

Proteomic characterisation of Estrogen Receptor (ER α) interactome in breast cancer



Evangelia Papachristou
Cancer Research UK Cambridge Institute
University of Cambridge
Clare College

This thesis is submitted for the degree of Doctor of Philosophy
December 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. In accordance with the guidelines of the Degree Committee for the Faculties of Clinical Medicine and Veterinary Medicine, this thesis does not exceed 60,000 words.

Evangelia Papachristou

December 2020

Summary

Proteomic characterisation of Estrogen Receptor (ER α) interactome in breast cancer

Evangelia Papachristou

Breast cancer is the most common cancer in women in the western world. It has been shown that estrogen receptor alpha (ER α) contributes to tumour formation in 75% of breast cancer cases and various drugs, such as 4-hydroxytamoxifen (OHT), have been used to inhibit estrogen-induced signals. Studies have highlighted the importance of interactions between ER α and various factors in regulating chromatin binding and gene transcription, underlying estrogen signalling in breast cancer. During the last decade, quantitative mass spectrometry (MS)-based proteomics has become a powerful tool for the study of protein interactions. However, technical challenges have restricted most applications to qualitative observations and the dynamics of protein interactomes have until recently been largely unexplored. Here, we describe the development and application of a quantitative multiplexed method, termed qPLEX-RIME, which integrates RIME with isobaric labelling for the study of protein interactome dynamics in a quantitative fashion with increased sensitivity. Using the qPLEX-RIME method we successfully identified endogenous ER α -associated proteins in human Patient Derived Xenograft (PDX) and primary tumours and we delineated the mechanistic temporal changes of the ER α interactome in MCF7 cells upon OHT treatment. Our pipeline was also successfully applied for the study of other factors, such as CBP, NCOA3 and RNA Polymerase II. The integration of the different interactome datasets, lead to the discovery of a transcription factor, termed ZNF207, that belongs to the zinc finger protein family. In MCF7 cells, ZNF207 interacts with ER α , known ER α interactors and components of the transcription machinery. Knockdown of ZNF207, affected cellular proliferation of various cancer and non-cancer cell models, suggesting a functional role in the majority of cellular contexts. Interactome and chromatin binding assays revealed that ZNF207 is important for the assembly and chromatin binding of the Mediator complex and general transcription factors, whilst RNA-sequencing and proteomic analysis showed a decrease in expression of genes linked to cell cycle followed ZNF207 knockdown. The effect was likely mediated by changes in the chromatin binding of the subunit MED1 and the transcription factor TAF3 to promoters of cell cycle genes. Thus, ZNF207 plays an important role in our model system and was studied in depth in this thesis by integrating various functional assays. Our results demonstrate that qPLEX-RIME offers a powerful tool for the in-depth characterisation of protein interactome dynamics, which is also

applicable to clinical samples. Our findings constitute the basis for a deeper mechanistic insight into the dynamics of ER α interaction network and its role in breast cancer along with the discovery of a transcription factor that may play a vital role in the initiation of transcription.

Acknowledgements

I would like to thank all the members of the Carroll group and the members of the CRUK Proteomics facility for the very useful discussions on my project, the guidance and training in new methods and their great support that allowed me to become a part time PhD student in parallel with my role in the proteomics facility.

I would especially like to express my gratitude to my supervisor, Professor Jason Carroll, for his continuous encouragement, guidance and support during my PhD. I am grateful for giving me the opportunity to work in such an inspiring field of research and learn new techniques.

I also would like to thank my second supervisor, Dr Clive D'Santos, for giving me the opportunity to work in the Proteomics facility and combine this role with a part time PhD. I am very grateful for his continuous support and encouragement.

A special thanks to Dr Rasmus Siersbaek for the invaluable scientific discussions, experimental advice, patience, support, encouragement and guidance, without which this work would not have been possible.

I would like to thank the CRUK-CI core facilities, in particular the Genomics Core, the Research Instrumentation & Cell Services and the Flow Cytometry for their valuable help in this work. Special thanks to the Bioinformatics core facility, especially Dr Matt Eldridge, Dr Kamal Kishore, Dr Ashley Sawle and Dr Chandra Sekhar Reddy Chilamakuri for the analysis of my data and for the development of customised software for the statistical analysis of the proteomics data. I further acknowledge Cancer Research UK for funding this study, and thank the Scientific Administration team and especially Ann Kaminski for all their help along the way.

I wish to thank my bioinformatic colleague Dr Igor Chernukhin for all the useful discussions and his help on the data analysis. Thank you also to Dr Stefan Schoenfelder for the training in a new technique and his input and help on the data interpretation.

Finally, I would like to thank my husband Theo for the invaluable scientific discussions and for being a constant source of love, support and encouragement.

Abbreviations

3C	Chromosome conformation capture
3D	Three-Dimensional
ACN	Acetonitrile
AD1	Activation domain 1
AD2	Activation domain 1
adj.p-value	adjusted P value
AF-1	Activation function 1
AF-2	Activation function 2
AI	Aromatase Inhibitor
AMBIC	Ammonium bicarbonate
AP	Affinity purification
AR	Androgen Receptor
AREG	Amphiregulin
BAZ2A	Bromodomain Adjacent to Zinc Finger Domain 2A
BioGRID	Biological General Repository for Interaction Datasets
BioID	Proximity-dependent Biotin Identification
bp	Base pairs
BRCA1/2	Breast cancer gene 1/2
BRG1	Brahma-Related Gene 1
BSA	Bovine Serum Albumin
CARM1	Coactivator-associated arginine methyltransferase 1
CBP	CREB-binding protein
CBX8	Chromobox homolog 8
CDK8/9	Cyclin Dependent Kinase 8/9
CHiCAGO	Capture HiC Analysis of Genomic Organisation
Chicdiff	Differential capture Hi-C
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation followed by high-throughput sequencing
CID	Collision-induced dissociation
CO ₂	Carbon dioxide
CORUM	Comprehensive resource of mammalian protein complexes
CRISPR	Clustered regularly interspaced short palindromic repeats

CRUK-CI	Cancer Research UK Cambridge Institute
CTD	Carboxy-terminal domain
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DBD	DNA binding domain
DDA	Data-Dependent Acquisition
DIA	Data-Independent Acquisition
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulphoxide
DNA	Deoxyribonucleic acid
DSG	Discuccinimidyl glutarate
E2	17 β -oestradiol
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
ER	Estrogen receptor
ERE	Estrogen response element
FA	Formaldehyde
FBS	Foetal bovine serum
FDR	False discovery rate
Fe-NTA	Ferric Nitrilotriacetate
FOXA1	Forkhead box A1
FO XK1	Forkhead box protein K1
GATA3	GATA binding protein 3
GFP	Green Fluorescent Protein
GREB1	Growth regulation by estrogen in breast cancer 1
GREB1L	Growth regulation by estrogen in breast cancer 1-like protein
GSEA	Gene Set Enrichment Analysis
HA	Hemagglutinin
HCD	Higher-energy collisional dissociation
HDAC	Histone deacetylase
HER2	Human epidermal growth factor receptor 2
HEY2	Hairy/enhancer-of-split related with YRPW motif protein 2
Hi-C	High-throughput chromosome conformation capture
HiCUP	Hi-C User Pipeline

HP1 γ	Heterochromatin protein 1 γ
IgG	Immunoglobulin G
IMAC	Immobilised Metal Affinity Chromatography
iTRAQ	isobaric tags for absolute and relative quantification
K238	Lysine 238
kb	Kilobase
kDa	Kilodaltons
LBD	Ligand Binding Domain
LC	Liquid Chromatography
LFQ	Label-Free Quantification
MACS	Model-based Analysis for ChIP sequencing
MED	Mediator
MMTS	methyl methanethiosulfonate
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MTA1/2	Metastasis Associated 1/2
MTG16	Myeloid Translocation Gene 16
NaCl	Sodium chloride
NaHCO ₃	Sodium bicarbonate
NCOA	Nuclear receptor co-activator
NCOR	Nuclear receptor co-repressor
NF- κ B	nuclear factor kappa-light-chain-enhancer of activated B cells
NHS	N-Hydroxysuccinimide
NIPBL	Nipped-B-like
NRIP1	Nuclear receptor-interacting protein 1
NuRD	Nucleosome Remodelling and Deacetylase
°C	Degrees Celsius
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDX	Patient-derived xenograft
PDZK1	PDZ Domain Containing 1
PGR	Progesterone receptor
PKIB	CAMP-Dependent Protein Kinase Inhibitor Beta

PLA	Proximity Ligation Assay
PRC1	Polycomb Repressive Complex 1
PRMT	Protein arginine methyltransferase
PTMs	Post-Translational Modifications
Q-Exactive	Quadrupole-Exactive
Q-Exactive HF	Quadrupole-Exactive High Field
qPCR	Quantitative polymerase chain reaction
qPLEX-RIME	Quantitative multiplexed rapid immunoprecipitation mass spectrometry of endogenous proteins
RARA	Retinoic acid receptor alpha
RIME	Rapid immunoprecipitation mass spectrometry of endogenous protein
RIPA	Radioimmunoprecipitation assay
RNAi	RNA interference
RPMI	Roswell Park Memorial Institute
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate
SEC	Super Elongation Complex
SequestHT	Sequest High-throughput
SERD	Selective oestrogen receptor degrader
SERM	Selective oestrogen receptor modulator
SIAH2	Siah E3 Ubiquitin Protein Ligase 2
SILAC	Stable isotope labelling by amino acids in cell culture
siRNA	Small interfering ribonucleic acid
SMC1A	Structural Maintenance of Chromosomes 1A
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SWATH-MS	Sequential window acquisition of all theoretical mass spectra
SWI/SNF	SWItch/Sucrose Non-Fermentable)
TAFs	TATA box binding protein associated factors
TAP	tandem affinity purification
TBS	Tris-Buffered Saline
TCEP	Tris-2-carboxymethyl phosphine
TE	Tris-EDTA (Ethylene diamine tetraacetic acid)
TEAB	Triethylammonium bicarbonate

TFF1/3	Trefoil factor 1 /3
TFIIA/B/H/D	Transcription factor IIA/B/H/D
Th	Thompon
TiO ₂	Titanium dioxide
TMT	Tandem mass tag
TNF	Tumor necrosis factor
TRIM24	Tripartite Motif Containing 24
Tris-HCl	Tris hydrochloride
TSS	Transcription start site
XBP1	X-box binding protein 1
XL-MS	Cross-Linking Mass Spectrometry
Y2H	yeast two-hybrid
ZNF	Zinc finger
GATAD2a/2b	GATA zinc finger domain containing 2A/2B
MBD2/3	Methyl-CpG Binding Domain Protein 2/3
RBAP46/48	Retinoblastoma-associated proteins 46/48
CHD3/4	Chromodomain-helicase-DNA-binding protein 3/4
ATP	adenosine triphosphate
AP-2 γ	Activator Protein 2 γ
JMJD6	Jumonji domain-containing protein 6
JmjC	Jumonji C

Contents

Chapter 1	1
1. Introduction.....	1
1.1 Breast Cancer	1
1.1.1 Classification of Breast Cancer	2
1.1.2 Experimental models in breast cancer research	3
1.1.3 The Proteomic Landscape of Breast Cancer	4
1.2 Estrogen receptors.....	6
1.2.1 The ERα receptor and its role in breast cancer	7
1.2.2 Mechanisms of ERα activation	8
1.3 ER α complex.....	10
1.3.1 ERα Coactivators and Corepressors	11
1.3.2 Pioneer factors	13
1.3.3 ERα and the general transcription apparatus	14
1.4 Treatment strategies for ER α positive breast cancer.....	17
1.4.1 Mechanisms of resistance to breast cancer therapies	18
1.5 Three-dimensional genome	20
1.6 Mass spectrometry-based proteomics	22
1.6.1 Approaches for studying protein-protein interactions	23
1.6.2 Mass spectrometry-based detection of Post-Translational Modifications	27
1.6.3 Relative quantification proteomics approaches	29
1.6.3.1 Metabolite labelling	30
1.6.3.2 Label-free quantification	31
1.6.3.3 Chemical isobaric labelling	32
1.7 Thesis objectives.....	34
Chapter 2.....	37
2. Materials and Methods.....	37
2.1 Cell lines and cell treatments	37
2.1.1 Whole cell lysate preparation and western blot analysis	37
2.1.2 Small interfering RNA (siRNA) assay	38
2.1.3 Cell proliferation assays	38
2.2 RNA-sequencing analysis	39
2.3 Proximity ligation assay.....	39
2.4 Immunofluorescence	40

