, *Sex steroids and the construction and conservation of the adult skeleton*. Endocrine reviews, 2002. **23**(3): p. 279-302.
19. McEwen, B.S. and S.E. Alves, *Estrogen actions in the central nervous system* 1. Endocrine reviews, 1999. **20**(3): p. 279-307.
20. Baker, L., K.K. Meldrum, M. Wang, R. Sankula, R. Vanam, A. Raiesdana, B. Tsai, K. Hile, J.W. Brown, and D.R. Meldrum, *The role of estrogen in cardiovascular disease*. Journal of surgical research, 2003. **115**(2): p. 325-344.
21. Auchus, M.L. and R.J. Auchus, *Human steroid biosynthesis for the oncologist*. Journal of Investigative Medicine, 2012. **60**(2): p. 495-503.
22. Speirs, V., G. Skliris, S. Burdall, and P. Carder, *Distinct expression patterns of ER α and ER β in normal human mammary gland*. Journal of clinical pathology, 2002. **55**(5): p. 371-374.
23. Fisher, C.R., K.H. Graves, A.F. Parlow, and E.R. Simpson, *Characterization of mice deficient in aromatase (ArKO) because of targeted disruption of the cyp19 gene*. Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6965-70.
24. Korach, K.S., J.F. Couse, S.W. Curtis, T.F. Washburn, J. Lindzey, K.S. Kimbro, E.M. Eddy, S. Migliaccio, S.M. Snedeker, D.B. Lubahn, D.W. Schomberg, and E.P. Smith, *Estrogen receptor gene disruption: molecular*

- characterization and experimental and clinical phenotypes*. Recent Prog Horm Res, 1996. **51**: p. 159-86; discussion 186-8.
25. Krege, J.H., J.B. Hodgin, J.F. Couse, E. Enmark, M. Warner, J.F. Mahler, M. Sar, K.S. Korach, J.A. Gustafsson, and O. Smithies, *Generation and reproductive phenotypes of mice lacking estrogen receptor beta*. Proc Natl Acad Sci U S A, 1998. **95**(26): p. 15677-82.
 26. Clarke, R.B., A. Howell, C.S. Potten, and E. Anderson, *Dissociation between steroid receptor expression and cell proliferation in the human breast*. Cancer Res, 1997. **57**(22): p. 4987-91.
 27. Shoker, B.S., C. Jarvis, R.B. Clarke, E. Anderson, J. Hewlett, M.P. Davies, D.R. Sibson, and J.P. Sloane, *Estrogen receptor-positive proliferating cells in the normal and precancerous breast*. Am J Pathol, 1999. **155**(6): p. 1811-5.
 28. Hartman, J., A. Strom, and J.A. Gustafsson, *Estrogen receptor beta in breast cancer--diagnostic and therapeutic implications*. Steroids, 2009. **74**(8): p. 635-41.
 29. Roger, P., M.E. Sahla, S. Makela, J.A. Gustafsson, P. Baldet, and H. Rochefort, *Decreased expression of estrogen receptor beta protein in proliferative preinvasive mammary tumors*. Cancer Res, 2001. **61**(6): p. 2537-41.
 30. Abdel-Hafiz, H.A. and K.B. Horwitz, *Role of epigenetic modifications in luminal breast cancer*. Epigenomics, 2015. **7**(5): p. 847-862.
 31. Garden, O.J. and S. Paterson-Brown, *Breast Surgery*. 5 ed. A COMPANION TO SPECIALIST SURGICAL PRACTICE, ed. J.M. Dixon. 2014, United Kingdom: Saunders Elsevier. 322.

32. Houssami, N., P. Macaskill, M.L. Marinovich, J.M. Dixon, L. Irwig, M.E. Brennan, and L.J. Solin, *Meta-analysis of the impact of surgical margins on local recurrence in women with early-stage invasive breast cancer treated with breast-conserving therapy*. Eur J Cancer, 2010. **46**(18): p. 3219-32.
33. Singletary, S.E., *Surgical margins in patients with early-stage breast cancer treated with breast conservation therapy*. Am J Surg, 2002. **184**(5): p. 383-93.
34. Morrow, M. and K.J. Van Zee, *Society of Surgical Oncology-American Society for Radiation Oncology-American Society of Clinical Oncology Consensus Guideline on Margins for Breast-Conserving Surgery with Whole-Breast Irradiation in Ductal Carcinoma In Situ*. 2016. **23**(12): p. 3801-3810.
35. Darby, S., P. McGale, C. Correa, C. Taylor, R. Arriagada, M. Clarke, D. Cutter, C. Davies, M. Ewertz, J. Godwin, R. Gray, L. Pierce, T. Whelan, Y. Wang, and R. Peto, *Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials*. Lancet, 2011. **378**(9804): p. 1707-16.
36. Group, E.B.C.C., E.R. Group, N. Bijker, P. Meijnen, J.L. Peterse, J. Bogaerts, I. Van Hoorebeeck, J.P. Julien, M. Gennaro, P. Rouanet, A. Avril, I.S. Fentiman, H. Bartelink, and E.J. Rutgers, *Breast-conserving treatment with or without radiotherapy in ductal carcinoma-in-situ: ten-year results of European Organisation for Research and Treatment of Cancer randomized phase III trial 10853--a study by the EORTC Breast Cancer Cooperative Group and EORTC Radiotherapy Group*. J Clin Oncol, 2006. **24**(21): p. 3381-7.
37. Fisher, B., S. Land, E. Mamounas, J. Dignam, E.R. Fisher, and N. Wolmark, *Prevention of invasive breast cancer in women with ductal carcinoma in situ:*

- an update of the National Surgical Adjuvant Breast and Bowel Project experience*. Semin Oncol, 2001. **28**(4): p. 400-18.
38. Wapnir, I.L., J.J. Dignam, B. Fisher, E.P. Mamounas, S.J. Anderson, T.B. Julian, S.R. Land, R.G. Margolese, S.M. Swain, J.P. Costantino, and N. Wolmark, *Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS*. J Natl Cancer Inst, 2011. **103**(6): p. 478-88.
 39. Houghton, J., W.D. George, J. Cuzick, C. Duggan, I.S. Fentiman, M. Spittle, U.K.C.C.o.C. Research, P. Ductal Carcinoma in situ Working, A. Dcis trialists in the Uk, and Z. New, *Radiotherapy and tamoxifen in women with completely excised ductal carcinoma in situ of the breast in the UK, Australia, and New Zealand: randomised controlled trial*. Lancet, 2003. **362**(9378): p. 95-102.
 40. Holmberg, L., H. Garmo, B. Granstrand, A. Ringberg, L.G. Arnesson, K. Sandelin, P. Karlsson, H. Anderson, and S. Emdin, *Absolute risk reductions for local recurrence after postoperative radiotherapy after sector resection for ductal carcinoma in situ of the breast*. J Clin Oncol, 2008. **26**(8): p. 1247-52.
 41. Goodwin, A., S. Parker, D. Ghera, and N. Wilcken, *Post-operative radiotherapy for ductal carcinoma in situ of the breast*. Cochrane Database Syst Rev, 2009(1): p. CD000563.
 42. Julien, J.P., N. Bijker, I.S. Fentiman, J.L. Peterse, V. Delledonne, P. Rouanet, A. Avril, R. Sylvester, F. Mignolet, H. Bartelink, and J.A. Van Dongen, *Radiotherapy in breast-conserving treatment for ductal carcinoma in situ: first results of the EORTC randomised phase III trial 10853*. EORTC Breast Cancer Cooperative Group and EORTC Radiotherapy Group. Lancet, 2000. **355**(9203): p. 528-33.

43. Beatson, G., *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): p. 162-165.
44. Palmieri, C., D.K. Patten, A. Januszewski, G. Zucchini, and S.J. Howell, *Breast cancer: current and future endocrine therapies*. Molecular and cellular endocrinology, 2014. **382**(1): p. 695-723.
45. 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. Petruzella, T. Pienkowski, R. Radhika, B. Rajan, M.T. Rubach, S. Tort, G. Urrutia, M. Valentini, Y. Wang, and R. Peto, *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, 2013. **381**(9869): p. 805-16.
46. Gray, R.G., D. Rea, K. Handley, S.J. Bowden, P. Perry, H.M. Earl, C.J. Poole, T. Bates, S. Chetiyawardana, J.A. Dewar, I.N. Fernando, R. Grieve, J. Nicoll, Z. Rayter, A. Robinson, A. Salman, J. Yarnold, S. Bathers, A. Marshall, M. Lee, and o.b.o.t.a.C. Group, *aTTom: Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years in 6,953 women with early breast cancer*. Journal of Clinical Oncology, 2013. **31**(18_suppl): p. 5-5.
47. Cuzick, J., L. Ambrosini, N. Davidson, R. Jakesz, M. Kaufmann, M. Regan, and R. Sainsbury, *Use of luteinising-hormone-releasing hormone agonists as adjuvant treatment in premenopausal patients with hormone-receptor-positive*

- breast cancer: a meta-analysis of individual patient data from randomised adjuvant trials*. Lancet, 2007. **369**(9574): p. 1711-23.
48. Gnant, M., B. Mlineritsch, W. Schippinger, G. Luschin-Ebengreuth, S. Postlberger, C. Menzel, R. Jakesz, M. Seifert, M. Hubalek, V. Bjelic-Radisic, H. Samonigg, C. Tausch, H. Eidtmann, G. Steger, W. Kwasny, P. Dubsky, M. Fridrik, F. Fitzal, M. Stierer, E. Rucklinger, R. Greil, and C. Marth, *Endocrine therapy plus zoledronic acid in premenopausal breast cancer*. N Engl J Med, 2009. **360**(7): p. 679-91.
 49. Baum, M., A.U. Budzar, J. Cuzick, J. Forbes, J.H. Houghton, J.G. Klijn, T. Sahmoud, and A.T. Group, *Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the ATAC randomised trial*. Lancet, 2002. **359**(9324): p. 2131-9.
 50. Baum, M., A. Buzdar, J. Cuzick, J. Forbes, J. Houghton, A. Howell, T. Sahmoud, and A.T. Group, *Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early-stage breast cancer: results of the ATAC (Arimidex, Tamoxifen Alone or in Combination) trial efficacy and safety update analyses*. Cancer, 2003. **98**(9): p. 1802-10.
 51. Coates, A.S., A. Keshaviah, B. Thurlimann, H. Mouridsen, L. Mauriac, J.F. Forbes, R. Paridaens, M. Castiglione-Gertsch, R.D. Gelber, M. Colleoni, I. Lang, L. Del Mastro, I. Smith, J. Chirgwin, J.M. Nogaret, T. Pienkowski, A. Wardley, E.H. Jakobsen, K.N. Price, and A. Goldhirsch, *Five years of letrozole compared with tamoxifen as initial adjuvant therapy for*

- postmenopausal women with endocrine-responsive early breast cancer: update of study BIG 1-98*. J Clin Oncol, 2007. **25**(5): p. 486-92.
52. Xu, H., S. Yu, Q. Liu, X. Yuan, S. Mani, R.G. Pestell, and K. Wu, *Recent advances of highly selective CDK4/6 inhibitors in breast cancer*. Journal of Hematology & Oncology, 2017. **10**: p. 97.
 53. Al-Ghazal, S.K., L. Fallowfield, and R.W. Blamey, *Comparison of psychological aspects and patient satisfaction following breast conserving surgery, simple mastectomy and breast reconstruction*. Eur J Cancer, 2000. **36**(15): p. 1938-43.
 54. Group, E.B.C.T.C., *Effects of Radiotherapy and Surgery in Early Breast Cancer — An Overview of the Randomized Trials*. New England Journal of Medicine, 1995. **333**(22): p. 1444-1456.
 55. Morris, A.D., R.D. Morris, J.F. Wilson, J. White, S. Steinberg, P. Okunieff, R. Arriagada, M.G. Le, M. Blichert-Toft, and J.A. van Dongen, *Breast-conserving therapy vs mastectomy in early-stage breast cancer: a meta-analysis of 10-year survival*. Cancer J Sci Am, 1997. **3**(1): p. 6-12.
 56. Schain, W.S., T.M. d'Angelo, M.E. Dunn, A.S. Lichter, and L.J. Pierce, *Mastectomy versus conservative surgery and radiation therapy. Psychosocial consequences*. Cancer, 1994. **73**(4): p. 1221-8.
 57. Al-Ghazal, S.K. and R.W. Blamey, *Cosmetic assessment of breast-conserving surgery for primary breast cancer*. Breast, 1999. **8**(4): p. 162-8.
 58. Group, E.B.C.T.C., *Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials*. The Lancet, 2005. **365**(9472): p. 1687-1717.

59. Forbes, J., J. Cuzick, A. Buzdar, A. Howell, J. Tobias, and M. Baum, *Arimidex, Tamoxifen, Alone or in Combination (ATAC) Trialists' Group. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 100-month analysis of the ATAC trial*. *Lancet Oncol*, 2008. **9**(1): p. 45-53.
60. Davies, C., H. Pan, J. Godwin, R. Gray, R. Arriagada, V. Raina, M. Abraham, V.H.M. Alencar, A. Badran, and X. Bonfill, *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*. *The Lancet*, 2013. **381**(9869): p. 805-816.
61. Fidler, I.J., *Tumor heterogeneity and the biology of cancer invasion and metastasis*. *Cancer Res*, 1978. **38**(9): p. 2651-60.
62. Marusyk, A., V. Almendro, and K. Polyak, *Intra-tumour heterogeneity: a looking glass for cancer?* *Nat Rev Cancer*, 2012. **12**(5): p. 323-34.
63. Hirsch, F.R., G. Ottesen, J. Podenphant, and J. Olsen, *Tumor heterogeneity in lung cancer based on light microscopic features. A retrospective study of a consecutive series of 200 patients, treated surgically*. *Virchows Arch A Pathol Anat Histopathol*, 1983. **402**(2): p. 147-53.
64. Fisher, R., L. Pusztai, and C. Swanton, *Cancer heterogeneity: implications for targeted therapeutics*. *Br J Cancer*, 2013. **108**(3): p. 479-85.
65. Vogelstein, B. and K.W. Kinzler, *Cancer genes and the pathways they control*. *Nat Med*, 2004. **10**(8): p. 789-99.
66. Rangarajan, A., S.J. Hong, A. Gifford, and R.A. Weinberg, *Species- and cell type-specific requirements for cellular transformation*. *Cancer Cell*, 2004. **6**(2): p. 171-83.

67. Ashworth, A., C.J. Lord, and J.S. Reis-Filho, *Genetic interactions in cancer progression and treatment*. Cell, 2011. **145**(1): p. 30-8.
68. Shackleton, M., E. Quintana, E.R. Fearon, and S.J. Morrison, *Heterogeneity in cancer: cancer stem cells versus clonal evolution*. Cell, 2009. **138**(5): p. 822-9.
69. Allinen, M., R. Beroukhi, L. Cai, C. Brennan, J. Lahti-Domenici, H. Huang, D. Porter, M. Hu, L. Chin, A. Richardson, S. Schnitt, W.R. Sellers, and K. Polyak, *Molecular characterization of the tumor microenvironment in breast cancer*. Cancer Cell, 2004. **6**(1): p. 17-32.
70. Tlsty, T.D. and L.M. Coussens, *Tumor stroma and regulation of cancer development*. Annu Rev Pathol, 2006. **1**: p. 119-50.
71. Bissell, M.J., P.A. Kenny, and D.C. Radisky, *Microenvironmental regulators of tissue structure and function also regulate tumor induction and progression: the role of extracellular matrix and its degrading enzymes*. Cold Spring Harb Symp Quant Biol, 2005. **70**: p. 343-56.
72. Egeblad, M., E.S. Nakasone, and Z. Werb, *Tumors as organs: complex tissues that interface with the entire organism*. Dev Cell, 2010. **18**(6): p. 884-901.
73. Anderson, A.R., A.M. Weaver, P.T. Cummings, and V. Quaranta, *Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment*. Cell, 2006. **127**(5): p. 905-15.
74. Chen, J., K. Sprouffske, Q. Huang, and C.C. Maley, *Solving the puzzle of metastasis: the evolution of cell migration in neoplasms*. PLoS One, 2011. **6**(4): p. e17933.

75. Goel, S., D.G. Duda, L. Xu, L.L. Munn, Y. Boucher, D. Fukumura, and R.K. Jain, *Normalization of the vasculature for treatment of cancer and other diseases*. *Physiol Rev*, 2011. **91**(3): p. 1071-121.
76. Farmer, P., H. Bonnefoi, P. Anderle, D. Cameron, P. Wirapati, V. Becette, S. Andre, M. Piccart, M. Campone, E. Brain, G. Macgrogan, T. Petit, J. Jassem, F. Bibeau, E. Blot, J. Bogaerts, M. Aguet, J. Bergh, R. Iggo, and M. Delorenzi, *A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer*. *Nat Med*, 2009. **15**(1): p. 68-74.
77. Finak, G., N. Bertos, F. Pepin, S. Sadekova, M. Souleimanova, H. Zhao, H. Chen, G. Omeroglu, S. Meterissian, A. Omeroglu, M. Hallett, and M. Park, *Stromal gene expression predicts clinical outcome in breast cancer*. *Nat Med*, 2008. **14**(5): p. 518-27.
78. Waddington, C.H., *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser*. 1957: London: George Allen & Unwin, Ltd. ix +-262 pp.
79. Hughes, T.R., M.J. Marton, A.R. Jones, C.J. Roberts, R. Stoughton, C.D. Armour, H.A. Bennett, E. Coffey, H. Dai, Y.D. He, M.J. Kidd, A.M. King, M.R. Meyer, D. Slade, P.Y. Lum, S.B. Stepaniants, D.D. Shoemaker, D. Gachotte, K. Chakraborty, J. Simon, M. Bard, and S.H. Friend, *Functional discovery via a compendium of expression profiles*. *Cell*, 2000. **102**(1): p. 109-26.
80. Whitesell, L. and S.L. Lindquist, *HSP90 and the chaperoning of cancer*. *Nat Rev Cancer*, 2005. **5**(10): p. 761-72.
81. Chaffer, C.L., I. Brueckmann, C. Scheel, A.J. Kaestli, P.A. Wiggins, L.O. Rodrigues, M. Brooks, F. Reinhardt, Y. Su, K. Polyak, L.M. Arendt, C. Kuperwasser, B. Bierie, and R.A. Weinberg, *Normal and neoplastic nonstem*

- cells can spontaneously convert to a stem-like state*. Proc Natl Acad Sci U S A, 2011. **108**(19): p. 7950-5.
82. 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, *Landscape of somatic mutations in 560 breast cancer whole-genome sequences*. Nature, 2016. **534**(7605): p. 47-54.
 83. Carroll, J.S., X.S. Liu, A.S. Brodsky, W. Li, C.A. Meyer, A.J. Szary, J. Eeckhoute, W. Shao, E.V. Hestermann, and T.R. Geistlinger, *Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1*. Cell, 2005. **122**(1): p. 33-43.

84. Carroll, J.S., C.A. Meyer, J. Song, W. Li, T.R. Geistlinger, J. Eeckhoute, A.S. Brodsky, E.K. Keeton, K.C. Fertuck, and G.F. Hall, *Genome-wide analysis of estrogen receptor binding sites*. Nature genetics, 2006. **38**(11): p. 1289-1297.
85. Ross-Innes, C.S., R. Stark, A.E. Teschendorff, K.A. Holmes, H.R. Ali, M.J. Dunning, G.D. Brown, O. Gojis, I.O. Ellis, and A.R. Green, *Differential oestrogen receptor binding is associated with clinical outcome in breast cancer*. Nature, 2012. **481**(7381): p. 389-393.
86. Magnani, L., E.B. Ballantyne, X. Zhang, and M. Lupien, *PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer*. PLoS genet, 2011. **7**(11): p. e1002368.
87. Magnani, L., M. Brunelle, N. Gevry, and M. Lupien, *Chromatin landscape and endocrine response in breast cancer*. Epigenomics, 2012. **4**(6): p. 675-83.
88. Jozwik, K.M. and J.S. Carroll, *Pioneer factors in hormone-dependent cancers*. Nat Rev Cancer, 2012. **12**(6): p. 381-5.
89. Stefansson, O.A. and M. Esteller, *Epigenetic modifications in breast cancer and their role in personalized medicine*. Am J Pathol, 2013. **183**(4): p. 1052-63.
90. Luger, K., A.W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond, *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. Nature, 1997. **389**(6648): p. 251-60.
91. Jenuwein, T. and C.D. Allis, *Translating the histone code*. Science, 2001. **293**(5532): p. 1074-80.
92. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.

93. Dawson, M.A. and T. Kouzarides, *Cancer epigenetics: from mechanism to therapy*. Cell, 2012. **150**(1): p. 12-27.
94. Mikkelsen, T.S., M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, and B.E. Bernstein, *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells*. Nature, 2007. **448**(7153): p. 553-60.
95. Magnani, L. and M. Lupien, *Chromatin and epigenetic determinants of estrogen receptor alpha (ESR1) signaling*. Mol Cell Endocrinol, 2014. **382**(1): p. 633-41.
96. Hnisz, D., B.J. Abraham, T.I. Lee, A. Lau, V. Saint-Andre, A.A. Sigova, H.A. Hoke, and R.A. Young, *Super-enhancers in the control of cell identity and disease*. Cell, 2013. **155**(4): p. 934-47.
97. Heintzman, N.D., R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, and B. Ren, *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome*. Nat Genet, 2007. **39**(3): p. 311-8.
98. Banerji, J., S. Rusconi, and W. Schaffner, *Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences*. Cell, 1981. **27**(2 Pt 1): p. 299-308.
99. Ernst, J. and M. Kellis, *Discovery and characterization of chromatin states for systematic annotation of the human genome*. Nat Biotechnol, 2010. **28**(8): p. 817-25.

100. Pott, S. and J.D. Lieb, *What are super-enhancers?* Nat Genet, 2015. **47**(1): p. 8-12.
101. Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nat Rev Genet, 2014. **15**(4): p. 272-86.
102. Lelli, K.M., M. Slattery, and R.S. Mann, *Disentangling the many layers of eukaryotic transcriptional regulation.* Annu Rev Genet, 2012. **46**: p. 43-68.
103. Deng, W., J. Lee, H. Wang, J. Miller, A. Reik, P.D. Gregory, A. Dean, and G.A. Blobel, *Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor.* Cell, 2012. **149**(6): p. 1233-44.
104. Tolhuis, B., R.J. Palstra, E. Splinter, F. Grosveld, and W. de Laat, *Looping and interaction between hypersensitive sites in the active beta-globin locus.* Mol Cell, 2002. **10**(6): p. 1453-65.
105. Bulger, M. and M. Groudine, *Functional and mechanistic diversity of distal transcription enhancers.* Cell, 2011. **144**(3): p. 327-39.
106. Visel, A., M.J. Blow, Z. Li, T. Zhang, J.A. Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, B. Ren, E.M. Rubin, and L.A. Pennacchio, *ChIP-seq accurately predicts tissue-specific activity of enhancers.* Nature, 2009. **457**(7231): p. 854-8.
107. Thurman, R.E., E. Rynes, R. Humbert, J. Vierstra, M.T. Maurano, E. Haugen, N.C. Sheffield, A.B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T.K. Canfield, M. Diegel, D. Dunn, A.K. Ebersol, T. Frum, E. Giste, A.K. Johnson, E.M. Johnson, T. Kuttyavin, B. Lajoie, B.K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E.D. Nguyen, H. Qu, A.P. Reynolds, V. Roach, A. Safi, M.E. Sanchez, A. Sanyal,

- A. Shafer, J.M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M.O. Dorschner, R.S. Hansen, P.A. Navas, G. Stamatoyannopoulos, V.R. Iyer, J.D. Lieb, S.R. Sunyaev, J.M. Akey, P.J. Sabo, R. Kaul, T.S. Furey, J. Dekker, G.E. Crawford, and J.A. Stamatoyannopoulos, *The accessible chromatin landscape of the human genome*. Nature, 2012. **489**(7414): p. 75-82.
108. *An integrated encyclopedia of DNA elements in the human genome*. Nature, 2012. **489**(7414): p. 57-74.
109. Johnson, D.S., A. Mortazavi, R.M. Myers, and B. Wold, *Genome-wide mapping of in vivo protein-DNA interactions*. Science, 2007. **316**(5830): p. 1497-502.
110. Barski, A., S. Cuddapah, K. Cui, T.Y. Roh, D.E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao, *High-resolution profiling of histone methylations in the human genome*. Cell, 2007. **129**(4): p. 823-37.
111. Ernst, J., P. Kheradpour, T.S. Mikkelsen, N. Shores, L.D. Ward, C.B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis, and B.E. Bernstein, *Mapping and analysis of chromatin state dynamics in nine human cell types*. Nature, 2011. **473**(7345): p. 43-9.
112. Rada-Iglesias, A., R. Bajpai, T. Swigut, S.A. Brugmann, R.A. Flynn, and J. Wysocka, *A unique chromatin signature uncovers early developmental enhancers in humans*. Nature, 2011. **470**(7333): p. 279-83.
113. Creyghton, M.P., A.W. Cheng, G.G. Welstead, T. Kooistra, B.W. Carey, E.J. Steine, J. Hanna, M.A. Lodato, G.M. Frampton, P.A. Sharp, L.A. Boyer, R.A. Young, and R. Jaenisch, *Histone H3K27ac separates active from poised*

- enhancers and predicts developmental state*. Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.
114. Zhu, J., M. Adli, J.Y. Zou, G. Verstappen, M. Coyne, X. Zhang, T. Durham, M. Miri, V. Deshpande, P.L. De Jager, D.A. Bennett, J.A. Houmard, D.M. Muoio, T.T. Onder, R. Camahort, C.A. Cowan, A. Meissner, C.B. Epstein, N. Shores, and B.E. Bernstein, *Genome-wide chromatin state transitions associated with developmental and environmental cues*. Cell, 2013. **152**(3): p. 642-54.
 115. Nord, A.S., M.J. Blow, C. Attanasio, J.A. Akiyama, A. Holt, R. Hosseini, S. Phouanavong, I. Plajzer-Frick, M. Shoukry, V. Afzal, J.L. Rubenstein, E.M. Rubin, L.A. Pennacchio, and A. Visel, *Rapid and pervasive changes in genome-wide enhancer usage during mammalian development*. Cell, 2013. **155**(7): p. 1521-31.
 116. Whyte, W.A., D.A. Orlando, D. Hnisz, B.J. Abraham, C.Y. Lin, M.H. Kagey, P.B. Rahl, T.I. Lee, and R.A. Young, *Master transcription factors and mediator establish super-enhancers at key cell identity genes*. Cell, 2013. **153**(2): p. 307-19.
 117. Hnisz, D., J. Schuijers, C.Y. Lin, A.S. Weintraub, B.J. Abraham, T.I. Lee, J.E. Bradner, and R.A. Young, *Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers*. Mol Cell, 2015. **58**(2): p. 362-70.
 118. Sur, I. and J. Taipale, *The role of enhancers in cancer*. Nat Rev Cancer, 2016. **16**(8): p. 483-93.

119. Loven, J., H.A. Hoke, C.Y. Lin, A. Lau, D.A. Orlando, C.R. Vakoc, J.E. Bradner, T.I. Lee, and R.A. Young, *Selective inhibition of tumor oncogenes by disruption of super-enhancers*. Cell, 2013. **153**(2): p. 320-34.
120. Chapuy, B., M.R. McKeown, C.Y. Lin, S. Monti, M.G. Roemer, J. Qi, P.B. Rahl, H.H. Sun, K.T. Yeda, J.G. Doench, E. Reichert, A.L. Kung, S.J. Rodig, R.A. Young, M.A. Shipp, and J.E. Bradner, *Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma*. Cancer Cell, 2013. **24**(6): p. 777-90.
121. Jang, M.K., K. Mochizuki, M. Zhou, H.S. Jeong, J.N. Brady, and K. Ozato, *The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription*. Mol Cell, 2005. **19**(4): p. 523-34.
122. Dey, A., F. Chitsaz, A. Abbasi, T. Misteli, and K. Ozato, *The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis*. Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8758-63.
123. Xiang, J.F., Q.F. Yin, T. Chen, Y. Zhang, X.O. Zhang, Z. Wu, S. Zhang, H.B. Wang, J. Ge, X. Lu, L. Yang, and L.L. Chen, *Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus*. Cell Res, 2014. **24**(5): p. 513-31.
124. Kwiatkowski, N., T. Zhang, P.B. Rahl, B.J. Abraham, J. Reddy, S.B. Ficarro, A. Dastur, A. Amzallag, S. Ramaswamy, B. Tesar, C.E. Jenkins, N.M. Hannett, D. McMillin, T. Sanda, T. Sim, N.D. Kim, T. Look, C.S. Mitsiades, A.P. Weng, J.R. Brown, C.H. Benes, J.A. Marto, R.A. Young, and N.S. Gray, *Targeting transcription regulation in cancer with a covalent CDK7 inhibitor*. Nature, 2014. **511**(7511): p. 616-20.

125. Knoechel, B., J.E. Roderick, K.E. Williamson, J. Zhu, J.G. Lohr, M.J. Cotton, S.M. Gillespie, D. Fernandez, M. Ku, H. Wang, F. Piccioni, S.J. Silver, M. Jain, D. Pearson, M.J. Kluk, C.J. Ott, L.D. Shultz, M.A. Brehm, D.L. Greiner, A. Gutierrez, K. Stegmaier, A.L. Kung, D.E. Root, J.E. Bradner, J.C. Aster, M.A. Kelliher, and B.E. Bernstein, *An epigenetic mechanism of resistance to targeted therapy in T cell acute lymphoblastic leukemia*. Nat Genet, 2014. **46**(4): p. 364-70.
126. Shi, J., W.A. Whyte, C.J. Zepeda-Mendoza, J.P. Milazzo, C. Shen, J.S. Roe, J.L. Minder, F. Mercan, E. Wang, M.A. Eckersley-Maslin, A.E. Campbell, S. Kawaoka, S. Shareef, Z. Zhu, J. Kendall, M. Muhar, C. Haslinger, M. Yu, R.G. Roeder, M.H. Wigler, G.A. Blobel, J. Zuber, D.L. Spector, R.A. Young, and C.R. Vakoc, *Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation*. Genes Dev, 2013. **27**(24): p. 2648-62.
127. Northcott, P.A., C. Lee, T. Zichner, A.M. Stutz, S. Erkek, D. Kawauchi, D.J. Shih, V. Hovestadt, M. Zapatka, D. Sturm, D.T. Jones, M. Kool, M. Remke, F.M. Cavalli, S. Zuyderduyn, G.D. Bader, S. VandenBerg, L.A. Esparza, M. Ryzhova, W. Wang, A. Wittmann, S. Stark, L. Sieber, H. Seker-Cin, L. Linke, F. Kratochwil, N. Jager, I. Buchhalter, C.D. Imbusch, G. Zipprich, B. Raeder, S. Schmidt, N. Diessl, S. Wolf, S. Wiemann, B. Brors, C. Lawerenz, J. Eils, H.J. Warnatz, T. Risch, M.L. Yaspo, U.D. Weber, C.C. Bartholomae, C. von Kalle, E. Turanyi, P. Hauser, E. Sanden, A. Darabi, P. Siesjo, J. Sterba, K. Zitterbart, D. Sumerauer, P. van Sluis, R. Versteeg, R. Volckmann, J. Koster, M.U. Schuhmann, M. Ebinger, H.L. Grimes, G.W. Robinson, A. Gajjar, M. Mynarek, K. von Hoff, S. Rutkowski, T. Pietsch, W. Scheurlen, J. Felsberg, G. Reifenberger, A.E. Kulozik, A. von Deimling, O. Witt, R. Eils, R.J. Gilbertson,

- A. Korshunov, M.D. Taylor, P. Lichter, J.O. Korbel, R.J. Wechsler-Reya, and S.M. Pfister, *Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma*. *Nature*, 2014. **511**(7510): p. 428-34.
128. Affer, M., M. Chesi, W.D. Chen, J.J. Keats, Y.N. Demchenko, K. Tamizhmani, V.M. Garbitt, D.L. Riggs, L.A. Brents, A.V. Roschke, S. Van Wier, R. Fonseca, P.L. Bergsagel, and W.M. Kuehl, *Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma*. *Leukemia*, 2014. **28**(8): p. 1725-35.
129. Walker, B.A., C.P. Wardell, A. Brioli, E. Boyle, M.F. Kaiser, D.B. Begum, N.B. Dahir, D.C. Johnson, F.M. Ross, F.E. Davies, and G.J. Morgan, *Translocations at 8q24 juxtapose MYC with genes that harbor superenhancers resulting in overexpression and poor prognosis in myeloma patients*. *Blood Cancer J*, 2014. **4**: p. e191.
130. Groschel, S., M.A. Sanders, R. Hoogenboezem, E. de Wit, B.A. Bouwman, C. Erpelinck, V.H. van der Velden, M. Havermans, R. Avellino, K. van Lom, E.J. Rombouts, M. van Duin, K. Dohner, H.B. Beverloo, J.E. Bradner, H. Dohner, B. Lowenberg, P.J. Valk, E.M. Bindels, W. de Laat, and R. Delwel, *A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia*. *Cell*, 2014. **157**(2): p. 369-81.
131. Maurano, M.T., R. Humbert, E. Rynes, R.E. Thurman, E. Haugen, H. Wang, A.P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kuttyavin, S. Stehling-Sun, A.K. Johnson, T.K. Canfield, E. Giste, M. Diegel, D. Bates, R.S. Hansen, S. Neph, P.J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S.R. Sunyaev, R. Kaul, and

- J.A. Stamatoyannopoulos, *Systematic localization of common disease-associated variation in regulatory DNA*. Science, 2012. **337**(6099): p. 1190-5.
132. Pasquali, L., K.J. Gaulton, S.A. Rodriguez-Segui, L. Mularoni, I. Miguel-Escalada, I. Akerman, J.J. Tena, I. Moran, C. Gomez-Marin, M. van de Bunt, J. Ponsa-Cobas, N. Castro, T. Nammo, I. Cebola, J. Garcia-Hurtado, M.A. Maestro, F. Pattou, L. Piemonti, T. Berney, A.L. Gloyn, P. Ravassard, J.L. Gomez-Skarmeta, F. Muller, M.I. McCarthy, and J. Ferrer, *Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants*. Nat Genet, 2014. **46**(2): p. 136-43.
 133. Trynka, G., C. Sandor, B. Han, H. Xu, B.E. Stranger, X.S. Liu, and S. Raychaudhuri, *Chromatin marks identify critical cell types for fine mapping complex trait variants*. Nat Genet, 2013. **45**(2): p. 124-30.
 134. Kundaje, A., W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M.J. Ziller, V. Amin, J.W. Whitaker, M.D. Schultz, L.D. Ward, A. Sarkar, G. Quon, R.S. Sandstrom, M.L. Eaton, Y.C. Wu, A.R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R.A. Harris, N. Shores, C.B. Epstein, E. Gjoneska, D. Leung, W. Xie, R.D. Hawkins, R. Lister, C. Hong, P. Gascard, A.J. Mungall, R. Moore, E. Chuah, A. Tam, T.K. Canfield, R.S. Hansen, R. Kaul, P.J. Sabo, M.S. Bansal, A. Carles, J.R. Dixon, K.H. Farh, S. Feizi, R. Karlic, A.R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T.R. Mercer, S.J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R.C. Sallari, K.T. Siebenthal, N.A. Sinnott-Armstrong, M. Stevens, R.E. Thurman, J. Wu, B. Zhang, X. Zhou, A.E. Beaudet, L.A. Boyer, P.L. De Jager, P.J. Farnham, S.J. Fisher, D. Haussler, S.J. Jones, W. Li, M.A. Marra, M.T. McManus, S. Sunyaev, J.A. Thomson, T.D. Tlsty, L.H. Tsai, W. Wang, R.A.