2.5 PDX propagation and tissue collection	41
2.5.1 Sample preparation of clinical tumour material	41
2.6 Immunohistochemistry	42
2.7 Flow cytometry	42
2.8 ChIP-Seq and RIME assays	42
2.8.1 ChIP-seq data analysis	44
2.8.2 ChIP-qPCR	46
2.9 Protein digestion and TMT labelling	47
2.9.1 Fe-NTA Phosphopeptide enrichment	48
2.9.2 LC-MS analysis	48
2.9.3 Data processing of mass spectrometry raw data	49
2.9.4 Analysis of proteomics data	50
2.10 Promoter Capture Hi-C sample preparation	50
2.10.1 PCHi-C data processing and interaction calling	52
Chapter 3	54
3. Results	54
3.1 Development of a quantitative tool	54
3.1.1 The qPLEX-RIME method	54
3.1.2 Characterisation of the ERα interactome in MCF7 cells	56
3.1.3 Comparison between qPLEX-RIME and non-quantitative RIME	57
3.1.4 Optimisation of crosslinking, starting material and antibody concentration	60
3.1.5 Evaluation of ChIP-seq and RIME grade antibodies for ERα	61
3.2 Different applications of the qPLEX-RIME method	66
3.2.1 Characterisation of different protein interactomes	66
3.2.2 Study of ERα complex dynamics upon OHT treatment	68
3.2.2.1 Dissociation and recruitment of co-factors upon OHT treatment	70
3.2.3 Identification of net changes in the ERα complex	74
3.2.4 Application of qPLEX-RIME in clinical tumour material	76
3.2.4.1 Characterisation of ERα interactome in vivo	76
3.2.4.2 ERα interactome in clinical material with variable expression profile of PR	80
3.3 Characterisation of a novel ER α -associated transcription factor	82
3.3.1 Identification of ZNF207 protein by qPLEX-RIME	82
3.3.2 Functional importance of ZNF207 for cell proliferation	85
3.3.3 Identification of the ZNF207 target gene program	89

3.3.4 Effect of ZNF207 knockdown on cell cycle progression	94
3.3.5 ZNF207 antibody testing using RIME and ChIP-seq	96
3.3.5.1 Characterisation of ZNF207 interactome	100
3.3.6 ZNF207 knockdown does not affect the assembly of the ER α complex	101
3.3.7 ZNF207 knockdown affects the assembly of the Mediator complex	103
3.3.8 ZNF207 knockdown affects the assembly of the Pol II complex	106
3.3.9 siZNF207 does not affect chromatin loops	108
3.3.10 Genomic characterisation of ZNF207 function	113
Chapter 4	123
4. Discussion	123
4.1 Supplementary Data	137
Supplementary Data 1: <i>Regulated proteins identified in ERα-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.</i>	137
Supplementary Data 2: <i>Regulated proteins identified in MED1-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.</i>	138
Supplementary Data 3: <i>Regulated proteins identified in MED12-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.</i>	139
Supplementary Data 4: <i>Regulated proteins identified in MED4-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.</i>	140
Supplementary Data 5: <i>Regulated proteins identified in MED1-qPLEX-RIME following ZNF207 knockdown in HEK293 cells.</i>	141
Supplementary Data 6: <i>Regulated proteins identified in MED14-qPLEX-RIME following ZNF207 knockdown in HEK293 cells.</i>	142
Supplementary Data 7: <i>Regulated proteins identified in Pol II-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.</i>	143
Chapter 5	144
5. References	144
Papers published during PhD	162
Data availability	162

List of Figures

Figure 1. ER α domains and mutations.	8
Figure 2. Schematic illustration of ER α signalling mechanisms.	10
Figure 3. Schematic illustration of the Mediator complex.	15
Figure 4. Mechanism of tamoxifen action.	18
Figure 5. RIME workflow.	27
Figure 6. Chemical labelling.	34
Figure 7. The qPLEX-RIME workflow.	55
Figure 8. Application of qPLEX-RIME for the identification of ER α -associated proteins.	57
Figure 9. Validation of novel ER α -associated cofactors using Proximity Ligation Assay (PLA).	59
Figure 10. Decrease in ER α protein levels upon Fulvestrant treatment.	59
Figure 11. Comparison of crosslinking approaches and titration of starting material.	61
Figure 12. A comparison between different ER α antibodies by ChIP-qPCR and ChIP-seq.	63
Figure 13. Comparison of RIME data between Santa Cruz, Millipore and Abcam antibodies.	65
Figure 14. A comparison between different ER α antibodies by ChIP-qPCR.	65
Figure 15. Application of qPLEX-RIME in CBP, NCOA3 and Pol II.	67
Figure 16. Comparison of the four interactomes (ER α , CBP, NCOA3, Pol II).	68
Figure 17. RNA-seq analysis and profile of ER α total protein levels.	69
Figure 18. Correction of quantitative results for the dependency on ER α target protein by linear regression.	70
Figure 19. Temporal profiling of the ER α interactome following treatment of MCF7 cells with OHT.	72
Figure 20. Validation of temporal changes in the ER α complex upon OHT treatment.	73
Figure 21. Comparison between qPLEX-RIME and RIME-SILAC data.	73
Figure 22. Comparison of qPLEX-RIME data with whole proteome and RNA-seq data.	75
Figure 23. Differentially regulated proteins and correlation with gene expression.	76
Figure 24. ER α qPLEX-RIME application in PDX tissues.	77
Figure 25. Characterisation of ER α interactors from in vivo samples.	79
Figure 26. Most frequently enriched ER α interactors.	80
Figure 27. Analysis of tumours with variable expression profile of PR.	82
Figure 28. Overview of ZNF207 discovery.	83
Figure 29. Association of ZNF207 with nuclear receptors.	84
Figure 30. Association of ZNF207 with co-activators and Pol II in MCF7 cells.	85
Figure 31. CRISPR screening in MCF7 cells and siRNA transfection conditions.	86
Figure 32. Proliferation assays.	88
Figure 33. Proliferation assays and analysis of DepMap data.	88

Figure 34. Human Protein Atlas RNA-seq data.....	89
Figure 35. RNA-seq analysis following ZNF207 knockdown.....	91
Figure 36. Sashimi plots.	91
Figure 37. Comparison of two RNA-seq datasets.....	92
Figure 38. Whole proteome analysis.....	93
Figure 39. Correlation between protein and gene expression data.	94
Figure 40. FACS and microscopy experiments.	96
Figure 41. Comparison of 6 antibodies specific for ZNF207 with RIME.	98
Figure 42. ZNF207 antibody testing using ChIP-seq in MCF7 cells.....	99
Figure 43. ZNF207 interactome and RIME for Mediator subunits.....	101
Figure 44. ER α qPLEX-RIME after ZNF207 knockdown.	103
Figure 45. Overview of qPLEX-RIME experiments after ZNF207 knockdown.	107
Figure 46. Comparison of qPLEX-RIME changes with full proteome.....	108
Figure 47. PCHi-C pre-sequencing quality controls.	110
Figure 48. Promoter capture Hi-C data.....	113
Figure 49. ChIP-seq data.....	115
Figure 50. Overlap between ZNF207, MED1, TAF3 and Pol II binding sites.	116
Figure 51. Overlap of ZNF207, ER α and FOXA1 binding sites and de novo motif analysis.	116
Figure 52. MED1 ChIP-qPCR.	118
Figure 53. Functional enrichment analysis.	119
Figure 54. Correlation between ChIP-seq and RNA-seq data.	120
Figure 55. Correlation between ChIP-seq, RNA-seq and PCHi-C.	121
Figure 56. Correlation between ChIP-seq experiments with or without spike-in control (<i>Drosophila melanogaster</i> chromatin).....	122
Figure 57. A proposed model of OHT mechanism.	127
Figure 58. A proposed model of ZNF207 functional role.	136

List of tables

Table 1. Antibodies used for RIME and ChIP-seq assays.	44
Table 2. Primer sequences for ChIP-qPCR at ER α binding sites.	47
Table 3. Primer sequences for quality control of human Hi-C libraries.	52

Chapter 1

1. Introduction

1.1 Breast Cancer

Breast cancer is the most common malignancy among women and the second most commonly diagnosed cancer worldwide with more than 2.1 million cases globally in 2018 (Bray et al., 2018). Nearly 55,000 women are diagnosed with breast cancer in the UK every year (<https://www.breastcancercare.org.uk>). Tumours can arise from either the ductal or lobular epithelial tissue of the breast, with ductal malignancies being the most common (Ciriello et al., 2015; Yoder et al., 2007). Metastatic breast cancer has a poor survival within 5-10 years and breast cancer recurrence divides into two groups; distant metastasis including bone, brain, liver, lung and locoregional relapse including breast, chest wall or regional lymph nodes (Yates et al., 2017). Breast cancer can also occur in men, although it is very rare as it accounts for less than 1% of all breast cancer cases worldwide (Korde et al., 2010).

Risk factors that are associated with breast cancer can be genetic or non-genetic (Mavaddat et al., 2010). Hereditary breast cancer is mainly caused by mutations in the tumour suppressor and DNA repair genes, BRCA1 and BRCA2 (Mavaddat et al., 2010). Most of the mutations on these two genes are small deletions or insertions that lead to the translation of a truncated protein. Germline mutations have also been detected in TP53, PTEN and ATM genes (Key et al., 2001). Additionally, genetic variants identified by genome-wide association (GWA) studies, have been linked with an increased risk of breast cancer (Turnbull et al., 2010; Zheng et al., 2009). Other, not genetic factors linked to breast cancer risk include age, diet, childbearing, obesity, alcohol consumption and hormone therapy (Key et al., 2001).

Large-scale genomic analyses in breast cancers patients have revealed the complex and heterogeneous mutational landscape of the disease. Driver mutations have been identified in PIK3CA, TP53, PTEN, CDH1, GATA3, ESR1 and genes involved in SWI/SNF signalling in primary and metastatic tumours (Yates et al., 2017). Interestingly most distant metastases have been shown to acquire additional driver mutations not observed in the primary tumour, suggesting that a more personalised therapy of breast cancer including biopsy and sequencing of metastases can have a significant effect on understanding the evolution of the disease and developing efficient treatments for breast cancer patients (Yates et al., 2017).

1.1.1 Classification of Breast Cancer

Histological type, grade, tumour size, lymph node involvement, expression of Estrogen Receptor alpha (ER α) and Progesterone Receptor (PR) or overexpression of Human Epidermal growth factor receptor 2 (HER2) can define the classification of breast cancer and can provide important information for prognosis, response to treatment and development of metastasis. Breast tumours are classified into five main intrinsic subtypes with distinct clinical outcomes: luminal A, luminal B, HER2-enriched, basal-like and normal-like tumours (Cancer Genome Atlas, 2012; Sorlie et al., 2001; Sorlie et al., 2003).

Luminal subtype A is characterised by a gene expression signature including estrogen receptor 1 (ESR1), GATA-binding protein 3 (GATA3), forkhead box protein A1 (FOXA1), B-cell chronic lymphocytic leukemia (CLL)/lymphoma 2 (BCL-2), X-box binding protein 1 (XBP1) and the myeloblastosis gene (MYB). Luminal B cancers express lower levels of the luminal gene signature, they have higher levels of proliferative genes and worse prognosis compared to subtype A (Sorlie et al., 2001). HER2-enriched subtype characterised by the overexpression of a number of receptor tyrosine kinases, including HER2 as well as genes within the HER2-amplicon. Interestingly, recent studies have observed an additional HER2-enriched subtype that has more luminal features, including high expression of GATA3 and ESR1 genes (Cancer Genome Atlas, 2012). The basal-like tumours are often referred as triple-negative breast cancers (TNBC), as they do not express any of the three receptors, ER α , PR and HER2. These tumours are characterised by the expression of the basal like signature that contains keratins 5, 6 and 17 and by the high expression of genes linked to cell cycle (Cancer Genome Atlas, 2012). Analysis has identified six TNBC subtypes that display unique gene expression and ontologies; Two basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), a mesenchymal stem-like (MSL) and a luminal androgen receptor (LAR) subtype (Lehmann et al., 2011). The normal-like tumours are distinguished by the high expression of genes characteristic of basal epithelial cells and adipose cells and the low expression of genes characteristic of luminal epithelial cells (Sorlie et al., 2001).

Each subtype has a unique somatic mutation spectrum (Cancer Genome Atlas, 2012); luminal A subtype has more frequent mutations in PIK3CA followed by GATA3, TP53 and CDH1, whereas luminal B reveals a high diversity of mutations, with TP53 and PIK3CA being the most frequent (Cancer Genome Atlas, 2012). Basal-like cancers display frequent mutations in TP53 gene; it has also been discovered that BRCA1 mutations are strongly associated with a basal tumour phenotype (Sorlie et al., 2003). The HER2-enriched subtypes have HER2

amplification and high frequency of TP53 and PIK3CA mutations, along with GATA3 mutations that were only detected in the HER2-enriched luminal subtype. Notably somatic mutations were found to be more diverse in Luminal A and Luminal B compared to basal-like and HER2-enriched subtypes, however the overall mutation rate was lowest in Luminal A and highest in basal-like and HER2-enriched subtypes (Cancer Genome Atlas, 2012).

More recently, studies have facilitated the further subclassification of breast cancer into eleven integrative Cluster (IntClust) subtypes, which are based on patterns of genomic copy-number alterations and gene expression (Curtis et al., 2012; Pereira et al., 2016; Rueda et al., 2019). The IntClust classification has illuminated differences in recurrence rates (Curtis et al., 2012; Rueda et al., 2019) that were obscured from immunohistochemistry approaches or the PAM50 model, which is based on a set of 50 genes to predict breast cancer outcomes (Parker et al., 2009). Overall, classification of breast cancer and identification of subtype-specific mutations have facilitated the discovery of molecular signatures that are associated with survival, disease relapse, development of rational therapies and treatment outcome (Cancer Genome Atlas, 2012; Curtis et al., 2012; Lam et al., 2014; Rueda et al., 2019; Sorlie et al., 2001).

1.1.2 Experimental models in breast cancer research

Various breast cancer cell lines have been widely used for breast cancer modelling, offering an unlimited and easy obtained source for tumour studies. Examples of the most used cell lines in breast cancer studies are MCF7, T47D, ZR-751 and MDA-MB-231 (Dai et al., 2017). The MCF7 cell line is ER α /PR-positive, belongs to the luminal A molecular subtype and has low metastatic potential (Comsa et al., 2015). The T47D cell line represents the luminal A subtype of breast cancer and is more susceptible to progesterone treatment compared to the MCF7 cell line (Yu et al., 2017). The ZR-751 has luminal A characteristics and MDA-MB-231 is a highly aggressive triple negative cell line (Subik et al., 2010). Although it is easier to work with cell lines, they do not recapitulate inter- and intra-tumor heterogeneity and may have been evolved differently from the primary tumours during the serial passaging (Burdall et al., 2003; Dai et al., 2017). Thus, it is important that any efforts to translate *in vitro* data into a clinical context are made carefully to avoid faulty data interpretations.

The development of various *in vivo* models can give new insights into the mechanisms of breast cancer progression and delineate the delivery of new improved therapies (Holen et al., 2017). Different models that capture various types and stages of the disease have been established and can be utilised to address specific scientific questions. One of the most common model systems

is based on engraftment of human cell lines into immunocompromised animals [cell-derived xenografts (CDX)]. Although CDX models have been used to study breast cancer genetics, they cannot capture the heterogeneity of human breast tumours and usually the cell lines that are used are derived from highly aggressive malignant tumours or pleural effusions, making them less attractive models for the study of the early events of breast cancer (Holen et al., 2017).

Alternative pre-clinical models that capture better features of the primary human tumours are the Patient-derived xenografts (PDXs). The development of these models includes transplantation of primary human cancer cells or tumour pieces into host mice (Holen et al., 2017). Although the implantation of tumour samples occur in immunodeficient mice, it has been reported that the morphological and molecular characteristics of the originating breast cancers are mostly preserved through the serial passaging in the mouse and more importantly PDXs can be used for drug screening to predict treatment responses (Bruna et al., 2016; Gao et al., 2015). Genetically engineered mouse models (GEMMs) have been essential tools for defining the role of genetic alterations in human breast cancer initiation, progression and metastasis. These transgenic models can be created by introducing or inactivating a gene of interest and they develop tumours in a natural immune-proficient microenvironment (Holen et al., 2017). GEMMs capture histopathological and molecular features of their human counterparts and they can progress toward a metastatic disease (Kersten et al., 2017). To conclude, the various *in vitro* and *in vivo* models have contributed to a better understanding of the breast cancer disease and further technological advances on breast cancer modeling may enhance our knowledge regarding the disease progression, treatment responses and development of resistance mechanisms.

1.1.3 The Proteomic Landscape of Breast Cancer

Studies based on genomic and transcriptomic characterisation are the main source for understanding the progression of breast cancer and defining the main intrinsic subtypes (Curtis et al., 2012; Rueda et al., 2019; Sorlie et al., 2001). However, it is less clear how the genomic changes drive the proteome and phosphoproteome variation and the phenotypic characteristics for each subtype. Additionally, studies have shown low correlation between the copy numbers of the gene in the genome and relative changes at the protein level, indicating that many genomic variations are not or only partially translated to changes linked to protein expression

(Liu et al., 2016; Zhang et al., 2014). This observation highlights the importance for proteomic analysis of tumour samples, that may better reflect changes in cellular functions and pathways. The first large scale proteomic characterisation of breast cancer tissues was conducted by Cancer Genome Atlas (TCGA) using reverse phase protein arrays (RPPA) for 171 cancer-related proteins and phospho-proteins on 403 breast tumours, leading to the identification of two potentially novel protein-defined subgroups (Cancer Genome Atlas, 2012). Although, the RPPA platform allows for the simultaneous quantitative analysis of a large number of samples and requires low sample amount, it is limited to antibody availability and specificity and lacks deep proteome coverage as only approximately 200 analytes can be profiled (Tibes et al., 2006). Quantitative proteomic studies based on mass spectrometry have also explored the proteomic landscape of breast cancer, revealing important features of the disease. Lawrence et al., characterised triple-negative breast cancer cell lines and tissues and correlated proteomics data with exome sequence resources and drug response data (Lawrence et al., 2015). Tyanova et al., analysed a panel of 40 breast cancer tumours leading to the quantification of more than 10,000 proteins and to the discovery of functional differences between breast cancer subtypes that were associated with energy metabolism, cell growth, mRNA translation and cell-cell communication (Tyanova et al., 2016a).

The most comprehensive proteogenomic study was performed by Mertins et al., where 105 breast cancer tissues from TCGA were analysed using quantitative mass spectrometry-based proteomic and phosphoproteomic analyses, resulting in the detection of three proteome subtypes of breast cancer; basal-enriched, luminal-enriched, and stromal-enriched clusters (Mertins et al., 2016). Notably, the well-defined HER2-enriched tumours by mRNA studies, were spread across these three proteomic subgroups. Pathway analysis of the first two subtypes demonstrated the enrichment of ESR1-driven gene sets (gene encodes for ER α) in the luminal-enriched subgroup and the enrichment for MYC target genes, cell cycle and DNA repair pathways in the basal-like subgroup. Interestingly, the analysis of the phosphoproteome data indicated four subgroups; subgroup 2 was similar to the stromal-enriched subgroup, the subgroup 3 recapitulated the luminal-enriched proteomic subgroups, the subgroup 4 included most of the basal-enriched subgroups as well as luminal-enriched samples, whereas subgroup 1 was a novel subgroup, mainly characterised by G protein, G-protein-coupled receptor and inositol phosphate metabolism signatures. This study also assessed changes in kinases driven by mutations in PIK3CA, TP53 or by the ERBB2 amplification, highlighting the discovery of kinases that can be therapeutic targets.

Although, the application of proteomics in clinical studies is limited by the relatively high amount of tumour tissues required for mass spectrometry analysis and the medium throughput, improvements in sample preparation protocols, mass spectrometry platforms and computational tools can offer the opportunity to connect the genome to the proteome for better understanding of the molecular landscape of breast cancer, with the ultimate goal to discover biomarkers and actionable targets for early detection and treatment.

1.2 Estrogen receptors

The Estrogen Receptors (ERs) are members of the nuclear receptor superfamily comprising 48 proteins (Hall and McDonnell, 2005). Nuclear receptors are activated in response to specific small ligands and are involved in many physiological processes, including metabolism, development and cell proliferation. Nuclear receptors are also characterised by highly conserved DNA- and ligand-binding domains (Hall and McDonnell, 2005). ERs are ligand-inducible transcription factors that in the absence of hormone, are segregated in a multiprotein inhibitory complex in either the cytoplasm or nuclear (Hall and McDonnell, 2005). There are two genetically distinct ER isoforms, Estrogen Receptor alpha (ER α) and Estrogen Receptor-beta (ER β); both isoforms bind estradiol (E2) with high affinity that leads to conformational changes in the receptors, enabling dissociation from inhibitory heat shock proteins and direct association with coactivator complexes (Deroo and Korach, 2006). The two isoforms display a high degree of sequence similarity in the central DNA- and C-terminal ligand-binding domains (DBD, 97% and LBD, 55%, respectively). Interestingly, the two isoforms are encoded by unique genes, ESR1 and ESR2, exhibit distinct expression patterns and biological functions (Deroo and Korach, 2006). ER α is widely expressed and is the predominant subtype in the breast, uterus and bone, whereas ER β is mainly expressed in ovary, prostate, testis, lung and localised areas of the brain (Couse et al., 1997). ER β has been proposed to oppose ER α activity in both normal development and cancer (Lazennec et al., 2001; Williams et al., 2008) and it has been found to be down-regulated during carcinogenesis of breast cancers (Iwao et al., 2000). Despite the potential role of ER β in breast cancer, of the two isoforms, ER α is the predominant subclass of ERs, that is expressed in the majority of breast cancer cells and has been the focus of many breast cancer studies.

1.2.1 The ER α receptor and its role in breast cancer

ER α is a key transcriptional regulator for development, bone integrity, neuronal tissues in both sexes and female reproductive organs and it is mainly activated by E2, which has an obligate role in the growth and development of female mammary and reproductive physiology, as well as in the skeletal and cardiovascular systems (Ali and Coombes, 2002). During E2 binding, structural changes on the receptor are induced that result in the transcriptional activation of estrogen-regulated genes (Ali and Coombes, 2002). ER α consists of six functional domains classified as A–F (Sommer and Fuqua, 2001): the amino-terminal A/B domain that contains the transcriptional activation function-1 (AF-1) and enables interaction with members of the transcription apparatus, the middle C domain which contains the DNA binding domain (DBD) that consists of two zinc finger motifs and mediate ER α binding to the estrogen response elements (EREs); the D domain, also referred to as hinge region and the domains E and F that contain the ligand-binding domain (LBD) and the second activation function domain, AF-2 that are involved in the receptor dimerisation and association with coregulatory factors. Importantly, both AF-1 and AF-2 domains, are key mediators of the interactions between ER α and co-regulator proteins, that are recruited by the receptor to regulate transcription of specific target genes. AF-1 and AF-2 domains can function either independently or synergistically in certain cells and promoter contexts to recruit coactivator proteins (Ali and Coombes, 2002). Notably, point mutations in ESR1 gene have been discovered within different domains, especially the LBD domain and have been mainly associated with endocrine resistance (Arnesen et al., 2020; Dhiman et al., 2018). The diagram of ER α domains and known mutations in specific residues are illustrated in **Figure 1**.

ER α -positive breast cancer is the most common breast cancer subtype; epidemiological studies and molecular characterisation of breast tumours have assessed the critical role of estrogens and ER α in the initiation, progression and treatment of breast cancers (Lin et al., 2004). ER α is central to the development of more than 75% of breast tumours and has been a valuable predictive and prognostic factor (Lin et al., 2004; Stender et al., 2017). Studies for the ER α chromatin binding and interactome in breast cancer have revealed the importance of spatial organisation in estrogen mediated transcription, development of breast cancer, drug resistance and clinical outcome (Bi et al., 2020; Mohammed et al., 2013; Mohammed et al., 2015; Ross-Innes et al., 2012). Overall, activation of ER α in breast tumour cells leads to alteration in transcriptional activity and expression profile of target genes. An upregulation of genes that are related to proliferation and cell cycle progression and a downregulation of pro-apoptotic

and anti-proliferative genes, has been reported (Frasor et al., 2003; Lin et al., 2004). Treatments that target ER α have been the major strategies for therapeutic intervention in breast cancer with dramatic effects on survival rates. A better insight into the role of ER α in breast cancer is of paramount importance and can lead to improvement of the survival rates of women with hormone-dependent breast cancer.

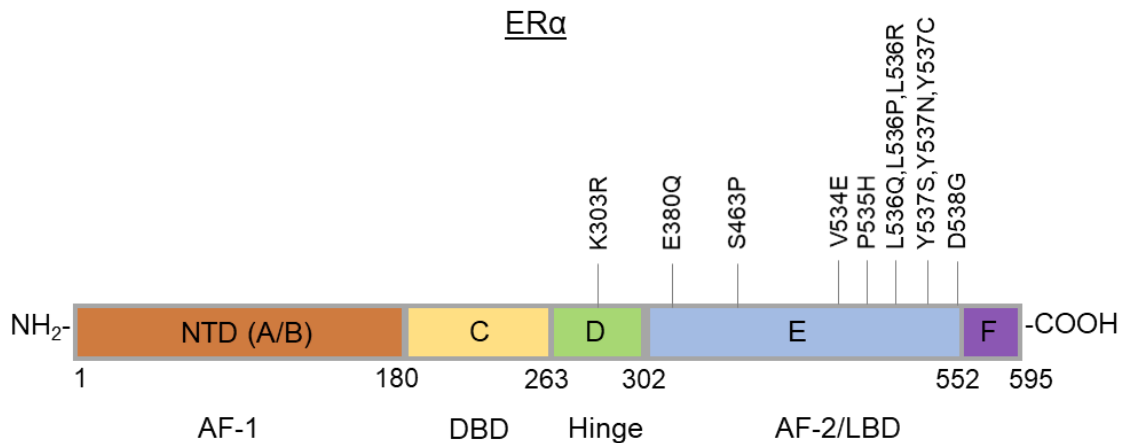


Figure 1. ER α domains and mutations.

Structural and functional domains of ER α : Different domains are highlighted in different colours: NTD (Amino Terminal Domain) in orange; DBD (DNA Binding Domain) in yellow; Hinge region in green; LBD (Ligand-Binding Domain) in blue; F region located towards the C-terminal end in purple. Amino acid sequence position is indicated for each domain. Known mutations in the ESR1 gene are highlighted (Figure adapted from Vineet K. Dhiman et. al.,2018).