- Waterland, M.Q. Zhang, L.H. Chadwick, B.E. Bernstein, J.F. Costello, J.R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J.A. Stamatoyannopoulos, T. Wang, and M. Kellis, *Integrative analysis of 111 reference human epigenomes*. *Nature*, 2015. **518**(7539): p. 317-30.
135. Soule, H.D., J. Vazquez, A. Long, S. Albert, and M. Brennan, *A human cell line from a pleural effusion derived from a breast carcinoma*. *J Natl Cancer Inst*, 1973. **51**(5): p. 1409-16.
136. Shaw, L.E., A.J. Sadler, D. Pugazhendhi, and P.D. Darbre, *Changes in oestrogen receptor-alpha and -beta during progression to acquired resistance to tamoxifen and fulvestrant (Faslodex, ICI 182,780) in MCF7 human breast cancer cells*. *J Steroid Biochem Mol Biol*, 2006. **99**(1): p. 19-32.
137. Keydar, I., L. Chen, S. Karby, F.R. Weiss, J. Delarea, M. Radu, S. Chaitcik, and H.J. Brenner, *Establishment and characterization of a cell line of human breast carcinoma origin*. *Eur J Cancer*, 1979. **15**(5): p. 659-70.
138. Engel, L.W., N.A. Young, T.S. Tralka, M.E. Lippman, S.J. O'Brien, and M.J. Joyce, *Establishment and characterization of three new continuous cell lines derived from human breast carcinomas*. *Cancer Res*, 1978. **38**(10): p. 3352-64.
139. Schmidt, D., M.D. Wilson, C. Spyrou, G.D. Brown, J. Hadfield, and D.T. Odom, *ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions*. *Methods*, 2009. **48**(3): p. 240-8.
140. Consortium, G.R., *UCSC GRCh37/Hg19 human genome build*. 2009: UCSC.
141. Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg, *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome*. *Genome Biol*, 2009. **10**(3): p. R25.

142. Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, and X.S. Liu, *Model-based analysis of ChIP-Seq (MACS)*. *Genome Biol*, 2008. **9**(9): p. R137.
143. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration*. *Brief Bioinform*, 2013. **14**(2): p. 178-92.
144. Liu, T., J.A. Ortiz, L. Taing, C.A. Meyer, B. Lee, Y. Zhang, H. Shin, S.S. Wong, J. Ma, Y. Lei, U.J. Pape, M. Poidinger, Y. Chen, K. Yeung, M. Brown, Y. Turpaz, and X.S. Liu, *Cistrome: an integrative platform for transcriptional regulation studies*. *Genome Biol*, 2011. **12**(8): p. R83.
145. Stark R, B.G., *DiffBind: Differential binding analysis of ChIP-Seq peak data*. 2011, Bioconductor. p. Differential binding analysis of ChIP-Seq peak data
146. S, A., *FastQC: a quality control tool for high throughput sequence data*. 2010, Babraham Bioinformatics. p. A quality control tool for high throughput sequence data.
147. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, *The Sequence Alignment/Map format and SAMtools*. *Bioinformatics*, 2009. **25**(16): p. 2078-9.
148. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features*. *Bioinformatics*, 2010. **26**(6): p. 841-2.
149. Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler, *The human genome browser at UCSC*. *Genome Res*, 2002. **12**(6): p. 996-1006.

150. Jee, J., J. Rozowsky, K.Y. Yip, L. Lochovsky, R. Bjornson, G. Zhong, Z. Zhang, Y. Fu, J. Wang, Z. Weng, and M. Gerstein, *ACT: aggregation and correlation toolbox for analyses of genome tracks*. Bioinformatics, 2011. **27**(8): p. 1152-4.
151. Institute, B., *Picard Tools*. Broad Institute. p. A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF.
152. Zhao, H., Z. Sun, J. Wang, H. Huang, J.P. Kocher, and L. Wang, *CrossMap: a versatile tool for coordinate conversion between genome assemblies*. Bioinformatics, 2014. **30**(7): p. 1006-7.
153. Consortium, G.R., *UCSC GRCh38 human genome build*. 2013: UCSC.
154. Heinz, S., C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, and C.K. Glass, *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities*. Mol Cell, 2010. **38**(4): p. 576-89.
155. Gregory R. Warnes, B.B., Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, Bill Venables, *gplots: Various R Programming Tools for Plotting Data*, G.R. Warnes, Editor. 2016, The Comprehensive R Archive Network. p. Various R Programming Tools for Plotting Data.
156. Faronato, M., V.T. Nguyen, D.K. Patten, Y. Lombardo, J.H. Steel, N. Patel, L. Woodley, S. Shousha, G. Pruneri, R.C. Coombes, and L. Magnani, *DMXL2 drives epithelial to mesenchymal transition in hormonal therapy resistant*

- breast cancer through Notch hyper-activation*. Oncotarget, 2015. **6**(26): p. 22467-79.
157. Stylianou, S., R.B. Clarke, and K. Brennan, *Aberrant activation of notch signaling in human breast cancer*. Cancer Res, 2006. **66**(3): p. 1517-25.
 158. Kandachar, V. and F. Roegiers, *Endocytosis and control of Notch signaling*. Curr Opin Cell Biol, 2012. **24**(4): p. 534-40.
 159. Le Borgne, R., *Regulation of Notch signalling by endocytosis and endosomal sorting*. Curr Opin Cell Biol, 2006. **18**(2): p. 213-22.
 160. Anthony W. Tolcher, R.C., Glenda Chambers, Villette Thorpe, Jakob Dupont, Dawn Hill, Lu Xu, Ann Kapoun, David C. Smith, *A first-in-human phase I study to evaluate the fully human monoclonal antibody OMP-59R5 (anti-Notch2/3) administered intravenously to patients with advanced solid tumors*. Journal of Clinical Oncology, 2012. **30**: p. suppl; abstr 3025.
 161. Doody, R.S., R. Raman, M. Farlow, T. Iwatsubo, B. Vellas, S. Joffe, K. Kieburtz, F. He, X. Sun, R.G. Thomas, P.S. Aisen, E. Siemers, G. Sethuraman, and R. Mohs, *A phase 3 trial of semagacestat for treatment of Alzheimer's disease*. N Engl J Med, 2013. **369**(4): p. 341-50.
 162. Ntziachristos, P., J.S. Lim, J. Sage, and I. Aifantis, *From fly wings to targeted cancer therapies: a centennial for notch signaling*. Cancer Cell, 2014. **25**(3): p. 318-34.
 163. Fan, X., L. Khaki, T.S. Zhu, M.E. Soules, C.E. Talsma, N. Gul, C. Koh, J. Zhang, Y.M. Li, J. Maciaczyk, G. Nikkhah, F. Dimeco, S. Piccirillo, A.L. Vescovi, and C.G. Eberhart, *NOTCH pathway blockade depletes CD133-positive glioblastoma cells and inhibits growth of tumor neurospheres and xenografts*. Stem Cells, 2010. **28**(1): p. 5-16.

164. Giachino, C. and V. Taylor, *Notching up neural stem cell homogeneity in homeostasis and disease*. Front Neurosci, 2014. **8**: p. 32.
165. Li, Y., J. Ma, X. Qian, Q. Wu, J. Xia, L. Miele, F.H. Sarkar, and Z. Wang, *Regulation of EMT by Notch signaling pathway in tumor progression*. Curr Cancer Drug Targets, 2013. **13**(9): p. 957-62.
166. Espinoza, I., R. Pochampally, F. Xing, K. Watabe, and L. Miele, *Notch signaling: targeting cancer stem cells and epithelial-to-mesenchymal transition*. Onco Targets Ther, 2013. **6**: p. 1249-59.
167. Ginnebaugh, K.R., A. Ahmad, and F.H. Sarkar, *The therapeutic potential of targeting the epithelial-mesenchymal transition in cancer*. Expert Opin Ther Targets, 2014. **18**(7): p. 731-45.
168. Artavanis-Tsakonas, S., M.D. Rand, and R.J. Lake, *Notch signaling: cell fate control and signal integration in development*. Science, 1999. **284**(5415): p. 770-6.
169. Bolos, V., E. Mira, B. Martinez-Poveda, G. Luxan, M. Canamero, A.C. Martinez, S. Manes, and J.L. de la Pompa, *Notch activation stimulates migration of breast cancer cells and promotes tumor growth*. Breast Cancer Res, 2013. **15**(4): p. R54.
170. Magnani, L., A. Stoeck, X. Zhang, A. Lanczky, A.C. Mirabella, T.L. Wang, B. Gyorffy, and M. Lupien, *Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer*. Proc Natl Acad Sci U S A, 2013. **110**(16): p. E1490-9.
171. Ji, Z., H. Mohammed, A. Webber, J. Ridsdale, N. Han, J.S. Carroll, and A.D. Sharrocks, *The forkhead transcription factor FOXK2 acts as a chromatin*

- targeting factor for the BAP1-containing histone deubiquitinase complex.* Nucleic Acids Res, 2014. **42**(10): p. 6232-42.
172. Yan, Y., N. Denef, and T. Schupbach, *The vacuolar proton pump, V-ATPase, is required for notch signaling and endosomal trafficking in Drosophila.* Dev Cell, 2009. **17**(3): p. 387-402.
 173. Sethi, N., Y. Yan, D. Quek, T. Schupbach, and Y. Kang, *Rabconnectin-3 is a functional regulator of mammalian Notch signaling.* J Biol Chem, 2010. **285**(45): p. 34757-64.
 174. Forgac, M., *Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology.* Nat Rev Mol Cell Biol, 2007. **8**(11): p. 917-29.
 175. Santen, R., M.H. Jeng, J.P. Wang, R. Song, S. Masamura, R. McPherson, S. Santner, W. Yue, and W.S. Shim, *Adaptive hypersensitivity to estradiol: potential mechanism for secondary hormonal responses in breast cancer patients.* J Steroid Biochem Mol Biol, 2001. **79**(1-5): p. 115-25.
 176. Brodie, A., D. Jelovac, G. Sabnis, B. Long, L. Macedo, and O. Goloubeva, *Model systems: mechanisms involved in the loss of sensitivity to letrozole.* J Steroid Biochem Mol Biol, 2005. **95**(1-5): p. 41-8.
 177. Martin, L.A., Z. Ghazoui, M.T. Weigel, S. Pancholi, A. Dunbier, S. Johnston, and M. Dowsett, *An in vitro model showing adaptation to long-term oestrogen deprivation highlights the clinical potential for targeting kinase pathways in combination with aromatase inhibition.* Steroids, 2011. **76**(8): p. 772-6.
 178. Magnani, L., E.B. Ballantyne, X. Zhang, and M. Lupien, *PBX1 genomic pioneer function drives ERalpha signaling underlying progression in breast cancer.* PLoS Genet, 2011. **7**(11): p. e1002368.

179. 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, *The pioneer factor PBX1 is a novel driver of metastatic progression in ERalpha-positive breast cancer*. Oncotarget, 2015. **6**(26): p. 21878-91.
180. Nguyen, V.T., I. Barozzi, M. Faronato, Y. Lombardo, J.H. Steel, N. Patel, P. Darbre, L. Castellano, B. Gyorffy, L. Woodley, A. Meira, D.K. Patten, V. Virchillo, M. Periyasamy, S. Ali, G. Frige, S. Minucci, R.C. Coombes, and L. Magnani, *Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion*. Nat Commun, 2015. **6**: p. 10044.
181. Xie, W., M.D. Schultz, R. Lister, Z. Hou, N. Rajagopal, P. Ray, J.W. Whitaker, S. Tian, R.D. Hawkins, D. Leung, H. Yang, T. Wang, A.Y. Lee, S.A. Swanson, J. Zhang, Y. Zhu, A. Kim, J.R. Nery, M.A. Urich, S. Kuan, C.A. Yen, S. Klugman, P. Yu, K. Suknuntha, N.E. Propson, H. Chen, L.E. Edsall, U. Wagner, Y. Li, Z. Ye, A. Kulkarni, Z. Xuan, W.Y. Chung, N.C. Chi, J.E. Antosiewicz-Bourget, I. Slukvin, R. Stewart, M.Q. Zhang, W. Wang, J.A. Thomson, J.R. Ecker, and B. Ren, *Epigenomic analysis of multilineage differentiation of human embryonic stem cells*. Cell, 2013. **153**(5): p. 1134-48.
182. Polo, J.M., E. Anderssen, R.M. Walsh, B.A. Schwarz, C.M. Nefzger, S.M. Lim, M. Borkent, E. Apostolou, S. Alaei, J. Cloutier, O. Bar-Nur, S. Cheloufi, M. Stadtfeld, M.E. Figueroa, D. Robinton, S. Natesan, A. Melnick, J. Zhu, S. Ramaswamy, and K. Hochedlinger, *A molecular roadmap of reprogramming somatic cells into iPS cells*. Cell, 2012. **151**(7): p. 1617-32.

183. Sharma, S.V., D.Y. Lee, B. Li, M.P. Quinlan, F. Takahashi, S. Maheswaran, U. McDermott, N. Azizian, L. Zou, M.A. Fischbach, K.K. Wong, K. Brandstetter, B. Wittner, S. Ramaswamy, M. Classon, and J. Settleman, *A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations*. Cell, 2010. **141**(1): p. 69-80.
184. Suva, M.L., E. Rheinbay, S.M. Gillespie, A.P. Patel, H. Wakimoto, S.D. Rabkin, N. Riggi, A.S. Chi, D.P. Cahill, B.V. Nahed, W.T. Curry, R.L. Martuza, M.N. Rivera, N. Rossetti, S. Kasif, S. Beik, S. Kadri, I. Tirosh, I. Wortman, A.K. Shalek, O. Rozenblatt-Rosen, A. Regev, D.N. Louis, and B.E. Bernstein, *Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells*. Cell, 2014. **157**(3): p. 580-94.
185. Malik, A.N., T. Vierbuchen, M. Hemberg, A.A. Rubin, E. Ling, C.H. Couch, H. Stroud, I. Spiegel, K.K. Farh, D.A. Harmin, and M.E. Greenberg, *Genome-wide identification and characterization of functional neuronal activity-dependent enhancers*. Nat Neurosci, 2014. **17**(10): p. 1330-9.
186. Stasevich, T.J., Y. Hayashi-Takanaka, Y. Sato, K. Maehara, Y. Ohkawa, K. Sakata-Sogawa, M. Tokunaga, T. Nagase, N. Nozaki, J.G. McNally, and H. Kimura, *Regulation of RNA polymerase II activation by histone acetylation in single living cells*. Nature, 2014. **516**(7530): p. 272-5.
187. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nat Methods, 2012. **9**(4): p. 357-9.
188. Salmon-Divon, M., H. Dvinge, K. Tammoja, and P. Bertone, *PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci*. BMC Bioinformatics, 2010. **11**: p. 415.

189. Yates, A., W. Akanni, M.R. Amode, D. Barrell, K. Billis, D. Carvalho-Silva, C. Cummins, P. Clapham, S. Fitzgerald, L. Gil, C.G. Giron, L. Gordon, T. Hourlier, S.E. Hunt, S.H. Janacek, N. Johnson, T. Juettemann, S. Keenan, I. Lavidas, F.J. Martin, T. Maurel, W. McLaren, D.N. Murphy, R. Nag, M. Nuhn, A. Parker, M. Patricio, M. Pignatelli, M. Rahtz, H.S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S.P. Wilder, A. Zadissa, E. Birney, J. Harrow, M. Muffato, E. Perry, M. Ruffier, G. Spudich, S.J. Trevanion, F. Cunningham, B.L. Aken, D.R. Zerbino, and P. Flicek, *Ensembl 2016*. Nucleic Acids Res, 2016. **44**(D1): p. D710-6.
190. Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen, *Primer3--new capabilities and interfaces*. Nucleic Acids Res, 2012. **40**(15): p. e115.
191. Park, P.J., *ChIP-seq: advantages and challenges of a maturing technology*. Nat Rev Genet, 2009. **10**(10): p. 669-80.
192. Lee, T.I., N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J.B. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford, and R.A. Young, *Transcriptional regulatory networks in Saccharomyces cerevisiae*. Science, 2002. **298**(5594): p. 799-804.
193. Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W.K. Sung, N.D. Clarke, C.L. Wei, and H.H. Ng, *Integration of external signaling pathways with*

- the core transcriptional network in embryonic stem cells*. Cell, 2008. **133**(6): p. 1106-17.
194. Nielsen, R., T.A. Pedersen, D. Hagenbeek, P. Moulos, R. Siersbaek, E. Megens, S. Denissov, M. Borgesen, K.J. Francoijs, S. Mandrup, and H.G. Stunnenberg, *Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis*. Genes Dev, 2008. **22**(21): p. 2953-67.
 195. Heintzman, N.D., G.C. Hon, R.D. Hawkins, P. Kheradpour, A. Stark, L.F. Harp, Z. Ye, L.K. Lee, R.K. Stuart, C.W. Ching, K.A. Ching, J.E. Antosiewicz-Bourget, H. Liu, X. Zhang, R.D. Green, V.V. Lobanenko, R. Stewart, J.A. Thomson, G.E. Crawford, M. Kellis, and B. Ren, *Histone modifications at human enhancers reflect global cell-type-specific gene expression*. Nature, 2009. **459**(7243): p. 108-12.
 196. Tolstorukov, M.Y., P.V. Kharchenko, J.A. Goldman, R.E. Kingston, and P.J. Park, *Comparative analysis of H2A.Z nucleosome organization in the human and yeast genomes*. Genome Res, 2009. **19**(6): p. 967-77.
 197. Ren, B., F. Robert, J.J. Wyrick, O. Aparicio, E.G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T.L. Volkert, C.J. Wilson, S.P. Bell, and R.A. Young, *Genome-wide location and function of DNA binding proteins*. Science, 2000. **290**(5500): p. 2306-9.
 198. Iyer, V.R., C.E. Horak, C.S. Scafe, D. Botstein, M. Snyder, and P.O. Brown, *Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF*. Nature, 2001. **409**(6819): p. 533-8.

199. Lieb, J.D., X. Liu, D. Botstein, and P.O. Brown, *Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association*. Nat Genet, 2001. **28**(4): p. 327-34.
200. Horak, C.E. and M. Snyder, *ChIP-chip: a genomic approach for identifying transcription factor binding sites*. Methods Enzymol, 2002. **350**: p. 469-83.
201. Weinmann, A.S., P.S. Yan, M.J. Oberley, T.H. Huang, and P.J. Farnham, *Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis*. Genes Dev, 2002. **16**(2): p. 235-44.
202. Landt, S.G., G.K. Marinov, A. Kundaje, P. Kheradpour, F. Pauli, S. Batzoglou, B.E. Bernstein, P. Bickel, J.B. Brown, P. Cayting, Y. Chen, G. DeSalvo, C. Epstein, K.I. Fisher-Aylor, G. Euskirchen, M. Gerstein, J. Gertz, A.J. Hartemink, M.M. Hoffman, V.R. Iyer, Y.L. Jung, S. Karmakar, M. Kellis, P.V. Kharchenko, Q. Li, T. Liu, X.S. Liu, L. Ma, A. Milosavljevic, R.M. Myers, P.J. Park, M.J. Pazin, M.D. Perry, D. Raha, T.E. Reddy, J. Rozowsky, N. Shores, A. Sidow, M. Slattery, J.A. Stamatoyannopoulos, M.Y. Tolstorukov, K.P. White, S. Xi, P.J. Farnham, J.D. Lieb, B.J. Wold, and M. Snyder, *ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia*. Genome Res, 2012. **22**(9): p. 1813-31.
203. Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O.L. Griffith, A. He, M. Marra, M. Snyder, and S. Jones, *Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing*. Nat Methods, 2007. **4**(8): p. 651-7.

204. Goren, A., F. Ozsolak, N. Shores, M. Ku, M. Adli, C. Hart, M. Gymrek, O. Zuk, A. Regev, P.M. Milos, and B.E. Bernstein, *Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA*. Nat Methods, 2010. **7**(1): p. 47-9.
205. Euskirchen, G.M., J.S. Rozowsky, C.L. Wei, W.H. Lee, Z.D. Zhang, S. Hartman, O. Emanuelsson, V. Stolc, S. Weissman, M.B. Gerstein, Y. Ruan, and M. Snyder, *Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies*. Genome Res, 2007. **17**(6): p. 898-909.
206. Ji, H., H. Jiang, W. Ma, D.S. Johnson, R.M. Myers, and W.H. Wong, *An integrated software system for analyzing ChIP-chip and ChIP-seq data*. Nat Biotechnol, 2008. **26**(11): p. 1293-300.
207. Rozowsky, J., G. Euskirchen, R.K. Auerbach, Z.D. Zhang, T. Gibson, R. Bjornson, N. Carriero, M. Snyder, and M.B. Gerstein, *PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls*. Nat Biotechnol, 2009. **27**(1): p. 66-75.
208. Qin, Z.S., J. Yu, J. Shen, C.A. Maher, M. Hu, S. Kalyana-Sundaram, J. Yu, and A.M. Chinnaiyan, *HPeak: an HMM-based algorithm for defining read-enriched regions in ChIP-Seq data*. BMC Bioinformatics, 2010. **11**: p. 369.
209. Patten D.K., Corleone G, Erdos E, Saiakhova A, Goddard K, Shousha A, Gyroffy B, Sandor L, Hadjiminias D.J., Scacheri P, Palmieri C, Barry P, Coombes, R.C., Pruner G, Magnani L, *Epigenetic mapping of breast cancer patients reveals principles and drivers of functional heterogeneity*. 2017, Imperial College London: United Kingdom.

210. Nowell, P.C., *The clonal evolution of tumor cell populations*. Science, 1976. **194**(4260): p. 23-8.
211. Greaves, M. and C.C. Maley, *Clonal evolution in cancer*. Nature, 2012. **481**(7381): p. 306-13.
212. Gerlinger, M., A.J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, I. Varela, B. Phillimore, S. Begum, N.Q. McDonald, A. Butler, D. Jones, K. Raine, C. Latimer, C.R. Santos, M. Nohadani, A.C. Eklund, B. Spencer-Dene, G. Clark, L. Pickering, G. Stamp, M. Gore, Z. Szallasi, J. Downward, P.A. Futreal, and C. Swanton, *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing*. N Engl J Med, 2012. **366**(10): p. 883-92.
213. Yates, L.R., M. Gerstung, S. Knappskog, C. Desmedt, G. Gundem, P. Van Loo, T. Aas, L.B. Alexandrov, D. Larsimont, H. Davies, Y. Li, Y.S. Ju, M. Ramakrishna, H.K. Haugland, P.K. Lilleng, S. Nik-Zainal, S. McLaren, A. Butler, S. Martin, D. Glodzik, A. Menzies, K. Raine, J. Hinton, D. Jones, L.J. Mudie, B. Jiang, D. Vincent, A. Greene-Colozzi, P.Y. Adnet, A. Fatima, M. Maetens, M. Ignatiadis, M.R. Stratton, C. Sotiriou, A.L. Richardson, P.E. Lonning, D.C. Wedge, and P.J. Campbell, *Subclonal diversification of primary breast cancer revealed by multiregion sequencing*. Nat Med, 2015. **21**(7): p. 751-9.
214. Skibinski, A. and C. Kuperwasser, *The origin of breast tumor heterogeneity*. Oncogene, 2015. **34**(42): p. 5309-16.
215. Williams, M.J., B. Werner, C.P. Barnes, T.A. Graham, and A. Sottoriva, *Identification of neutral tumor evolution across cancer types*. Nat Genet, 2016. **48**(3): p. 238-44.

216. Tirosh, I., A.S. Venteicher, C. Hebert, L.E. Escalante, A.P. Patel, K. Yizhak, J.M. Fisher, C. Rodman, C. Mount, M.G. Filbin, C. Neftel, N. Desai, J. Nyman, B. Izar, C.C. Luo, J.M. Francis, A.A. Patel, M.L. Onozato, N. Riggi, K.J. Livak, D. Gennert, R. Satija, B.V. Nahed, W.T. Curry, R.L. Martuza, R. Mylvaganam, A.J. Iafrate, M.P. Frosch, T.R. Golub, M.N. Rivera, G. Getz, O. Rozenblatt-Rosen, D.P. Cahill, M. Monje, B.E. Bernstein, D.N. Louis, A. Regev, and M.L. Suva, *Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma*. Nature, 2016. **539**(7628): p. 309-313.
217. Cohen, A.J., A. Saiakhova, O. Corradin, J.M. Luppino, K. Lovrenert, C.F. Bartels, J.J. Morrow, S.C. Mack, G. Dhillon, L. Beard, L. Myeroff, M.F. Kalady, J. Willis, J.E. Bradner, R.A. Keri, N.A. Berger, S.M. Pruett-Miller, S.D. Markowitz, and P.C. Scacheri, *Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome*. Nat Commun, 2017. **8**: p. 14400.
218. Cowper-Salari, R., X. Zhang, J.B. Wright, S.D. Bailey, M.D. Cole, J. Eeckhoute, J.H. Moore, and M. Lupien, *Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression*. Nat Genet, 2012. **44**(11): p. 1191-8.
219. Magnani, L., G. Frige, R.M. Gadaleta, G. Corleone, S. Fabris, H. Kempe, P.J. Verschure, I. Barozzi, V. Viricillo, S.P. Hong, Y. Perone, M. Saini, A. Trumpf, G. Viale, A. Neri, S. Ali, M.A. Colleoni, G. Pruneri, and S. Minucci, *Acquired CYP19A1 amplification is an early specific mechanism of aromatase inhibitor resistance in ERalpha metastatic breast cancer*. Nat Genet, 2017. **49**(3): p. 444-450.
220. 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. Graf, G. Ha, G. Haffari, A.

- Bashashati, R. Russell, S. McKinney, A. Langerod, A. Green, E. Provenzano, G. Wishart, S. Pinder, P. Watson, F. Markowitz, L. Murphy, I. Ellis, A. Purushotham, A.L. Borresen-Dale, J.D. Brenton, S. Tavaré, C. Caldas, and S. Aparicio, *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups*. *Nature*, 2012. **486**(7403): p. 346-52.
221. *Comprehensive molecular portraits of human breast tumours*. *Nature*, 2012. **490**(7418): p. 61-70.
222. Györfy, B., A. Lanczky, A.C. Eklund, C. Denkert, J. Budczies, Q. Li, and Z. Szallasi, *An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients*. *Breast Cancer Res Treat*, 2010. **123**(3): p. 725-31.
223. Wang, S., H. Sun, J. Ma, C. Zang, C. Wang, J. Wang, Q. Tang, C.A. Meyer, Y. Zhang, and X.S. Liu, *Target analysis by integration of transcriptome and ChIP-seq data with BETA*. *Nat Protoc*, 2013. **8**(12): p. 2502-15.
224. Neph, S., J. Vierstra, A.B. Stergachis, A.P. Reynolds, E. Haugen, B. Vernot, R.E. Thurman, S. John, R. Sandstrom, A.K. Johnson, M.T. Maurano, R. Humbert, E. Rynes, H. Wang, S. Vong, K. Lee, D. Bates, M. Diegel, V. Roach, D. Dunn, J. Neri, A. Schafer, R.S. Hansen, T. Kutayavin, E. Giste, M. Weaver, T. Canfield, P. Sabo, M. Zhang, G. Balasundaram, R. Byron, M.J. MacCoss, J.M. Akey, M.A. Bender, M. Groudine, R. Kaul, and J.A. Stamatoyannopoulos, *An expansive human regulatory lexicon encoded in transcription factor footprints*. *Nature*, 2012. **489**(7414): p. 83-90.
225. Kittler, R., J. Zhou, S. Hua, L. Ma, Y. Liu, E. Pendleton, C. Cheng, M. Gerstein, and K.P. White, *A comprehensive nuclear receptor network for breast cancer cells*. *Cell Rep*, 2013. **3**(2): p. 538-51.

226. Magnani, L., J. Eeckhoute, and M. Lupien, *Pioneer factors: directing transcriptional regulators within the chromatin environment*. Trends Genet, 2011. **27**(11): p. 465-74.
227. Jeon, Y. and J.T. Lee, *YY1 tethers Xist RNA to the inactive X nucleation center*. Cell, 2011. **146**(1): p. 119-33.
228. Sigova, A.A., B.J. Abraham, X. Ji, B. Molinie, N.M. Hannett, Y.E. Guo, M. Jangi, C.C. Giallourakis, P.A. Sharp, and R.A. Young, *Transcription factor trapping by RNA in gene regulatory elements*. Science, 2015. **350**(6263): p. 978-81.
229. Tang, Z., O.J. Luo, X. Li, M. Zheng, J.J. Zhu, P. Szalaj, P. Trzaskoma, A. Magalska, J. Wlodarczyk, B. Ruszczycki, P. Michalski, E. Piecuch, P. Wang, D. Wang, S.Z. Tian, M. Penrad-Mobayed, L.M. Sachs, X. Ruan, C.L. Wei, E.T. Liu, G.M. Wilczynski, D. Plewczynski, G. Li, and Y. Ruan, *CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription*. Cell, 2015. **163**(7): p. 1611-27.
230. Hurtado, A., K.A. Holmes, C.S. Ross-Innes, D. Schmidt, and J.S. Carroll, *FOXA1 is a key determinant of estrogen receptor function and endocrine response*. Nat Genet, 2011. **43**(1): p. 27-33.
231. Pendharkar, N., A. Gajbhiye, K. Taunk, S. RoyChoudhury, S. Dhali, S. Seal, A. Mane, S. Abhang, M.K. Santra, K. Chaudhury, and S. Rapole, *Quantitative tissue proteomic investigation of invasive ductal carcinoma of breast with luminal B HER2 positive and HER2 enriched subtypes towards potential diagnostic and therapeutic biomarkers*. J Proteomics, 2016. **132**: p. 112-30.
232. Uhlen, M., L. Fagerberg, B.M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K.

- Edlund, E. Lundberg, S. Navani, C.A. Szgyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Ponten, *Proteomics. Tissue-based map of the human proteome*. Science, 2015. **347**(6220): p. 1260419.
233. Uhlen, M., P. Oksvold, L. Fagerberg, E. Lundberg, K. Jonasson, M. Forsberg, M. Zwahlen, C. Kampf, K. Wester, S. Hober, H. Wernerus, L. Bjorling, and F. Ponten, *Towards a knowledge-based Human Protein Atlas*. Nat Biotechnol, 2010. **28**(12): p. 1248-50.
234. Brooks, M.D., M.L. Burness, and M.S. Wicha, *Therapeutic Implications of Cellular Heterogeneity and Plasticity in Breast Cancer*. Cell Stem Cell, 2015. **17**(3): p. 260-71.
235. G, S., *MMDiff: Statistical Testing for ChIP-Seq data*. 2012, Bioconductor. p. Statistical Testing for ChIP-Seq data sets.
236. Shah, S.P., R.D. Morin, J. Khattra, L. Prentice, T. Pugh, A. Burleigh, A. Delaney, K. Gelmon, R. Guliany, J. Senz, C. Steidl, R.A. Holt, S. Jones, M. Sun, G. Leung, R. Moore, T. Severson, G.A. Taylor, A.E. Teschendorff, K. Tse, G. Turashvili, R. Varhol, R.L. Warren, P. Watson, Y. Zhao, C. Caldas, D. Huntsman, M. Hirst, M.A. Marra, and S. Aparicio, *Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution*. Nature, 2009. **461**(7265): p. 809-13.
237. Wan, M., W. Huang, T.E. Kute, L.D. Miller, Q. Zhang, H. Hatcher, J. Wang, D.B. Stovall, G.B. Russell, P.D. Cao, Z. Deng, W. Wang, Q. Zhang, M. Lei, S.V. Torti, S.A. Akman, and G. Sui, *Yin Yang 1 plays an essential role in*

- breast cancer and negatively regulates p27*. Am J Pathol, 2012. **180**(5): p. 2120-33.
238. Shi, Y., E. Seto, L.S. Chang, and T. Shenk, *Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein*. Cell, 1991. **67**(2): p. 377-88.
 239. Shi, Y., J.S. Lee, and K.M. Galvin, *Everything you have ever wanted to know about Yin Yang 1*. Biochim Biophys Acta, 1997. **1332**(2): p. F49-66.
 240. Sui, G., *The regulation of YY1 in tumorigenesis and its targeting potential in cancer therapy*. Molecular and Cellular Pharmacology, 2009. **1**(3): p. 157-176.
 241. Zhang, Q., D.B. Stovall, K. Inoue, and G. Sui, *The oncogenic role of Yin Yang 1*. Crit Rev Oncog, 2011. **16**(3-4): p. 163-97.
 242. Thomas, M.J. and E. Seto, *Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key?* Gene, 1999. **236**(2): p. 197-208.
 243. Gordon, S., G. Akopyan, H. Garban, and B. Bonavida, *Transcription factor YY1: structure, function, and therapeutic implications in cancer biology*. Oncogene, 2006. **25**(8): p. 1125-1142.
 244. Atchison, L., A. Ghias, F. Wilkinson, N. Bonini, and M.L. Atchison, *Transcription factor YY1 functions as a PcG protein in vivo*. The EMBO journal, 2003. **22**(6): p. 1347-1358.
 245. Sankar, N., S. Baluchamy, R. Kadeppagari, G. Singhal, S. Weitzman, and B. Thimmapaya, *p300 provides a corepressor function by cooperating with YY1 and HDAC3 to repress c-Myc*. Oncogene, 2008. **27**(43): p. 5717-5728.

246. Zhou, Q., R.W. Gedrich, and D.A. Engel, *Transcriptional repression of the c-fos gene by YY1 is mediated by a direct interaction with ATF/CREB*. Journal of Virology, 1995. **69**(7): p. 4323-4330.
247. Begon, D.Y., L. Delacroix, D. Vernimmen, P. Jackers, and R. Winkler, *Yin Yang 1 cooperates with activator protein 2 to stimulate ERBB2 gene expression in mammary cancer cells*. Journal of Biological Chemistry, 2005. **280**(26): p. 24428-24434.
248. Saponaro, C., A. Malfettone, T.S. Dell'Endice, A.E. Brunetti, P. Achimas-Cadariu, A. Paradiso, and A. Mangia, *The prognostic value of the Na⁺/H⁺ exchanger regulatory factor 1 (NHERF1) protein in cancer*. Cancer Biomarkers, 2014. **14**(2-3): p. 177-184.
249. Weinman, E.J., D. Steplock, Y. Wang, and S. Shenolikar, *Characterization of a protein cofactor that mediates protein kinase A regulation of the renal brush border membrane Na⁺ (+)-H⁺ exchanger*. Journal of Clinical Investigation, 1995. **95**(5): p. 2143.
250. Cao, T.T., H.W. Deacon, D. Reczek, A. Bretscher, and M. von Zastrow, *A kinase-regulated PDZ-domain interaction controls endocytic sorting of the β 2-adrenergic receptor*. Nature, 1999. **401**(6750): p. 286-290.
251. Lazar, C.S., C.M. Cresson, D.A. Lauffenburger, and G.N. Gill, *The Na⁺/H⁺ exchanger regulatory factor stabilizes epidermal growth factor receptors at the cell surface*. Molecular biology of the cell, 2004. **15**(12): p. 5470-5480.
252. Li, J.-G., C. Chen, and L.-Y. Liu-Chen, *Ezrin-radixin-moesin-binding phosphoprotein-50/Na⁺/H⁺ exchanger regulatory factor (EBP50/NHERF) blocks U50, 488H-induced down-regulation of the human κ opioid receptor by*

- enhancing its recycling rate*. Journal of Biological Chemistry, 2002. **277**(30): p. 27545-27552.
253. Wang, B., Y. Yang, and P.A. Friedman, *Na/H Exchange Regulatory Factor 1, a Novel AKT-associating Protein, Regulates Extracellular Signal-regulated Kinase Signaling through a B-Raf–Mediated Pathway*. Molecular biology of the cell, 2008. **19**(4): p. 1637-1645.
254. Voltz, J.W., E.J. Weinman, and S. Shenolikar, *Expanding the role of NHERF, a PDZ-domain containing protein adapter, to growth regulation*. Oncogene, 2001. **20**(44): p. 6309.