1.2.2 Mechanisms of ER α activation

ER α can be activated and regulate gene expression by a number of distinct mechanisms (Bjornstrom and Sjoberg, 2005) (**Figure 2**). The classical model of ER α activation involves estrogen binding to the receptor that induces conformational changes. Upon binding to estrogen, inhibitory proteins are released (heat shock proteins) and the receptor dimerises and translocates to the nucleus where it binds to conserved estrogen response elements (EREs), small palindromic DNA motifs that are commonly identified at the centre of the ER α binding sites (Bjornstrom and Sjoberg, 2005). These sites are mostly located distal to promoters of target genes, as studies have shown that ER α binds mainly to enhancers and only 4% of ER α binding sites mapped to 1kb promoter-proximal regions (Carroll et al., 2006). This binding leads to recruitment of co-factors and chromatin modifying enzymes, that influence positively

or negatively the transcription of target genes (Bjornstrom and Sjoberg, 2005; Hall and McDonnell, 2005).

ER α can also regulate the expression of particular genes without direct binding on the DNA, but through protein-protein interactions with other DNA-binding transcription factors (ERE-independent genomic actions), forming complexes in alternative sites on chromatin (Bjornstrom and Sjoberg, 2005; Welboren et al., 2009). ER α mainly associates with activating protein 1 (AP-1) and specificity protein 1 (Sp1) transcription factor complexes and their respective binding sites (Welboren et al., 2009). The Sp1 transcription factor plays an important role in proliferation, differentiation, survival and angiogenesis and binds with high affinity to GC-rich motifs, which are present in many estradiol responsive promoters (Welboren et al., 2009). The two major families of AP-1 factors are the Fos and Jun transcription factors; the AP1 complex binds to promoters of genes associated with development, growth and differentiation (Welboren et al., 2009). A third mechanism involves a ligand independent activation of ER α through phosphorylation of various serine and tyrosine residues in the functional domains AF-1 and AF-2 (Stender et al., 2017). Activation of protein-kinase cascades by growth factors leads to phosphorylation and activation of nuclear ER α at EREs. Lastly, ER α exerts rapid non-genomic actions that are possibly mediated by membrane associated ER α , which activates protein-kinase cascades that lead to eNOS activation (Bjornstrom and Sjoberg, 2005).

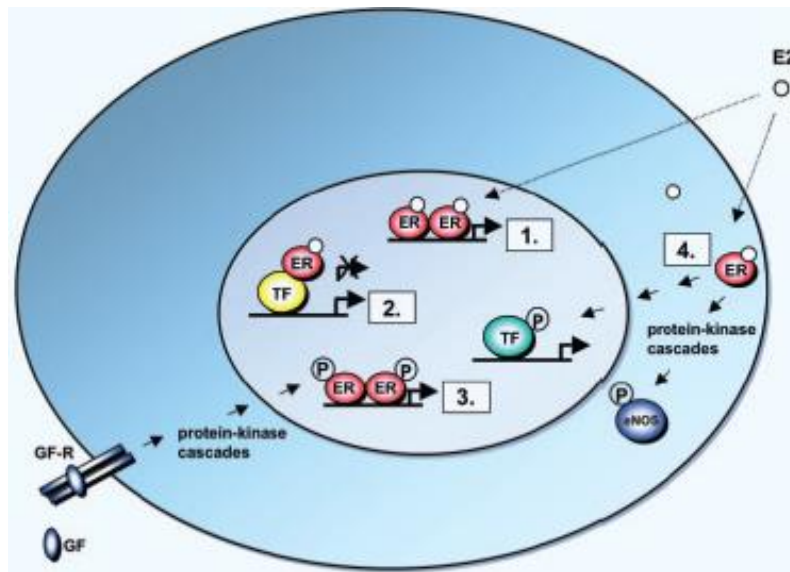


Figure 2. Schematic illustration of ER α signalling mechanisms.

(1) Classical model of ER α activation that involves direct binding to EREs. (2) ERE-independent genomic actions that involve protein-protein interactions. (3) Ligand-independent genomic actions through phosphorylation. (4) Non-genomic actions where membrane associated ER α activates protein-kinase cascades (source: Bjornstrom, L. & Sjoberg, M, 2005).

1.3 ER α complex

ER α coregulators and the various combinations of coregulator recruitment (coactivators or corepressors) to ER α , play a critical role in chromatin remodelling and in the resulting gene expression patterns (Shang et al., 2000). It has become clearer that the recruitment of coregulatory proteins is required for ER α -mediated transcriptional and biological activities. These diverse ER α complexes enable ER α to respond in a proper manner (a) to hormones or pharmacological ligands, (b) translate extra- and intra-cellular signals, (c) regulate chromatin accessibility and (d) transmit signals to the general transcription apparatus at target gene promoters (Hall and McDonnell, 2005).

A wide range of coactivators, corepressors and chromatin remodelling enzymes are recruited by ER α to target promoter and enhancer regions (Kong et al., 2011). Specifically, about 700 proteins have been catalogued in the BioGRID database (<https://thebiogrid.org/>) as ER α interactors. Chromatin immunoprecipitation experiments have indicated that ER α along with coactivators associate with estrogen responsive promoters, following estrogen treatment, in a cyclic fashion, highlighting the complex and dynamic nature of the ER α complex on chromatin (Metivier et al., 2003; Shang et al., 2000). These experiments showed the different role of

particular coregulators in the assembly of ER α transcription complexes (Metivier et al., 2003; Shang et al., 2000). Additionally, changes at the relative expression levels of some of these coregulatory proteins affect ER α -mediated physiological processes and have been implicated in enhanced estrogen-induced growth stimulation and development of endocrine resistance (Sommer and Fuqua, 2001).

Additionally, it has also been shown that ER α cooperates with other transcription factors for binding on specific sites on genome (also known as cisomes) to build active enhancer elements, a finding that revealed the dynamic nature of the ER α cisomes and how it is altered during breast cancer progression and in response to therapy (Bi et al., 2020). The ChIP-seq method in which chromatin immunoprecipitation (ChIP) assays are combined with sequencing, has been a powerful technique for identifying genome-wide DNA binding sites for transcription factors and other proteins. Many studies using this method have raised the importance of interactions between ER α and co-factors on regulating enhancer reprogramming that can be linked to endocrine therapy and clinical outcome (Bi et al., 2020; Ross-Innes et al., 2012).

These observations highlight the essentiality of the cofactors in ER α activity and the critical role of these interactions. Moreover, the diversity of combinations of coregulators highlights the complexity of the regulatory mechanisms and activation of ER α and the need for further investigation of the interactions between ER α and the different co-factors. Such associations are especially important in the context of dynamically regulated cellular environment (Kong et al., 2011) and can be targeted to develop new therapeutics. Additionally, most of the studies have been focused on specific genes and limited number of promoters, highlighting the need for a more detailed study of the different complex assemblies in an unbiased manner.

1.3.1 ER α Coactivators and Corepressors

Coactivators are proteins that interact with ER α and enhance gene transcription; the most characterised family of ER α co-activators is the p160/steroid receptor coactivator family (SRC), that consists of three members; SRC1 (NCOA1), SRC2 (NCOA2 or Tif2 or GRIP1) and SRC3 (NCOA3 or AIB1) (Zwart et al., 2011). The SRC co-activators are among the first factors to be recruited to ER α and this occurs via a short motif in co-activators, LXXLL (where L is leucine and X is any amino acid), which forms an α -helix and it is essential for the coactivator function toward nuclear receptors (Heery et al., 1997). The p160 coactivators can recruit and interact with secondary molecules that have an important role as chromatin modifiers, such as the histone acetyltransferases p300 and CREB-binding protein (CBP) and

the protein arginine methyltransferases (PRMTs), including coactivator-associated arginine methyltransferase 1 (CARM1) and PRMT1 (Acevedo and Kraus, 2003), in a ligand dependent manner. Members of the p160 family have been linked to breast cancer and response to endocrine therapy. SRC3 is frequently amplified in breast cancer and its increased expression in combination with ERBB2 expression has been correlated with poor tamoxifen response (Osborne et al., 2003), whereas SRC1 has been associated with promotion and execution of breast cancer metastasis and mediation of resistance to endocrine therapies (Browne et al., 2018; Walsh et al., 2012).

ER α can also interact directly with subunits of the SWI/SNF multiprotein complex including, BRG1, BRM and BAF57 (Belandia et al., 2002); these interactions lead to chromatin accessibility in an ATP-dependent manner that results in the subsequent binding of transcription factors and transcriptional activation. The BAF57 subunit interacts directly with ER α and the p160 family and studies support that BAF57 is required to target SWI/SNF complexes to estrogen-responsive promoters and enable p160 coactivators to enhance the ER α transcriptional activity (Belandia et al., 2002). Inhibition of BAF57 causes a reduction in the expression of endogenous ER α target genes and blocks the ER α -dependent cellular proliferation (Belandia et al., 2002). Experiments have also shown that defective BRG1 prevents ER α -mediated transcriptional activation (Belandia et al., 2002). ARID1A, one of the most well-studied subunits of the SWI/SNF complex, has been found to be frequently mutated in primary and recurrent breast cancers (Pereira et al., 2016; Yates et al., 2017) and notably the expression levels of ARID1A have been linked to tumour growth and clinical outcome in MCF7 breast cancer cells (Nagarajan et al., 2020; Xu et al., 2020). The findings from the inhibition of different SWI/SNF subunits and the detection of mutations in breast cancer models, highlight the essential role of this complex in ER α transcription, breast cancer and drug treatment response.

The interaction with factors that function as negative regulators and decrease transcription, provide an additional level of complexity in ER α action. The interaction of ER α with the nuclear receptor corepressor (NCoR) and the co-repressor silencing mediator for retinoid X receptor and thyroid hormone receptor (SMRT), mainly happens via LXXML motifs (Shiau et al., 1998), lead to the recruitment of large repressor complexes, including histone deacetylases (HDACs), that repress gene activity by causing chromatin condensation (Liu and Bagchi, 2004). Additionally, NCoR proteins interact with the nucleosome remodelling and histone deacetylation (NuRD) complex (Liu and Bagchi, 2004). The NuRD complex is a highly conserved multi-subunit protein complex, that combines chromatin remodeling and histone

deacetylation activities and consists of many different protein subunits, including histone deacetylases HDAC1/2, ATP-dependent remodelling enzymes CHD3/4, histone chaperones RBAP46/48, CpG-binding proteins MBD2/3, the GATAD2a (p66 α) and/or GATAD2b (p66 β) and specific DNA-binding proteins MTA1/2/3 (Torchy et al., 2015). It has been shown that components of the NuRD complex are essential factors, required for sustained cell growth of MCF7 cells, displaying the important role of these components (Serandour et al., 2018). In the absence of hormone or upon treatment with antagonists such as tamoxifen, NCoRs, HDACs and the nucleosome remodelling complex NuRD are recruited in a sequential manner, leading to a repressive chromatin state and inhibition of ER α target genes (Liu and Bagchi, 2004). Corepressors have also been implicated in breast cancer treatment; reduced levels of NCoR have been linked to tamoxifen resistance (Ali and Coombes, 2002). Additionally, a coding mutation in NCoR1 has been detected in primary breast cancer tumours that contributes to tumour progression (Pereira et al., 2016).

The complexity of ER α -associated factors has also been increased with the discovery of coregulators that display both coactivator and corepressor activity. CARM1 has been involved in repression of gene transcription via methylation of SRC3 or EP300 (Feng et al., 2006), whereas BRM and BRG1 are required for both ligand-dependent transcriptional activation and repression of the same set of genes (Zhang et al., 2007). Despite the plethora of proteins that have been discovered to be associated with ER α and play important roles in ER α -mediated signalling, a deeper understanding of the underlying molecular mechanisms is crucial to illuminate these interactions and reveal novel therapeutic targets for treating ER α -dependent breast cancers.

1.3.2 Pioneer factors

Studies have showed the important role of pioneer transcription factors in the ER α recruitment to DNA, as well as in ER α mediated transcriptional regulation of target genes. Pioneer factors can actively initiate the binding of regulatory factors on the DNA by either opening-up the local chromatin directly or by recruiting other chromatin modifiers and co-regulators (Zaret and Carroll, 2011). A well-known protein that functions as a pioneer factor for ER α is FOXA1. FOXA1 binding is necessary for gene regulation by ER α , as well as for the chromatin binding of ER α in the presence of both an agonist (estrogen) and an antagonist (tamoxifen) (Hurtado et al., 2011; Zaret and Carroll, 2011). It has been shown that FOXA1 binds to approximately 50% of all ER α -binding regions and knockdown of FOXA1 expression results in decreased

association of ER with chromatin, decreased cofactor recruitment and reduced expression of ER α target genes, demonstrating the important role of FOXA1 in mediating an estrogen response in breast cancer cells (Carroll et al., 2005; Hurtado et al., 2011). Notably, FOXA1 upregulation via gene amplification, has been reported in estrogen receptor-positive endocrine-resistant metastatic breast cancer to lead to genome-wide enhancer reprogramming and activation of prometastatic transcriptional programs that are associated with poor clinical outcome (Fu et al., 2019).