Appendix

DMXL2 drives epithelial to mesenchymal transition in hormonal therapy resistant breast cancer through Notch hyper-activation

Monica Faronato¹, Van T.M. Nguyen¹, Darren K. Patten¹, Ylenia Lombardo¹, Jennifer H. Steel¹, Naina Patel¹, Laura Woodley¹, Sami Shousha¹, Giancarlo Pruneri², R. Charles Coombes¹, Luca Magnani¹

¹Department of Surgery and Cancer, Imperial College London, London, UK

²Division of Pathology, European Institute of Oncology and University of Milan, School of Medicine, London, UK

Correspondence to:

Luca Magnani, e-mail: l.magnani@imperial.ac.uk

Keywords: DMXL2, breast cancer, Notch, endocrine therapy, EMT

Received: March 23, 2015

Accepted: May 22, 2015

Published: June 04, 2015

ABSTRACT

The acquisition of endocrine therapy resistance in estrogen receptor α (ER α) breast cancer patients represents a major clinical problem. Notch signalling has been extensively linked to breast cancer especially in patients who fail to respond to endocrine therapy. Following activation, Notch intracellular domain is released and enters the nucleus where activates transcription of target genes. The numerous steps that cascade after activation of the receptor complicate using Notch as biomarker. Hence, this warrants the development of reliable indicators of Notch activity. DMXL2 is a novel regulator of Notch signalling not yet investigated in breast cancer. Here, we demonstrate that DMXL2 is overexpressed in a subset of endocrine therapy resistant breast cancer cell lines where it promotes epithelial to mesenchymal transition through hyper-activation of Notch signalling via V-ATPase dependent acidification. Following DMXL2 depletion or treatment with Bafilomycin A1, both EMT targets and Notch signalling pathway significantly decrease. We show for the first time that DMXL2 protein levels are significantly increased in ER α positive breast cancer patients that progress after endocrine therapy. Finally, we demonstrate that DMXL2 is a transmembrane protein with a potential extra-cellular domain. These findings identify DMXL2 as a novel, functional biomarker for ER α positive breast cancer.

INTRODUCTION

Notch signalling is frequently deregulated in estrogen receptor positive breast cancer and represents a challenging but important therapeutic opportunity [1]. At least in some respect, difficulties in targeting Notch arise from the complex molecular features of this pathway [2]. In mammals, there are four Notch receptors and five Notch ligands. The activity of both is regulated by endocytic trafficking, although the detailed mechanisms are yet to be fully defined [3]. Data suggest that defective endosomes maturation could cause aberrant Notch signalling by interfering with γ -secretase (GS) mediated Notch cleavage [4]. At present, Notch can be targeted in the clinic with γ -secretase inhibitors (GSI), a series of compounds designed to block Notch cleavage and nuclear translocation of Notch intracellular domain. Furthermore,

antibodies against Notch are currently being tested [5] although patients reported side effects similar to the ones due to the use of GSIs [6]. Notch plays a pivotal role in many processes including embryonic development, carcinogenesis, cancer stem cells maintenance, angiogenesis, and epithelial to mesenchymal transition (EMT) [7–13].

Aberrant Notch signalling has been linked to breast cancer formation and progression [14]. Breast cancer remains a major clinical problem worldwide, accounting for over for 23% of all cancer deaths [15]. Nearly 70% of breast cancers express estrogen receptor α (ER α) and are treated with adjuvant endocrine therapy designed to block ER α driven transcriptional programs. Alas, around 40% of patients evolve resistance to endocrine therapies and progress towards more advanced disease [16–18]. The mechanisms behind drug resistance have not yet

been fully elucidated, although work from our and other groups proposes that epigenetic reprogramming leading to increased Notch signalling may contribute to it [19–21]. Notch involvement in endocrine therapy resistance makes it attractive for developing targeted therapies. Thus, there is a strong need for Notch based biomarkers, especially with potential applicability for clinical assays. This prompted us to investigate DMXL2, which was recently presented as a novel Notch modulator. The *DMXL2* gene has been recently described as a novel player in Notch signalling, regulating the acidification of intracellular compartments through the vacuolar protonic pump (V-ATPase) both in *Drosophila* and in mammalian systems [22, 23]. The V-ATPase pump plays an important role in vesicular trafficking alongside endocytic and exocytic traffic [24]. We investigated the role of DMXL2 in the context of endocrine therapy resistance using patients' derived clinical specimens and long-term estrogen deprived LTED cells [25–28]. Our results demonstrate that DMXL2 regulates Notch cleavage and chromatin recruitment, epithelial to mesenchymal transition, invasion and migration of endocrine therapy resistant breast cancer cells. Hence, DMXL2 can be identified as novel biomarker in ER α positive breast cancer patients.

RESULTS

Notch signalling is dependent on notch accumulation in the chromatin

We have previously shown that Notch signalling plays a significant role in the development of endocrine therapy resistance [18, 19, 29, 30]. Notch signalling is partly dependent on the absolute amounts of Notch receptors present in the cells. However, there is substantial debate on the importance of different Notch receptors in endocrine resistance (reviewed in [31]). We reasoned that the rate of Notch activation, via its cleavage, is equally important to downstream signalling controlling tumor cell growth and survival. To test this, we first quantified total Notch levels using two isogenic cell lines: MCF7 and long-term estrogen deprived (LTED). The latter were derived from MCF7 following one year estrogen withdrawal to mimic endocrine therapy resistant breast cancer [28]. Notably, we find that all the receptors are downregulated at the protein level (Figure 1A). Notch1 is also downregulated at the mRNA level, whereas Notch3 and Notch4 mRNA levels are not (Figure 1B). Paradoxically, LTED cells contain higher protein levels of the Notch target Hey2, a well-characterized estrogen

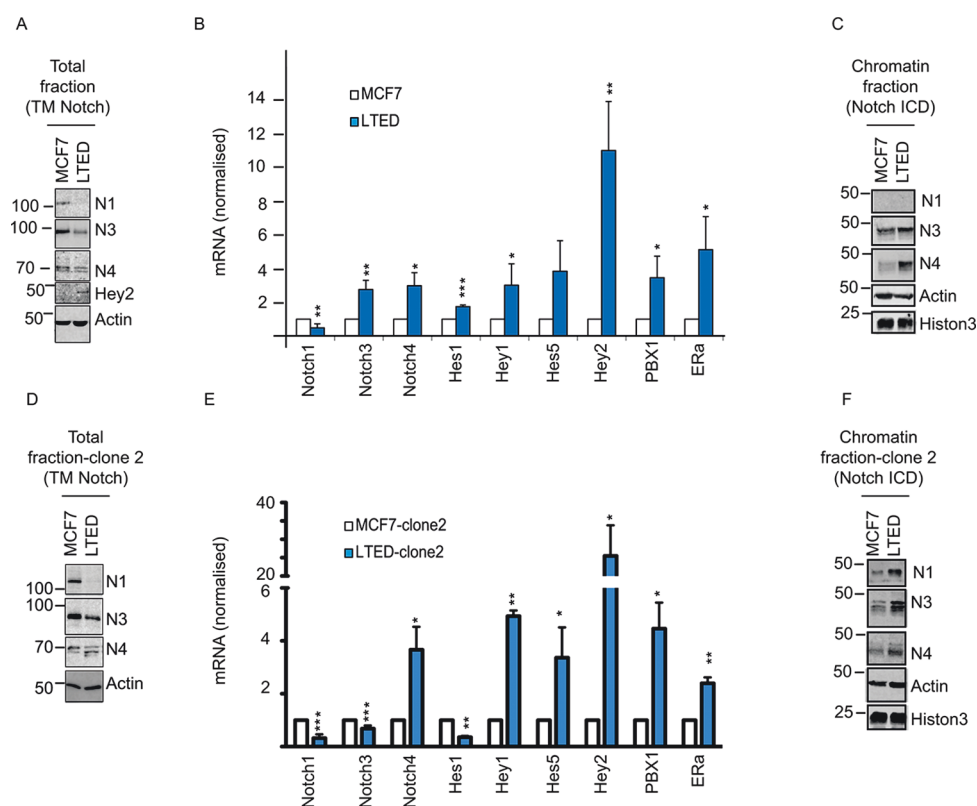


Figure 1: Notch pathway is upregulated in LTED and is dependent on DMXL2 overexpression. A–D. Representative western blot showing (reported) Notch transmembrane domain (TM) and Notch target Hey2 in MCF7 and LTED cells (clone 1 and 2 respectively) Membranes were blotted with commercially available antibodies against Notch total level B–E, q(RT) PCR mRNA normalised to 28S of Notch receptors and Notch targets C–F. Cells were fractionated and the chromatin fraction was loaded. Membranes were blotted with in house-developed antibodies against the intracellular domain (Notch ICD).

target gene [32] suggesting that receptor levels alone could not accurately predict Notch signalling (Figure 1A). LTED cells have impaired ER α signalling as a consequence of estradiol starvation, thus suggesting that Hey2 may be uniquely regulated by Notch in endocrine therapy resistant cells. Furthermore, we find that numerous additional Notch targets are also transcriptionally upregulated in LTED cells despite a general reduction in Notch receptor (Figure 1B). The discrepancy between total Notch protein levels and Notch target genes might be explained by hyper-activated Notch signalling via increased cleavage rates. To test this hypothesis, we measured Notch ICDs by chromatin fractionation and western blot, using ICD specific antibodies, which were designed and fully characterised in house (Supplementary Figure S1A–1B). Interestingly, we find that the LTED chromatin fraction presents greater enrichment for the Notch3 and Notch4 ICDs (Figure 1C). Importantly, we also screened a second LTED model, derived in a different lab (LTED clone 2) [10]. We confirm that the total amount of Notch is downregulated at protein level (Figure 1D) whereas Notch ICDs are upregulated in the chromatin fraction (Figure 1F). We also find Notch target genes are transcriptionally upregulated in the second clone (Figure 1E) thus confirming hyper-activated Notch signalling in endocrine therapy resistant cells.

DMXL2 is over-expressed in metastatic breast tumors and is epigenetically activated in endocrine therapy resistant (ETR) cell lines

The recent discovery of DMXL2 as novel Notch regulator, prompted us to look into clinical setting. We built a case set including biopsies from ER α positive breast cancer patients treated with endocrine therapies. For each patient we measured DMXL2 in untreated primary breast cancer samples and matched metastatic samples resistant to endocrine therapy using IHC which revealed that DMXL2 had a membrane-cytoplasmic staining (Figure 2A). Using a pair-wise comparison we found a significant increase in DMXL2 staining in the metastatic samples (Figure 2B), suggesting that DMXL2 is upregulated in endocrine therapy resistant cells *in vivo*. Based on our findings *in vivo*, we decided to explore DMXL2 protein levels using a comprehensive panel of endocrine resistant cells including MCF7 derived tamoxifen or fulvestrant resistant (MCF7-TAMR, MCF7-FULVR), as well as MCF7 cultured in the absence of Estrogen (LTED). In addition, we have analysed two LTED-derived models, which also developed sequential resistance (LTED-TAMR and LTED-FULVR) [28]. Our data indicate that DMXL2 is upregulated in all the LTED models (Figure 2C) while MCF7-TAMR and MCF7-FULVR do not exhibit increased DMXL2 expression suggesting endocrine therapy specific modulation of DMXL2. In addition, we confirmed that changes in DMXL2 expression are not related to different culturing

conditions (e.g. chronic drug exposure). LTED maintained high levels of DMXL2 even in the presence of parental MCF7 media (supplemented with E2), thus confirming that DMXL2 is not estradiol dependent but it is epigenetically stably overexpressed (Figure 2D). We also show that DMXL2 is significantly upregulated in the second clone (Figure 2E). DMXL2 overexpression is accompanied by increased chromatin marks typical of active regulatory elements (H3K27ac) at the DMXL2 promoter and nearby putative enhancers (Figure 2F). Similarly, the chromatin near DMXL2 regulatory elements is more accessible to transcription factor regulation in LTED compared to MCF7 lines as shown by DNaseI hypersensitivity assays followed by next generation sequencing. (DHS-seq, Figure 2F). All the following experiments were performed using LTED clone 1 (henceforth identified as LTED).

DMXL2 depletion impairs notch signalling in endocrine therapy resistant cells

Previous evidence suggests that DMXL2 might regulate Notch in breast cancer cells and that this regulation is dependent on the interaction between DMXL2 and the V-ATPase pump. [23, 33]. Initially, we postulated that high levels of DMXL2 in endocrine resistant cells lines might result in increased Notch signalling. To test the role of DMXL2 in Notch signalling in the resistant cells, we measured Notch targets following DMXL2 depletion. As expected, siRNA mediated DMXL2 knock-down (Supplementary Figure S2A) significantly impairs Notch downstream signalling (Figure 3A). Furthermore, DMXL2 depleted LTED display less nuclear Hey2 protein levels demonstrating that DMXL2 is required in Notch signalling (Figure 3B). Importantly, following DMXL2 depletion we show augmented trans-membrane (TM) Notch1 and Notch4 at protein and mRNA levels (Figure 3C). Strikingly, this is associated with a decrease in the chromatin bound Notch intracellular domains (NICDs) (Figure 3D). These data strongly suggest that DMXL2 regulates Notch cleavage in endocrine therapy resistant breast cancer cell lines and mediates NICD chromatin accumulation.

V-ATPase inhibition mimics DMXL2 depletion

DMXL2 has been described as a regulator of the V-ATPase pump in drosophila and mammalian systems. Furthermore, there is a growing body of evidence linking oncogenic Notch signalling to V-ATPase pump in mammals. [23, 34, 35]. First, we wanted to confirm that DMXL2 is required for V-ATPase activity. To do so, we performed DMXL2 depletion and measured pH by LysoTracker staining [36]. As expected, LTED cells depleted for DMXL2 show reduced LysoTracker staining confirming that DMXL2 is a pump regulator (Figure 4A). To directly inhibit the pump, we employed Bafilomycin A1 [37] and measured pH changes in the intracellular

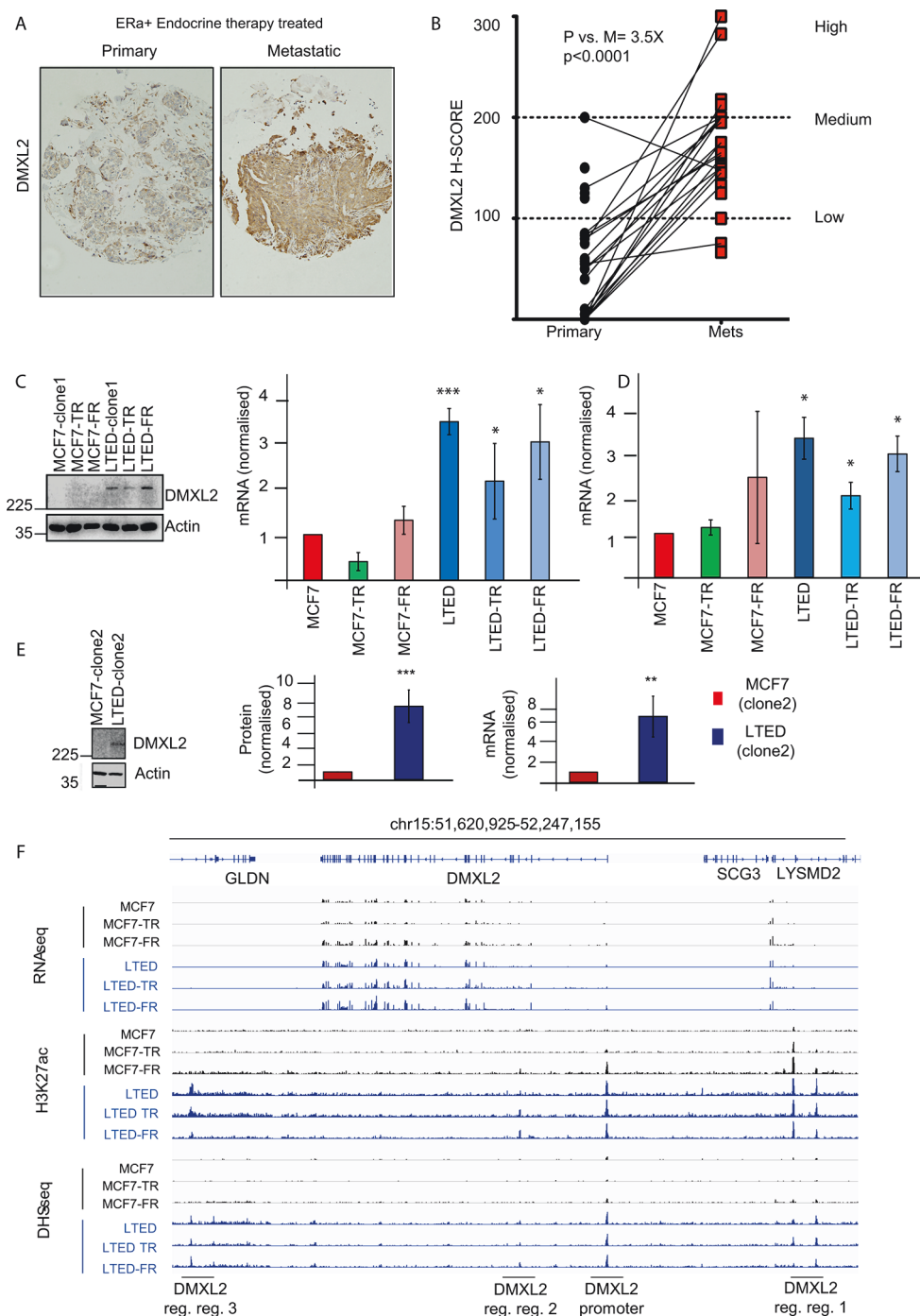


Figure 2: DMXL2 is overexpressed in metastatic tumors resistant to endocrine therapy. **A.** Primary and matched Metastatic samples were processed using DMXL2 IHC **B.** H scores were plotted. Pair-wise *T*-test between the average score (3 independent scorers, duplicate sections) was used to establish significance. DMXL2 is overexpressed in ETR cells and it is epigenetically activated. **C.** Representative western blot showing DMXL2 is overexpressed in LTED resistant cells. Graph shows protein normalised to loading control (Actin). q(RT) PCR mRNA levels normalised to 28S. In (C) cells were cultured as described in material and methods. In **D.** cells were cultured in MCF7 media (3 days). (*) $P < 0.05$ (**) $P < 0.01$ (***) $P < 0.001$ **E.** Different clone of MCF7 and LTED showing DMXL2 upregulation. **F.** Snapshots from DNaseI hypersensitivity site (DHS-seq) and Chip-seq for K27ac histone mark near DMXL2 promoter across all 6 breast cancer cell models. Chip-seq tracks display the active regulatory elements near DMXL2 promoter are uniquely found in estrogen deprived cell lines (LTED cells), indicating the activation of DMXL2 gene in LTED cells, but not MCF7 cells.

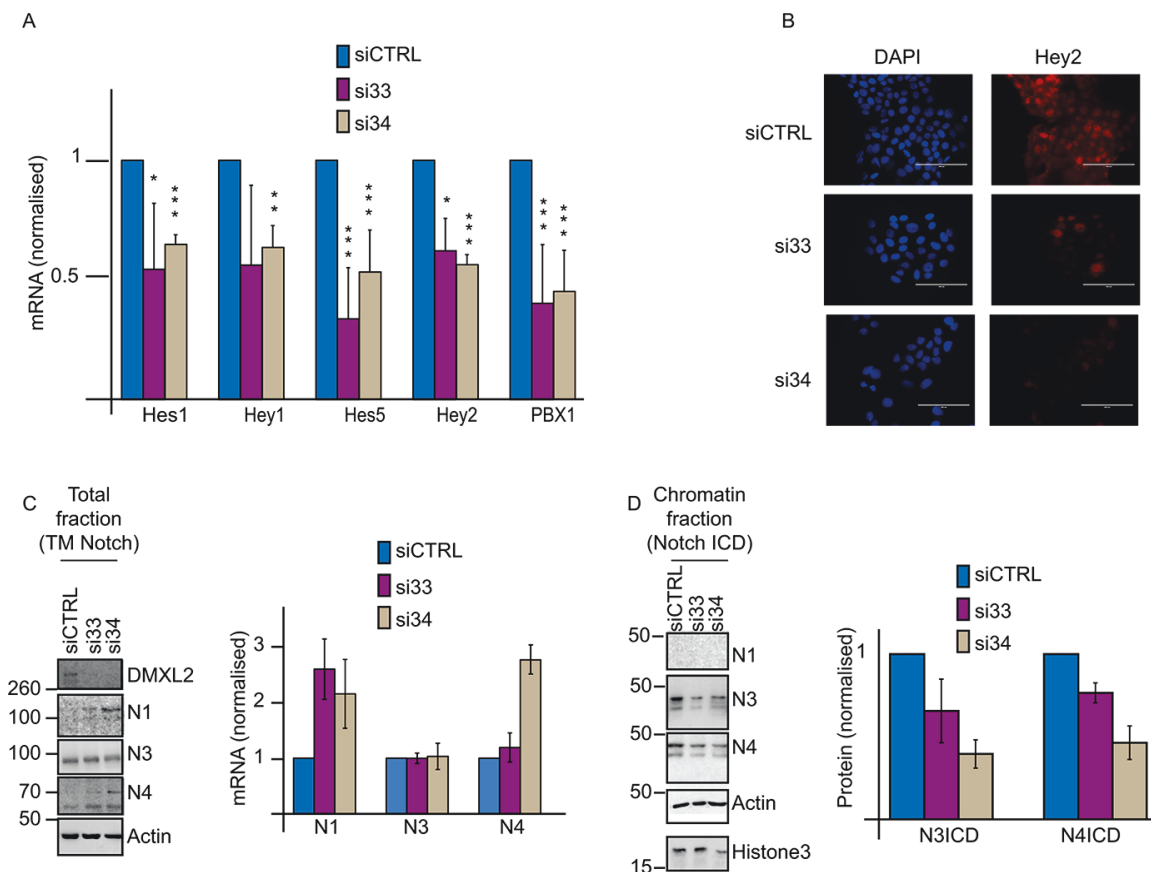


Figure 3: DMXL2 regulates Notch pathway. **A.** q(RT) PCR mRNA levels of Notch targets following DMXL2 depletion **B.** Representative immunofluorescence images of Hey2 following siDMXL2 showing Hey2 nuclear signal was impaired. Bars represent 400 μ m **C.** Representative western blot showing reported TM Notch accumulates following DMXL2 depletion. Antibodies were used as in 1A **D.** Representative blot showing Notch3 and 4ICD reduced chromatin accumulation following siDMXL2. Antibodies were used as in 2C. Knock down was performed for 48 or 72 hrs (for mRNA or protein extraction respectively). (*) $P < 0.05$ (**) $P < 0.01$ (***) $P < 0.001$. Every experiment is an average of three independent experiments. Bars represent standard deviation.

compartment. Bafilomycin A1 disrupts V-ATPase function, as shown by the reduction in LysoTracker staining (Figure 4B). These experiments were performed using 125 nM Bafilomycin A1, a dose that does not affect cell proliferation (data not shown). To confirm that Notch signalling is dependent on the V-ATPase activity in endocrine therapy resistant breast cancer cells, we monitored Notch intracellular domain (NICD) accumulation into the chromatin following Bafilomycin A1 treatment. We demonstrate that Notch3 and 4-ICD stop accumulating in the nucleus in a time dependent manner following Bafilomycin A1 treatment (Figure 4C) mirroring the effect of DMXL2 depletion (Figure 3D). In agreement, Notch targets are transcriptionally silenced at the same rate (Figure 4D). To further confirm that DMXL2 is upstream of Notch signalling, we performed a rescue experiment. We used Bafilomycin A1 to down-regulated Notch targets and then over-expressed Notch3 and Notch4ICDs. Notably, we were able to completely rescue their expression (Figure 4E). Overall, these data strongly suggest that increased Notch signalling in LTED

is linked to V-ATPase acidification and that this is likely to be mediated by DMXL2 overexpression.

DMXL2 over-expression correlates with acquisition of EMT phenotype

Within our panel of endocrine therapy specific cells, we noticed that LTED were uniquely associated with migratory and invasive phenotypes (Nguyen *et al.* under revision). Importantly, Notch signalling has been linked to the epithelial to mesenchymal transition (EMT) phenotype [29, 38]. Transcriptional and protein assays confirm that key EMT genes (ZEB1, IQGAP1, CDC42, and CD44) are significantly upregulated in LTED cells at mRNA levels and at protein levels (ZEB1 and IQGAP1). Moreover, loss of E-Cadherin and ZO1, a tight junction marker [39], were downregulated thus confirming LTED cells underwent EMT (Figure 5A–5B). Recently, Tuttle and colleagues proposed that DMXL2 controls migration in Zebrafish Neural Crest [40] through its associated binding partner V-ATPase pump. Therefore we hypothesized that overexpressed DMXL2 could play a role

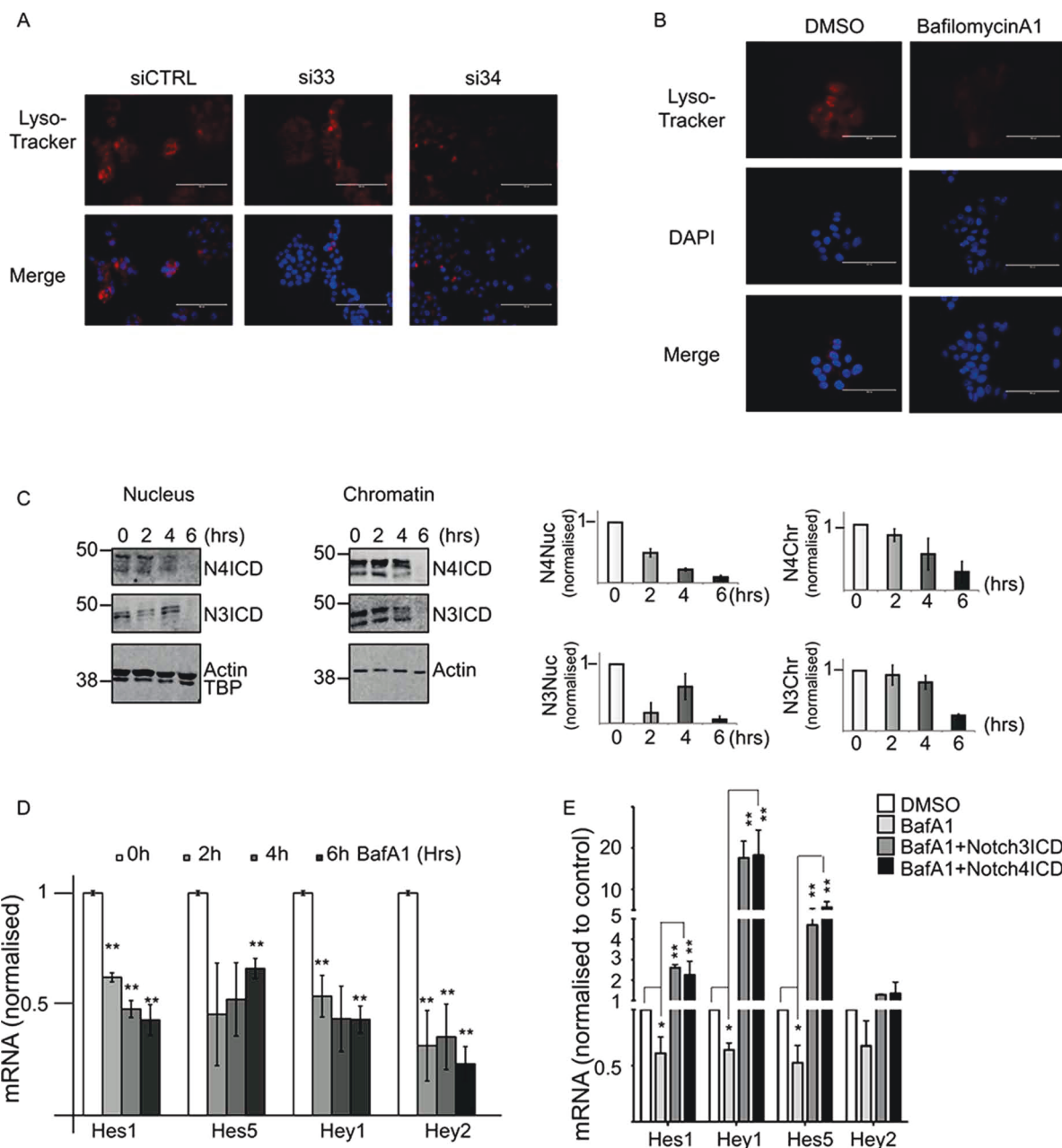


Figure 4: DMXL2 depletion is mirrored by *V-ATPase* inhibition. **A.** Representative immunofluorescence showing reduced LysoTracker staining following 72 hours DMXL2 depletion. Bars represent 400 μ M. **B.** LysoTracker staining of acidic intracellular compartments in LTED cells following Bafilomycin A1 treatment (6 hours). Bars represent 400 μ M. **C.** Representative western blot showing Notch ICDs stop accumulating in the nucleus and the chromatin fraction following Bafilomycin A1 (BafA1) treatment in a time dependent manner. **D.** q(RT) PCR mRNA of Notch targets levels following BafA1 treatment. In B, C, and D cells were treated with 125 pM of Bafilomycin A1. **E.** Notch targets are rescued by ectopic expression of Notch ICD. Cells were pre-treated for 6 hrs with 500 pM of BafA1 followed by 24 hrs Notch ICD over-expression. Every experiment is an average of three independent experiments. Bars represent standard deviation (* $P < 0.05$ (** $P < 0.01$ (***) $P < 0.001$).

in determining the invasive behaviour of LTED cell lines. DMXL2 depletion leads to a significant downregulation of the majority of EMT genes (IQGAP1 5C, D and ZEB1, CD44, and RAC2 in Figure 5D) whilst tight junction marker ZO1 was upregulated (Figure 5C–5D). Strikingly, DMXL2 depletion also resulted in a significant decrease in real time migration and invasion cells capacity (Figure 5E–5F).

These data suggest that DMXL2 is involved in the aggressive phenotype exhibited by LTED cells. Notably, DMXL2 depletion did not affect cell viability (Supplementary Figure S2B). Next, we confirmed that DMXL2 siRNA mediated knock down using a 3D invasion assay. Remarkably, DMXL2 depletion significantly impairs the formation of pseudopodia and the invasion of the

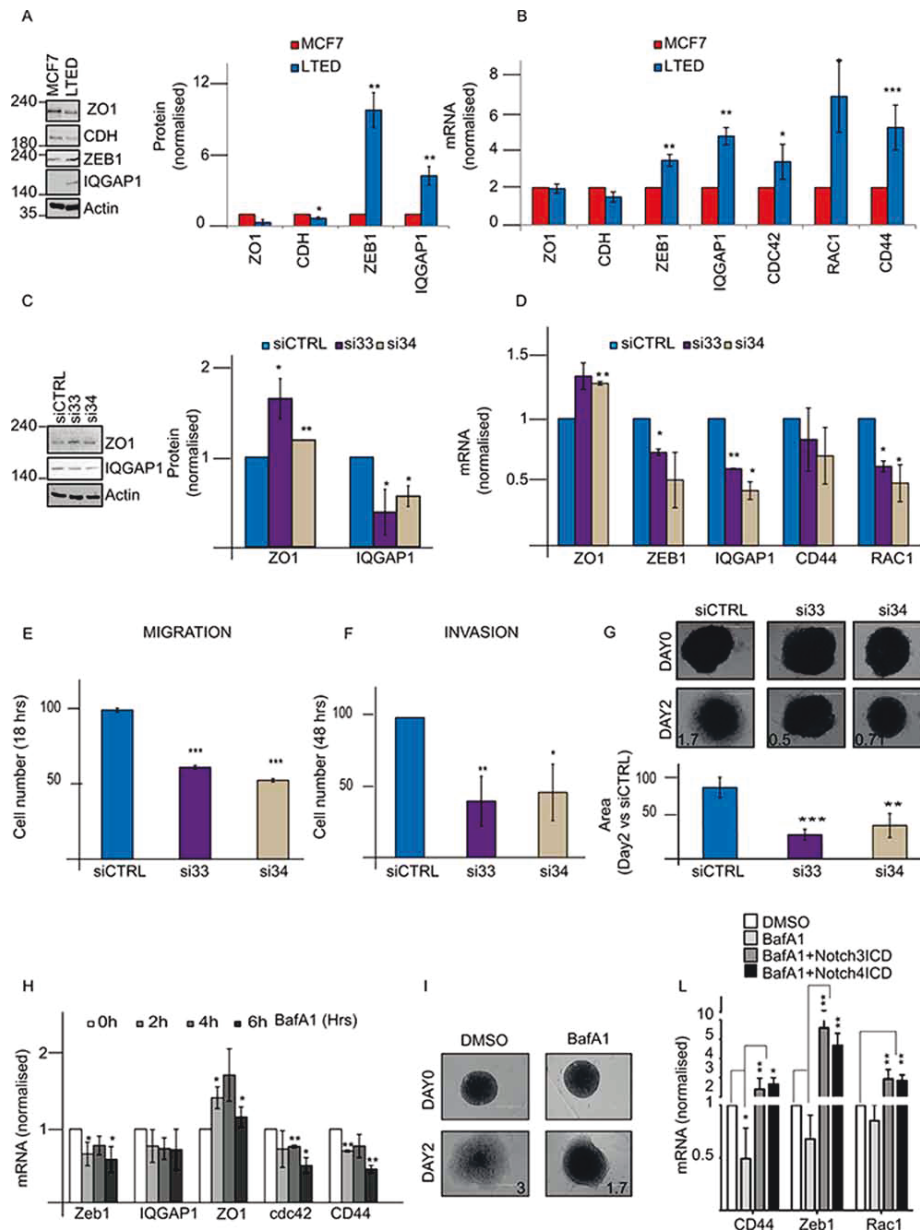


Figure 5: DMXL2 regulates mesenchymal switch in LTED cells. **A.** Representative western blot showing EMT targets genes in LTED cells compared to MCF7 **B.** q(RT) PCR mRNA levels of EMT targets normalised to 28S **C.** Representative western blot showing regulation of EMT genes following DMXL2 knock down **D.** q(RT) PCR mRNA of EMT genes after siDMXL2 **E, F.** Knock down was performed on 6 well plate for 24/48 hours. Cells were harvested and counted. 40,000 were seeded on 16 wells ECM plates with/without matrigel (for invasion and migration respectively) and left migrating or invading for 48/24 hours respectively. The graphs represent the difference in cell number compared to siControl **G.** 3D Invasion assay organoids. Representative pictures of organoids embedded in matrigel after 5 days knock down (performed during the hanging time) at day 0 and 2. The graph represents the difference in the area between day 0 (cells just embedded) and day 2 (cells left invading. Bars represent 400 uM. **H.** Bafilomycin A1 impairs cell invasion. Cells were treated as in 4C. qRT-PCR mRNA for EMT genes is shown. **I.** Representative pictures of organoids. Medium containing 500 pM BafA1 was added to the medium for 48 hours after embedding the organoids in matrigel. Day 0 and day 2 images are shown. **L.** EMT targets are rescued by ectopic expression of Notch ICD. Cells were pre-treated for 6 hrs with 500 pM of BafA1 followed by 24 hours Notch ICD over-expression. Every experiment is an average of three independent experiments. Bars represent standard deviation (*) $P < 0.05$ (**) $P < 0.01$ (***) $P < 0.001$.

extra-cellular matrix in LTED cells (Figure 5G). We then wanted to confirm these findings using Bafilomycin A1 since we have shown that direct inhibition of the V-ATPase pump phenocopy DMXL2 depletion. Interestingly, main EMT targets are significantly reduced upon Bafilomycin A1

treatment and the tight junction marker ZO1 is upregulated (Figure 5H). Furthermore, Bafilomycin A1 treatment significantly impairs invasion in our 3D invasion assay (Figure 5I). Moreover, in agreement with our previous results (Figure 4E) Notch ICD overexpression was sufficient

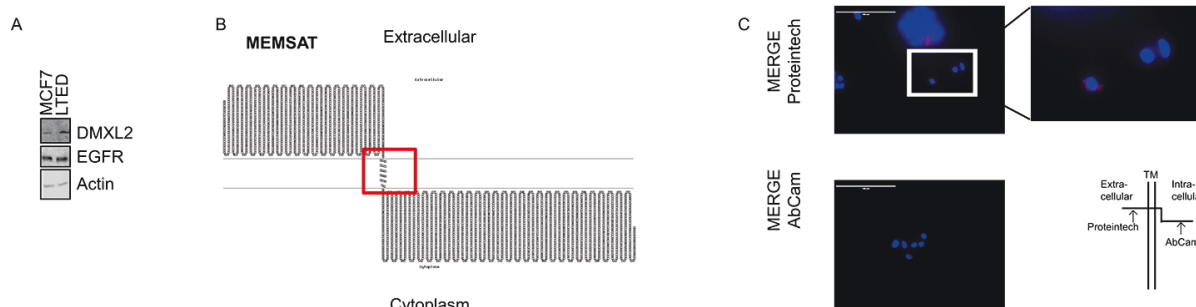


Figure 6: DMXL2 has a potential extracellular domain. **A.** Western blot showing cells labeled with biotin and processed as described in material and methods. EGFR and Actin are used as controls **B.** Graphs showing DMXL2 putative transmembrane domain **C.** Representative live immunofluorescence images showing DMXL2 is detected only with the antibody interacting with the potential extracellular domain (AbCam antibody). Bars represent 400 μ M.

to rescue EMT genes expression in Bafylomycin treated cells (Figure 5L). Altogether, these evidences strongly support DMXL2 as a major player in determining mesenchymal switch in our endocrine resistant cell line model.