Additional analysis has uncovered other putative pioneer factors that contribute to ER α -chromatin interactions, such as GATA factors, PBX1 and AP-2 γ (Zaret and Carroll, 2011). GATA factors can interact with condensed chromatin and moderately influence chromatin accessibility (Cirillo et al., 2002). One of the members of GATA family, GATA3, has been implicated in tumour differentiation and ESR1 enhancer accessibility and has been previously shown to bind in MCF7 breast cancer cells in a ligand-independent manner (Kouros-Mehr et al., 2008; Theodorou et al., 2013). ER α , FOXA1 and GATA3 are frequently colocalised (Takaku et al., 2020) and it has been suggested that they form a network that is necessary for the full repertoire of cancer-associated effects of the ER α (Theodorou et al., 2013). Interestingly the expression of these three proteins defines ER α positive, luminal breast cancers and notably GATA3 is the third most commonly mutated gene in luminal breast cancer, following PIK3CA and TP53 (Lacroix and Leclercq, 2004; Sorlie et al., 2003).

PBX1 overlaps with approximately 50% of all ER α -binding events in the MCF7 breast cancer genome. It has been shown that PBX1 guides ER α recruitment to a specific subset of sites, within the genome, that promote breast cancer progression and can be used as a prognostic factor to discriminate ER α breast cancer outcomes (Magnani et al., 2011). The enrichment of AP-2 motifs in ER α binding sites led to the discovery of the transcription factor AP-2 γ . AP-2 γ has been implicated in breast cancer oncogenesis and it binds to ER α binding sites in a ligand-independent manner (Tan et al., 2011). Perturbation of AP-2 γ expression affects ER α DNA binding and expression of ER α target genes (Tan et al., 2011). It is clear that pioneer factors are required for ER α binding on DNA and their implication in drug resistance and prognosis of clinical outcome makes them attractive therapeutic drug targets.

1.3.3 ER α and the general transcription apparatus

Nuclear receptors activate gene expression by recruiting components of the basal transcription machinery to promoters of responsive genes. An important ER α interactor that communicate

the regulatory signals from DNA-binding transcription factors at enhancers directly to RNA polymerase II at core promoters, is the Mediator complex (Zhang et al., 2005). The Mediator complex is a multi-subunit complex that consists of 33 subunits in mammals and is organised into four distinct modules; head, middle, tail and CDK8 kinase (Soutourina, 2018; Yin and Wang, 2014) (**Figure 3**). Interactions are mainly mediated by the Mediator complex subunit 1 (MED1), that has been previously shown to directly interact with ER α and other nuclear receptors in a ligand-dependent manner (Kang et al., 2002; Zhang et al., 2005). This direct interaction is involved in mammary gland development, luminal cell differentiation and cell growth (Hasegawa et al., 2012). Interestingly, Mediator subcomplexes that do not contain the MED1 subunit can be stable and functional, however in nuclear receptor-mediated transcriptional activation, MED1 is the subunit mainly enriched in the pre-initiation complex (PIC), which comprises the six TFIIA to F complexes and RNA polymerase II on the promoter (Hasegawa et al., 2012). ER α can also interact directly with basal transcription factors such as TFIIB and subunits of TFIID, providing evidence for the direct role of ER α in transcriptional activation (Wu et al., 1999).

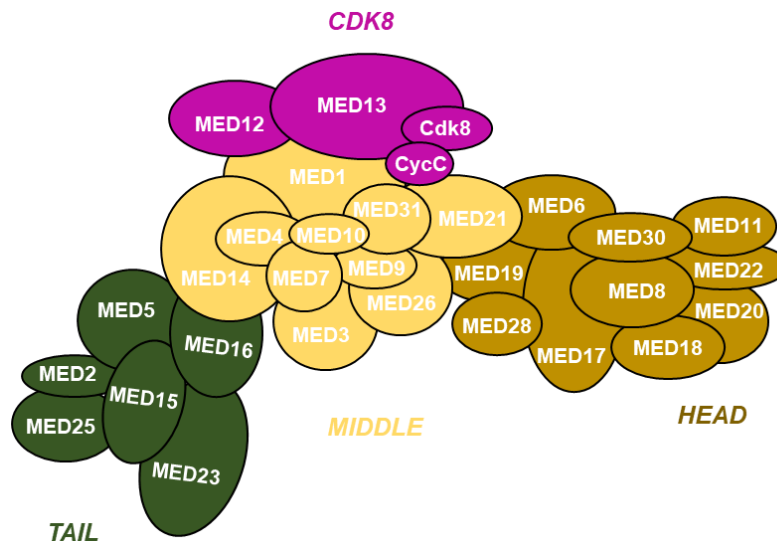


Figure 3. Schematic illustration of the Mediator complex.

The subunit composition of the mammalian Mediator complex and the four distinct modules (tail, middle, head and CDK8). The figure was adapted from Yin and Wang, 2014.

Overall, the Mediator complex has been detected at both enhancers and core promoters of actively transcribed genes, supporting the observations that this complex bridges interactions

between transcription factors at enhancers and the transcription initiation apparatus at core promoters. The Mediator complex functions as a bridge between ER α and basal transcriptional machinery, including Pol II and general transcription factors (GTFs; including transcription initiation factor IIA (TFIIA), TFIIB, TFIID, TFIIIE, TFIIF and TFIIH), promoting the assembly of the PIC complex on ligand-activated promoters (Soutourina, 2018) and enhance the transcriptional activity of ligand-activated ER α (Kang et al., 2002; Zhang et al., 2005). The Mediator complex can transfer signals from coactivators or corepressors to affect changes in the expression of ER α target genes (Malik and Roeder, 2010). Importantly, the Mediator complex integrates into complexes with known ER α co-activators, introducing multiple layers of cross-talk between cofactors. Studies have shown that the Mediator complex and p300/CBP-SRC are recruited in a specific order and work synergistically to promote a stable transcription pre-initiation complex, that is crucial for the initiation of transcription (Acevedo and Kraus, 2003; Shang et al., 2000). Recently it was highlighted that JMJD6, a member of the large family of JmjC-domain-containing proteins is responsible for the recruitment of Mediator in ER α -bound active enhancers via modulating the binding between MED12 subunit and CARM1 protein (Gao et al., 2018). Additionally, Mediator stimulates the phosphorylation of the Pol II carboxyl-terminal domain (CTD) by the CDK7 kinase subunit of TFIIH complex that has an important role in the initiation of transcription and in the transition from initiation to elongation (Sogaard and Svejstrup, 2007). Following the proper assembly of PIC, Pol II breaks contact with Mediator and the PIC complex leaves the promoter and moves to elongation. Interactions between Mediator and transcription elongation factors have been emerged; it is has been found that the super elongation complex, which is created from the combination of different Pol II elongation factors [eleven-nineteen Lys-rich leukaemia (ELL) proteins, positive transcription elongation factor b (P-TEFb) and several frequent mixed lineage leukaemia (MLL) translocation partners], is recruited via subunits of the Mediator complex, highlighting the involvement of the Mediator complex in Pol II transcription elongation (Luo et al., 2012).

The MED1 subunit is amplified and overexpressed in aggressive breast cancer and has been correlated with poor survival and resistance to endocrine therapies (Nagalingam et al., 2012; Nagpal et al., 2018; Zhang et al., 2013; Zhang et al., 2005). To this end, analysis of clinical samples indicated that high expression of MED1 was significantly associated with tamoxifen resistance in breast cancer (Nagalingam et al., 2012). Moreover, it has been shown that MED1 is required for estrogen-dependent growth of breast cancer cells. The MED1 role in drug resistance and cell growth raises the possibility of alternative therapeutic strategies that target MED1 for the treatment of breast cancer (Zhang et al., 2005). Additionally, mutations in other

Mediator subunits have been associated with developmental diseases or cancer (Yin and Wang, 2014), indicating the important role of the Mediator complex in gene transcription regulation. Further study of the Mediator complex and its association with ER α and other transcription factors may help to understand better the function of the different subunits and the distinct subcomplexes, how they are regulated and how the complex is implicated in various human diseases.

1.4 Treatment strategies for ER α positive breast cancer

ER α -positive breast cancers largely depend on estrogen and endocrine therapies have remained the mainstay of ER α -positive breast cancer treatment for several decades, leading to decreased cancer recurrence and mortality (Ali and Coombes, 2002; Hanker et al., 2020). There are two main classes of antiestrogens based on their chemical structure and tissue specific functions. The first class includes pure antiestrogens, known as selective ER α downregulators (SERDs), such as ICI 164, 384 and its derivative ICI 182,780 (fulvestrant), that act primarily as antagonists in all tissues and prevent activation of AF1 and AF2, decreasing the half-life of the ER α (Dauvois et al., 1992; Guan et al., 2019). In the second class, the non-steroidal antiestrogens (known as selective ER α modulators-SERMs), such as tamoxifen and raloxifene, have a dual function with selective estrogenic activities in certain tissues and anti-estrogenic activities in others. These drugs affect ER α activity by inhibiting AF2 activation and not the AF1; because the ER α activity in breast epithelium is based mainly on AF2 domain, tamoxifen function as an antagonist in breast cells, whereas in other tissues, such as the uterus, AF1 activity is more crucial, resulting in agonistic activity (Ali and Coombes, 2002).

The anti-estrogen tamoxifen was initially used for the treatment of metastatic breast cancer leading to disease regression in approximately 30% of the cancers, but it is restricted to patients with ER α -positive breast cancer (Ali and Coombes, 2002). The active metabolite of tamoxifen, 4-hydroxytamoxifen, competes with E2 for ER α binding and induces conformational changes that have been suggested to prevent coactivator recruitment and facilitate the recruitment of corepressor complexes, such as NCoR and NuRD, that cause the abrogation of gene transcription (**Figure 4**). Structural studies have revealed that when tamoxifen binds to LBD, its carboxy-terminal-most helix is positioned over the coactivator-binding pocket, thereby preventing their recruitment and transcription activation by AF2 (Shiau et al., 1998). The differential coactivator-corepressor recruitment in certain cell and promoter contexts mediate

the activities of SERMs (Shang et al., 2000). However, despite the clinical utility of tamoxifen, the exact mechanism of action still remains unclear.

In addition to anti-estrogen therapies, patients with ER α -positive breast cancer have also been treated with aromatase inhibitors (anastrozole, letrozole and exemestane), that function by blocking estrogen production (Chumsri et al., 2011; Lonning and Eikesdal, 2013). Estrogen is synthesised through conversion from androgen by aromatase enzyme, which is a member of the cytochrome P450 family and the product of the CYP19A1 gene (Chumsri et al., 2011). Aromatase inhibitors interact competitively with the active site of aromatase, preventing androgen binding and the conversion of androgens to estrogens (Chumsri et al., 2011). In postmenopausal women, aromatase inhibitors lead to a 98% decrease in circulating levels of estrogen and are considered to be the standard of care for postmenopausal women with hormone receptor-positive breast cancer (Lonning and Eikesdal, 2013). Patients that have been treated with anti-estrogens or aromatase inhibitors can develop resistance via different mechanisms and interestingly patients with resistance to aromatase inhibitors often respond to anti-estrogen therapies (Ingle et al., 2006). The molecular mechanisms that lead to endocrine resistance and disease recurrence in ER α -positive breast cancer have been the main focus of several studies and continues to be an active area of research.

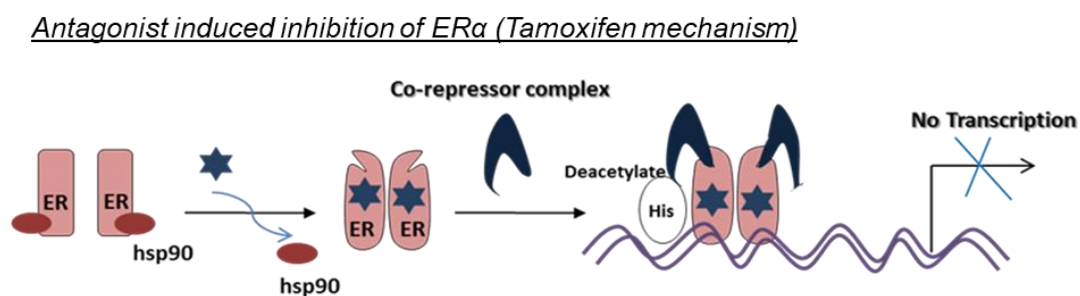


Figure 4. Mechanism of tamoxifen action.

The binding of tamoxifen to the ER α receptor causes conformational changes that are distinct from those that are induced by agonists. The changes lead to the recruitment of a corepressor complex that causes deacetylation of histones thereby abrogating the transcription of ER α target genes.