DXML2 has a potential extracellular domain

The V-ATPase pump complex consists of two different domains of which one is a peripheral membrane sub-complex. DXML2 is thought to control the modulation of the pump assembly/disassembly process [41, 42]. We speculated that DMXL2 could also be a peripheral membrane protein. To test this, we first employed EZ-Link Sulfo-NHS-LC-LC-Biotin to label cell surface proteins and confirm an increase of signal from DMXL2 protein outside LTED cells membrane (Figure 6A). Next, we used several publicly available informatics predictive algorithms (e.g. MEMSAT, PSIPRED, TMHMM, TMPRED [43–46] to test if DMXL2 carries potential transmembrane domains. Notably, several algorithms predicted the same potential transmembrane domain (Figure 6B and Supplementary Figure S3). We further investigated this possibility, using two different antibodies: one mapping onto the potential outer domain, and one mapping onto the putative cytoplasmic end (They were both fully characterised in Supplementary Figure S2A–2D). Live immunofluorescence without cell permeabilization shows that only the antibody interacting with the potential extra-cellular domain could detect DMXL2 (Figure 6C). Overall, these data strongly suggest that DMXL2 could be a peripheral membrane protein with a potentially targetable extracellular domain.

DISCUSSION

Here we demonstrate that DMXL2 overexpression drives Notch signalling and EMT transition in endocrine therapy resistant breast cancer cells. DMXL2 activity impinges on the V-ATPase pump and can be antagonized using siRNA or chemical inhibitors. More importantly, DMXL2 increases *in vivo* in patients who relapse following treatment with endocrine therapy.

Notch is involved in numerous pathways and plays a significant role in tumor progression and metastasis [10]. However, there is still a lack of understanding of the molecular details of Notch signalling, especially in cancer cells. On this ground, we investigated DMXL2 as it was recently discovered as a novel potential Notch regulator [22, 23]. We show that DMXL2 expression significantly correlates with the development of endocrine resistance. However we cannot exclude that a resistant pre-existent population with high levels of DMXL2 emerge during chronic exposure to treatment. At any rate our data provide novel insights that explain apparent contradictory results such as that ETR cells have enhanced Notch signalling while having lower total Notch receptors. This suggests the processing rate of the Notch receptors cleavage might be more accurate in predicting actual Notch signalling compared the total unprocessed protein levels. DMXL2 appears to be essential in this process as direct DMXL2 suppression or blockage of the V-ATPase activity result in an important reduction of NICD accumulation in the chromatin fraction.

Drug resistance is a major hurdle for patients with ER α breast cancer. A spectrum of published research supports Notch as a key player in drug resistance [18, 19, 29, 47]; therefore targeting this pathway is becoming more and more attractive [48, 49]. The mechanisms behind drug resistance are not completely understood but increasing evidence suggests that resistant cells may epigenetically activate alternative pathways [19, 50]. Our model suggests that DMXL2 is upregulated in response to endocrine therapy treatment and contributes to a the more aggressive phenotype observed in these cells.

A recent work from Tuttle and colleagues uncovered a potential role of DMXL2 and V-ATPase in EMT/migration of neuronal crest in Zebrafish models. Based on these finding and on the highlight of our recent work [29], we wanted to explore the role of DMXL2 in driving the mesenchymal switch in therapy resistant breast cancer. We reveal for the first time that DMXL2 is responsible for driving the mesenchymal switch in ETR cells and that blocking DMXL2 significantly impairs both migration and invasion. We also demonstrate that DMXL2 is a V-ATPase interacting partner

and the pump acidification controls Notch signalling. Indeed, blocking V-ATPase using Bafilomycin A1 mirrors DMXL2 depletion and impairs Notch binding to the chromatin as well as invasion. DMXL2 overexpression in ETR cell lines, and its implication in metastatic development, gave us the idea of looking in the clinical setting. To our knowledge, this is the first report indicating that DMXL2 protein is present in patients with ER α breast cancer. Notably, DMXL2 increases dramatically during breast cancer progression, in agreement with our cell line model.

V-ATPase inhibition, through Bafilomycin A1, causes tumor regression and metastasis inhibition *in vivo* and *in vitro* [51]. Based on these facts, we reasoned that manipulating Notch via DMXL2 inhibition could be an exciting route that warrants further investigation. At the moment, the approach taken to target Notch signalling had been through the use of GSIs that inhibit signalling by all four Notch receptors [52, 53]. Our recent works showed that using monoclonal antibodies to target the extra-cellular domain of Nicastrin, the catalytic subunit of the GS complex, greatly reduced the tumor bulk *in vivo* and limits invasion and migration *in vitro* [29, 54]. Here, we show that DMXL2 likely has an extracellular portion. This warrants further investigation focusing on the development of monoclonal antibodies targeting DMXL2 as an alternative option to block Notch in breast cancer.

MATERIALS AND METHODS

Cell culture

Cells were originated as previously described [28]. MCF7 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS). LTED were maintained in phenol-red free DMEM containing 10% charcoal stripped fetal calf serum (SFCS). Both media were supplemented with 2 mM L-glutamine, 100 units/mL penicillin, 0.1 mg/mL. 10^{-8} M Estradiol (E2758 Sigma) was added routinely to MCF7. Cells used in Figures 1 and 2 and named "clone 2", were independently derived [19]. 10^{-7} Tamoxifen (H7904 Sigma) and Fulvestrant (I4409 Sigma) were added to TAMR and FULVR cells.

Migration and invasion assay

This assay was performed as previously described (Nguyen et al., under revision). Briefly, real time cell analysis (RTCA) was carried on using xCELLigence technology (Cambridge Bioscience), and 40,000 cells were seeded on a well of 16-transwell CIM-PLATES with/without matrigel to monitor migration and invasion respectively. Prior to each experiment, cells were subjected to 48/24 hrs DMXL2 knock down, harvested, counted, and then allowed to migrate or invade for 24–48 hours respectively (total knock down 72 hrs). Serum free medium was used to resuspend and seed the cells, and 10% FCS or DC/FCS full medium was used as chemo-attractant.

RNA extraction and real time RT-PCR

RNA extraction and cDNA preparation were performed as previously described [29]. Briefly, QIAGEN columns were used for RNA extraction. 2 μ g were retro-transcribed using High Capacity cDNA Reverse Transcription (AB 4368814 Applied Biosystem). 2X SYBR GREEN mix (4309155 Invitrogen) was used to perform quantitative real time RT-PCR. The PCR amplification was as follows: 95°C (10 min), 95°C (10 sec), 60°C (30 sec), 72°C (30 seconds), 40 cycles followed by melting curve analysis. Melting curves were analysed after 40 cycles. Expression levels of each gene were evaluated using comparative threshold cycle (Ct) method using the $2^{-\Delta\Delta C_t}$ method with normalization to 28S housekeeping gene. Primers sequences are available upon request.

siRNA

Silencers Select Small interfering RNA (siRNA) against DMXL2 were purchased from Invitrogen (s23533, s23534). Silencer Select Negative 1 was used as control (4404020). Briefly, 300,000 cells were seeded on a 6 well plate. The following day Lipofectamine 2000 (11668) was used according to the manufacturer protocol. 5nM siRNA were used as final concentration. RNA and proteins were analysed 24 and 72 hours later respectively.

Overexpression

Notch1, Notch3 and Notch4 ICDs were a kind gift from Keith Brennan. Plasmids were fully sequenced. Lipofectamine 2000 was used according to the instructions. 300,000 cells were seeded on day 1. The following day, 1 μ g of DNA was used for transfection. Cells were harvested and lysed 24 hrs later.

Rescue experiment

LTED cells (200,000) were seeded on 6 well plates. The following day 500pM Bafilomycin A1 was added to the medium. After 6 hours, cells were transfected with 1 μ g of Notch3, 4 plasmid using Lipofectamine 2000 according to the manufacturer instructions. 24 hours later, mRNA was extracted according to the manufacturer instructions (QIAGEN). DMSO was used as control.

3D organoids assay (hanging drop assay)

A 3D assay was performed as previously described (Nguyen et al. under revision). Briefly, 250,000 cells were resuspended in 1ml of medium. Drops (20 μ l) were placed onto the lid of a 10 cm dish, then the lid was flipped onto of the dish, containing 5 ml of media to prevent evaporation. The dish was then incubated for 5 days at 37°C. On day 5, the newly formed organoids were mounted on 10ul phenol red free matrigel (356237) on a coverslip

in a 24 well plate. Matrigel was left for 1 hr at 37°C in the incubator to solidify, then 500 ul of full medium was added. Pictures were taken at day 0 and day 2. In knock down experiments, siDMXL2 or siControl were added to the cells before making the drops. In Bafilomycin A1 experiments, the drug was added to the medium on day 0 for 48 hours.

Immunofluorescence

Briefly, cells were seeded on coverslips on a 6 well plate. On the final day, cells were washed twice with room temperature PBS and 4% PFA/PBS was added for 15 minutes. Cells were washed twice with PBS and NH₄Cl was added as quencher. 1% Triton/PBS was added for 5 minutes. 10% BSA/PBS was used as blocking reagent. 5% BSA/PBS was used to dilute antibodies and Alexa-fluor 488 or 594labelled anti-rabbit or anti-mouse secondary antibodies. Cells were mounted in mounting medium (MOWIOL) containing 4',6-diamidino-2-phenylindole. For live immunofluorescence, no fixative was used. Pictures were acquired using EVOS microscope system (Advanced Microscopy Group, Bothell, WA, USA) or Leica SP5 confocal microscope.

Antibodies-drugs

We used two antibodies against DMXL2 (24415-1-AP from Proteintech) for western blot and immunofluorescence and ab122552 (abCam) for IHC and immunofluorescence. Antibodies against Notch1 and Notch3 were purchased from Cell Signalling (3608, and 2889) and Notch4 was purchased from Santa Cruz (Sc-5594 H-225). Actin and V5 antibodies were purchased from abCam (ab6276 and ab9116). ZEB1 (3396), ZO1 (8193), E-Cadherin (3195), and IQGAP1 (2293) antibodies were purchased from Cell Signalling. Bafilomycin A1 was purchased from Sigma (B1793). TBP was purchased from Santa Cruz. LysoTracker (LysoTracker® Red DND-99 Life Technologies) was used according to the manufacturer protocol for labelling acidic organelles in live cells following Bafilomycin A1 and DMXL2 siRNA treatments. In knock down experiments, the LysoTracker was added after 72 hours. Estradiol for culturing MCF7 was also purchased from Sigma and diluted in 100% Ethanol in a final concentration of 10⁻⁷M. Biotinylation assay was performed according to the manufacturer protocol (21435 Thermo Scientific) GSI was purchased from Pfizer, Inc (New York, NY, USA) (PF03084014). Dll4 ligand was purchased from abCam (108557).

Western blot and fractionation

RIPA (Sigma) was used to prepare total lysate. Fractionation method was performed as previously described [55]. Bicinchoninic Acid (BCA) assay (Thermo Scientific) was used to determine protein concentration. Equal amounts of lysates were subjected to immunoblotting on SDS-PAGE. IRDyes IR680, IR680-LT,

or IR800 (LI-COR) were used as secondary antibodies. LI-COR Odyssey 2.1 system was used to visualise proteins. 16-bit images were analyzed and quantified using Image J.

Tissue microarray (TMA) of paired primary and secondary breast cancers

Twenty primary breast carcinomas with a paired metastasis were acquired from the pathology archives of Charing Cross Hospital, London, UK. A tissue microarray was constructed using a manual microarray (Beecher) and 0.6 mm punches. The tissue microarray was immunohistochemically profiled for DMXL2 and other biological antibodies as previously described (ref?). DMXL2 rabbit antibody (ab122552, abCam) was optimized to a working concentration (1:500 in PBS); utilizing tissue sections (5 µm). Antigen retrieval was performed using 0.01M citrate buffer, pH 6.0 followed by blocking in 0.3% hydrogen peroxide in PBS, then in normal goat serum (20 µl per ml) for 30 min. The primary antibody was incubated overnight at 4°C, and then detected using anti-rabbit secondary antibody (Vector Laboratories), Vectastain Elite peroxidase ABC kit, and ImmPACT DAB kit (Vector Laboratories). Subsequently, 4 µm TMA sections were immuno-stained using the optimized staining protocol, including negative controls (omission of the primary antibody). Staining was scored based on the H-score and Allred Quick score (LM, MF). Cores were scored using an H-score by three independent investigators (including one consultant pathologist) blinded to the clinico-pathological characteristics of patients. $H = (3 \times \% \text{ of strongly stained cells}) + (2 \times \% \text{ of moderately stained cells}) + (1 \times \% \text{ of weakly stained cells}) + (0 \times \% \text{ of cells without staining})$. This gave a score ranging from total absence of DMXL2 in the tumor compartment (H-score 0) to DMXL2 expression in tumor cells equivalent to surrounding normal and stromal cells (H-score 300).

Statistical analysis

Two-tailed paired *T*-test was used to generate *P*-values (indicated as **P* = 0.05, ***P* = 0.01, and ****P* = 0.001). Anova test with Tukey's post test was used to generate *P*-values for the rescue experiment (**P* = 0.05, ***P* = 0.01).

ACKNOWLEDGMENTS

The Imperial College Healthcare NHS Trust Tissue Bank provided tissue samples. Other investigators may have received samples from these same tissues. The research [was funded / supported by] the National Institute for Health Research (NIHR) Biomedical Research Centre based and Experimental Cancer Medicine Centre, Cancer Research UK Centre at Imperial College Healthcare NHS Trust Imperial College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We wish to thank Philippa Darbre and Keith Brennan for sharing reagents and Simak Ali for critical assessment of the manuscript.

CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed

FINANCIAL SUPPORT

This work was supported by Imperial Junior Research Fellowship and Breast Cancer Campaign Pilot (P46869) to LM, the Wellcome Trust [103034/Z/13/Z] to DKP and CRUK (P28302) to MF and YL.

Authors' contributions

Conception and design: LM conceived the study, LM, MF, CRC designed the experiments

Development of methodology: LM, MF

Acquisition of data: LW and NP helped with Imperial College tissue collection. MF, VN, DP, YL, acquired data

Analysis and interpretation of data: SS, and GP helped with TMA scoring. LM, MF analysed the experiments

Writing, review and/or revision of the manuscript: LM, MF wrote the paper. All authors reviewed and approved the manuscript

Administrative, technical, or material support NA

Study supervision: LM

REFERENCES

1. Stylianou S, Clarke RB, Brennan K. Aberrant activation of notch signaling in human breast cancer. *Cancer research*. 2006; 66:1517–25.
2. Andersson ER, Lendahl U. Therapeutic modulation of Notch signalling--are we there yet?. *Nature reviews Drug discovery*. 2014; 13:357–78.
3. Kandachar V, Roegiers F. Endocytosis and control of Notch signaling. *Current opinion in cell biology*. 2012; 24:534–40.
4. Le Borgne R. Regulation of Notch signalling by endocytosis and endosomal sorting. *Current opinion in cell biology*. 2006; 18:213–22.
5. Tolcher AW, Chugh R, Chambers G, Thorpe V, Dupont J, Hill D, et al. A first-in-human phase I study to evaluate the fully human monoclonal antibody OMP-59R5 (anti-Notch2/3) administered intravenously to patients with advanced solid tumors. *Journal of Clinical Oncology*. 2012; 30.
6. Doody RS, Raman R, Farlow M, Iwatsubo T, Vellas B, Joffe S, et al. A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *The New England journal of medicine*. 2013; 369:341–50.
7. Ntziachristos P, Lim JS, Sage J, Aifantis I. From fly wings to targeted cancer therapies: a centennial for notch signaling. *Cancer cell*. 2014; 25:318–34.
8. Fan X, Khaki L, Zhu TS, Soules ME, Talsma CE, Gul N, et al. NOTCH pathway blockade depletes CD133-positive glioblastoma cells and inhibits growth of tumor neurospheres and xenografts. *Stem cells*. 2010; 28:5–16.
9. Giachino C, Taylor V. Notching up neural stem cell homogeneity in homeostasis and disease. *Frontiers in neuroscience*. 2014; 8:32.
10. Li Y, Ma J, Qian X, Wu Q, Xia J, Miele L, et al. Regulation of EMT by Notch signaling pathway in tumor progression. *Current cancer drug targets*. 2013; 13:957–62.
11. Espinoza I, Pochampally R, Xing F, Watabe K, Miele L. Notch signaling: targeting cancer stem cells and epithelial-to-mesenchymal transition. *OncoTargets and therapy*. 2013; 6:1249–59.
12. Ginnebaugh KR, Ahmad A, Sarkar FH. The therapeutic potential of targeting the epithelial-mesenchymal transition in cancer. *Expert opinion on therapeutic targets*. 2014; 18:731–45.
13. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. *Science*. 1999; 284:770–6.
14. Bolos V, Mira E, Martinez-Poveda B, Luxan G, Canamero M, Martinez C, et al. Notch activation stimulates migration of breast cancer cells and promotes tumor growth. *Breast Cancer Research*. 2013; 15.
15. Donepudi MS, Kondapalli K, Amos SJ, Venkateshan P. Breast cancer statistics and markers. *Journal of cancer research and therapeutics*. 2014; 10:506–11.
16. Musgrove EA, Sutherland RL. Biological determinants of endocrine resistance in breast cancer. *Nature reviews Cancer*. 2009; 9:631–43.
17. Osborne CK, Schiff R. Mechanisms of endocrine resistance in breast cancer. *Annual review of medicine*. 2011; 62:233–47.
18. Austreid E, Lonning PE, Eikesdal HP. The emergence of targeted drugs in breast cancer to prevent resistance to endocrine treatment and chemotherapy. *Expert opinion on pharmacotherapy*. 2014; 15:681–700.
19. Magnani L, Stoeck A, Zhang X, Lanczky A, Mirabella AC, Wang TL, et al. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:E1490–9.
20. Magnani L, Brunelle M, Gevry N, Lupien M. Chromatin landscape and endocrine response in breast cancer. *Epigenomics*. 2012; 4:675–83.
21. Ji ZL, Mohammed H, Webber A, Ridsdale J, Han N, Carroll JS, et al. The forkhead transcription factor FOXK2 acts as a chromatin targeting factor for the BAP1-containing histone deubiquitinase complex. *Nucleic Acids Research*. 2014; 42:6232–42.
22. Yan Y, Deneff N, Schupbach T. The Vacuolar Proton Pump, V-ATPase, Is Required for Notch Signaling and Endosomal Trafficking in *Drosophila*. *Developmental Cell*. 2009; 17:387–402.
23. Sethi N, Yan Y, Quek D, Schupbach T, Kang YB. Rabconnectin-3 Is a Functional Regulator of Mammalian

- Notch Signaling. The Journal of Biological Chemistry. 2010; 285:34757–64.
24. Forgac M. Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. Nature reviews Molecular cell biology. 2007; 8:917–29.
 25. Santen R, Jeng MH, Wang JP, Song R, Masamura S, McPherson R, et al. Adaptive hypersensitivity to estradiol: potential mechanism for secondary hormonal responses in breast cancer patients. The Journal of Steroid Biochemistry. 2001; 79:115–25.
 26. Brodie A, Jelovac D, Sabnis G, Long B, Macedo L, Goloubeva O. Model systems: Mechanisms involved in the loss of sensitivity to letrozole. The Journal of Steroid Biochemistry. 2005; 95:41–8.
 27. Martin LA, Ghazoui Z, Weigel MT, Pancholi S, Dunbier A, Johnston S, et al. An *in vitro* model showing adaptation to long-term oestrogen deprivation highlights the clinical potential for targeting kinase pathways in combination with aromatase inhibition. Steroids. 2011; 76:772–6.
 28. Shaw LE, Sadler AJ, Pugazhendhi D, Darbre PD. Changes in oestrogen receptor- α and - β during progression to acquired resistance to tamoxifen and fulvestrant (Faslodex, ICI 182,780) in MCF7 human breast cancer cells. The Journal of steroid biochemistry and molecular biology. 2006; 99:19–32.
 29. Lombardo Y, Faronato M, Filipovic A, Viricillo V, Magnani L, Coombes RC. Nicastrin and Notch4 drive endocrine therapy resistance and epithelial to mesenchymal transition in MCF7 breast cancer cells. Breast cancer research: BCR. 2014; 16:R62.
 30. Zheng LH, Zhao YH, Feng HL, Liu YJ. Endocrine resistance in breast cancer. Climacteric: the journal of the International Menopause Society. 2014; 17:522–8.
 31. Espinoza I, Miele L. Notch inhibitors for cancer treatment. Pharmacology & therapeutics. 2013; 139:95–110.
 32. Bourdeau V, Deschenes J, Laperriere D, Aid M, White JH, Mader S. Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. Nucleic Acids Research. 2008; 36:76–93.
 33. Einhorn Z, Trapani JG, Liu Q, Nicolson T. Rabconnectin3 α promotes stable activity of the H⁺ pump on synaptic vesicles in hair cells. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2012; 32:11144–56.
 34. Kobia F, Duchi S, Deflorian G, Vaccari T. Pharmacologic inhibition of vacuolar H⁺ ATPase reduces physiologic and oncogenic Notch signaling. Molecular oncology. 2014; 8:207–20.
 35. Rath S, Liebl J, Furst R, Vollmar AM, Zahler S. Regulation of endothelial signaling and migration by v-ATPase. Angiogenesis. 2014; 17:587–601.
 36. Chazotte B. Labeling lysosomes in live cells with LysoTracker. Cold Spring Harbor protocols 2011; 2011:prot5571.
 37. Huss M, Wieczorek H. Inhibitors of V-ATPases: old and new players. The Journal of experimental biology. 2009; 212:341–6.
 38. Gonzalez DM, Medici D. Signaling mechanisms of the epithelial-mesenchymal transition. Science signaling. 2014; 7:re8.
 39. Gumbiner B, Lowenkopf T, Apatira D. Identification of a 160-Kda Polypeptide That Binds to the Tight Junction Protein-Zo-1. Proceedings of the National Academy of Sciences of the United States of America. 1991; 88:3460–4.
 40. Tuttle AM, Hoffman TL, Schilling TF. Rabconnectin-3a regulates vesicle endocytosis and canonical Wnt signaling in zebrafish neural crest migration. PLoS biology. 2014; 12:e1001852.
 41. Kane PM. Targeting Reversible Disassembly as a Mechanism of Controlling V-ATPase Activity. Current Protein & Peptide Science. 2012; 13:117–23.
 42. Tata B, Huijbregts L, Jacquier S, Csaba Z, Genin E, Meyer V, et al. Haploinsufficiency of Dmx12, encoding a synaptic protein, causes infertility associated with a loss of GnRH neurons in mouse. PLoS biology. 2014; 12:e1001952.
 43. Jones DT. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics. 2007; 23:538–44.
 44. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research. 2013; 41:W349–57.
 45. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology. 1999; 292:195–202.
 46. Moller S, Croning MDR, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics. 2001; 17:646–53.
 47. Hsu KT, Yu XM, Audhya AW, Jaume JC, Lloyd RV, Miyamoto S, et al. Novel Approaches in Anaplastic Thyroid Cancer Therapy. The oncologist. 2014; 19:1148–55.
 48. McAuliffe SM, Morgan SL, Wyant GA, Tran LT, Muto KW, Chen YS, et al. Targeting Notch, a key pathway for ovarian cancer stem cells, sensitizes tumors to platinum therapy. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:E2939–48.
 49. Liu JT, Mao ZF, Huang J, Xie SP, Liu TS, Mao ZF. Blocking the NOTCH pathway can inhibit the growth of CD133-positive A549 cells and sensitize to chemotherapy. Biochemical and Biophysical Research Communications. 2014; 444:670–5.
 50. Yashiro-Ohtani Y, Wang H, Zang C, Arnett KL, Bailis W, Ho Y, et al. Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:E4946–53.
 51. Graham RM, Thompson JW, Webster KA. Inhibition of the vacuolar ATPase induces Bnip3-dependent death of

cancer cells and a reduction in tumor burden and metastasis. *Oncotarget*. 2014; 5:1162–73.

52. Wei P, Walls M, Qiu M, Ding R, Denlinger RH, Wong A, et al. Evaluation of selective gamma-secretase inhibitor PF-03084014 for its antitumor efficacy and gastrointestinal safety to guide optimal clinical trial design. *Molecular cancer therapeutics*. 2010; 9:1618–28.
53. Fouladi M, Stewart CF, Olson J, Wagner LM, Onar-Thomas A, Kocak M, et al. Phase I trial of MK-0752 in children with refractory CNS malignancies: a pediatric brain tumor consortium study. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2011; 29:3529–34.
54. Filipovic A, Lombardo Y, Faronato M, Abrahams J, Aboagye E, Nguyen QD, et al. Erratum to: Anti-nicatin monoclonal antibodies elicit pleiotropic anti-tumour pharmacological effects in invasive breast cancer cells. *Breast cancer research and treatment*. 2014; 148:463.
55. Faronato M, Patel V, Darling S, Dearden L, Clague MJ, Urbe S, et al. The deubiquitylase USP15 stabilizes newly synthesized REST and rescues its expression at mitotic exit. *Cell cycle*. 2013; 12:1964–77.

The pioneer factor PBX1 is a novel driver of metastatic progression in ER α -positive breast cancer

Luca Magnani¹, Darren K. Patten¹, Van T.M. Nguyen¹, Sung-Pil Hong¹, Jennifer H. Steel¹, Naina Patel¹, Ylenia Lombardo¹, Monica Faronato¹, Ana R. Gomes¹, Laura Woodley¹, Karen Page², David Guttery², Lindsay Primrose², Daniel Fernandez Garcia², Jacqui Shaw², Patrizia Viola³, Andrew Green⁴, Christopher Nolan⁴, Ian O. Ellis⁴, Emad A. Rakha⁴, Sami Shousha¹, Eric W.-F. Lam¹, Balázs Györfy⁵, Mathieu Lupien^{6,7}, R. Charles Coombes¹

¹Department of Surgery and Cancer, Imperial College London, London, UK

²Department of Cancer Studies, University of Leicester, Leicester, UK

³Laboratory of Medicine, Histopathology Department, Royal Brompton Hospital, London, UK

⁴Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, Budapest, HU

⁵MTA TTK Lendület Cancer Biomarker Research Group, 2nd Department of Pediatrics, Semmelweis University and MTA-SE Pediatrics and Nephrology Research Group, Budapest, HU

⁶The Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada

⁷Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

Correspondence to:

Luca Magnani, e-mail: l.magnani@imperial.ac.uk

Keywords: breast cancer, estrogen receptor, drug resistance, metastasis, pioneer factors

Received: February 18, 2015

Accepted: June 02, 2015

Published: June 15, 2015

ABSTRACT

Over 30% of ER α breast cancer patients develop relapses and progress to metastatic disease despite treatment with endocrine therapies. The pioneer factor PBX1 translates epigenetic cues and mediates estrogen induced ER α binding. Here we demonstrate that PBX1 plays a central role in regulating the ER α transcriptional response to epidermal growth factor (EGF) signaling. PBX1 regulates a subset of EGF-ER α genes highly expressed in aggressive breast tumours. Retrospective stratification of luminal patients using PBX1 protein levels in primary cancer further demonstrates that elevated PBX1 protein levels correlate with earlier metastatic progression. In agreement, PBX1 protein levels are significantly upregulated during metastatic progression in ER α -positive breast cancer patients. Finally we reveal that PBX1 upregulation in aggressive tumours is partly mediated by genomic amplification of the PBX1 locus. Correspondingly, ER α -positive breast cancer patients carrying PBX1 amplification are characterized by poor survival. Notably, we demonstrate that PBX1 amplification can be identified in tumor derived-circulating free DNA of ER α -positive metastatic patients. Metastatic patients with PBX1 amplification are also characterized by shorter relapse-free survival. Our data identifies PBX1 amplification as a functional hallmark of aggressive ER α -positive breast cancers. Mechanistically, PBX1 amplification impinges on several critical pathways associated with aggressive ER α -positive breast cancer.

INTRODUCTION

Over 30% of ER α -positive breast cancer patients treated with endocrine therapies eventually relapse and progress into metastatic disease [1]. Previous studies

strongly suggest that the endocrine therapy resistance and metastatic progression occur in a multi-step manner and are driven by genetic and epigenetic alterations [2–4]. Several evidences indicate that breast cancer growth becomes ultimately independent from ER α [5–8].

Nonetheless this may be occurring toward the final stages of tumour progression. For instance, recent studies have discovered the existence of ER α activating mutations specifically in metastatic patients [9–12]. Additionally, ER α can be activated in the absence of estrogen by alternative mitogens including the epidermal growth factor (EGF) [13, 14]. EGF promotes ER α binding to thousands of regulatory elements and activates the transcription of a set of genes commonly over-expressed in the HER2 subtype and associated with disease progression [15].

We have recently described the role of PBX1 as a novel pioneer factor in ER α breast cancer [16]. PBX1 is critical to maintain chromatin accessibility and allow estrogen induced ER α recruitment to thousands of regulatory elements [16]. More importantly, we demonstrated that PBX1 is necessary for the transcriptional regulation of a subset of estrogen target genes directly associated with poor outcome. Furthermore, we have demonstrated that PBX1 upregulation is involved in the development of endocrine therapy resistance *in vitro* and that mRNA levels of genes controlled via Notch-PBX1 crosstalk are associated with poor prognosis in ER α patients [5]. Whether PBX1 contributes to estrogen-independent, EGF driven activation and recruitment of ER α is not yet known. We have thus investigated the role of PBX1 in mediating EGF signalling in ER α -positive breast cancer cell lines using genome wide-analyses including microarray and ChIP-seq. Our results demonstrate that PBX1 is required to direct EGF-ER α signalling at the chromatin level in breast cancer cells. PBX1 is also required for the expression of genes associated with tumour progression. PBX1 protein levels were significantly upregulated *in vivo* in metastatic samples from patients treated with endocrine therapies. Analysis of an independent historical collection of primary samples demonstrated that PBX1 protein levels correlate with shorter time to progression in ER α patients. Finally, we present evidences suggesting that PBX1 may be amplified in over 10% of ER α -positive breast cancer patients and amplification is associated with shorter survival. PBX1 amplification was confirmed in the circulating free DNA (cfDNA) of metastatic patients further suggesting the potential for using PBX1 to monitor breast cancer progression.

RESULTS

PBX1 mediates EGF signaling in ER α -positive breast cancer cells

We recently demonstrated that PBX1 is a pioneer factor in ER α -positive breast cancer, occupying the chromatin prior to ER α binding following its activation with estrogen and regulating the expression of estrogen dependent genes associated with aggressive progression

[16]. In addition we showed that ER α controls a different transcriptional program when activated by growth factors [15]. Specifically, EGF can activate ER α via serine phosphorylation and promotes transcription of genes associated with aggressive HER2 positive tumours [15]. Therefore, we reasoned that PBX1 might also play a role in estrogen independent EGF driven ER α induced transcription. ER α -positive MCF7 cells proliferate in response to EGF when cultured in the absence of estrogen and this phenotype is dependent on ER α [15]. PBX1 depletion using two siRNA molecules is sufficient to abrogate the proliferative response induced by EGF suggesting that PBX1 regulates EGF-ER α signaling (Fig. 1A–1B). Similar results were also obtained using T47-D, a second independent ER α -positive breast cancer cell line (Fig. S1A). We and others have previously showed that PBX1 is a direct Notch target that can be potentially antagonized with gamma secretase inhibitors (GSI, MRK003 and PF03084014) [5, 17]. In agreement, treatment with MRK003 is sufficient to downregulate PBX1 at mRNA and protein levels (Fig. S1B–S1C) and suppresses EGF induced proliferation in MCF7 cells (Fig. S1D–S1E). EGF activates ER α by promoting receptor phosphorylation at several key residues via AKT signaling pathway [15], including S118 [18]. Notably, PBX1 silencing did not impair ER α phosphorylation suggesting that defects in EGF induced proliferation did not stem from defect in non-canonical activation of ER α [19] (Fig. S2A). However, growth deficit in response to GSI treatment might stem from a combination of PBX1 downregulation and direct inhibition of EGF signalling (Fig. S2A).