1.4.1 Mechanisms of resistance to breast cancer therapies

Endocrine therapy has led to a significant reduction in recurrence and mortality of breast cancer, however many patients either fail to respond or develop drug resistance. More than

25% of patients diagnosed with early stage breast cancer, develop resistance and relapse with metastatic and incurable disease. Drug resistance can emerge through various mechanisms (Ali and Coombes, 2002; Jeselsohn et al., 2015; Shiino et al., 2016):

- A small fraction of metastatic tumours (15-20%) lose ER α expression. Studies have described the discordance of ER α levels between primary and recurrent tumours (Shiino et al., 2016).
- Truncated ER α variant isoforms; increased expression of the variant ER- α 36, that lacks AF-1 and a large portion of the LBD, has been associated with anti-estrogen resistance in breast cancer cells (Thomas and Gustafsson, 2015).
- Point mutations in ESR1 gene have been discovered in 30% of metastatic ER α positive breast cancers. These mutations are clustered within the LBD domain and lead to constitutive ligand-independent activity and endocrine resistance (Dhiman et al., 2018; Sommer and Fuqua, 2001). It has been suggested that these mutations promote metastatic phenotype by changing gene expression through constant ER α activity or indirectly via additional transcription factors (Arnesen et al., 2020; Jeselsohn et al., 2018). Two of the most common mutations, Y537S and D538G, account for approximately 70% of all ESR1 mutations identified in patients with metastatic breast cancer.
- Growth factor-driven mitogenic and survival pathways (i.e. phosphatidylinositol 3-kinase [PI3K]/mammalian target of rapamycin [TOR], RAS/RAF/MEK/ERK, CDK4/6-cyclin D1-dependent inactivation of Retinoblastoma Protein (RB) and de-repression of E2F transcription factors) that induce ER α phosphorylation and promote transcription of ER α -regulated genes in a ligand-independent manner. These resistant tumours are typically dependent upon both the aberrantly activated survival pathway and ER α .
- Changes in cofactors expression levels: Overexpression of co-activators that have a major role in mediating transcription activation by ER α or lower expression of corepressor activity that affects the inhibitory action of tamoxifen, are involved in anti-estrogen resistance (Ali and Coombes, 2002).

The different classes of endocrine drugs affect ER α via distinct mechanisms and patients that have developed resistance to one drug, can be treated sequentially with different endocrine agents. Alternative strategies that have been developed to overcome endocrine resistance, include the Proteolysis-targeting chimeras (PROTACs), which are small molecules that recruit

an E3-ubiquitin ligase to the target protein and can degrade both wild type and mutant forms of ER α (Hanker et al., 2020). Additionally, therapeutic strategies that target the crosstalk between ER α and other signalling pathways have shown promising results (Ignatiadis and Sotiriou, 2013). The combination of CDK4/6 inhibitors (palbociclib, ribociclib and abemaciclib) with antiestrogens or aromatase inhibitors, that target the ligand-independent activity of ER α and cell-cycle signalling, have increased the patient survival and is currently the standard first-line treatment for metastatic ER α -positive breast cancer (Ignatiadis and Sotiriou, 2013). Interestingly, combined inhibition of mTORC1/2, CDK4/6 and ER α have revealed significant regression in breast cancer cell lines and xenografts (Michaloglou et al., 2018). However, patients develop resistance to CDK4/6 inhibitor therapy and the mechanisms that have been suggested to be related to acquired resistance is retinoblastoma loss, amplification of CDK6 or cyclin E (CCNE1), alterations in DNA repair and IL6/STAT3 pathways (Kettner et al., 2019). Studies have shown that patients with resistance to CDK4/6 inhibitors may still be sensitive to mTORC1/2 inhibitors together with antihormonal therapy (Michaloglou et al., 2018) or to a combination of STAT3 and PARP inhibitors that target both IL6/STAT3 and DNA repair pathways (Kettner et al., 2019). To conclude, more efforts to elucidate the mechanisms that underlie resistance might lead to different combination of drugs or to the development of alternative breast cancer therapeutic strategies that can decrease the number of patients that develop resistance, associated with metastatic and incurable disease.

1.5 Three-dimensional genome

The three-dimensional (3D) organisation of mammalian genomes holds a fundamental role in many cellular functions by allowing genomic elements that are located very remotely to contact and regulate each other. Vital cellular processes including transcription, replication, DNA damage and DNA repair have been linked to the way chromosomes are folded in the 3D space (McCord et al., 2020). It has been previously reported that ER α binding sites are located within enhancer regions and the contact with promoter regions happens through the formation of chromatin loops (Fullwood et al., 2009). ER α mainly binds at distal enhancers to regulate the expression of its target genes and dictate cell growth and endocrine therapy; differential binding of ER α has been associated with endocrine resistance and clinical outcome in breast cancer, highlighting the importance of the ER α cistrome (Bi et al., 2020). The long-range looping interactions between enhancers and promoters appear to be crucial for driving high-level and cell type-specific gene expression. Enhancer deletions or ectopic enhancer-promoter

interactions along with ‘forced looping’ experiments have highlighted the importance of promoter-enhancer interactions and how vital is to regulate these interactions to prevent inappropriate gene expression (Deng et al., 2012; Lupianez et al., 2015; Sur et al., 2012).

The study of the three-dimensional genome has been facilitated by the development of techniques, such as chromosome conformation capture (3C) and its variants (Dekker et al., 2002; McCord et al., 2020). These techniques have highlighted the high complexity of the three-dimensional promoter-enhancer architecture, with promoters skipping enhancers in close proximity and often interact with multiple enhancers (Schoenfelder et al., 2015). The most powerful of these methods, Hi-C (high throughput chromosome conformation capture), has been applied for the global detection of interactions between individual restriction fragments (such as those containing a promoter or enhancer). The Hi-C method can capture all interactions across the whole genome at the same time, without depending on immunoprecipitation steps, in which the properties of the antibody can affect the library complexity (Denker and de Laat, 2016). The main limitation of Hi-C method is the complexity of the libraries and the need for deep-sequencing that can significantly increase the cost of the process (Schoenfelder et al., 2018). Recently, Peter Fraser’s lab developed a technique termed Promoter Capture Hi-C (PCHi-C) to specifically enrich for promoter-containing ligation products from Hi-C libraries, reducing the complexity of the libraries and increasing the resolution of the assay (Schoenfelder et al., 2018). Focusing on promoter interactions can be very informative, as promoter-enhancer contacts have been shown to be essential for proper gene expression levels (Deng et al., 2012; Lupianez et al., 2015; Sur et al., 2012). Additionally, promoters are highly conserved along different cell types, thus the same capture bait system can be used to study the same interactions across multiple cell types and conditions (Schoenfelder et al., 2018).

Although, approaches to study long-range interactions have been developed, for most of the enhancers their target genes are still unknown. Thus, assigning transcriptional enhancers to their target genes remains a challenge in decoding mammalian gene expression control. Additionally, how these loops are formed and stabilised remains uncertain; initial studies indicated a colocalisation and an important interplay between Mediator and cohesin complexes for gene activation through formation and stabilisation of enhancer-promoter DNA looping (Kagey et al., 2010). The core cohesin complex consists of a heterodimer of the proteins SMC1 (structural maintenance of chromosomes 1-produced from the SMC1A gene) and SMC3, in combination with SCC1 and SCC3 and it is crucial for genome stability, cell division, transcription and chromatin organisation (Pombo and Dillon, 2015). The interaction between

the DNA-binding protein CCCTC-binding factor (CTCF), which is a conserved zinc finger transcription factor and the cohesin complex, has also been suggested to contribute to the establishment of chromatin loops and the regulation of gene expression in mammalian cells, as studies have reported loss of DNA looping following depletion of either cohesin or CTCF (Li et al., 2020b; Nora et al., 2017; Rao et al., 2017). Notably, a role for CTCF in ER α biology has been shown, as CTCF binding events overlap with ER α and FOXA1 (Ross-Innes et al., 2011). Recently published data indicate that the Mediator complex along with transcription factors, coactivators and Pol II creates a functional bridge between promoters and enhancers, but the Mediator complex is not required for the formation of DNA loops, as chromatin architecture was largely unchanged after Mediator depletion (El Khattabi et al., 2019). In the same study the importance of the cohesin complex was highlighted, as its depletion decreased the promoter-enhancer contacts (El Khattabi et al., 2019). These data were supported from one more recent study, which showed that the loss of Mediator affected transcription of super enhancer-driven, cell-type-specifying genes and had a mild impact on overall transcriptional output and genome architecture (Jaeger et al., 2020). Linking distal regulatory elements to specific promoters, genome-wide, is important for understanding how the expression of multiple genes is modulated and modified across multiple diseases. Additionally, further study of the spatial regulatory circuitry can unravel the mechanistic role of the Mediator complex, cohesin, CTCF and other novel transcription factors in controlling and stabilising enhancer-promoter looping.

1.6 Mass spectrometry-based proteomics

Like ER α , proteins often function as part of different complexes to carry out specific biological activities. Protein-protein interactions or interactions with non-protein molecules, such as DNA or RNA are critical for cellular functions (Bludau and Aebersold, 2020; Gingras et al., 2005). Especially, the composition of complexes in the regulatory region of different genes determines the activation or repression of genes (Wierer and Mann, 2016). A better understanding and characterisation of complex assembly can offer important insights into protein function and its role in the development of pathological states. Moreover, a more comprehensive map of the protein interaction data can enable further opportunities for molecular characterisation of human diseases and sensitivity to drugs.

Proteomics, the large-scale analysis of proteins, utilising high-resolution mass spectrometry has been a powerful technology for the study of protein complexes and entire cellular

proteomes with increasingly improved depth and sensitivity (Wierer and Mann, 2016). Affinity purification in combination with high resolution mass spectrometry (MS)-based proteomics is now the method of choice for the characterisation of soluble protein complexes, especially chromatin associated interactomes (Gingras et al., 2007). In addition, the combination with quantitative MS techniques can be applied to discriminate between true interactors and noise, offering a more dynamic study of the regulation of the protein interactions in a cell, tissue or organism. In a typical mass spectrometry-based protein identification analysis, peptides generated by proteolytic digestion of proteins enter the spectrometer and their mass/charge ratios are measured in real time (MS1 event). Then, the most intense peptides are automatically isolated and fragmented in a collision cell, which contains an inert gas, followed by disruption of amine bonds. This event is known as MS/MS (MS2 event) and provides a mass spectrum of the fragments that can be used for peptide sequencing and subsequently protein identification. Specialised peptide identification algorithms are finally utilised to assign the collected MS/MS spectra to theoretical spectra predicted from known protein sequences (Gingras et al., 2005). Mass spectrometry-based proteomics can be a powerful tool for the rapid and unbiased identification of protein interactomes and a wide variety of MS based-methods has been developed, yielding invaluable insights into protein function. Yet improvements in sample preparation protocols, chromatographic separations, bioinformatics data analysis tools and particularly mass spectrometry instrumentation continue to advance the field, increasing its application for the deeper study of protein networks.

1.6.1 Approaches for studying protein-protein interactions

Several methods have been developed to understand how proteins are organised into modules and the role of these interactome networks into driving cellular functions. One of the first methods that was used to map protein-protein interactions, was the yeast two-hybrid approach (Y2H) that provides a quick and simple mechanism to identify proteins that interact physically *in vivo*, yielding large Y2H interaction maps (Rolland et al., 2014; Stelzl et al., 2005). In a Y2H assay, both the specific bait and the variable prey are exogenous fusion proteins expressed in specific compartments of the yeast cells. When there is a direct interaction between bait and prey, a selectable readout is generated, permitting identification of the prey based on cDNA sequence analysis (Stelzl et al., 2005). Although, it has been successfully used to study protein function of different model organisms, the Y2H approach only measures binary and often direct

interactions, losing more complex structural interactions. Additionally, is prone to high rates of false positive and negative results (Parrish et al., 2006).

The human protein interactome maps generated using yeast two-hybrid, have been recently complemented by studies utilising large-scale affinity purification of epitope-tagged proteins followed by mass spectrometry (AP-MS), that have identified proteins associated with baits from many protein families, facilitating the study of large protein assemblies (Hein et al., 2015; Huttlin et al., 2015). An example of a complex with conformational and compositional heterogeneity is the Mediator complex, which has been mainly studied with AP-MS, revealing the locations and orientations of Mediator subunits (Robinson et al., 2015). Huttlin et al, performed an MS profiling of C-terminally FLAG-HA-tagged baits, using lentiviral expression, from 13,000 proteins and generated a network that contained 23,744 interactions among 7,668 proteins, revealing known complexes and many novel interactions that had not been described before (Huttlin et al., 2015). In the same context, Hein et al, studied 28,000 protein-protein interactions by expressing 1,125 GFP-tagged proteins in HeLa cell lines (Hein et al., 2015). In the same study authors combined AP-MS with label-free quantification to enable the study of stoichiometric changes in protein complexes. Another method that has been applied to study protein-protein interactions is tandem affinity purification (TAP) combined with mass spectrometry, which involves cloning tags onto the gene of interest and the introduction of the construct into the host cell, resulting in expression of a fused protein (Rigaut et al., 1999). This approach has been successfully used to characterise protein associations including components of the TNF- α /NF- κ B pathway (Bouwmeester et al., 2004) and ER α complex (Ambrosino et al., 2010; Gigantino et al., 2020). Although the application of tagging approaches followed by mass spectrometry has allowed the characterisation of multiple interactomes with increased sensitivity, a few limitations should be considered before its application (Bauer and Kuster, 2003; Zheng et al., 2020). The tagging strategy may mask areas of the bait protein that have a crucial role in the complex assembly and can have negative effects on the protein structure of the target protein. One more aspect that should be considered in the tag-based AP-MS approach is that the overexpression of the bait protein can alter the stoichiometry of protein complexes, introducing some artificial results. Additionally, the transient expression of bait proteins can be toxic for some cell lines or can compete with the expression of the endogenous untagged version, affecting the different protein networks and their physiologic relevance. One approach to overcome this challenge is to silence the endogenous genes by RNA interference or apply the method on biological systems that do not express the target protein (Forler et al., 2003). Additionally, a growing number of studies have

adopted CRISPR/Cas9 technologies to keep the expression of the tagged bait protein at the endogenous level, overcoming the effect of the protein overexpression and increasing the specificity of the AP-MS results (Vandemoortele et al., 2019).