Next, we investigated the transcriptional effect of PBX1 silencing on the EGF-ER α transcriptional program using qRT-PCR. As expected, silencing PBX1 in MCF7 cells impaired the transcriptional activation of a known set of EGF-ER α target genes [15] (Fig. 1C). These results were confirmed and expanded using expression microarrays (Fig. 1D). Our data indicate that over 60% (252 out of 383) of the genes differentially expressed in response to EGF signaling require PBX1 for their regulation (Fig. 1D and Table S1). These data are strongly reminiscent of our previous finding in which a subset of estrogen responsive genes were dependent on PBX1 [16]. PBX1 dependent genes were enriched for several important ontological terms associated with breast cancer and endocrine therapy resistance, including Notch signaling, pro-invasive signaling and epithelial and mammary carcinoma (Fig. 2). Interestingly, genes that failed to be induced by EGF in PBX1 depleted cells are also significantly enriched in several independent genes datasets obtained from patients characterized by aggressive breast cancers (Fig. S2B). Overall, these data suggest that PBX1 underlies the expression of EGF dependent genes involved in aggressive tumour progression.

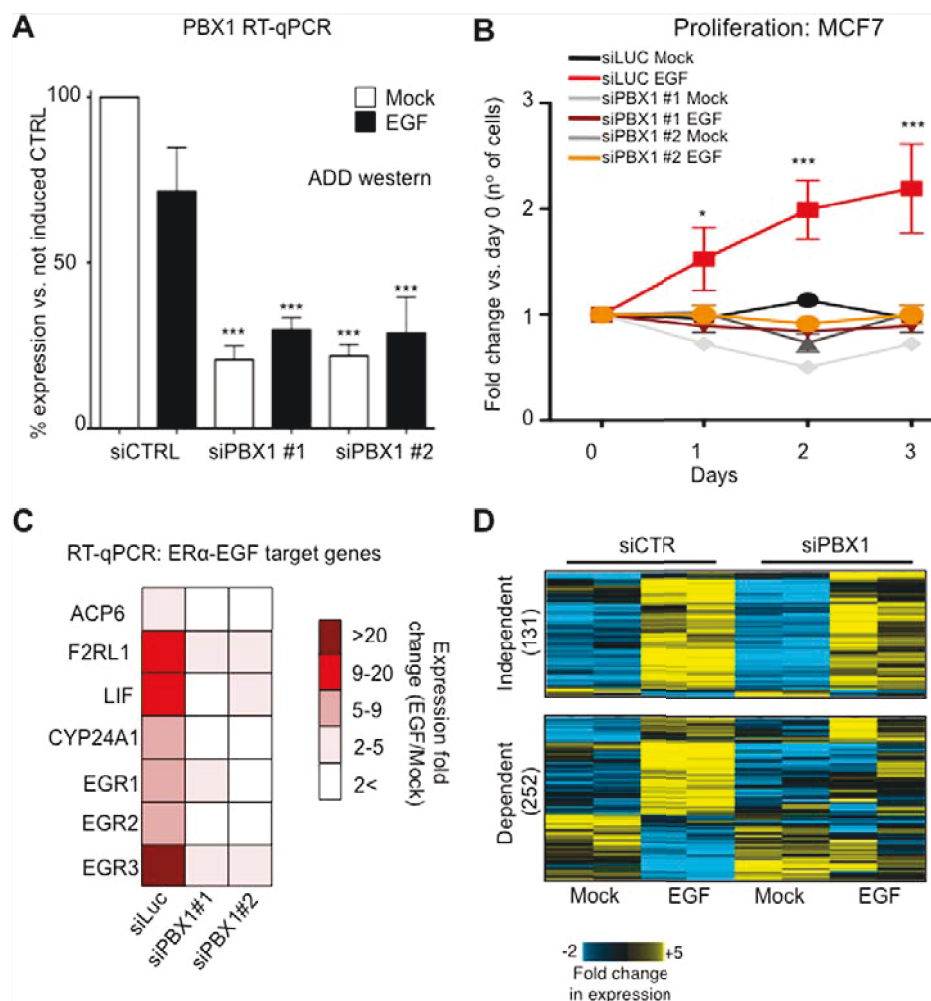


Figure 1: PBX1 is required to activate EGF signaling in ERα breast cancer cells. **A.** PBX1 silencing using two independent siRNAs **B.** Proliferation curves in response to EGF stimulation (100 ng/ml) or mock control in MCF7 cells. **C.** Transcriptional activation of several EGF target genes is lost upon PBX1 depletion. The fold induction was calculated averaging three independent experiments. All comparison between siRNA and control are statically significant (Student's *t*-test, $p < 0.05$) **D.** Microarray analysis of PBX1 depleted cells demonstrates that PBX1 control a large portion of EGF dependent genes. Asterisks identify significant differences between treatments.

PBX1 is pre-loaded at regulatory elements recruiting ERα during EGF signaling

Pioneer factors such as PBX1 and FOXA1 bind to the chromatin prior to ERα recruitment where they are thought to mediate or maintain chromatin accessibility [20–22]. To determine if PBX1 could serve as a pioneer factor for ERα under EGF stimulation we conducted ChIP-seq assays in MCF7 cells. We then compared the EGF-ERα cistrome (e.g. ERα genome wide binding sites induced upon EGF stimulation) with the estrogen-ERα cistrome and previously published PBX1 cistrome. In agreement with previous reports [15], most ERα chromatin binding is unique to specific stimuli (EGF-ERα vs. E2-ERα) (Fig. 3A and S3A). PBX1 occupies approximately 30% of ERα binding sites, regardless of the stimulus (Fig. 3A). PBX1 binding occurs prior and is not affected by EGF stimulation as demonstrated by direct ChIP-qPCR assays

at selected binding sites in proximity of EGF regulated genes (Fig. 3B). This process is strongly reminiscent of PBX1 behaviour in the context of estrogen stimulation [16]. We next considered if EGF induced ERα binding sites that overlap with PBX1 have a functional relationship with the EGF-ERα transcriptional program. To do this we have combined transcriptional and ChIP-seq data of MCF-7 cells treated with EGF. We have analysed genes whose expression changed in response to EGF stimulation (all upregulated genes) for enriched binding patterns near their transcriptional starting site (up to 20 kb upstream of the TSS). We focused on ERα binding in response to estrogen (17 beta-estradiol, E2) or EGF with or without concurrent PBX1 and FOXA1 binding. Our data show that genes that respond to EGF are significantly enriched for PBX1-EGF-ERα binding but lack PBX1-E2-ERα binding sites (Fig. S3B, red dots). FOXA1 binding was less enriched near these genes (Fig. S3B, light-blue dots).

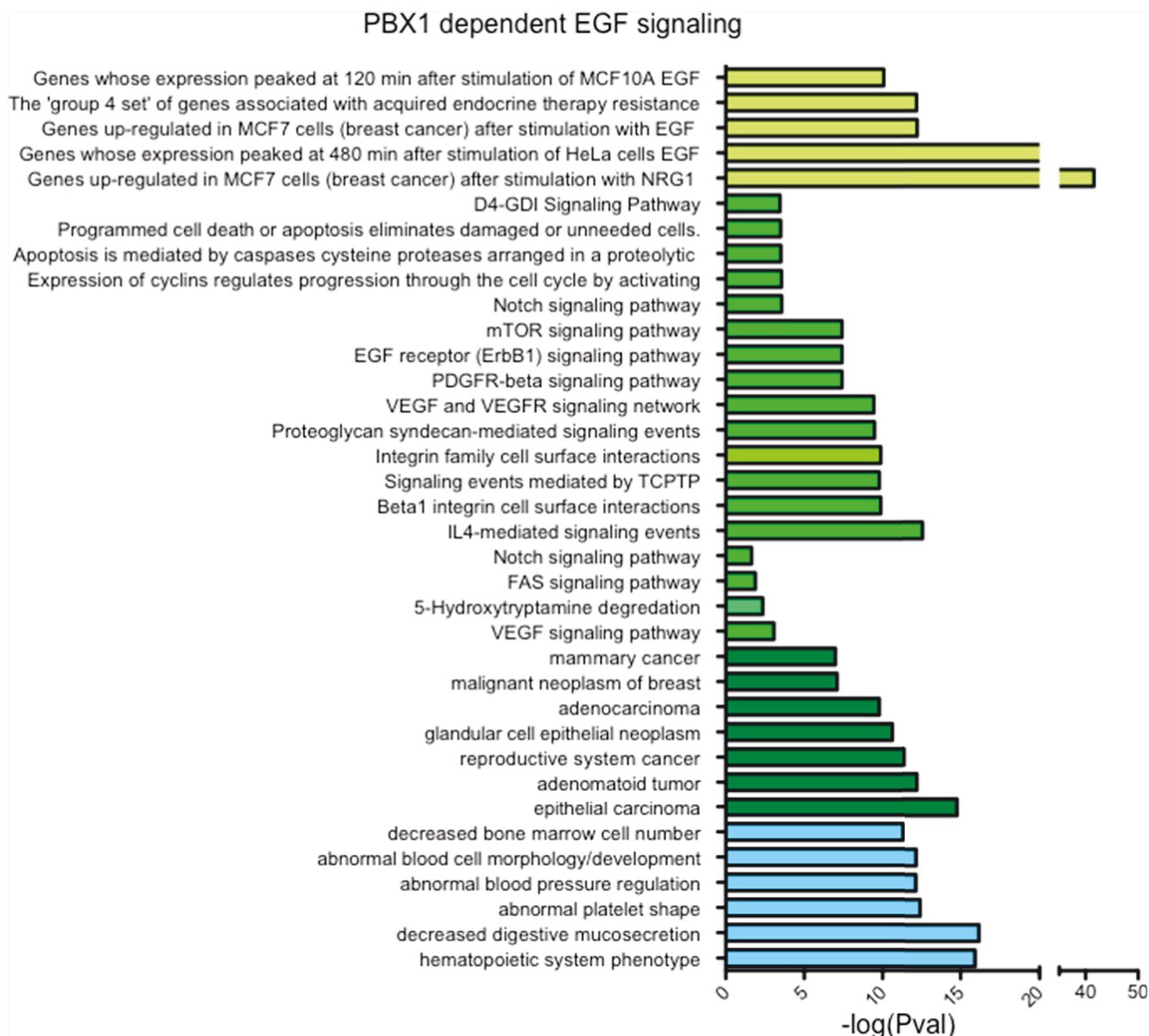


Figure 2: PBX1 controls the expression of genes associate with breast cancer progression. Genes whose failed to be upregulated by EGF upon PBX1 silencing ($n = 147$) were used for ontology analysis using GREAT [48].

EGF responsive genes are also significantly enriched for PBX1-FOXA1-non specific ER α binding, suggesting that co-activation may also play an important role to regulate this subset of genes (Fig. S3, purple dots). As expected, regions pre-bound by pioneer factors that did not recruit ER α lacked any association with transcription. Collectively, these analyses confirm that PBX1 binding is directly involved in regulating EGF-ER α signaling in MCF-7 cells.

We have previously demonstrated that H3K18ac can be used to detect stimuli specific co-activation (EGF vs. E2) at regulatory regions in which ER α binding occurs under both treatment [15]. Interestingly PBX1 depletion directly suppressed H3K18ac signal at several putative elements normally recruiting ER α (Fig. 3C). Globally,

PBX1-EGF-ER α regions are characterized by open chromatin status as demonstrated by FAIRE-seq and DNaseI-hypersensitivity-seq assays (Fig. 3D). However, PBX1 depletion results in a significant decrease in chromatin openness (as demonstrated by reduced FAIRE signal) at several regions associated with EGF-ER α target genes suggesting that PBX1 modulates ER α signaling by influencing the chromatin environment

PBX1 is a novel luminal breast cancer prognostic biomarker

Our mechanistic study revealed that PBX1 controls the expression of a subset of ER α target genes stimulated

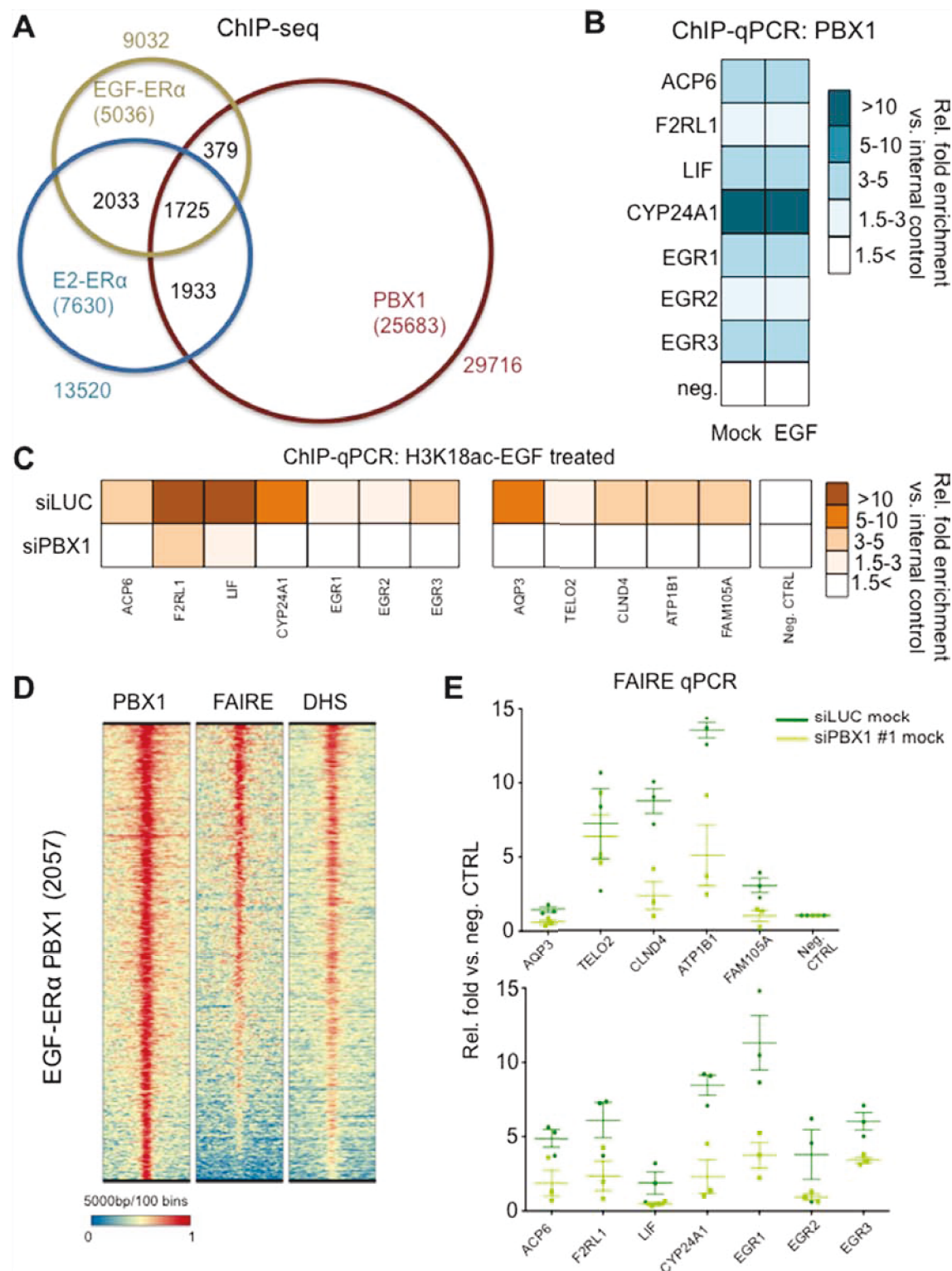


Figure 3: PBX1 regulate chromatin status at EGF-ERα binding sites. **A.** ChIP-seq analysis of PBX1 and ERα in MCF7 cells **B.** ChIP-qPCR validation of PBX1 binding at potential regulatory EGF-ERα binding sites in the absence of EGF stimulation. CTRL indicate stimulation using mock. The fold enrichment was calculated averaging three independent experiments. All comparison between siRNA and control are not statically significant (Student's *t*-test, $p < 0.050$) **C.** PBX1 silencing abrogates epigenetic co-activation of EGF-ERα regulatory elements as measured by induced H3K18 acetylation. The fold enrichment was calculated averaging three independent experiments. All comparison between siRNA and control are statically significant (Student's *t*-test, $p < 0.050$) **D.** Genome wide chromatin accessibility analyses at PBX1 and EGF-ERα shared sites. **E.** PBX1 depletion negatively impact chromatin accessibility at EGF-ERα binding sites as identified by reduced FAIRE signal.

by EGF *in vitro* and may be linked to breast cancer progression. Furthermore we previously demonstrated that PBX1 displays an analogous function by controlling a subset of estrogen induced genes strongly associated with poor prognosis [16]. In light of these observations we hypothesized that PBX1 expression levels could contribute to breast cancer progression in ER α -positive breast cancer patients. To test this hypothesis we performed meta-analysis on METABRIC (Illumina arrays) and The Cancer Genome Atlas (TCGA, RNA-seq data) studies stratifying ER α -positive breast cancer patients in high and low PBX1 expressors. In both datasets high PBX1 mRNA levels correlate with shorter survival (Fig. S4A). For the METABRIC dataset, PBX1 expression retained significance in a multivariate analysis including lymph node status, grade and tumor size (HR = 1.27, p = 0.0173). For the TCGA datasets the TNM status was published for each patient. In the multivariate analysis including PBX1, T, N and M, PBX1 reached the highest significance (p = 0.041, HR = 7.82), M also reached significance p = 0.045, HR = 2.89) while T and N were not significant. These correlations were not significant in ER α -negative breast cancer patients further supporting the notion that PBX1 is a key regulator of ER α activity (Fig. S4A). Notably, higher PBX1 levels were significantly associated with poor prognosis in a subset of patients treated with endocrine therapy (Fig. S4B). Again, PBX1 retained significance in a multivariate analysis including lymph node status, grade, tumor size and PBX1 expression (HR = 1.27, p = 0.0173). To elucidate further the potential role of PBX1 as a breast cancer biomarker, we tested the association between its expression and metastatic progression in an additional six independent studies (Affymetrix arrays). Overall, PBX1 expression was associated with increased risk in metastatic progression (HR 1.47, p < 0.008, n = 674). We found more variability within the individual cohorts, however, statistical significance was reached only for datasets in which PBX1 represents a poor prognostic marker (Fig. S5). Additionally, multivariate analysis in the grouped cohort indicates that PBX1 prognostic power is independent from Ki67 (p < 0.0001).

To confirm these findings we measured PBX1 protein levels using IHC in longitudinal clinical samples. We collected 20 biopsies from primary ER α -positive breast cancers and compared PBX1 protein levels with matched relapses (all endocrine treated, average time to relapse = 44.4 ± 35 months). We stained all sections using a validated antibody (Fig. S1C and S2A) [5, 16] and scored each sample based on PBX1 nuclear intensity (Fig. 4A). Remarkably, metastatic samples had a significant increase in H score compared to the original primary biopsies confirming PBX1 upregulation in drug-resistant metastatic patients (Fig. 4A). Finally, we investigated PBX1 protein levels in the Tenovus Nottingham cohort of FFPE primary breast cancers (n = 1650) [23]. In agreement with cell

line data, we found that PBX1 protein levels significantly correlate with ER α , FOXA1 and GATA3 (another important ER α pioneer factor [22]) thus demonstrating that PBX1 is strongly associated with ER α -positive, luminal breast cancer subtypes (Fig. 4B). We therefore restricted further analysis to ER α -positive breast cancer patients. Since PBX1 protein levels were dramatically increased in metastatic samples, we hypothesized that high levels of PBX1 in primary samples might identify patient at higher risk of relapse. Our results show that ER α -positive breast cancer patients with elevated PBX1 staining intensity at diagnosis (over 90% positive staining) develop metastatic disease significantly faster than patients with lower staining intensity (Fig. 4C). Overall this data strongly suggest that PBX1 protein levels in primary breast cancer are significantly associated with breast cancer progression.

PBX1 locus is amplified in patients with aggressive ER α breast cancer

Our analysis using PBX1 transcriptional and protein levels strongly supports the role of PBX1 as a functional biomarker associated with ER α breast cancer progression (Figs. 4 and S4–S5). PBX1 transcripts are significantly overexpressed in breast cancer when compared to normal tissues (Fig. S6). For example, PBX1 ranks in the top 50 most overexpressed genes in ductal and lobular carcinoma compared to normal tissue in TCGA cohort (dataset 30 in Fig. S6) [24]. Similarly, PBX1 ranks in the top 4% of the most overexpressed transcripts in the METABRIC cohort (Invasive Ductal vs. Normal, dataset 7, Fig. S6) [25]. To the best of our knowledge, there is no indication in the literature that PBX1 is required for normal breast development. In agreement, we found that normal ductal cells adjacent to tumor tissue have very low PBX1 staining (stronger stained cells are myoepithelial cells, Fig. S7). Thus, PBX1 expression seems to be a major determinant for the development and progression of luminal breast cancer patients.

Recent tumour sequencing efforts indicate that PBX1 locus is potentially amplified in 13% of primary breast cancers (TCGA provisional) (Fig. S8A) [26] supporting the hypothesis that genomic amplification may partially explain PBX1 overexpression in breast cancer (TCGA provisional) (Fig. S8B). Of note, within the published TCGA dataset [24], PBX1 potential amplification strongly correlates with higher mRNA levels in ER α -positive but not in ER α -negative breast cancer patients (Fig. S9). These data suggest that PBX1 copy number variation (CNV) might then be functional only in ER α -positive breast cancer patients. We then identified the TCGA patients (TCGA provisional) with ER α -positive disease that had at least 3 copies of PBX1 (95/1145, 8.3%) and analysed the proportion of patients with lymph-node micro-metastasis. Strikingly, we found that 60% of patients with PBX1 amplification had at

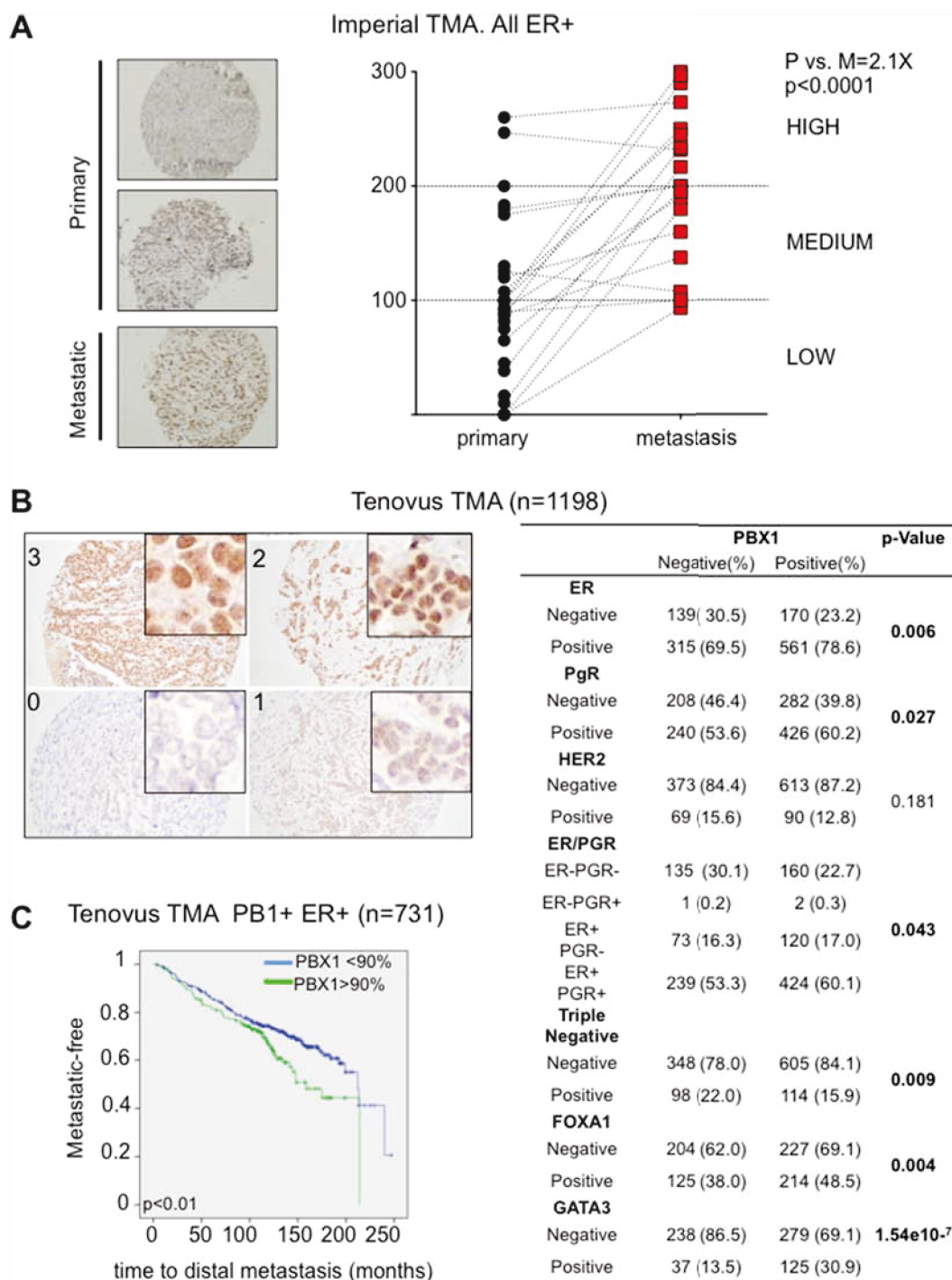


Figure 4: PBX1 is a novel prognostic biomarker for ER α breast cancer. **A.** Primary and matched Metastatic samples from the Imperial College TMA were processed using PBX1 IHC and H score were plotted. Pair-wise *t*-test between the average score (3 independent scorer, duplicate sections) was used to establish significance **B.** Core primary biopsies in the Tenovus Nottingham cohort show different level of staining by PBX1 IHC. PBX1 correlates with several important clinico-pathological parameters including hormone status and ER α associated pioneer factors. **C.** Survival analysis (Log-rank Mantel-Cox test) was performed looking at patients with very high PBX1 levels versus the remaining patients. Analysis was restricted to the ER α -positive population.

least one positive lymphnodes vs. 38.7% of the control group (Fisher exact test 7.3×10^{-5}). Conversely, PBX1 amplification correlates, although not significantl , with

a lower chance of metastatic disease in the limphnodes in ER α -negative breast cancer patients (45% vs. 57% respectively).

We then restricted our analysis to ER α -positive breast cancer patients from the published TCGA data to exploit the more complete clinical annotations [24] and found that 28/364 (8%) carried potential PBX1 amplification/overexpression (Fig. 5A). Patients with

putative PBX1 amplification had a median survival of 30.98 months compared to 113.74 (PBX1 Amplified vs. Not Amplified in Luminal patients, $P < 10^{-5}$) (5A). On the other hand, PBX1 amplification did not reach prognostic significance in ER α -negative patients (however, only

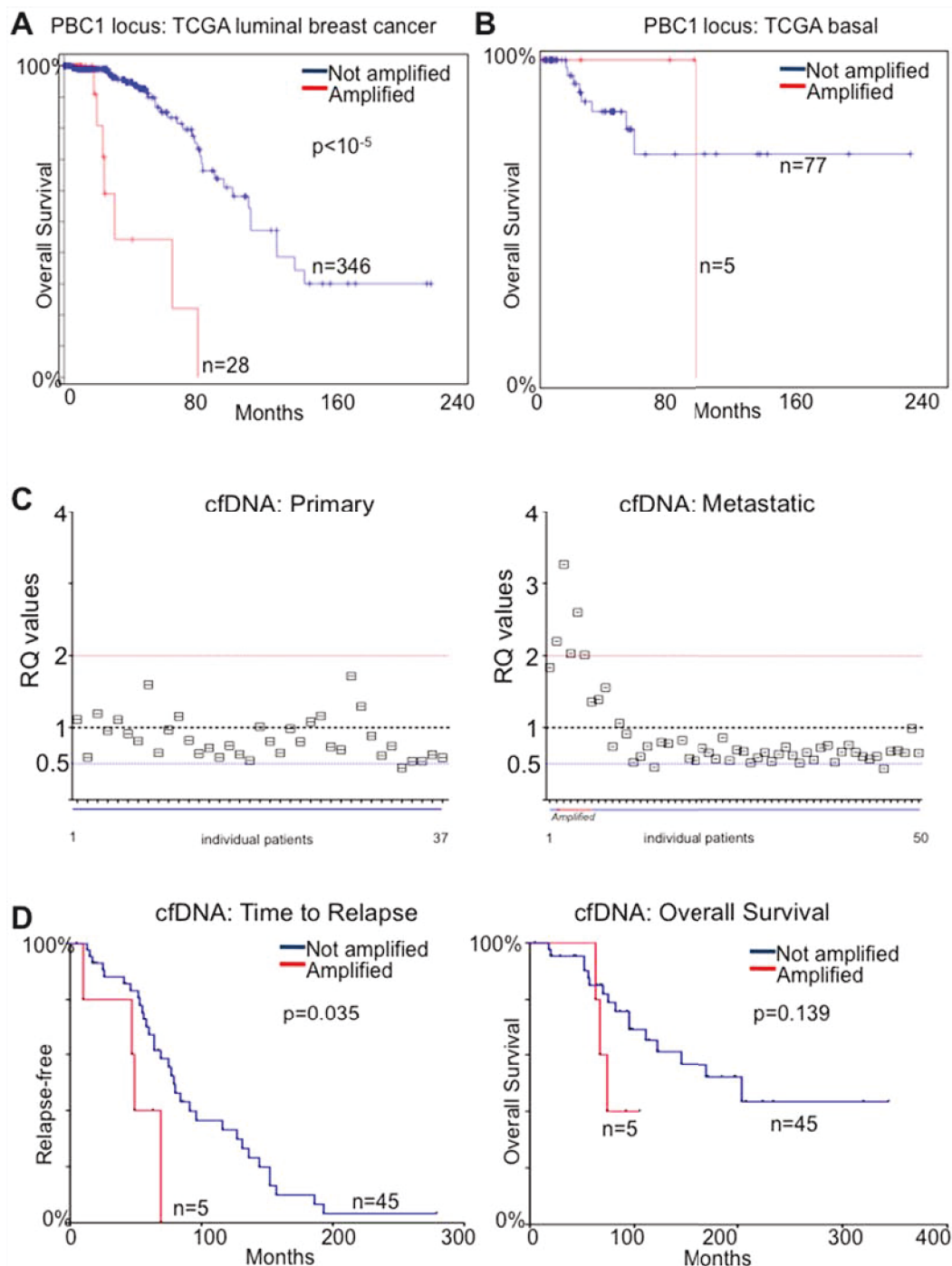


Figure 5: PBX1 amplification is a prognostic biomarker for ER α breast cancer. **A.** TCGA luminal patients [24] were stratified based on putative PBX1 amplification and overall survival was calculated **B.** The same analysis was repeated in basal and claudin low patients **C.** qPCR measurement of PBX1 genomic DNA in cfDNA from primary and metastatic ER α breast cancer patients **E–F.** Metastatic patients from panel C were stratified based on PBX1 amplification and Time to relapse and Overall survival were calculated. Groups were compared using a Log-rank Mantel-Cox test.

6% of patients carries amplification) (Fig. 5B). When we looked at genes potentially co-amplified with PBX1 (chr1q23.3) [27] we found that very few candidates were also over-expressed in breast cancer (Fig. S10A). More importantly, PBX1 was one of the only two transcripts (the other being PVRL4) carrying evidence of prognostic significance in the ER α -positive breast cancer subtype (METABRIC) (Fig. S10B).

We then considered the possibility of using PBX1 CNV as a proxy for PBX1 protein levels. Potentially, this would allow for monitoring PBX1 levels by measuring its amplification in DNA derived from cancer cells using non-invasive methods. The tumour derived fraction of circulating free-DNA (cfDNA) originating from tumour cells can be used to estimate the genomic status of breast cancer lesions [28–30]. We then measured PBX1 amplification using a qPCR assay in cfDNA from 37 patients with ER α -positive primary disease and 50 patients with ER α -positive metastatic disease treated with endocrine therapy (Table S2). We could not identify significant CNV in blood from primaries ER α -positive breast cancer patients. However, we identified 5/50 (10%) patients with PBX1 amplification (RQ > 2) and two patients with probable allelic loss (4%) in the metastatic cohort (Fig. 5C). We then reasoned that the emergence of PBX1 CNV may have contributed to tumor progression and used PBX1 amplification in cfDNA as a classifier to stratify patients based on time to relapse or overall survival. In this small cohort, PBX1 amplification was significantly associated with earlier metastatic relapse ($p < 0.035$, HR 5.95, Cox- Mantel-Test) (Fig. 5D). These data are in strong agreement with our protein and transcriptional tumour profiling and coupled with our mechanistic study demonstrate that PBX1 plays a crucial role in ER α -positive breast cancer development and progression.

DISCUSSION

Our work has identified PBX1 as a novel functional biomarker in ER α -positive breast cancer. We have identified a mechanistic association between PBX1 and ER α recruitment in response to EGF signaling. We also established the significance of PBX1 mRNA and protein level as a prognostic biomarker in several independent large cohorts of primary breast cancers. Notably, PBX1 levels are strongly associated with metastatic progression. Finally, we demonstrate that the longitudinal increment of PBX1 levels observed throughout cancer progression are partly due to genomic events that could be potentially used to monitor disease progression in patients by non-invasive assays.

The number of signaling molecules potentially involved in breast cancer is rapidly increasing. In this context, ER α remains a central hub required to regulate and integrate signaling at the chromatin level [31, 32].

EGF signaling cascade partially converge on ER α in breast cancer cells where it provides for an additional survival pathways. ER α target genes induced via classical estrogen signaling and non-canonical EGF signalling are quite distinct [14, 15]. Nonetheless, PBX1 remains a common denominator of these two pathways via its conserved interaction with ER α . It will be interesting to investigate the role of pioneers factors in alternative ER α pathways, including metabolic induced ER α signaling [33, 34]

Genetic aberration targeting chromatin remodelers and other proteins involved in epigenetic processes are now frequently recognized as major drivers in several types of cancers [35]. Synergy between epigenetic and genetic drivers may be particularly important in breast cancer considering the essential role of ER α in these malignancies and the intimate crosstalk between ER α and the chromatin [21, 22, 36]. For example, it is possible that PBX1 overexpression may lead to an increased number of chromatin accessible sites thereby allowing ER α binding near oncogenes [37]. While several other pioneer factors play a central role in modulating ER α binding in breast cancer cells [16, 22, 36, 38, 39], our data suggest that PBX1 has some unique properties. For example, FOXA1 [39] and GATA3 [40] play a central role in mammary and ductal morphogenesis and their interaction with ER α can be regarded as developmentally conserved [41, 42]. On the other hand PBX1 expression is restricted to ER α transformed cells and practically absent in normal ductal cells. In addition, FOXA1 and GATA3 are frequently targeted by mutations in breast cancer [43] (Fig. S11), while PBX1 is rarely affected by somatic mutations (V117M and H425N, 2 out of 892 patients). Furthermore FOXA1 and GATA3 are frequently downregulated in basal tumour but rarely over-expressed or amplified in luminal cancer (Fig. S11) [44] while PBX1 is among the top upregulated genes and is frequently amplified in luminal cancers (Figs. 4–5, S6, S8, S10 and S11).

Despite the impossibility to elucidate the temporal details of PBX1 amplification/overexpression, our data strongly suggest that PBX1 upregulation may contribute to breast cancer progression and metastatic development. Clonal evolution is thought to contribute significantly to cancer progression [45]. In this context, we can envision two possible models: in the first, cells carrying PBX1 amplification may be more adapted to escape the primary site and colonize distal tissues. On the other hand, it is possible that endocrine therapies exert positive selective pressure on cells that incur in Chr1q23.3 gains. In both scenarios, it is tempting to speculate that PBX1 would contribute to epigenetic reprogramming by interpreting or shaping the histone modifications landscape [21]. Ultimately, these changes would modulate ER α signaling and allow breast cancer cells to have an increased fitness in response to a diverse array of signalling pathways.

MATERIALS AND METHODS

Cell lines

MCF-7 and T47D cells were cultured and maintained as previously described [16]. EGF stimulation was performed as previously described [15]. Proliferation curves were performed as previously described [16]. Treatment with GSI (10 μ M of MRK003 or PF03084014) was performed as previously described [5]. Experiments were repeated independently at least five times

RT-qPCR and microarray

RNA samples from siControl or siPBX1 treated MCF7 in the presence or absence of EGF were extracted with RNAeasy (Quiagen) and retro-transcribed with iSCRIPT (Biorad). For microarrays, RNAs were hybridized on HT12 human beads array (Illumina Inc.). Analyses were performed using BRB- Array Tools Version 3.8.1. Raw intensity data were log₂ transformed, median normalized and filtered to remove non- detected spots as determined by Illumina Software. The normalization was performed by computing a gene-by-gene difference between each array and the median (e.g. reference) array, and subtracting the median difference from the log intensities on that array, so that the gene-by-gene difference between the normalized array and the reference array is zero. Two class non-paired comparison analyses were performed by computing a *t*-test for each gene using normalized log-intensities. Differentially expressed genes were determined at a significance level of *q* less than 0.01. A four class ANOVA at *p* less than 0.01 was also performed to identify genes expressed differentially across the four groups. Hierarchical clustering was employed by calculating Euclidean distance to generate heat maps for subsets of significant genes using the open source software Cluster/Treeview.

ChIP and ChIP-seq

ChIP assay were conducted as described previously described using PBX1 (Abnova, M01 clone 4A2) or ER α (Santa Cruz biotechnology) [16] using 5–10 $\times$ 10⁶ cells. Library preparation for next-generation sequencing was performed according to manufacturer's instruction starting with 5 ng of material (Illumina Inc.). Single paired libraries were sequenced using the HT2500 (Illumina Inc). Over 20 million 50 bp reads were generated through for the ER α ChIP and Input samples. Of those, over 92% were aligned to the human reference genome. These reads were aligned using bowtie [46] (default setting). The MACS peak-calling algorithm was used to call significantly enriched peaks using default settings (*P* < 10⁻⁵) [47].

Ontology analysis and heatmap analysis

Ontology analyses were performed using GREAT [48]. Heatmaps were generated using CHASE [49].

Imperial tissue microarray of paired primary and secondary breast cancers

Twenty primary breast carcinomas with a paired metastasis were acquired from the pathology archives of Charing Cross Hospital, London, UK. A tissue microarray was constructed using a manual microarrayer (Beecher) and 0.6 mm punches. The tissue microarray was immunohistochemically profiled for PBX1 and other biomarkers as previously described [16]. PBX1 monoclonal antibody (Abnova, (M01), clone 4A2) was optimized to a working concentration using tissue sections (5 μ m). Antigen retrieval was performed using 10mM Tris-HCl with 1 mM EDTA (pH 9.0), heated in a water bath at 95°C for 30 min followed by cooling at room temp for 30 min. Blocking was done using 0.3% hydrogen peroxide in PBS, followed by normal goat serum (20 μ l per ml) for 30 min. The primary antibody (1:50 in PBS) was incubated overnight at 4°C, then detected using anti-rabbit secondary antibody, Vectastain Elite peroxidase ABC kit, and ImmPACT DAB kit (all from Vector Laboratories). Subsequently, immunohistochemistry was performed on TMA sections (4–5 μ m) using the optimized staining protocol, including negative controls (omission of the primary antibody). PBX1 immunoreactivity was localized in the nucleus. Staining was scored based on the H-score and Allred Quick score (LM, DKP and PV).

Nottingham tenovus primary breast cancer series

Primary operable breast cancer cases (*n* = 1650) from the Nottingham Tenovus Primary Breast Carcinoma Series were used and were utilized for immunohistochemistry. Clinical data were maintained on a prospective basis with a median follow-up of 126 months [23, 50]. The tissue microarrays and full-face sections from the Nottingham Tenovus Primary breast cancer series were immunohistochemically profile for PBX1 and other biological antibodies as previously described. PBX1 mouse monoclonal antibody (Abnova, (M01), clone 4A2 [16] was optimized to a working concentration, utilizing μ m matched full- face excisional tissue sections. Antigen retrieval was performed using Leica ER2 (pH 9.0) retrieval solution, water bath at 95C for 35 minutes followed by TBS at 50C for 10minutes. Blocking was done using Thermo Fisher UltraV block 5mins. Antibody concentration was 1:50 overnight at 4°C. Subsequently, 4 μ m TMA sections were immunostained using the optimized staining protocol. Detection

was achieved using the Novalink Polymer Detection kit (Leica Micro- systems Inc., USA). Negative controls were performed by omission of the primary antibody. IHC revealed that PBX1 had a nuclear location (4B). Nuclear staining was scored based on the H-score and Allred Quick score. Determination of the optimal cut-offs was performed using histograms and confirmed using X-tile bio-informatics software (Yale University, USA) [51] where data were split into training and validation sets. A total of 1198 tumours were suitable for analysis (460 PBX1 negative and 738 PBX1 positive). The tumour cores were evaluated by two of the coauthors (DKP and PV) blinded to the clinico-pathological characteristics of patients. There was substantial intra- and inter- observer agreement (k [0.768; Cohen's j and multi-rater j tests, respectively).

Survival analysis using transcriptomics data

Positive PBX1 expression was dichotomised to 90% using distant metastases as an outcome and determined through X-tile software (University of Yale, Yale, USA) [51]. Prognostic analyses were undertaken using Kaplan-Meier curves using the Log-rank test. A *p*-value < 0.05 was considered significant

Survival analysis using transcriptomics data

Database construction and survival analysis was performed as described previously [52]. For the expression of the genes, the median expression was used as the cut-off in a Cox regression analysis. Kaplan-Meier survival plot, and hazard ratio with 95% confidence intervals and logrank *P* value were calculated and plotted in R using Bioconductor packages.

Construction of METABRIC microarray database

The Metabric project employed Illumina microarrays to measure gene expression across all genes. The raw gene chip files were downloaded from the European Genome-phenome Archive (EGA) (<https://www.ebi.ac.uk/ega/>) [25]. The entire cohort contains 1988 samples, the average overall survival is 8.07 years, 76% of the samples are ER positive and 47.3% are node positive. Due to batch effects between the Metabric training and validation sets we have not used the pre-normalized tables but have re-run the complete pre-processing for all arrays. In this, raw data was first imported into R and summarized using the beadarray package [53]. For annotation, the Illumina Humanv3 database of Bioconductor was used (<http://www.bioconductor.org>). Finally, quantile normalization was performed using the preprocessCore package [54].

FAIRE and DNaseI-hypersensitivity-seq

FAIRE-seq and DHS-seq analysis were performed as previously described [16, 55] with minor modifications

Circulating free DNA analysis

DNA was extracted from blood cell pellets and 1ml plasma as described previously [30]. To confirm amplification of PBX1, 10ng cfDNA was subjected to 5 cycles of preamplification with a mix of primers (including PBZ1 and reference genes). Each sample was then analysed in triplicate by real-time quantitative PCR (qPCR) in a 10 µl reaction volume. Reactions were run on an Applied Biosystems thermal cycler (Step One Plus) and analysed with Step One v2.1 software and Microsoft Excel.

Primers

Primers sequences (RT-qPCR, ChIP-qPCR) are available upon requests. Primers to study PBX1 amplification were as follow: Primers PBX1-F: AGCCCACTCATCTTACGTGAC PBX1-R: ACGAAATTCCACTCCAACCTCCA Probe PBX1 GCTCAGGCCTATCTTCTGGA FAM-MGB.

ACKNOWLEDGMENTS AND FUNDING

This work was supported by an Imperial Junior Research Fellowship and a Breast Cancer Campaign Pilot Grant (P46869) to LM, Wellcome Trust [103034/Z/13/Z] to DKP and the OTKA K108655 grant to BG. The Imperial College Healthcare NHS Trust Tissue Bank provided tissue samples. Other investigators may have received samples from these same tissues' The research [was funded / supported by] the National Institute for Health Research (NIHR) Biomedical Research Centre based and Experimental Cancer Medicine Centre, Cancer Research UK Centre at Imperial College Healthcare NHS Trust Imperial College London. M.L. holds a young investigator award from the Ontario Institute for Cancer Research (OICR) and a new investigator salary award from the Canadian Institute of Health Research (CIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

Authors' contribution

LM designed the study. LM, EWL, BG, ML and CC performed bioinformatic analysis and interpreted the data. LM, BG, ML and CC wrote the manuscript. LM, DP, VN,

SPH, DFG, YL, MF and ARG performed lab experiments. LM, DP, MF, JHS, NP, PV, AG, CN, IOE, SS and EAR performed and interpreted the pathological analysis. CC, LW, KP, DG and LP recruited patients and collected and prepared clinical specimens. LM, DFG and JS performed and analyzed cfDNA data.

Abbreviations

ER α , Estrogen Receptor α ; EGF, Epidermal Growth Factor; TF, Transcription factor; cfDNA, circulating free-DNA FAIRE, Formaldehyde assisted isolation of regulatory elements; CNV, copy number variation.

REFERENCES

1. Early Breast Cancer Trialists Collaborative Group EBCTCG Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, McGale P, Pan HC, Taylor C, Wang YC, Dowsett M, Ingle J, Peto R. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet*. 2011; 378:771–84.
2. Tam WL, Weinberg RA. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med*. Nature Publishing Group 2013; 19:1438–49.
3. You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell*. 2012; 22:9–20.
4. Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and evolving paradigms. *Cell*. 2011; 147:275–92.
5. Magnani L, Stoeck A, Zhang X, Lanczky A, Mirabella AC, Wang TL, Györfy B, Lupien M. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proc Natl Acad Sci USA*. 2013.
6. Di Leo A, Jerusalem G, Petruzelka L, Torres R, Bondarenko IN, Khasanov R, Verhoeven D, Pedrini JL, Smirnova I, Lichinitser MR, Pendergrass K, Garnett S, Lindemann JPO, Sapunar F, Martin M. Results of the CONFIRM phase III trial comparing fulvestrant 250 mg with fulvestrant 500 mg in postmenopausal women with estrogen receptor-positive advanced breast cancer. *J Clin Oncol*. 2010; 28:4594–600.
7. Perey L, Paridaens R, Hawle H, Zaman K, Nolé F, Wildiers H, Fiche M, Dietrich D, Clément P, Köberle D, Goldhirsch A, Thürlimann B. Clinical benefit of fulvestrant in postmenopausal women with advanced breast cancer and primary or acquired resistance to aromatase inhibitors: final results of phase II Swiss Group for Clinical Cancer Research Trial (SAKK 21/00). *Ann Oncol*. 2007; 18:64–9.
8. Chia S, Gradishar W, Mauriac L, Bines J, Amant F, Federico M, Fein L, Romieu G, Buzdar A, Robertson JFR, Brufsky A, Possinger K, Rennie P, Sapunar F, Lowe E, Piccart M. Double-blind, randomized placebo controlled trial of fulvestrant compared with exemestane after prior nonsteroidal aromatase inhibitor therapy in postmenopausal women with hormone receptor-positive, advanced breast cancer: results from EFACT. *J Clin Oncol*. 2008; 26:1664–70.
9. Toy W, Shen Y, Won H, Green B, Sakr RA, Will M, Li Z, Gala K, Fanning S, King TA, Hudis C, Chen D, Taran T, Hortobagyi G, Greene G, Berger M, Baselga J, Chandarlapaty S. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet*. 2013; 45:1439–45.
10. Robinson DR, Wu Y-M, Vats P, Su F, Lonigro RJ, Cao X, Kalyana-Sundaram S, Wang R, Ning Y, Hodges L, Gursky A, Siddiqui J, Tomlins SA, Roychowdhury S, Pienta KJ, Kim SY, Roberts JS, Rae JM, Van Poznak CH, Hayes DF, Chugh R, Kunju LP, Talpaz M, Schott AF, Chinnaiyan AM. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet*. 2013; 45:1446–51.
11. Barone I, Iacopetta D, Covington KR, Cui Y, Tsimelzon A, Beyer A, Andò S, Fuqua SAW. Phosphorylation of the mutant K303R estrogen receptor alpha at serine 305 affects aromatase inhibitor sensitivity. *Oncogene*. 2010; 29:2404–14.
12. Lupien M, Jeyakumar M, Hébert E, Hilmi K, Cotnoir-White D, Loch C, Auger A, Dayan G, Pinard G-A, Wurtz JM, Moras D, Katzenellenbogen J, Mader S. Raloxifene and ICI182, 780 increase estrogen receptor-alpha association with a nuclear compartment via overlapping sets of hydrophobic amino acids in activation function 2 helix 12. *Mol Endocrinol*. 2007; 21:797–816.
13. Bunone G, Briand PA, Miksicek RJ, Picard D. Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *EMBO J*. 1996; 15:2174–83.
14. Osborne CK, Schiff R. Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer. *Breast*. 2003; 12:362–7.
15. Lupien M, Meyer CA, Bailey ST, Eeckhoutte J, Cook J, Westerling T, Zhang X, Carroll JS, Rhodes DR, Liu XS, Brown M. Growth factor stimulation induces a distinct ER(alpha) cistrome underlying breast cancer endocrine resistance. *Genes Dev*. 2010; 24:2219–27.
16. Magnani L, Ballantyne EB, Zhang X, Lupien M. PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLoS Genet*. 2011; 7:e1002368.
17. Park JT, Shih I-M, Wang T-L. Identification of Pbx1, a potential oncogene, as a Notch3 target gene in ovarian cancer. *Cancer Res*. 2008; 68:8852–60.
18. Martin L-A, Farmer I, Johnston SRD, Ali S, Dowsett M. Elevated ERK1/ERK2/estrogen receptor cross-talk enhances estrogen-mediated signaling during long-term estrogen deprivation. *Endocr Relat Cancer*. 2005; 12:S75–84.

19. Tyson JJ, Baumann WT, Chen C, Verdugo A, Tavassoly I, Wang Y, Weiner LM, Clarke R. Dynamic modelling of oestrogen signalling and cell fate in breast cancer cells. *Nat Rev Cancer*. 2011; 11:523–32.
20. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet*. 2010; 43:27–33.
21. Magnani L, Eeckhoutte J, Lupien M. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet*. 2011; 27:465–74.
22. Jozwik KM, Carroll JS. Pioneer factors in hormone-dependent cancers. *Nat Rev Cancer*. 2012; 12:381–5.
23. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991; 19:403–10.
24. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490:61–70.
25. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, METABRIC Group Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012; 486:346–52.
26. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2:401–4.
27. Sarajlić A, Filipović A, Janjić V, Coombes RC, Pržulj N. The role of genes co-amplified with nicastrin in breast invasive carcinoma. *Breast Cancer Res Treat*. 2014; 143:393–401.
28. Page K, Hava N, Ward B, Brown J, Guttery DS, Ruangpratheep C, Blighe K, Sharma A, Walker RA, Coombes RC, Shaw JA. Detection of HER2 amplification in circulating free DNA in patients with breast cancer. *Br J Cancer*. 2011; 104:1342–8.
29. Whale AS, Huggett JF, Cowen S, Speirs V, Shaw J, Ellison S, Foy CA, Scott DJ. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic Acids Res*. 2012; 40:e82.
30. Shaw JA, Page K, Blighe K, Hava N, Guttery D, Ward B, Brown J, Ruangpratheep C, Stebbing J, Payne R, Palmieri C, Cleator S, Walker RA, Coombes RC. Genomic analysis of circulating cell free DNA infers breast cancer dormancy. 2011.
31. Umetani M, Domoto H, Gormley AK, Yuhanna IS, Cummins CL, Javitt NB, Korach KS, Shaul PW, Mangelsdorf DJ. 27-Hydroxycholesterol is an endogenous SERM that inhibits the cardiovascular effects of estrogen. *Nat Med*. 2007; 13:1185–92.
32. Blair RM, Fang H, Branham WS, Hass BS, Dial SL, Moland CL, Tong W, Shi L, Perkins R, Sheehan DM. The estrogen receptor relative binding affinities of 188 natural and xenochemicals: structural diversity of ligands. *Toxicol Sci*. 2000; 54:138–53.
33. Nelson ER, Wardell SE, Jasper JS, Park S, Suchindran S, Howe MK, Carver NJ, Pillai RV, Sullivan PM, Sondhi V, Umetani M, Geradts J, McDonnell DP. 27-Hydroxycholesterol links hypercholesterolemia and breast cancer pathophysiology. *Science*. 2013; 342:1094–8.
34. Wu Q, Ishikawa T, Sirianni R, Tang H, McDonald JG, Yuhanna IS, Thompson B, Girard L, Mineo C, Brekken RA, Umetani M, Euhus DM, Xie Y, Shaul PW. 27-Hydroxycholesterol promotes cell-autonomous, ER-positive breast cancer growth. *Cell Rep*. 2013; 5:637–45.
35. Suva ML, Riggi N, Bernstein BE. Epigenetic Reprogramming in Cancer. *Science*. 2013; 339:1567–70.
36. Magnani L, Lupien M. Chromatin and Epigenetic Determinants of Estrogen Receptor Alpha (ESR1) Signaling. *Mol Cell Endocrinol*. 2013.
37. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, Ali S, Chin S-F, Palmieri C, Caldas C, Carroll JS. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012; 481:389–93.
38. Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*. 2010; 143:470–84.
39. Lupien M, Eeckhoutte J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*. 2008; 132:958–70.
40. Theodorou V, Stark R, Menon S, Carroll JS. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res*. 2012.
41. Chou J, Provot S, Werb Z. GATA3 in development and cancer differentiation: cells GATA have it! *J Cell Physiol*. 2010; 222:42–9.
42. Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, Godwin AK, Korach KS, Visvader JE, Kaestner KH, Abdul-Karim FW, Montano MM, Keri RA. FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. *Development*. 2010; 137:2045–54.

43. Polyak K, Metzger Filho O. SnapShot: Breast Cancer. *Cancer Cell*. 2012; 22:562–562.e1.
44. Brunner AL, Li J, Guo X, Sweeney RT, Varma S, Zhu SX, Li R, Tibshirani R, West RB. A shared transcriptional program in early breast neoplasias despite genetic and clinical distinctions. *Genome Biol*. 2014; 15:R71.
45. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012; 481:306–13.
46. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10:R25.
47. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008; 9:R137.
48. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010; 28:495–501.
49. Nielsen CB, Younesy H, O'Geen H, Xu X, Jackson AR, Milosavljevic A, Wang T, Costello JF, Hirst M, Farnham PJ, Jones SJM. Spark: a navigational paradigm for genomic data exploration. *Genome Res*. 2012; 22:2262–9.
50. Ellis IO, Galea M, Broughton N, Locker A, Blamey RW, Elston CW. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. *Histopathology*. 1992; 20:479–89.
51. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bioinformatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res*. 2004; 10:7252–9.
52. Györfy B, Lanczky A, Szallasi Z. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocr Relat Cancer*. 2012; 19:197–208.
53. Dunning MJ, Smith ML, Ritchie ME, Tavaré S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics*. (Oxford, England): 2007; 23:2183–4.
54. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. Oxford, England: 2003; 19:185–93.
55. Ling G, Sugathan A, Mazor T, Fraenkel E, Waxman DJ. Unbiased, genome-wide *in vivo* mapping of transcriptional regulatory elements reveals sex differences in chromatin structure associated with sex-specific liver gene expression. *Mol Cell Biol*. 2010; 30:5531–44.

ARTICLE

Received 20 Feb 2015 | Accepted 28 Oct 2015 | Published 27 Nov 2015

DOI: 10.1038/ncomms10044

OPEN

Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion

Van T.M. Nguyen^{1,*}, Iros Barozzi^{2,*}, Monica Faronato¹, Ylenia Lombardo¹, Jennifer H. Steel¹, Naina Patel¹, Philippa Darbre³, Leandro Castellano¹, Balázs Györfy^{4,5}, Laura Woodley⁶, Alba Meira¹, Darren K. Patten¹, Valentina Vircillo⁷, Manikandan Periyasamy¹, Simak Ali¹, Gianmaria Frige², Saverio Minucci², R. Charles Coombes¹ & Luca Magnani¹

Endocrine therapies target the activation of the oestrogen receptor alpha (ER α) via distinct mechanisms, but it is not clear whether breast cancer cells can adapt to treatment using drug-specific mechanisms. Here we demonstrate that resistance emerges via drug-specific epigenetic reprogramming. Resistant cells display a spectrum of phenotypical changes with invasive phenotypes evolving in lines resistant to the aromatase inhibitor (AI). Orthogonal genomics analysis of reprogrammed regulatory regions identifies individual drug-induced epigenetic states involving large topologically associating domains (TADs) and the activation of super-enhancers. AI-resistant cells activate endogenous cholesterol biosynthesis (CB) through stable epigenetic activation *in vitro* and *in vivo*. Mechanistically, CB sparks the constitutive activation of oestrogen receptors alpha (ER α) in AI-resistant cells, partly via the biosynthesis of 27-hydroxycholesterol. By targeting CB using statins, ER α binding is reduced and cell invasion is prevented. Epigenomic-led stratification can predict resistance to AI in a subset of ER α -positive patients.

¹Department of Surgery and Cancer, Imperial College London, London W12 0NN, UK. ²IFOM-IEO Campus, European Institute of Oncology, Milan 20139, Italy. ³School of Biological Science, University of Reading, Reading RG6 6LA, UK. ⁴MTA TTK Lendület Cancer Biomarker Research Group, 2nd Department of Pediatrics, Semmelweis University, Budapest H-1117, Hungary. ⁵MTA-SE Pediatrics and Nephrology Research Group, 2nd Department of Pediatrics, Semmelweis University, Budapest H-1117, Hungary. ⁶ECMC, Charing Cross Hospital, London W120nn, UK. ⁷Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, Arcavacata di Rende (CS) 87036, Italy. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to L.M. (email: l.magnani@imperial.ac.uk).

Over 70% of all breast cancers are characterized by the expression of the nuclear receptor oestrogen receptor alpha (ER α)¹. Patients with ER α -positive disease are routinely treated with adjuvant endocrine therapies (ETs) after surgery. ETs include a series of compounds designed to interfere with ER α activation, including selective oestrogen receptor modulators (for example, tamoxifen), aromatase inhibitors (AIs, for example, Letrozole and Anastrozole) and selective oestrogen receptor downregulators (Faslodex/Fulvestrant)². Tamoxifen competitively bind to ER α in the place of estrogens such as estradiol (E2). AIs are designed to deplete the estrogens in circulation, while Fulvestrant irreversibly bind to ER α leading to ER α degradation. Over 40% of ER α breast cancers eventually relapse from ETs becoming progressively refractory to further treatments¹. It remains unclear whether resistance to ETs involves drug-specific mechanisms. Since ETs are characterized by different mechanisms^{2,3}, it is conceivable that chronic exposure to therapy may impart specific selective pressure and elicit different response mechanisms to induce stable phenotypic changes.

During development and differentiation, cell identity is established by the epigenetic activation of cell-type-specific distal regulatory elements mediated by combinatorial patterns of histone modifications^{4,5}. These regions are often embedded in chromatin-accessible sites and contain DNA sequence motifs for cell-type-specific transcription factors (TFs)⁶. The epigenome retains some degree of plasticity, as cell identity can be epigenetically reprogrammed by exposure to external stimuli such as the ectopic expression of TFs⁶. Cancer cells can also remodel histone modifications and chromatin accessibility in response to chronic exposure to therapeutic agents^{7–9}. For example, adaptation to oestrogen deprivation in ER α -positive breast cancer cells invokes extensive epigenetic reprogramming especially in distal regulatory regions such as enhancers⁷. Regulatory elements directly associated with active transcription, and more complex regulatory regions (that is, super-enhancers (SEs)) can be mapped by measuring the histone modification histone 3-lysine 27 (H3K27ac)^{10–14}. In this study, we have investigated the nature of drug adaptation to a diverse range of ETs in isogenic ER α -positive breast cancer cells using integrative and orthogonal epigenomic approaches.

Our data demonstrate that breast cancer cells evolve individual epigenomic and transcriptomic profiles in response to treatment. Both *in vitro* and *in vivo*, AI-resistant cells acquire the most aggressive phenotype and develop invasive characteristics in two-dimensional and three-dimensional (3D). Integrative analyses of AI-resistant cells identify stable upregulation of the entire cholesterol biosynthesis (CB) pathways including genes involved in 27-hydroxyl-cholesterol (27HC) biosynthesis. 27HC stimulation is sufficient to promote oestrogen-independent ER α binding to thousands of putative regulatory regions. In agreement with this finding, chromatin immunoprecipitation sequencing (ChIP-seq) analysis of the chromatin of AI-resistant cells confirmed extensive ER α binding despite oestrogen-deprived conditions. Furthermore, CB blockers (statins) can reduce ER α binding to DNA and abrogate cell invasion. CB upregulation also occurs *in vivo* during breast cancer progression. Finally, we demonstrate that a CB-based signature might be used to improve the stratification of ER α breast cancer patients before adjuvant treatment.

Results

Adaptation to AI treatment leads to *de novo* invasiveness. ETs are designed to block oestrogen-driven proliferation by interfering with one specific TF (for example, ER α). However, we

hypothesized that the development of resistance may follow distinct routes and generate alternative phenotypes through the different molecular mechanisms specific to each agent². To test this hypothesis, we used a series of isogenic cell lines resistant to single agents or a combination of agents (endocrine therapy (ET)-resistant ETR cells, Fig. 1a)¹⁵. Our aim was to understand the connection between the acquisition of drug-resistance and breast cancer progression, particularly metastatic development. We then carried out a real-time, impedance-based assay to monitor the migratory and invasion behaviour of ETR cells. These assays demonstrated that long term estrogen deprived (LTED) cells (mimicking AI resistance¹⁶) specifically develop migration and invasion properties, while MCF7- and TAM/Fulvestrant-resistant cells (MCF7T and MCF7F) do not (Fig. 1b; Supplementary Fig. 1A). It is worth noting that sequential resistance did not increase these traits (LTEDT and LTEDF versus LTED). To corroborate these findings, we developed a 3D invasion assay (organoids assay) in which the cells were allowed to form 3D bodies and then embedded in Matrigel. In agreement with the 2D assay, AI-resistant cells (LTEDs) spread remarkably through the Matrigel environment, while the MCF7s do not (Fig. 1c; Supplementary Movie 1A–D). Finally, we developed a metastatic mouse model to validate our findings *in vivo* by engineering two red-fluorescent protein reporter (mCherry) lines (MCF7-FRP and LTED-RFP). After injection into the tail vein of NOD-SCID mice (Fig. 1c), only LTED cells colonize several sites including the lymph nodes and bones (Fig. 1d). Histology analyses of putative metastatic tissues from the lymph nodes and bone marrow confirms that the invading cells are of human origin and contains breast cancer protein markers¹⁷ (HNA and PBX1 (ref. 18), Fig. 1d; Supplementary Fig. 1B). Overall, these data demonstrate that AI-resistant cells acquire traits common to aggressive breast cancers. Tamoxifen- or Fulvestrant-resistant cells, although becoming resistant to the cytostatic effect of the agent, do not evolve into highly invasive cells thus supporting the notion of agent-specific reprogramming.

ETR cells follow distinct reprogramming routes. To decipher the molecular changes induced by drug-specific resistance, we developed an unbiased integrative approach combining RNA sequencing (RNA-seq) and H3K27ac ChIP-seq in ETR cells. All ETR cells express significantly lower messenger RNA (mRNA) levels for several oestrogen receptor target genes, indicating that ET treatments still exert negative pressure on ER α signalling. As expected, MCF7F and LTEDF cells acquire an ER α -negative status, while the rest of the panel remained ER α positive (Supplementary Fig. 2A). Interestingly, other chromatin components of ER α signalling¹⁹ (for example, ER α pioneer factors such as FoxA1, PBX1 and GATA3 (refs 20,21)) remain expressed at similar levels in all cell lines (Supplementary Fig. 2B,C). Cell lines with sequential resistance are characterized by a greater number of differentially regulated genes (Supplementary Fig. 3A). Correspondingly, these cells display increased numbers of potential regulatory regions as demonstrated by the significant increase in H3K27ac-positive loci (Supplementary Fig. 3B). This increase could not simply be attributed to general changes in H3K27ac distribution (Supplementary Fig. 3C).

We then carried out a gene ontology analysis on our RNA-seq data to identify potential pathways responsible for the invasive phenotype characterizing LTED cells. To increase the power of these analyses, we clustered together all the genes differentially regulated in invasive (LTED–LTEDT–LTEDF) and non-invasive (MCF7T–MCF7F) cells (Supplementary Fig. 4A). Strikingly, we found that LTED cells activated metabolic pathways and lipid metabolism (for example, super-pathway of

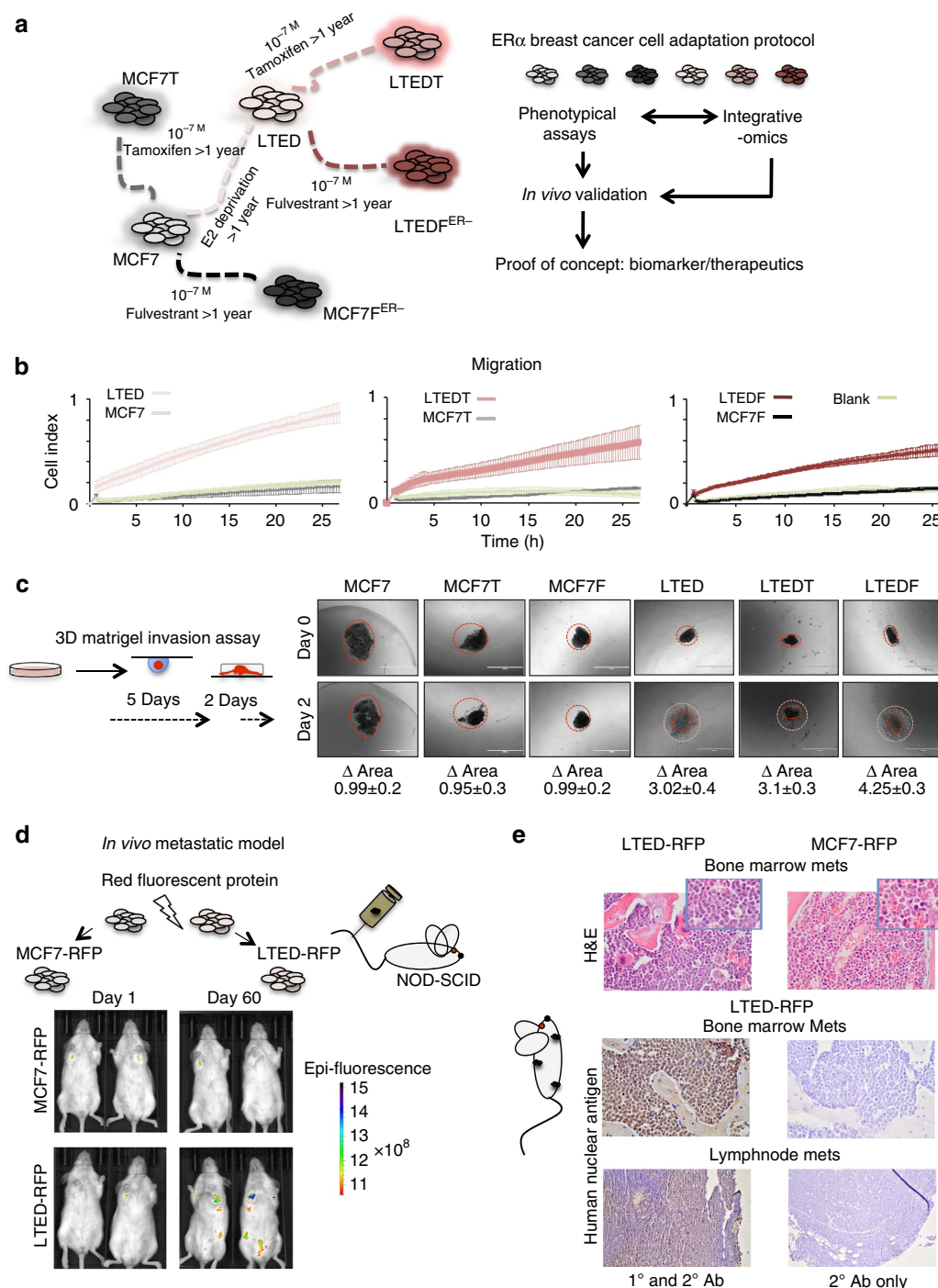


Figure 1 | Adaptation to AI treatment is associated with cell invasion. (a) *In vitro* adaptation experimental protocol: ER α -positive breast cancer cells (MCF7) were chronically exposed to endocrine therapies to generate ETR lines¹⁵. Each cell type was screened for the development of invasive potential. ETR cells were also profiled using integrative epigenomics and the results were cross-validated using clinical samples. (b) Real-time monitoring of cell migration demonstrates increased motility of LTEDs cells compared with other resistant lines. (c) 3D organoids assay was used to demonstrate invasive potential under physical constraints (Matrigel). Δ Area is calculated using ImageJ (area day 2/area day 0, $n=5$). The horizontal white bar indicates a distance of 1,000 μ m. (d) LTED cells can form metastases *in vivo* at various sites. Cells were stably transfected with the mCherry construct, injected in the tail of the mice and followed using *in vivo* imaging. (e) Snapshots of histological analysis of tissues from mice injected with LTED or MCF7 cells. The first two panels represent haematoxylin and eosin (H&E) staining on bone marrow (insets taken at a higher magnification). The bottom four panels represent immunohistochemistry on tissue from mouse bone marrow or lymph nodes showing invasion by cells immunoreactive for human nuclear antigen (left panels) or negative controls with no primary antibody (right panels).