The BioID (Proximity-dependent Biotin Identification) method followed by mass spectrometry has been commonly used for protein interaction screening (Roux et al., 2018). Proteins that are in close proximity to the target protein are biotinylated and isolated with standard biotin-affinity capture, before they are analysed using mass spectrometry. The BioID approach has been beneficial for the study of insoluble and membrane-associated proteins, however an exogenous fusion protein needs to be expressed and potential interactions need to have accessible primary amines for biotinylation (Roux et al., 2018). Recently, an alternative computational approach has been described for the prediction of protein-protein interactions; studies have used coregulation analysis to map protein-protein associations and their role in breast and colorectal cancer. The study by Roumeliotis et al (Roumeliotis et al., 2017), produced a deep representation of the functional network and the molecular structure underlying the heterogeneity of colon cancer cells, whereas the study by Lapek et al. (Lapek et al., 2017), revealed protein co-regulations that had an impact on cellular fitness of breast cancer cell lines.

More recently, the use of chemical crosslinking combined with mass spectrometry (XL-MS) has provided information about endogenous protein assemblies not only at the level of selected protein complexes, but also at a proteome-wide scale (Yu and Huang, 2018). In a general XL-MS workflow, proteins are reacted with bifunctional crosslinking reagents, that are used to covalently bind two or more protein molecules, capturing their spatial interactions, followed by digestion and analysis of the resultant peptides via liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis. Quantitative XL-MS strategies have also been developed to dissect the interaction and structural dynamics of protein complexes. Improvements and innovations in crosslinkers, mass spectrometry and data processing have made the XL-MS technology a powerful structural tool that offers information about spatial orientation, protein connectivity and physical contacts between protein complexes, elucidating protein complex architecture (Yu and Huang, 2018). Nevertheless, challenges do exist with regard to data processing and sensitivity that need to be addressed to enable the application of this approach more routinely.

The combination of modern MS with Chromatin Immunoprecipitation (ChIP) protocols has enabled the sensitive and accurate characterisation of chromatin associated complexes (Berggard et al., 2007; Rafiee et al., 2016). Among the different approaches described in

literature, the method termed RIME (Rapid immunoprecipitation mass spectrometry of endogenous proteins), developed in the Carroll lab, concentrates several advantages for the study of the composition of different target complexes (Mohammed and Carroll, 2013). The workflow starts with formaldehyde crosslinking, which helps to maintain transient protein-protein interactions, followed by chromatin sonication and immunoprecipitation using an antibody against a target protein. Finally, the eluents are digested with trypsin and subjected to mass spectrometry for analysis (**Figure 5**). The RIME method has been extensively applied for the study of endogenous protein-protein interactions from our lab (Mohammed et al., 2013; Mohammed et al., 2015; Nagarajan et al., 2020), as well as from other laboratories (Das Gupta et al., 2020; Shu et al., 2016), as it is easy adaptable. Usually, appropriate negative controls, such as IgGs are included to discriminate between bona fide components of the complex and background contaminants. Additionally, the RIME method has been combined with quantitative approaches to define the dynamics of interactions upon different treatments (Mohammed et al., 2013; Mohammed et al., 2015). However, this application was limited to the number of different conditions that could be analysed within one experiment. Also, protein interactions could not be captured in tissues; RIME was combined with a targeted proteomic approach to quantify peptides from selected proteins (Mohammed et al., 2013). To conclude, a wide range of approaches has been developed to study the complex protein networks and their dynamics on a global scale and further improvements in the pipelines can be incredibly informative and reveal insight into pathways and regulatory mechanisms.

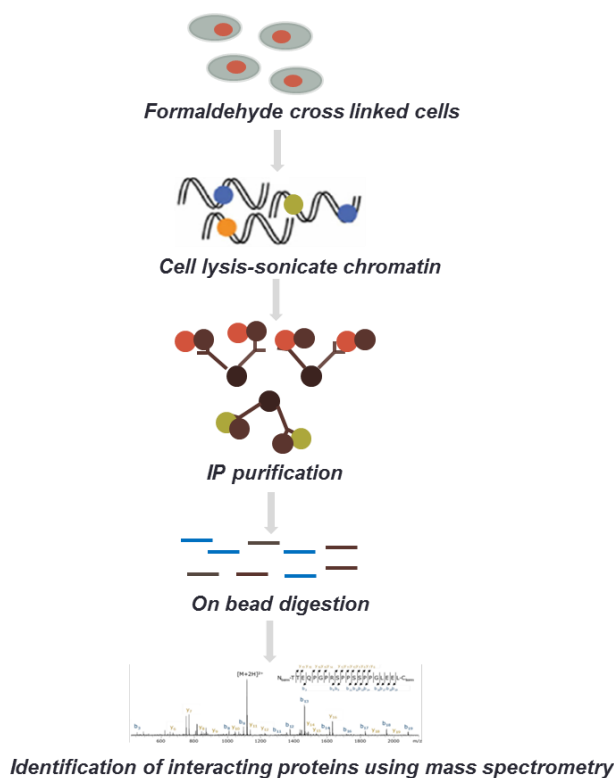


Figure 5. RIME workflow.

RIME starts with crosslinking followed by cells lysis and chromatin sonication. Then an antibody is used to pull down the protein of interest along with its interactors, followed by on bead digestion and identification using mass spectrometry.

1.6.2 Mass spectrometry-based detection of Post-Translational Modifications

For the better understanding of signalling processes, high resolution mass spectrometry combined with quantitative approaches has been used for the analysis of post-translational modifications (PTMs) including acetylation, ubiquitination and phosphorylation (Paulo et al., 2015; Rose et al., 2016; Svinkina et al., 2015). PTMs can modulate the function of a protein, its structure and subcellular localisation, as well as the interactions with other proteins, emphasising the need for their in-depth characterisation (Bludau and Aebersold, 2020; Csizmok and Forman-Kay, 2018). Notably, the analysis of post-translational modifications (PTMs) by mass spectrometry can be challenging as it depends on many aspects: (a) the mass shift in the peptide molecular weight, (b) the abundance of the modified peptide, (c) the stability of the modification during the mass spectrometry analysis and (d) the impact of the modification on the peptide's ionisation efficiency (Parker et al., 2010). Overall, the identification of PTMs has been a challenge for the proteomics research, mainly because of the high diversity and heterogeneity of the modified gene products.

Acetylation is common on N-terminal and amino groups of lysine and arginine; it is a well-known regulatory PTM in the context of nuclear signalling, specifically for regulating gene expression via modification of histones (Parker et al., 2010). Non-histone acetylation plays a role in protein stability, DNA binding, gene expression, localisation, mRNA stability, enzymatic activity and protein interactions (Spange et al., 2009). An example is the interaction between TFIIB and TFIIF that is increased via autoacetylation of TFIIB at K238 leading to strong activation of transcription (Choi et al., 2003). Histone acetylation has been studied using metabolic labelling prior to mass spectrometry analysis (Everitts et al., 2013), whereas for the detection of acetylated lysines, strategies based on the enrichment of acetylated peptides using antibodies have been developed (Svinkina et al., 2015). Ubiquitination regulates numerous cellular processes, including protein degradation, signal transduction, DNA repair and cell division, as well as stability, function and localisation of various proteins (Danielsen et al., 2011). The study of protein ubiquitination is difficult because of the large size of the modification and the low abundance of ubiquitinated proteins that can be overcome with large amount of starting material (Parker et al., 2010). Recent strategies, including antibody-based methods combined with quantitative mass spectrometry have improved the coverage of the ubiquitinome, leading to the identification and quantification of 19,000 ubiquitylation sites (Kim et al., 2011).

Phosphorylation is the most frequent protein modification in cells and has a vital role in important regulatory processes involved in cell signalling and cancer (Parker et al., 2010). Notably, it has been reported that phosphorylation on specific residues can affect protein interactions (Bauer et al., 2003; McFarland et al., 2008). The analysis of phospho-sites by mass spectrometry can be challenging as the stoichiometry of phosphorylation is relatively low and an enrichment step is required for better coverage of phospho-sites and the discovery of important phosphorylation motifs (Mann et al., 2002; Paulo et al., 2015). Additionally, the phosphorylated sites on proteins may vary and one protein can exist in several phosphorylated forms increasing the complexity of the phosphoproteome analysis. Furthermore, the quantification of phosphorylation and other post-translational modifications requires correction by whole protein quantification to avoid misinterpretations arising from changes in protein expression over the course of an experiment (Mann et al., 2002). Methods that are widely used for the enrichment of phosphosites, on serine and threonine residues, are the miniaturised immobilised metal affinity chromatography columns (IMAC) and the titanium dioxide (TiO₂) columns. IMAC enrichment is based on the high affinity of phosphate groups, that are negatively charged, towards a metal-chelated stationary phase, especially Fe³⁺ and Ga³⁺, that

are positively charged, followed by release of phosphopeptides from the IMAC resin using alkaline buffers (Thingholm et al., 2009). In the TiO₂ enrichment method, samples are loaded into the columns in an acidic environment and the phosphopeptides are eluted using an ammonia solution at high pH (Thingholm et al., 2009).

Due to the above technical challenges, the modification landscape of proteins and especially protein interactomes remain largely unexplored. However, with continued development of techniques with improved sensitivity and specificity, the study of PTMs and their role in protein-protein interactions can be facilitated.

1.6.3 Relative quantification proteomics approaches

The latest advances in mass spectrometry in terms of sensitivity, speed, mass accuracy and resolution have facilitated the development of robust quantitative proteomic approaches that allow the unbiased study of the functional role of proteins in a plethora of biological contexts. In addition to mass spectrometry, the development of software packages, such as Proteome Discoverer (Lemeer et al., 2012) and MaxQuant (Tyanova et al., 2016b), that are compatible with high-throughput proteomics, had a great impact on the development of quantitative proteomics, as they have enabled the analysis of large quantitative datasets within hours. Relative quantification measurements can now be collected for complex proteomes, in terms of both numbers of proteins and dynamic expression range, for the analysis of protein expression patterns and post-translational modification states, as well as for the study of biological systems, where sample availability is limited (Rauniyar and Yates, 2014).

Among the many formats for quantitative proteomics, isotope-labelled (TMT or iTRAQ and SILAC) and label-free, have facilitated the quantification of whole proteomes and more recently interactomes (Hein et al., 2015; Huttlin et al., 2015; Mohammed et al., 2013; Roumeliotis et al., 2017). Stable isotope labelling of proteins or peptides prior to analysis is one of the most popular methods for relative quantification combined with mass spectrometry as it is characterised by high reproducibility and proteome coverage. The different quantitative methods (isotope-labelled or label-free) have distinct advantages and disadvantages and the choice of quantitative proteomics method depends primarily on the type of the experiment and the resources that are available to each researcher (Gingras et al., 2007).

The combination of these techniques with affinity purification has enabled the study of the dynamics of complexes and the discovery of proteins that are significantly enriched over mock pull-downs or between target protein pull-downs in different conditions (Gingras et al., 2007).

The quantitative study of interactomes is a growing field and the development of more robust pipelines can enable the comprehensive study of interactome dynamics and stoichiometry, providing answers to open biological questions.

1.6.3.1 Metabaling labelling

Stable Isotope Labelling by Amino acids in Cell culture (SILAC) is the most popular metabolic labelling method, that utilises the translational machinery of the cell in order to incorporate heavy versions of specified isotopic labelled amino acids added in the growth medium (Ong et al., 2002). Protein extracts from differentially labelled cells are mixed and digested with a proteolytic enzyme prior to mass spectrometry analysis. Depending on the labelling scheme, the light and heavy forms of a given peptide simultaneously appear as duplets or triplets on the mass spectrum and quantification is retrieved from the relative peak intensities by comparing heavy/light peptide pairs (Gingras et al., 2007). The SILAC approach has been successfully used as a main quantitative strategy in many studies (Blagoev et al., 2003; Everley et al., 2006; Mohammed et al., 2013; Mohammed et al., 2015; Wiley et al., 2019) and the combination of the samples at the earliest step of the workflow has minimised variability and systematic errors generated during the sample preparation. Although, SILAC is a simple and robust method, it is still limited in the number of samples that can be combined in a single analysis run. Additionally, this approach can only be applied for cell lines that can grow with SILAC culture media, as there are cell lines that their growth rates can be affected when their culture media is changed (Gingras et al., 2007).