CB, Fig. 2a; Supplementary Fig. 4B; Supplementary Data 1). More importantly, these pathways were not active in non-invasive MCF7T and MCF7F cells (grey versus red bars), suggesting

differential transcriptional reprogramming (Fig. 2a). These changes are not transiently imposed by the culture condition, but rather represent stably engrained transcriptional. In point of

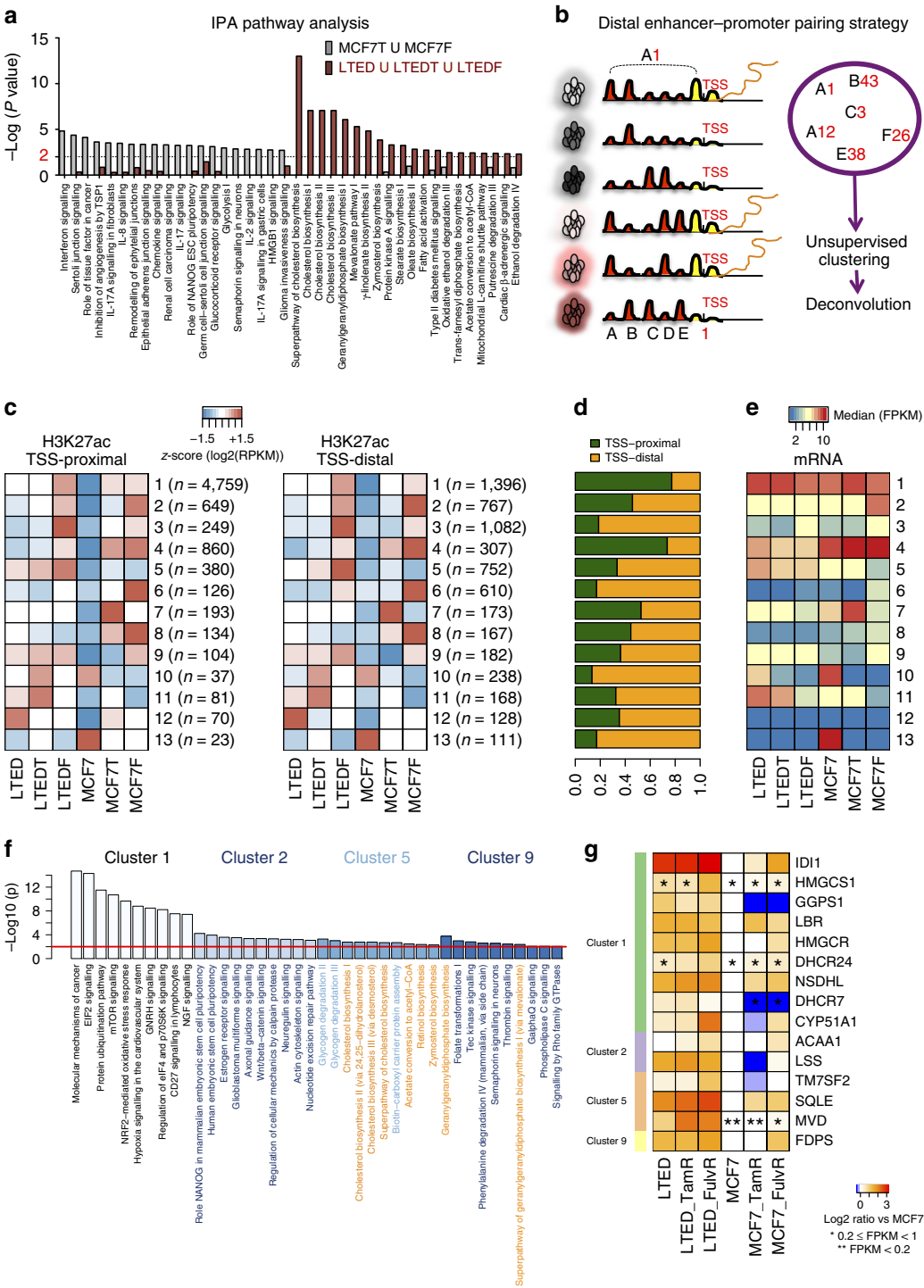


Figure 2 | Drug adaptation and epigenetic reprogramming. (a) IPA pathway analysis of differentially upregulated genes (for details of the analysis, see Supplementary Fig. 4A). The top 20 IPA pathways for invasive and non-invasive cells are plotted with their respective *P* values. Metabolic pathways are in bold. The dotted line is set at a *pVal* = 0.01. (b) Diagrammatic image of the bioinformatics strategy adopted for the epigenetic analysis in the six cell lines. For more details, please see the Methods (c) Heatmaps showing the mean H3K27ac level in each cluster for each of the six cell lines. Regions were split into TSS-proximal and distal and shown as separate heatmaps. Total number of regions in brackets. (d) The proportion of TSS-proximal and distal regions in each cluster is shown. (e) Heatmap showing the median of the linear FPKM for each cluster along the cell lines. (f) IPA canonical pathways significantly enriched considering the genes putatively regulated by the TSS-proximal elements in enhancer clusters. The 10 most enriched terms were extracted and their log10 (*P* value) (Fisher’s exact test) are shown as bar plots. The red line is set at a *pVal* = 0.01. (g) FPKM values of genes belonging to the ‘Superpathway of Cholesterol Biosynthesis’ (IPA canonical pathways) identified using H3K27ac patterns alone but also showing a significant deregulation in RNA-seq compared with MCF7. Transcripts showing low expression levels are indicated with asterisks.

fact, we could not re-establish the original transcriptional profile (for example, MCF7) by culturing resistant cells in MCF7 culture conditions and the absence of endocrine treatment (Supplementary Fig. 4C).

We hypothesized that the reprogrammed transcriptomes evolved by invasive LTEDs cells might play a central role in driving aggressive breast cancer *in vivo*. To test this, we devised five individual cell signatures containing the top 10% of upregulated genes specific to each individual cell line (Supplementary Data 2) and split patients into high and low expressors (Supplementary Fig. 5A). Analysis of the METABRIC²² data set (restricting it to the ER α -positive patients) demonstrates that only LTED-based signatures significantly predict poor survival (Supplementary Fig. 5B,C). Collectively, these data demonstrate that AI-resistant cell lines evolve transcriptional programmes typical of patients developing more aggressive tumours.

Drug resistance invokes activation of regulatory elements. We reasoned that chronic drug exposure probably acts as a constant selective pressure within our cell line system and allow for selective activations of regulatory regions. Indeed, we demonstrated that these changes become heritable and eventually independent from media conditions. We then investigated the epigenetic circuitry that drives the cell-type-specific transcription in our ETR models. By comparing our epigenomic maps with publically available H3K27ac profiles, we confirmed the common ancestry of ETR cells (data from ENCODE⁵), suggesting that adaptation to therapy does not entirely upset the epigenetic landscape (Supplementary Fig. 6A). Next, we devised a strategy to pinpoint the distal regulatory enhancers actively involved in regulating cell-type-specific gene expression (Fig. 2b). We first identified all the loci that showed changes in H3K27ac by comparing ETR cells with MCF7s using MACS²³. We then pooled and clustered all the regions showing a significant increase or decrease in at least one cell line according to the co-variation of their patterns across the six cell lines (Fig. 2b and extended Methods). Using this strategy, we extrapolate genome-wide groups of potentially interacting TSS and enhancer pairs sharing the same pattern of activity across the lines (13 different clusters). Each cluster is then de-convoluted to TSS-proximal and TSS-distal (canonical enhancers) elements (Fig. 2c,d). We then match the promoter-proximal H3K27-acetylated regions to RefSeq genes and annotate the genes according to our RNA-seq analysis. These analyses demonstrate that H3K27ac at distal enhancers is highly consistent with H3K27ac at promoter and mRNA transcription (Fig. 2e). Of note, our analysis provides a solid alternative for studying distal regulatory regions in the absence of detailed interaction maps.

We then used Ingenuity Pathway Analysis to identify cell-type-specific pathways based on canonical enhancers (Supplementary Data 3). Cluster 1 contains mainly promoter-proximal elements (Fig. 2d) and shows an increase in H3K27ac signals in all resistant cells (Fig. 2c). In agreement, pathways significantly enriched in this cluster have been previously associated with ET resistance (mTOR, EIF2, NGF, EGF and HER2 signalling^{22,24}) (Fig. 2f). Clusters 2 and 3 are directly associated to the loss of ER α , thus demonstrating the central role of this TF in breast cancer cells. Despite over half of the regions in these groups mapping to distal enhancers, coding regions unambiguously assigned to TSS-proximal elements are enriched for pluripotency genes (Fig. 2f), a feature typical of basal, stem-like ER α -negative breast cancers²⁵. Focusing on LTED-specific regulatory elements, cluster 5 and cluster 9 are almost exclusively enriched for genes belonging to

the super-pathway of CB confirming our transcriptional analysis (Fig. 2f). Overall, these unbiased epigenomics analyses strongly support our initial findings, and substantiate the hypothesis that AI-resistant cells epigenetically activate distinct pathways compared with other resistant cells.

Drug-resistance-specific activation of SEs. Our integrative epigenomic analyses suggest that endogenous activation of CB is one of the key pathways epigenetically activated in AI-resistant cells. Epigenetic activation of the CB pathway appears to occur through a multi-step process. For example, many regulatory elements associated with genes belonging to CB are included in cluster 1 (common to all resistant cells, Supplementary Data 3), indicating that these genes are active also in Tamoxifen- or Fulvestrant-resistant cells (Fig. 2g). Nevertheless, the rate-limiting enzymes of this pathway and their associated regulatory regions (for example, HMGCR and SQLE) are ultimately activated/upregulated only in oestrogen-deprived LTED cells (Fig. 2g). We therefore decided to investigate whether epigenomic activation of CB involves more complex epigenetic features such as SEs. Recent publications^{11,26,27} have identified dense clusters of distal regulatory regions that are strongly associated with transcription and particularly susceptible to perturbation, termed SEs (Fig. 3a). Adapting a previously published approach²⁷, we identified a set of 710 SEs with different types of distribution across ETR cells (Supplementary Fig. 6B). Initial clustering analysis using SEs reveals that LTEDs cells share extensive epigenetic features. Instead, MCF7F cells maintain fewer connections with parental MCF7 further demonstrating the marked effect of ER α loss (Supplementary Fig. 6C). As expected, SEs activation strongly reflects cell unique traits, for example, the SEs associated with oestrogen/oestrogen receptor target gene¹⁷ EGR3 was identified only in MCF7 cells (Fig. 3a). Next, we clustered SEs to obtain 10 distinct groups using a similar approach-based on co-variance (Figs 2b and 3b and extended materials). IPA analysis of LTED-specific SE clusters (clusters 3 and 5) identified once more metabolic and signalling pathways. It must be pointed out that the acetylation of SEs was strongly associated with the expression of nearby genes, in agreement with previous findings²⁸ (Supplementary Fig. 6D; Supplementary Data 4). Finally, we identified the TF motifs buried within peak-valley-peak structures to gain further insights into the regulatory circuitry underlying these SEs. Running IPA analysis to identify pathways coordinated by TFs that could potentially bind these SE valleys, we found that LTED-specific SE clusters 3, 5 and 6 were enriched for several metabolic nuclear receptor-driven pathways (Fig. 3c, the TFs identified include FOXO1, FOXA1, FOXA2, NR5A2 for cluster 3 and RXRG, VDR, RXRB, RXRA for cluster 5). Finally, we investigated the accessible chromatin landscape to gain more detailed insights into global changes in TF occupancy using high-depth DHS-seq²⁹. Footprinting analysis to scan for putative binding sites using DHS-seq identifies a massive increase in SREBP1 and SREBP2, the two master regulators of cholesterol homeostasis in mammals³⁰, in LTED cells (Fig. 3d). Significantly, SREBP1/e are also the top candidates transcriptional regulators in LTED for our RNA-seq-based analysis (see Supplementary Data 1, upstream regulators). SREBP1 activates cholesterol genes by binding their promoter after nuclear translocation in response to intracellular cues^{31,32}. SREBP1 nuclear translocation can be inhibited using a novel series of compounds called fatostatin³³. To test whether SREBP1 translocation was effectively one of the key events in CB activation, we treated MCF7 and LTED cells with fatostatin. Fatostatin treatment induced a significant reduction for several key genes in the CB pathway including the two rate-limiting enzymes HMGCR and SQLE specifically in LTED cells (Fig. 3e).

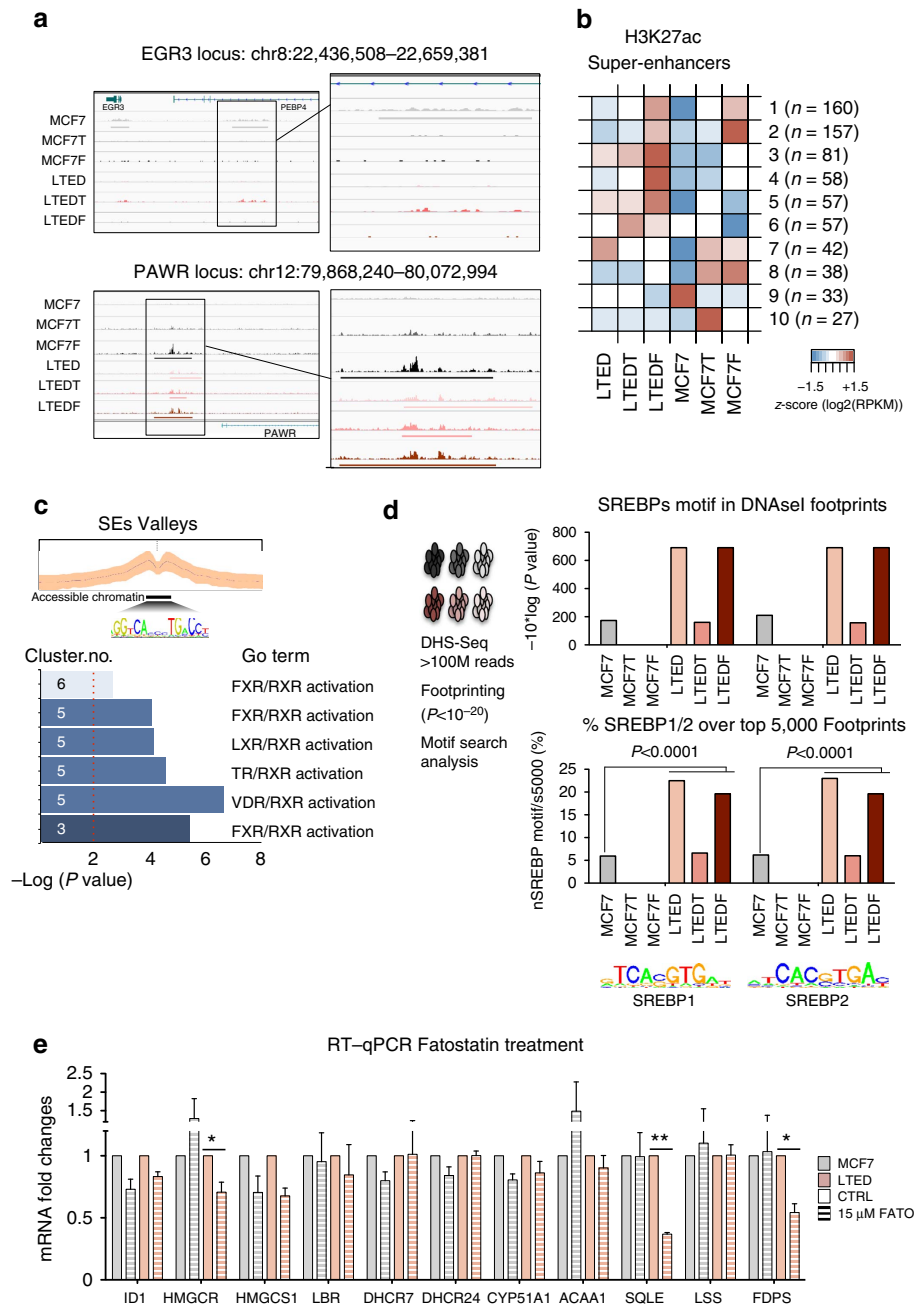


Figure 3 | Activation of super-enhancers during epigenetic reprogramming. (a) Snapshot representing SE peaks and associated raw data for two genomic locus underlying cell-type-specific phenotypes (EGR3 and PAWR). (b) Heatmaps showing the mean H3K27ac level in each cluster of SE regions for each one of the six cell lines. Total number of regions in brackets. (c) DNA motif analysis (HOMER) was carried out in the potential regulatory regions between H3K27ac-marked nucleosomes for specific clusters of SEs. Gene Ontology analysis based on potential interacting TFs is shown in the bottom panel. The vertical dashed red line indicates $P=0.01$. (d) Pipeline for the analysis of DHS-set Footprinting data sets. SREBP1 and SREBP2 footprints are uniquely enriched within the accessible chromatin landscape of LTEDs cells. The left panel indicates P values from Cistrome analysis (Sitepro). The right panel displays the number of motifs within the top 5,000 footprints (expressed as a percentage). Statistical analyses were carried out using a Pearson's χ^2 -test. (e) Treatment with a SREBP1 inhibitor (Fatostatin) induces downregulation of key cholesterol biosynthesis transcripts specifically in LTED cells. Error bars represent s.e.m. calculated on three independent experiments. Asterisks indicate significant statistical differences comparing treated with mock-treated cells (T -test, $P<0.05$).

In summary, SEs and chromatin accessibility analyses further demonstrate a role for cholesterol-driven signalling centred on the potential crosstalk between nuclear receptors and cholesterol-specific TFs. These data suggest that epigenetic reprogramming might induce new transcription factor dependencies.

Epigenetic reprogramming involves larger chromatin structures. Recent evidence suggests that SEs are regulated in the context of TADs²⁸, with relatively insulated effects within CTCF-defined regions. TADs are physically defined genomic regions that shows cohesive gene regulation during development³³. Several

studies indicate that TAD boundaries are generally resistant to differentiation³⁴. We observed that in some instances, clusters of genes physically located in the same genomic region (for example, keratin locus, chr12q13.13) are differentially regulated as a single transcriptional unit (Supplementary Data 4). More specifically, the KRT locus (KRT7, KRT80, KRT81, KRT83 and KRT86) is upregulated in all resistant cells and its putative SE belongs to cluster 1 (Fig. 3b). This raises the possibility of coarse-grained epigenetic reprogramming occurring at the level of entire genomic compartments. To test this hypothesis, we sorted 2,283 TADs (previously identified in IMR90 cells³³) according to the ratio of their H3K27ac levels in LTED versus MCF7 (Fig. 4a). The TAD encompassing the KRT genes (chr12:52,513,733 – 53,913,733) ranked in the top 5%, showing a 1.42-fold increase in H3K27ac in LTED cells (Fig. 4a), confirming that epigenetic reprogramming might target sets of genes simply based on their physical proximity.

We then investigated whether analogously to KRTs, CB genes are regulated in response to coarse-grain activation of their respective SE-TAD structures. Genes in the CB pathways were found scattered over different genomic compartments with no coherent changes in H3K27ac, suggesting a TAD-independent, punctuated regulation (Fig. 4a). Thus, while some CB genes might be passively regulated in response to widespread epigenetic changes, others may be controlled by more precise epigenetic changes, often counteracting the local TAD changes. For example, at the SQLE locus the local increase in acetylation (LTED versus MCF7) substantially overshadows the modest increase in TAD acetylation (Fig. 4b, yellow arrow). Strikingly, genes such as HMGCR and CYP51A1 gain H3K27ac signals in their putative regulatory regions despite locally losing acetylation at the TAD level (Fig. 4b, black arrows). We obtained similar results when comparing global changes in RNA transcription within the TAD to the local changes for single CB genes. These data strongly suggest that CB genes activated in non-invasive cells (Fig. 2g) might be passengers of TAD-wide epigenetic activation, while rate-limiting CB genes are specifically targeted in AI-resistant cells only.

To further investigate the role of CB in AI-resistant cells, we transcriptionally profiled the entire cholesterol superpathway using reverse transcription (RT)–quantitative (q)PCR. The CB pathway is made up of 24 genes coding for enzymes necessary to build cholesterol molecules from acetyl-derived carbons (Fig. 4c; Supplementary Fig. 7A). Strikingly, 22/24 genes in the CB superpathway are upregulated in LTED cells, while 2/24 and 3/24 are upregulated in MCF7T and MCF7F cells, respectively (Fig. 4c). To rule out cell-type-specific bias in the activation of CB, we then analysed a second ER α -positive cell lines (in T47D breast cancer cells³⁵). In agreement with MCF7 data, we observe a marked upregulation of CB transcripts in T47D-LTED cells as compared with naive T47D (21/24, Fig. 4d). On the other hand, non-tumorigenic MCF10A have much lower mRNA levels for all CB genes compared with non-invasive T47D or MCF7 cells (average $14 \pm 11\%$).

To confirm the epigenetic nature of CB activation in T47D-LTED cells, we assessed the H3K27ac status of several regulatory elements comparing resistant and naive cells. In agreement with MCF7 ETR cells, T47D-LTED exhibit a strong increase in K27acetylation at enhancers associated with CB genes (Fig. 4e). Finally, we investigated the epigenetic activation of distal enhancers associated with CB directly in clinical specimens. In agreement with our *in vitro* data, ChIP–qPCR analysis demonstrate a significant increase in H3K27ac signal in a biopsy derived from a metastatic deposit (AI resistant) compared with a sample derived from a primary, drug-sensitive ER α breast cancer patient (Fig. 4g).

Altogether, these data demonstrate that the selective pressure operated by chronic drug exposure is counteracted using very

distinctive epigenetic mechanisms. More specifically, these data strongly support the centrality of CB epigenetic activation in AI-resistant cells.

CB promotes ER α activation. ER α and other nuclear receptors can be activated by a vast array of ligands including sexual hormones (estrogens) and other cholesterol-derived compounds. Recent evidence has shown that cholesterol derivatives such as 25 and 27 hydroxycholesterol can also promote the transcription of ER α target genes in breast cancer cells^{36,37}. Interestingly, 27HC can also increase metastatic invasion in mice xenografted with MCF7 cells³⁸. We then hypothesized that one of the possible consequence of epigenetic activation of CB in AI-resistant cells might be the activation of autocrine signalling via *de novo* synthesis of 27HC³⁹. To test this, we initially investigated the expression levels of CYP27A1 and CYP7B1, the two enzymes involved in the synthesis and catabolism of 27HC from cholesterol precursors³⁹. Analogously to CB genes, CYP27A1 is significantly upregulated in LTEDs compared with MCF7 and other resistant cells (Fig. 4c). Similarly, T47D-LTED exhibited an increase in CYP27A1 compared with parental T47D cells (Fig. 4d). On the other hand, we could not identify CYP7B1 transcripts in our cells (Fig. 4c,d). CYP27A1 is highly expressed in macrophages but also in ER α -positive breast cancer cells *in vivo*³⁸. For example, we could identify strong expression of CYP27A1 in two newly established ER α metastatic cell line models derived from pleural effusion of AI-resistant patients (Supplementary Fig. 7B). These data suggest that LTED cells can convert the cholesterol obtained via augmented CB into 27HC in an autocrine pattern³⁹.

Next, we performed ChIP-seq analysis to determine whether 27HC could induce ER α chromatin binding in a way similar to other well-characterized ligands such as estradiol (E2)³⁸ and EGF⁴⁰. We performed these experiments using parental MCF7, in which endogenous CB is modest compared with LTED, to increase the signal to noise ratio and get robust ER α induction. 27HC stimulation (1 μ M)³⁸ led to ER α binding to 5,174 regions. Interestingly, over 70% of 27HC-ER α -binding sites were occupied by E2-ER α or EGF-ER α (Fig. 5a). Shared ER α -binding sites were enriched for H3K27Ac, H3K4me1, PBX1 and FOXA1, two TFs central to ER α signalling¹⁹ and were highly accessible (as shown by DHS-seq) while depleted of repressive histone marks (K9me3 and K27me3) (Fig. 5b). Shared ER α -binding sites had over 70% overlap with the core-ER α -binding sites found *in vivo*⁴¹. We thus concluded that 27HC-bound ER α can be recruited to regulatory regions commonly associated with oestrogen signalling. In addition, we also identified a set of binding sites unique to 27HC activation (Fig. 5c). These sites appeared to be weaker and less enriched for K27ac but still under selective constraints (Phastcons method).

We then examined ER α recruitment in LTED cells cultured in the complete absence of oestrogen. We hypothesized that endogenous CB activity might be sufficient to promote ER α binding to a large set of regulatory elements. Focusing on 27HC-specific ER α -binding sites, we found that oestrogen-depleted LTED cells have a higher average ER α binding compared with MCF7 cells grown in the same conditions (Fig. 5d). When we expanded this analysis to the entire repertoire of ER α -binding sites (E2, EGF and 27HC in MCF7 plus LTED, total 51,974 regions), LTED cells still exhibited stronger genome-wide recruitment of ER α compared with their parental MCF7 cells grown in the absence of oestrogen (Fig. 5e), suggesting that CB activation might be sufficient to replace oestrogen in LTED cells. Blocking endogenous CB using Lovastatin (HMGCR inhibitor) or Terbinafine (SQLE inhibitor) induces a modest

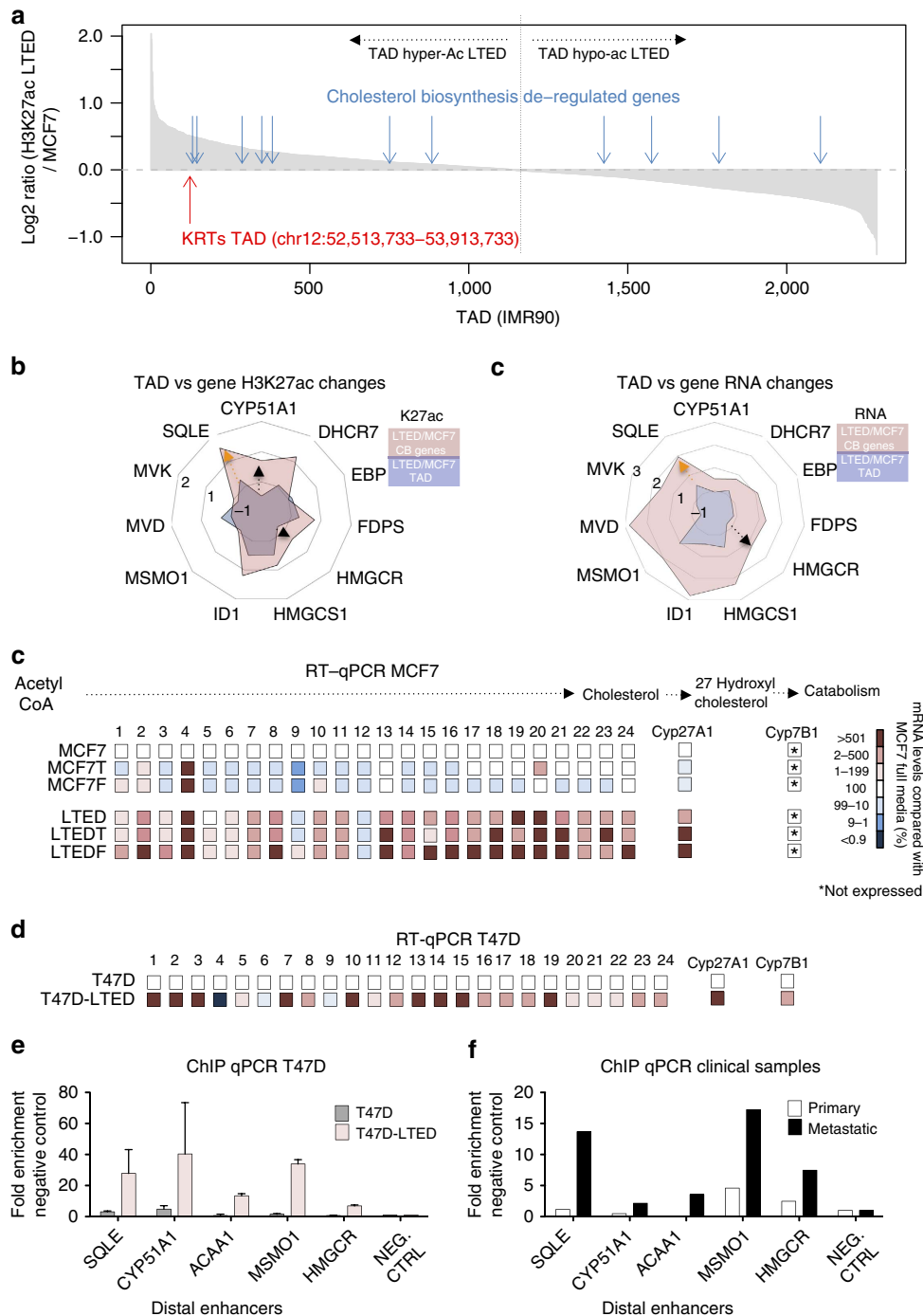


Figure 4 | Epigenetic activation of cholesterol biosynthesis. (a) In all, 2,283 topologically associating domains (TADs) (identified in IMR90 cells³³ and showing acetylation in at least one out of the six cell lines) were sorted according to the ratio of their H3K27ac levels in LTED versus MCF7. The red arrow highlights the TADs encompassing the KRT genes (chr12:52,513,733 – 53,913,733) while the light-blue arrows highlights the TADs encompassing the deregulated genes in the cholesterol biosynthesis pathway. (b) Radar plots of global (TAD) versus local (CB genes) changes in H3K27ac and RNA transcription. Data are represented as ratios (Log₂ RPKM LTED/ RPKM MCF7). (c) RT-qPCR analysis of all the enzymes involved in cholesterol and 27HC biosynthesis (identifiers in Supplementary Fig. 8A). RNA fold differences were calculated comparing MCF7 with individual resistant cells averaging at least three independent experiments. All comparisons with different colours were statistically significant (T-test). (d) RT-qPCR analysis of all the enzymes involved in cholesterol and 27HC biosynthesis (identifiers in Supplementary Fig. 8A). RNA fold differences were calculated comparing T47D with individual resistant cells averaging at least three independent experiments. (e) ChIP-qPCR for H3K27ac demonstrates strong enrichment at distal regulatory regions controlling CB genes in AI-resistant T47D cells. Error bars represent s.e.m. calculated on three independent experiments. (f) ChIP-qPCR for H3K27ac demonstrates strong enrichment at distal regulatory regions controlling CB genes in metastatic clinical samples.

although significant reduction in ER α recruitment at 27HC unique sites in LTED cells (Fig. 5f), further suggesting that CB contributes to cell-autonomous ER α recruitment in LTED cells.

Finally, we reasoned that epigenetic CB activation in AI-resistant LTED cells might contribute to the invasive phenotype displayed by these cells. We therefore tested the possibility of

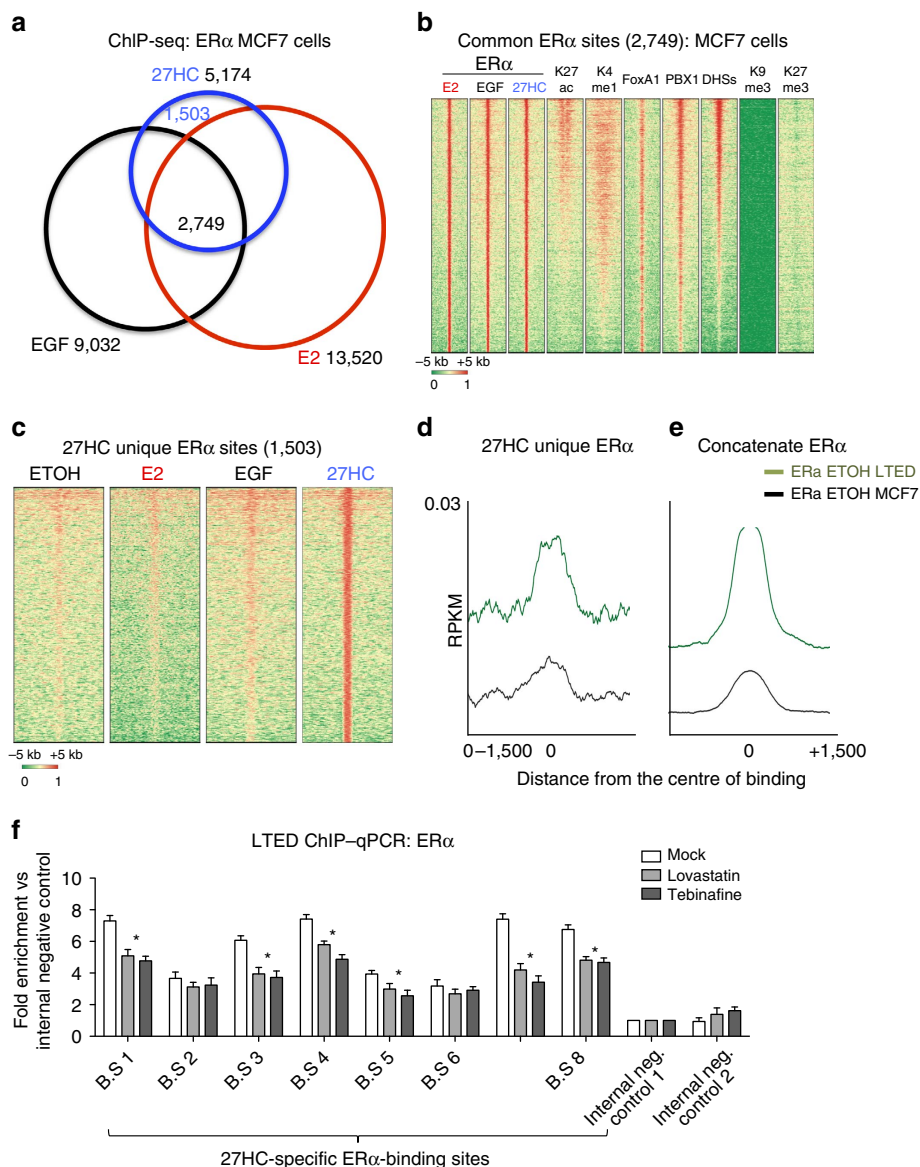


Figure 5 | 27-Hydroxycholesterol promotes ER α binding. (a) Venn diagrams of ligand-specific ER α cistromes. (b) Heatmaps showing the epigenetic features of ER α -binding sites common to all stimuli. (c) Heatmaps showing ER α binding at 27HC-specific loci. (d) Comparison between ER α ChIP-seq signals from LTED (green) and MCF7 (black) grown in the absence of estrogens at 27HC-specific sites. (e) Same analysis as in d focusing on all ER α -binding sites (right panel). (f) ChIP-qPCR analysis shows reduced ER α binding at 27HC-specific ER α -binding sites upon treatment with CB pathway inhibitors (48 h treatment). Error bars represent s.e.m. calculated on three independent experiments. Asterisks indicate significant differences compared with mock-treated cells (analysis of variance, with Dunnett's t -test $P < 0.05$).

blocking cellular invasion by targeting CB. As expected, LTED cells tended to be more sensitive to inhibition of CB as demonstrated by their respective half-maximal inhibitory concentration (IC_{50}) (MCF7: 89 and 90 μ M, LTED: 22 and 42 μ M, Lovastatin and Terbinafine, respectively). More importantly, inhibition of CB or short interfering RNA-mediated depletion of SQLE was sufficient to block the invasion of AI-resistant LTED cells in 3D Matrigel assay (Fig. 6a; Supplementary Movie 5A–D). Conversely, treatment with mevalonate (the byproduct of HMGCR) significantly increased the invasive potential of non-invasive MCF7 cells (Fig. 6b). SQLE depletion also resulted in the significant repression of several genes involved in cellular invasion (Fig. 6c). Altogether, these data strongly suggest that epigenetically activated CB contributes to the invasive phenotype exhibited by AI-resistant LTED cells.

CB serves as a biomarker in AI-treated breast cancer patients.

Our integrative analysis of breast cancer cell lines resistant to individual ETs identified epigenetic activation of CB as a potentially important resistance mechanism specific to AI-treated patients. We therefore investigated whether our findings had translational potential and could be used to stratify patients before endocrine treatment. We initially developed a gene signature (Supplementary Data 1) to retrospectively stratify TCGA ER α -positive breast cancer patients that expressed high or low levels of CB gene mRNA levels. We used TCGA data since it was possible to retrieve treatment information from these patients (AI $n = 144$, Tamoxifen $n = 127$). ER α -positive patients with high expression of our CB-based signature at diagnosis consistently display shorter recurrence- and metastatic-free survival and are characterized by poor survival (Supplementary Fig. 8A,B). The

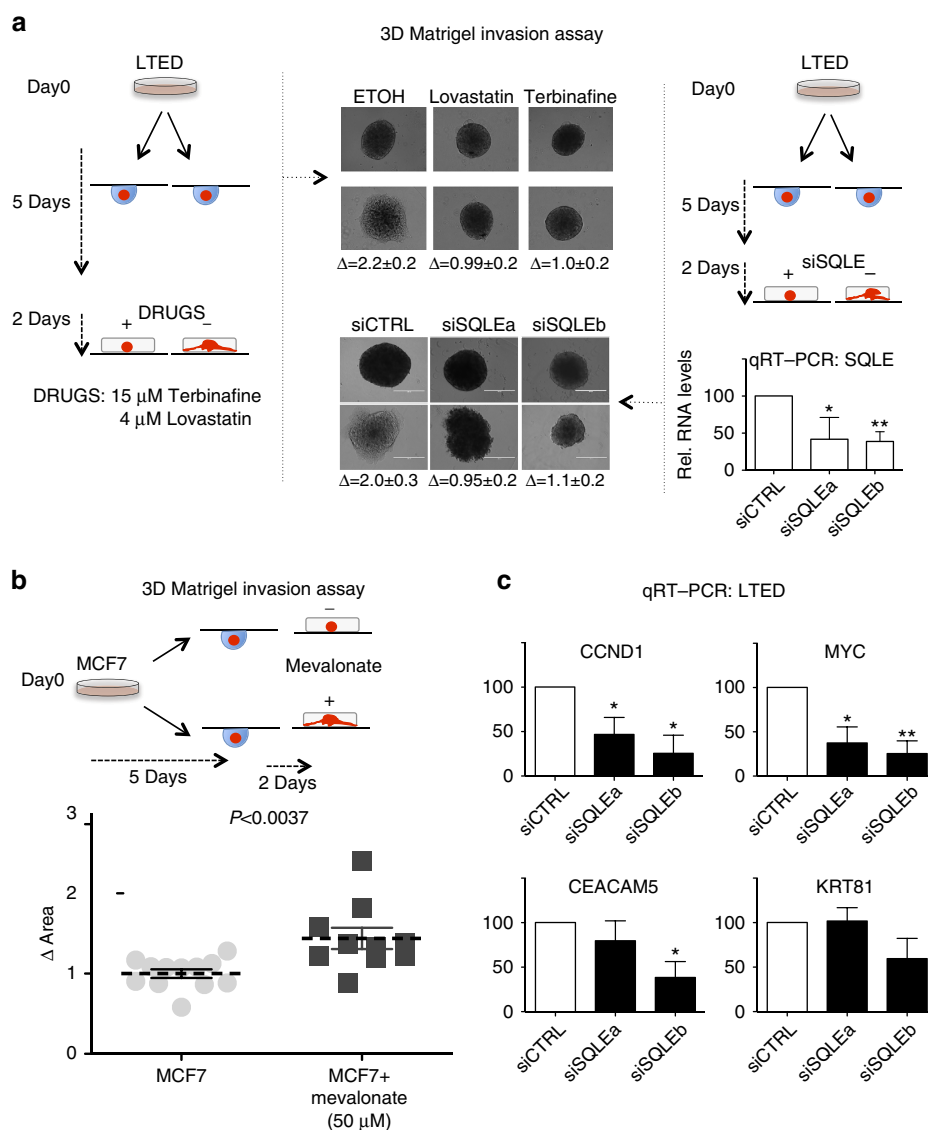


Figure 6 | Blocking cholesterol biosynthesis reduces cellular invasion. (a) Blocking cholesterol biosynthesis via treatment with CB inhibitors (left panel) or direct depletion of SQLE (right panel) is sufficient to arrest LTED invasion in Matrigel organoids assays. The horizontal white bar indicate a distance of 400 μ m. Δ Area is calculated using ImageJ (area day 2/area day 0, $n = 5$). SQLE RNA fold changes were determined comparing two independent short interfering RNA (siRNA) (a,b) versus a non-targeting control. Bars represent the mean with the s.d. of three independent experiments. Asterisks indicate significant difference (analysis of variance (ANOVA), $P < 0.05$). (b) Treatment with Mevalonate increases the invasive potential of MCF7 cells in 3D organoid assays. Δ Area is calculated using ImageJ (area day 2/area day 0). P value was calculated using a Student's T -test. (c) qRT-PCR analysis of genes involved in invasion/migration in LTED cells upon SQLE depletion. KRT81 is shown as a control. RNA fold changes were determined comparing two independent siRNAs versus a non-targeting control. Error bars represent s.e.m. calculated on three independent experiments. Asterisks indicate significant differences (ANOVA, $P < 0.05$).

same signature cannot be used to identify ER α -negative patients, suggesting a significant degree of specificity (Supplementary Fig. 8C). There are no significant changes in performance when correcting for luminal A or luminal B (Supplementary Fig. 9). CB can be directly upregulated in response to p53 mutations⁴²; however, MCF7 cells are p53^{wt} and LTED do not acquire p53 mutations. On the other hand, T47-LTED incur in significant CB activation despite T47D being already p53^{mut}, suggesting a degree of independence between the genetic status of p53 and the activation of CB. Finally, we still observed a strong prognostic significance within the ER α -p53-mutated subpopulation, suggesting the existence of additional mechanisms (Supplementary Fig. 9).

Most importantly, our CB signature was particularly efficient in assessing overall survival in AI-treated patients but not in

Tamoxifen-treated patients (Fig. 7a) again suggesting that endocrine resistance is driven by distinct mechanisms. Examining the expression of genes involved in CB in an independent cohort (FEMARA trial⁴⁵), we find that MSMD1, MVD and SQLE mRNA levels are statistically higher in patients that did not respond to neo-adjuvant (pre-surgery) AI therapy (Fig. 7b). Overall, these data suggest that patients with high CB are less likely to benefit from AI treatment and are in substantial agreement with our *in vitro* results.

SQLE is the second rate-limiting enzyme in the CB pathway and one of the genes epigenetically activated specifically in LTED cells (Figs 2g and 4b). As expected, SQLE-encoded protein is strongly upregulated in all LTED models (Fig. 7c). We therefore quantified mRNA levels for HMGCR and SQLE in an ER α breast cancer patient treated with AI for whom we had

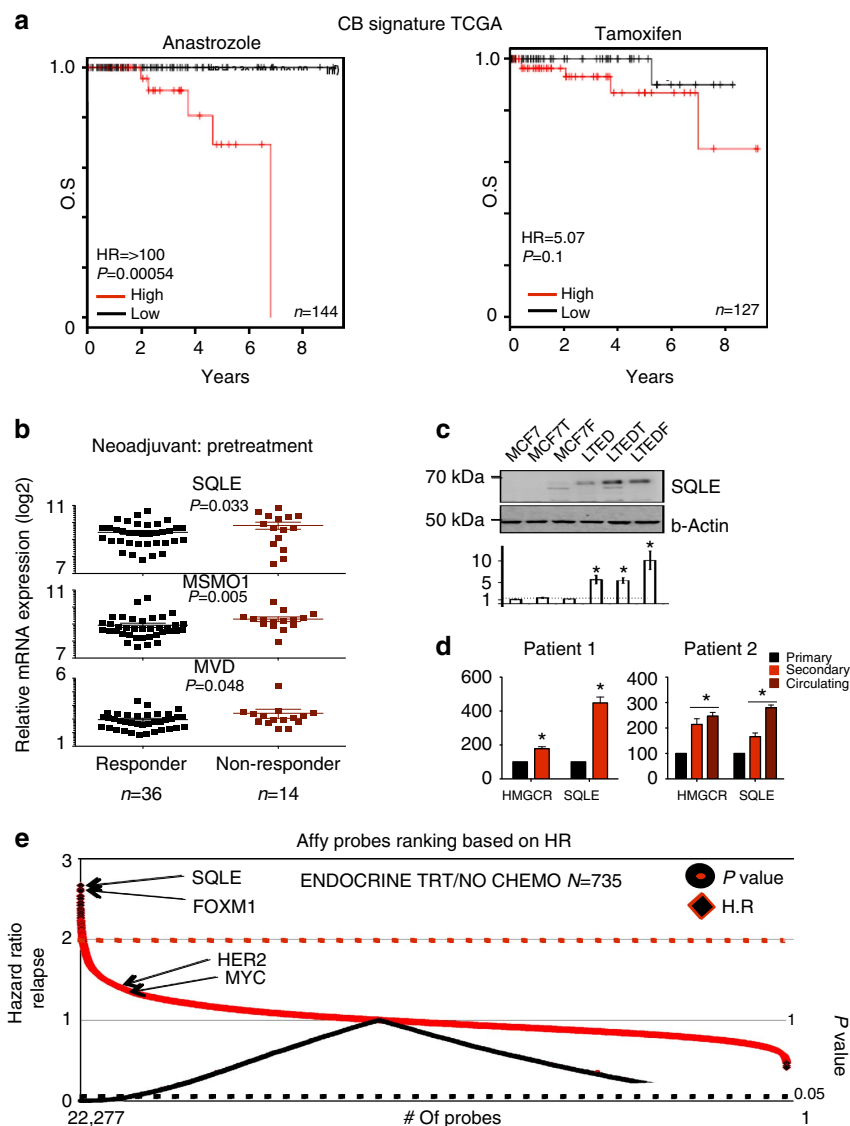


Figure 7 | CB-guided stratification of AI-resistant patients. (a) Kaplan-Meier survival plots using averaged expression of differentially regulated enzymes (RNA-seq) involved in cholesterol biosynthesis in specific subgroups of ET-treated patients. (b) Pretreatment mRNA levels (FEMARA Trial, see text) of key enzymes can help identify AI-resistant patients. *P* values were calculated using a Mann-Whitney *U*-test. (c) Western blot analysis of SQLE and protein quantification. Error bars represent s.e.m. calculated on four independent experiments. Asterisks indicate significant difference comparing each cell type with MCF7 cells (analysis of variance, *P*<0.05). (d) HMGCR and SQLE mRNA levels were quantified in longitudinal samples from breast cancer patients. mRNA was extracted from FFPE slides marked by a registered pathologist. Error bars represent s.e.m. calculated on three independent experiments. Asterisks indicate significant differences compared with primary samples (*T*-test, *P*<0.05). (e) Manhattan plot for the prognostic potential of all human genes contained in the affy-chip U133. Both hazard ratio and *P* values are represented. Red and black dotted lines represent hazard ratio = 1 and *P*<0.05, respectively.

longitudinal data (RNA obtained at diagnosis, first relapse and in circulating metastatic cells from pleural effusion). In agreement with our previous finding, we found that key CB genes are progressively upregulated during cancer progression (Fig. 7d). Overall, SQLE emerged from these analyses as a novel candidate biomarker to identify AI-resistant breast cancer patients. This was further confirmed by our unbiased analysis of the entire human transcriptome. When we sorted all 22,277 probes present in the U133A gene chip according to their potential to predict relapse in a panel of 724 ER α -positive breast cancer patients treated with ET, SQLE (probe set 209218_at) ranked 1st (hazard ratio = 2.67) (Fig. 7e; Supplementary Data 5). Finally, we speculated that genetic-epigenetic interactions might facilitate SQLE expression in ER α -positive breast

cancer. Remarkably, SQLE is amplified in almost 10% of ER α -positive patients and overexpressed in an additional 13%⁴⁴ (Supplementary Fig. 10A). SQLE amplification directly correlates with increased mRNA expression in primary tumours (Supplementary Fig. 10B). SQLE emerges also as one of the top-ranking overexpressed genes in 10 breast cancer-independent data sets (invasive ductal carcinoma (IDC) versus normal breast, Supplementary Data 6). More importantly, SQLE amplification or mRNA overexpression is sufficient to stratify outcome in ER α -positive breast cancer patients from the TCGA cohort (Supplementary Fig. 10C). Altogether, our data strongly support the prediction generated using drug-specific resistant cells and confirm a functional role for epigenetic activation of CB as a mechanism of resistance to AI therapy *in vivo*.

Discussion

In addition to clonal selection⁴¹, adaptation to drug treatment may involve heritable transcriptional changes that allows cells to survive in the presence of externally imposed selective pressure^{7,8}. Using ER α -dependent breast cancer cells as a working model, we demonstrate that adaptation to therapeutic agents sharing the same target follow distinct routes depending on the specific mechanism of action. More specifically, the appearance of oestrogen-independent growth corresponds to the epigenetic activation of CB. CB functionally contributes to the invasive phenotype AI-resistant cells and CB activation can be used to predict resistance to AI before treatment. Drug-resistance, however, does not invariably coincide with the development of more aggressive phenotypes, suggesting that in some instances the transition to metastatic progression requires additional reprogramming. Using recently developed paradigms, our data demonstrate that individual drugs impose very specific epigenetic changes that target both local regulatory elements such as individual enhancers and large topological domains and SEs¹¹. This is surprising considering that endocrine treatments impinge on the same TF (for example, ER α). It will be important to evaluate whether a similar process occurs also in other malignancies.

TF motif analyses of epigenetically reprogrammed loci show that deregulation of the core network of TFs, which lies at the basis of cell-type specification (for example, FoxA1)⁴⁵ might lead to aberrant rearrangements of entire chromosomal territories. These rearrangements may include surrounding genes that are central to functional reprogramming, drug resistance and invasion. At the same time, specific selective pressures might trigger epigenetic activation of tailored pathways (for example, CB), irrespective of the broad chromatin context in which single genes are nested.

Our data support a model in which chronic exposure to AI constitutively activates CB. Intra-tumour activation of CB may then lead to the local accumulation of metabolic ligands for other nuclear receptors. Promisingly, we find that AI-resistant SEs are enriched with DNA motifs of metabolic nuclear receptors including FXR and LXR. It is well established that 27HC is also an LXR agonist³⁸, therefore we cannot exclude that LXR might also contribute to AI resistance in combination with endogenous ER α activation⁴⁶. In addition, endogenously produced ligands, including 27HC might act on additional nuclear receptors and allow the activation of an alternative cohort of regulatory elements (Supplementary Fig. 11). Of note, 40/48 nuclear receptors (NRs) have increased expression in LTED cells (FXR > 300 fold, LXR > 1.8 fold) and 9 of these are markedly increased in ER α -negative LTEDF (PPAR γ > 500-fold) strongly suggesting that other nuclear receptors might contribute to the invasive phenotype observed in LTEDF cells. In the same way, autocrine metabolic signalling might also contribute to ER α binding found in metastatic breast cancer *in vivo*⁴¹.

Epigenetic activation of the CB pathway might also be supported by specific genetic aberrations. For example, SQLE amplification occurs in a significant proportion of cancers⁴⁷, but the functional consequences of SQLE amplification have not previously been appreciated. SQLE locus is just 2 Mb away from the oncogene *Myc* in a hotspot for genomic rearrangements⁴⁴. However, we could not find any clear correlation between SQLE and MYC expression, suggesting that the poor outcome typical of breast cancer patients with high SQLE does not simply reflect MYC activation. Notably, our model could partially explain recently published epidemiological observations. While the protective role of anti-cholesterol statin towards breast cancer is still debated, a nation-wide study found that statin users with ER α -positive breast cancer are less likely to develop relapses⁴⁸. In

agreement with our findings, no effect on relapse rates was noted in patients with ER α negative⁴⁸.

Recent clinical trials recommend a switch from Tamoxifen to AI with ovarian suppression in pre-menopausal women⁴⁹. Our data suggest that Tamoxifen and AI exposure do not have an equivalent effect on breast cancer cells; thus ER α patients should be assigned to a specific endocrine treatment depending on the molecular profile of their tumour. Moreover, our data offer novel molecular insights and warrants for additional clinical trials combining statins and AI in patients with metastatic disease⁴⁷. Pharmacodynamics studies show no influence of statins on E2 levels or Anastrozole metabolism⁵⁰. Interestingly, statins were shown to interfere with ovarian cancer and multiple myeloma growth in pre-clinical models^{51,52}. Our work therefore offers a new paradigm in the context of how epigenomes evolve in response to interrelated exogenous pressures while paving the way for future studies designed to help us understand the specific effects of therapeutic intervention during metastatic progression.

Methods

Cell lines and clinical samples. Cell lines were cultured as follows MCF7 and T47D: DMEM cell growth medium (DMEM with 10% of fetal bovine serum (FBS) and 100 U penicillin/0.1 mg ml⁻¹ streptomycin plus 10⁻⁸ 17- β -estradiol (SIGMA E8875). LTED cells were passaged in stripped DMEM cell growth medium (Phenol-free DMEM with 10% of charcoal-stripped FBS (CDT) and 100 U penicillin/0.1 mg ml⁻¹ streptomycin MCF7T and LTEDT were cultured as MCF7 and LTED with the addition of 10⁻⁷ 4-hydroxytamoxifen (SIGMA H7904). MCF7F and LTEDF were cultured as MCF7 with the addition of 10⁻⁷ Fulvestrant (SIGMA I4409)¹⁵. Tissues (primary, secondary and pleural effusions) were obtained from the Imperial Tissue Bank. The Tissue Management Committee reviewed and approved all documents. Imperial Tissue bank released stores the samples under an HTA Licence (12275). The appropriate regulatory committees approved all the relevant documents. Each patient signed an informed consent to donate tissue to Imperial Tissue Bank. All patients' notes are kept within locked filing cabinets within locked offices in research. This is to maintain confidentiality. Tissues used for qRT-PCR were manually enriched for tumour cells by the resident pathologist. Pleural effusions were cultured in DMEM, 10% FBS and phenol for a brief time before RNA extraction. Cells were tested, genotyped and characterized as ER α positive (LGC standards). Circulating tumour cells were characterized as mammary gland based on short tandem repeat similarity with HTB-22 (LGC standards).

Invasion assays. Boyden chamber and 3D Matrigel invasion assays were performed as previously described⁵³. Briefly, cells (5 $\times$ 10⁴ in 200 μ l of α -MEM) were plated in the Matrigel-coated upper chambers of the 24-well Transwell invasion assay plate (Corning). Each condition was represented in triplicate. Plates were incubated at 37 °C for 24 h for HCC1806 cells. Cells in the lower chamber (including those attached to the under surface of the membrane) were trypsinized and counted using a Casy 1 counter (Sharfe System). Mean and s.e.m. of independent experiments were calculated. Statistical analysis was performed using a two-tailed Student's *t*-test to determine the statistical significance of the differences observed. A *P* value below 0.05 was considered significant.

Fatostatin and mevalonate treatments. Cell lines were cultured in the appropriate conditions until 60–70% confluence and then treated with 15 μ M of Fatostatin (SIGMA F8932) for 6 h before RNA extraction and qRT-PCR analysis. For mevalonate treatment, organoids were allowed to form and embedded in Matrigel (see Hanging-drop assay) then treated with 50 μ M mevalonate⁵⁴ (SIGMA 90469).

Mouse models. MCF7-mCherry-Puro and LTED-mCherry-Puro cells (1 $\times$ 10⁶) were injected into the tail vein of NOD-SCID mice (*n* = 3 per group). Images were taken of the mice weekly from dorsal and ventral views for 8 weeks to monitor the development of metastases. The mice were assessed weekly using whole-body imaging to quantify the relative amounts of tumour burden. Imaging was performed with a highly sensitive, cooled charge-coupled device camera mounted in a light-tight specimen box (IVIS; Xenogen). The imaging and quantification of signals were controlled by acquisition and analysis software. At termination, the lymph nodes, lungs, kidney, bones and livers were harvested by dissection. The lymph nodes, lungs, kidneys, liver and bones were fixed for 24 h in neutral buffered formalin, and then rinsed in 70% ethanol. Following fixation, the bones were decalcified in 10% EDTA in 0.1 M Tris-HCl pH 7.4 for 1 week, re-fixed in formalin and then rinsed in 70% ethanol. All tissues were dehydrated using ethanol, cleared with Histoclear and embedded in paraffin wax. Paraffin sections (5 μ m) were

stained with haematoxylin and eosin to reveal their tissue morphology and the presence of infiltrating cells. The animal trials were carried out under London Home Office license authority and London Home Office Ethics Committee guidelines.

Immunohistochemistry and western blot. Paraffin sections were dewaxed and antigen retrieval was carried out with citrate buffer at pH 6.0 or in 10 mM Tris-HCl, pH 9.0. Tissue sections were pretreated using 0.3% H₂O₂ in PBS, rinsed in PBS and then incubated with 20 µl ml⁻¹ normal goat serum. Primary antibodies were diluted in PBS and incubated overnight at 4 °C, along with negative control sections where the primary antibody was omitted. All the bound antibodies were detected using biotinylated anti-mouse or anti-rabbit secondary antibodies, detected using the Vectastain Elite peroxidase ABC kit and the ImmPACT DAB kit (Vector Laboratories). Stained sections were counterstained with haematoxylin. Antibody dilutions: PBX1 1:50 (Abnova H00005087-M01), Human nuclear antigen 1:25 (Abcam ab191181), Pan-cytokeratin 1:100 (SIGMA C5992). For western blot, SQLE antibody (SIGMA HPA018038) was used at 1:250 and ERα (Santa Cruz sc-543) was used at 1:1,000.

Hanging-drop assay. Briefly, cells were trypsinized and counted. In all, 250,000 cells were resuspended in 1 ml of phenol-red or phenol-red-free DMEM with the relevant concentrations of Estradiol, Fulvestrant and Tamoxifen. Twenty-microlitre drops were dispensed on to a 10-cm dish lid. Five millilitre of medium were added to the bottom of the dish to prevent evaporation. The lid was inverted and the drops were left hanging for 5 days. The newly formed organoids were then transferred into 10 µl Matrigel drops (BD 356237, BD 356234) and moved to 24- or 48-well plates after which immunofluorescence or real-time monitoring (respectively) was carried out. Two hundred microlitre of medium with the relevant concentration of drugs were added to the wells.

Real-time invasion and migration assays. Cell invasion and migration were assessed in real time using the xCELLigence system CIM-16 wells plates. The bottom chamber was filled with 160 µl of full medium (phenol-red or phenol-red-free DMEM with the relevant concentrations of Estradiol, Fulvestrant and Tamoxifen) and acted as a chemoattractant. The upper chamber was then secured on top and 30 µl of serum-free medium were added to avoid the membrane drying out. The plate was loaded on the machine cradle and left for 1 h to allow membrane equilibration. After 1 h, the baseline measurement was taken. Hundred microlitre of serum-free medium (again phenol-red or phenol-red-free DMEM with the relevant concentration of Estradiol, Fulvestrant and Tamoxifen) containing 40,000 cells were subsequently added to each well, and left for 30 min at room temperature to allow for the attachment of the cells. The impedance value for each well was recorded every 15 min for 20 h for migration experiments and 48 h for invasion and expressed as Cell Index value. For invasion, the CIM plates were coated with 30 µl of a 1:40 dilution of Matrigel (BD Biosciences) and allowed to set for 4 h at 37 °C before proceeding as per the migration experiment. Four wells per condition were used in each experiment and every experiment was repeated four times.

Chromatin immunoprecipitation (ChIP). Cells were crosslinked with 1% formaldehyde and processed according to Schmidt *et al.*⁵⁵ Briefly, chromatin extract were sonicated using a Diagenode sonicator using 20 cycles (30 s on and 30 s off) at maximum intensity. Purified chromatin was then immuno precipitated using 4 µg of K27ac antibodies (Abcam ab4729) per ChIP. Non-immunoprecipitated chromatin was used as Input control. Chromatin was then decrosslinked and sonication efficiencies tested on a 1.5% agarose gel. Before the construction of ChIP-seq libraries (NEB, see Supplementary Methods), we determined the enrichment using positive and negative controls. We then used 10 ng of IPed chromatin and 10 ng of Input for library preparation.

Bioinformatic analyses. Detailed protocols on Peak Calling, enhancers-promoters Peak-pairing, Clustering analysis, IPA analysis and RNA-seq differential expression analysis can be found in the Supplementary Methods.

RT-qPCR and ChIP-qPCR. RT-qPCR and ChIP-qPCR were performed and analysed as follows. Briefly, reactions were carried on in 10 µl volume containing 5 µl of SYBR mix (ABI 4472918), 0.5 µl of primers (2.5 and 5 µM final concentration, respectively), 2.5 µl of complementary DNA or DNA and 2 µl of water. We used a three-step cycle programme with melting analysis. The cycles were as follow: 10 s at 95 °C, 30 s at 60 °C and 30 s at 72 °C, repeated 40 times⁷.

SRB and IC₅₀ assay. SRB and IC₅₀ assays were performed as follow: briefly, the sulphorhodamine B (SRB) assay was used to monitor the effects of statin or Terbinafine treatment cell on cell proliferation in monolayer cultures. Cells were seeded in flat-bottomed 96-well plates (3 × 10³ cells per well). The cells were allowed to adhere overnight. One plate was assayed at this time point (day 0) and further plates were assayed at 2-day intervals. The cells were fixed by adding 100 µl per well of ice-cold 40% (vol/vol) TCA to each well for 60 min. The plates were

washed five times in running tap water and stained with 100 µl per well of the SRB reagent (0.4% wt/vol SRB in 1% (vol/vol) acetic acid) for 30 min. The plates were washed five times in 1% (vol/vol) acetic acid and allowed to dry overnight. SRB was solubilized with 100 µl per well 10 mM Tris-base, shaken for 30 min and the optical density was measured at 492 nm.

Primers. The sequences of primers (RT-qPCR, ChIP-qPCR) are listed in Supplementary Data 7.

References

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG) *et al.* 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 (2011).
2. Palmieri, C., Patten, D. K., Januszewski, A., Zucchini, G. & Howell, S. J. Breast cancer: current and future endocrine therapies. *Mol. Cell. Endocrinol.* **382**, 695–723 (2014).
3. Musgrove, E. A. & Sutherland, R. L. Biological determinants of endocrine resistance in breast cancer. *Nat. Rev. Cancer*, 1–13, doi:10.1038/nrc2713 (2009).
4. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
5. Consortium, T. E. P. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **488**, 57–74 (2012).
6. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012).
7. Magnani, L. *et al.* Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proc. Natl Acad. Sci. USA* **110**, E1490–E1499 (2013).
8. Sharma, S. V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80 (2010).
9. Suvà, M. L. *et al.* Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 1–15 (2014).
10. Ernst, J. & Kellis, M. Nat Biotechnol 2010 Ernst. *Nat. Biotechnol.* **28**, 817–838 (2010).
11. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
12. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
13. Malik, A. N. *et al.* Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat. Neurosci.* **17**, 1330–1339 (2014).
14. Stasevich, T. J. *et al.* Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **516**, 272–275 (2014).
15. Shaw, L. E., Sadler, A. J., Pugazhendhi, D. & Darbre, P. D. Changes in oestrogen receptor-α and -β during progression to acquired resistance to tamoxifen and fulvestrant (Faslodex, ICI 182,780) in MCF7 human breast cancer cells. *J. Steroid Biochem. Mol. Biol.* **99**, 19–32 (2006).
16. Martin, L. A. *et al.* An in vitro model showing adaptation to long-term oestrogen deprivation highlights the clinical potential for targeting kinase pathways in combination with aromatase inhibition. *Steroids* **76**, 772–776 (2011).
17. Magnani, L., Ballantyne, E. B., Zhang, X. & Lupien, M. PBX1 genomic pioneer function drives ERα signaling underlying progression in breast cancer. *PLoS Genet.* **7**, e1002368 (2011).
18. Magnani, L. *et al.* The pioneer factor PBX1 is a novel driver of metastatic progression in ERα-positive breast cancer. *Oncotarget* **6**, 21878–21891 (2015).
19. Magnani, L., Eeckhoutte, J. & Lupien, M. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet.* **27**, 465–474 (2011).
20. Magnani, L. & Lupien, M. Chromatin and epigenetic determinants of estrogen receptor alpha (ESR1) signaling. *Mol. Cell. Endocrinol.* **382**, 633–641 (2014).
21. Jozwik, K. M. & Carroll, J. S. Pioneer factors in hormone-dependent cancers. *Nat. Rev. Cancer* **12**, 381–385 (2012).
22. Tyson, J. J. *et al.* Dynamic modelling of oestrogen signalling and cell fate in breast cancer cells. *Nat. Rev. Cancer* **11**, 523–532 (2011).
23. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
24. Miller, W. R. & Larionov, A. Changes in expression of oestrogen regulated and proliferation genes with neoadjuvant treatment highlight heterogeneity of clinical resistance to the aromatase inhibitor, letrozole. *Breast Cancer Res.* **12**, R52 (2010).
25. Bertucci, F. Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers. *Cancer Res.* **66**, 4636–4644 (2006).
26. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
27. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

28. Downen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
29. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **488**, 83–90 (2012).
30. Jeon, T.-I. & Osborne, T. F. SREBPs: metabolic integrators in physiology and metabolism. *Trends Endocrinol Metab* **23**, 65–72 (2012).
31. Yokoyama, C. *et al.* SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low density lipoprotein receptor gene. *Cell* **75**, 187–197 (1993).
32. Briggs, M. R., Yokoyama, C., Wang, X., Brown, M. S. & Goldstein, J. L. Nuclear protein that binds sterol regulatory element of low density lipoprotein receptor promoter. I. Identification of the protein and delineation of its target nucleotide sequence. *J. Biol. Chem.* **268**, 14490–14496 (1993).
33. Kamisuki, S. *et al.* A small molecule that blocks fat synthesis by inhibiting the activation of SREBP. *Chem. Biol.* **16**, 882–892 (2009).
34. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
35. Sanchez, C. G. *et al.* Preclinical modeling of combined phosphatidylinositol-3-kinase inhibition with endocrine therapy for estrogen receptor-positive breast cancer. *Breast Cancer Res.* **13**, R21 (2011).
36. Umetani, M. *et al.* 27-Hydroxycholesterol is an endogenous SERM that inhibits the cardiovascular effects of estrogen. *Nat. Med.* **13**, 1185–1192 (2007).
37. Phimister, E. G., Warner, M. & Gustafsson, J.-Å. On estrogen, cholesterol metabolism, and breast cancer. *N. Engl. J. Med.* **370**, 572–573 (2014).
38. Carroll, J. S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33–43 (2005).
39. Wu, Q. *et al.* 27-Hydroxycholesterol promotes cell-autonomous, ER-positive breast cancer growth. *Cell Rep.* **5**, 637–645 (2013).
40. Lupien, M. *et al.* Growth factor stimulation induces a distinct ER cistrome underlying breast cancer endocrine resistance. *Genes Dev.* **24**, 2219–2227 (2010).
41. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
42. Freed-Pastor, W. A. *et al.* Mutant p53 disrupts mammary tissue architecture via the mevalonate pathway. *Cell* **148**, 244–258 (2012).
43. Miller, W. R. *et al.* Gene expression profiles differentiating between breast cancers clinically responsive or resistant to Letrozole. *J. Clin. Oncol.* **27**, 1382–1387 (2009).
44. Parris, T. Z. *et al.* Frequent MYC coamplification and DNA hypomethylation of multiple genes on 8q in 8p11-p12-amplified breast carcinomas. *Oncogenesis* **3**, e95–e98 (2014).
45. Yamaguchi, N. *et al.* FoxA1 as a lineage-specific oncogene in luminal type breast cancer. *Biochem. Biophys. Res. Commun.* **365**, 711–717 (2008).
46. Lin, M.-L. *et al.* Expression profiling of nuclear receptors in breast cancer identifies TLX as a mediator of growth and invasion in triple-negative breast cancer. *Oncotarget* **6**, 21685–21703 (2015).
47. Ahern, T. P., Lash, T. L., Damkier, P., Christiansen, P. M. & Cronin-Fenton, D. P. Statins and breast cancer prognosis: evidence and opportunities. *Lancet Oncol.* **15**, e461–e468 (2014).
48. Ahern, T. P. *et al.* Statin prescriptions and breast cancer recurrence risk: a Danish Nationwide Prospective Cohort Study. *J. Natl Cancer Inst.* **103**, 1461–1468 (2011).
49. Pagani, O. *et al.* Adjuvant exemestane with ovarian suppression in premenopausal breast cancer. *N. Engl. J. Med.* **371**, 107–118 (2014).
50. Bao, T., Blackford, A. L. & Stearns, V. Effect of simvastatin on the pharmacokinetics of anastrozole. *Breast Cancer Res. Treat.* **131**, 709–711 (2011).
51. Kobayashi, Y. *et al.* Mevalonate pathway antagonist inhibits proliferation of serous tubal intraepithelial carcinoma and ovarian carcinoma in mouse models. *Clin. Cancer Res.* doi: 10.1158/1078-0432.CCR-14-3368 (2015).
52. Clendening, J. W. *et al.* Exploiting the mevalonate pathway to distinguish statin-sensitive multiple myeloma. *Blood* **115**, 4787–4797 (2010).
53. Lombardo, Y. *et al.* Nicastrin regulates breast cancer stem cell properties and tumor growth in vitro and in vivo. *Proc. Natl Acad. Sci. USA* **109**, 16558–16563 (2012).
54. Warita, K. *et al.* Statin-induced mevalonate pathway inhibition attenuates the growth of mesenchymal-like cancer cells that lack functional E-cadherin mediated cell cohesion. *Sci. Rep.* **4**, 7593 (2014).
55. Schmidt, D. *et al.* ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240–248 (2009).

Acknowledgements

This work was supported by an Imperial Junior Research Fellowship to L.M. and a Wellcome Trust grant (103034/Z/13/Z) to D.K.P. and the OTKA K108655 grant to B.G. We wish to thank Mrs Lorna Watson for her help with the manuscript. The Imperial College Healthcare NHS Trust Tissue Bank provided tissue samples. Other investigators may have received samples from these same tissues. The research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre, the Experimental Cancer Medicine Centre and the Cancer Research UK Centre at Imperial College Healthcare NHS Trust, Imperial College, London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Author contributions

L.M. designed the study. L.M., I.B., B.G., S.M., L.C. and R.C.C. carried out the bioinformatic analysis and interpreted the data. L.M. wrote the manuscript. V.T.M.N., M.F., Y.L., D.K.P., G.F., P.D., A.M. and J.H.S. carried out laboratory experiments. R.C.C., L.W. and N.P. recruited patients, collected and prepared clinical specimens.

Additional information

Accession codes: RNA-seq can be accessed at ENA (<http://www.ebi.ac.uk/ena>) with the code PRJEB11448. ChIP-seq data can be found in the GEO archive under number GSE60517.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Nguyen, V. T. M. *et al.* Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion. *Nat. Commun.* 6:10044 doi: 10.1038/ncomms10044 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>