A recent development has expanded the application of SILAC *in vivo*, using SILAC-labeled cells for the quantification of proteins from Patient-Derived xenografts (PDX) or human tissue (Tyanova et al., 2016a). This method is termed super-SILAC and it uses a mixture of different SILAC-labelled cell lines as an internal standard for quantifying proteins in tissue (Geiger et al., 2010). Although, this approach offers great opportunities for the quantitative study of clinical material it can only be comprehensive and accurate when the reference standard contains all proteins that are detected in the tissue samples (Geiger et al., 2010). Even if SILAC has a few limitations, the low cost of the technique, its easy implementation and the high efficiency of the labelling has made it one of the first choices especially for the study of culture cells or SILAC-compatible organisms.

1.6.3.2 Label-free quantification

The label-free quantification (LFQ) approach provides quantification of peptides and proteins without the use of stable-isotope labels and can be divided into two main approaches, that are commonly used (Zhu et al., 2010): (1) an approach that utilise measurements of ion intensity changes, such as peptide peak areas or peak intensities in the mass spectrometer, as a quantitative measure and (2) spectrum count method that utilise the number of peptide-to-spectrum matches (PSMs), obtained for each protein to measure protein quantity. Both approaches are based on measurements from individual LC-MS/MS runs and any changes at protein levels are calculated via a direct comparison between different analyses. They have been very popular for studies where large-scale sets of experimental data need to be compared due to its simplicity. The label-free approach can be the first choice for the quantitative study of cell lines that do not grow under the SILAC media. An important experimental requirement for label free-based experiments is a robust LC-MS setup that minimises the shift on retention times and changes in chromatographic peaks, thus leading to fewer false assignments and more accurate quantification.

Several software with unique algorithms have been developed for the analysis of label-free mass spectrometry data such as MaxLFQ, which is implemented in the MaxQuant platform (Cox et al., 2014). Recently, a label-free quantification data analysis algorithm, called Minora, that utilises peak area or intensity has been implemented in the new versions of Proteome Discoverer software (Glont et al., 2019b), allowing an easy label-free quantification in large datasets. The Minora algorithm has already been used for precursor ion quantification, however it was recently adjusted to detect and quantify isotopic clusters, whether they are associated or not with a peptide spectral match, increasing the coverage of the quantified proteome. A comparison between Minora and the spectral counting approach, revealed that the precision of the ratios was better using Minora and a higher number of quantified proteins was detected in the Minora-based quantification (<http://tools.thermofisher.com/content/sfs/posters/PN-64792-Label-Free-Proteome-Discoverer-ASMS2016-PN64792-EN.pdf>).

Most quantification approaches follow a data-dependent acquisition (DDA) strategy, where precursors with the highest intensity are selected for fragmentation during the mass spectrometry analysis. An alternative approach for label-free quantification is the sequential window acquisition of all theoretical mass spectra (SWATH-MS), where all ionised compounds of a given sample, that fall within a specified mass range, are fragmented in a systematic and unbiased fashion [data-independent acquisition (DIA)] (Ludwig et al., 2018).

SWATH-MS can be used for large number of samples, offering increased quantitative consistency and accuracy, avoiding a stochastic and not reproducible precursor ion selection. However, it requires prior knowledge about the chromatographic and mass spectrometric behaviour of peptides in a form of custom spectral libraries (Ludwig et al., 2018). This can increase the complexity of data analysis and limited the total protein identifications to the ones that are only included in the libraries.

Overall, the main limitation of the label-free based quantification approaches is that the individual LC-MS/MS analysis for each sample can introduce variability in peptide detection, as an ion selected for fragmentation in one LC-MS/MS run may not be selected consistently in subsequent runs; this is more pronounced for low intensity ions corresponding to low abundant peptides. This can result in missing observations across the different samples that can increase the complexity of the data analysis and reduce the power of the downstream statistical analysis, affecting more significantly the quantification of lower abundance proteins. Additionally, the separate analysis of many samples can increase the mass spectrometry run time leading to an increased cost of the experiment (Li et al., 2012). Although, label-free based quantification approaches can be an excellent alternative to labelling techniques, some important limitations of the approach should be considered before the application on samples with limited amount of protein or high heterogeneity.

1.6.3.3 Chemical isobaric labelling

The two most widely used chemical labelling methods are tandem mass tags (TMTs) and isobaric tags for absolute and relative quantification (iTRAQ), both of which target primary amines and are so-called isobaric tags (Rauniyar and Yates, 2014). Stable isotope labelling with iTRAQ or TMT is the technology of choice, when simultaneous multiple comparison analysis of clinical samples or temporal multi-time point profiling is required. In this approach, peptides from up to 8 or 11 different conditions are chemically modified with isobaric molecules through an amine-reactive NHS-ester group before they are mixed in a single sample (**Figure 6a**). These isobaric tags contain different combinations of stable isotopes on a spacer arm and a complementary reporter group, which are designed to maintain a constant total mass for all the different reagents. Tandem MS fragmentation gives rise to up to 8 or 11 different low mass reporter ion peaks with intensities proportional to the peptide relative abundance. Fragmentation also generates peptide fragment ion peaks that can be used for peptide and subsequently protein identifications. Biological and technical replicates can be accommodated

in the same analysis with no additional sample preparation and analysis time, therefore experimental designs can be improved in terms of throughput and statistical power. Hyperplexing by the combinatorial use of isobaric and metabolic labelling has also been described that enabled the simultaneous quantification of 18 samples and provided a tool for studying protein turnover (Dephoure and Gygi, 2012).

The main limitation of isobaric labelling-based experiments is the presence of co-eluted peptides within the isolation window that has been applied for the selection and fragmentation of the peptide of interest. The co-isolation and co-fragmentation of interfering peptides of similar mass-to-charge, compromise protein quantification and result in an underestimation or compression of actual protein abundance differences in the analysed sample set (Rauniyar and Yates, 2014). The latest developments in mass spectrometry and the additional isolation and fragmentation event (MS3) scan can eliminate the ratio distortions. Ting et al., used a multi-proteome model (mixture of human and yeast proteins) in a 6-plex isobaric labelling system to demonstrate the interference effect and how this can be eliminated with the multistage MS3-based approach. Although, this approach reduced the contribution of any interfering signals, it also affected the overall sensitivity and data acquisition speed (Ting et al., 2011). This occurs because only the most intense fragment of a precursor is selected for subsequent interrogation per cycle, which results in only a small percentage of MS1 precursor ions to be converted into the MS3 reporter ions. Recently, McAlister et al. developed a method called MultiNotch MS3, that uses isolation waveforms with multiple frequency notches for synchronous precursor selection (SPS) of five or ten MS2 fragment ions, thereby increasing the number of reporter ions in the MS3 spectrum by 10-fold compared to the standard MS3 method (McAlister et al., 2014). The MultiNotch MS3 approach in combination with the development of advanced mass spectrometers, such as Orbitrap Fusion Lumos and Orbitrap Eclipse have addressed the interference issue and have now made the use of isobaric labels a powerful tool for the quantitative study of proteomes and interactomes (Myers et al., 2019; Yu et al., 2020).

Recently, an increased sample multiplexing capability was introduced with the synthesis of the next generation TMT reagents that are based on an isobutyl-proline immonium ion reporter structure (TMTpro) (**Figure 6b**) and enable the analysis of up to 16 samples simultaneously (Li et al., 2020a; Thompson et al., 2019) (**Figure 6c**). A comparison between the TMT (TMT 10-plex or TMT-11plex) reagents and the TMTpro label tags revealed the same functionality and similar proteome coverage and quantitative precision for both reagents (Li et al., 2020a). The TMTpro application can be more beneficial, when more replicates are required or sample amount is limited as the mixture of 16 samples within one experiment can increase significantly

the sensitivity of the analysis (Thompson et al., 2019). Additionally, with the application of TMTpro, the number of different TMT experiments can be reduced, as more samples or replicates can be included in a single TMTpro experiment. This will result in increased high throughput and improved data quality as the missing quantitative values among samples and within replicates will be reduced. Recently, a 27plex approach was introduced, where 11-plex and 16-plex TMT approaches were successfully combined (Wang et al., 2020). In this approach the peptides were labeled by the two different TMT sets; the two types of tags display different mass and hydrophobicity and could be separated in LC-MS/MS, offering high multiplexity capacity. Overall, multiplexed isobaric labelling methods can remarkably improve the throughput of quantitative mass spectrometry and can be applied widely to study different functional biological systems.

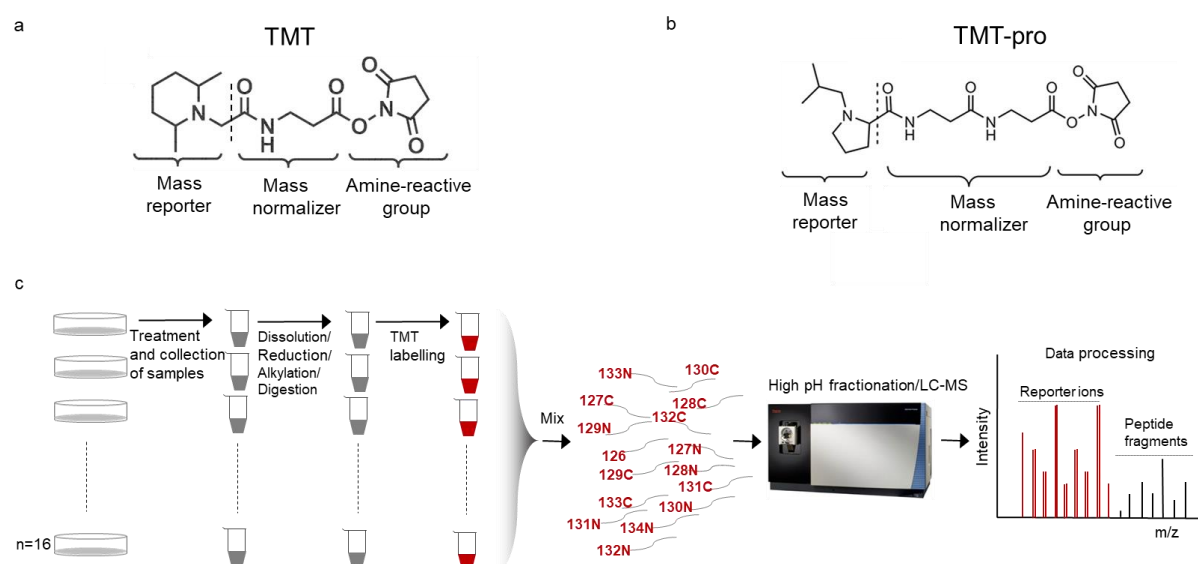


Figure 6. Chemical labelling.

a) Chemical formula of TMT-6plex reagents. b) Chemical formula of TMTpro tags. c) Peptides from up to sixteen different samples are labelled using the different TMTpro tags and are mixed and fractionated prior to mass spectrometry analysis. At the MS2 or MS3 level the fragmentation will generate the reporter ion peaks and the peak intensity will give the relative quantification of this particular peptide across the different conditions.

1.7 Thesis objectives

The first aim of the thesis is the development and optimisation of a quantitative method for the study of the dynamics of protein complexes. For this purpose, we will combine the well-established RIME approach with the TMT isobaric labelling technique, which provides a

powerful tool for global protein quantification with increased multiplexity and sensitivity. Our new approach will be compared with non-quantitative RIME and SILAC-RIME that have been previously applied for the characterisation of different protein interactomes. The second aim of the thesis is the application of the new quantitative method to characterise the interactome of ER α and well known ER α -coactivators using breast cancer cell lines as a model. In parallel, we aim to validate a selection of novel interactions, that will be found in the cellular models, in PDX and human clinical tumour samples and evaluate the application of our quantitative pipeline *in vivo*. Novel interactions will be validated using the proximity ligation assay (PLA) approach. The third aim of the thesis is the study of the effects of 4-hydroxytamoxifen, the active metabolite of tamoxifen, on the ER α interactome. To this end we will perform a time course experiment in a breast cancer cell line to explore the temporal effects of 4-hydroxytamoxifen on the composition of the ER α complex. We will combine the quantitative interactome assays with RNA-sequencing to validate the efficiency of the drug treatment as well as with whole proteome analysis to discriminate total protein changes induced by 4-hydroxytamoxifen treatment from changes specific for the ER α complex assembly. The application of our quantitative pipeline will enable the functional study of a drug that is widely used for the treatment of ER α positive pre-menopausal breast cancer patients, in an unbiased manner, without exclusively focusing on the effect of the assembly of ER α complex on specific genes. This can be an excellent tool to study ER α transcriptional repression and validate the application of our new developed quantitative method. The fourth aim of the study is the integration of all the quantitative interactome data, performed *in vitro* and *in vivo*, for the discovery of novel ER α -associated proteins with clinical potential. The functional role of the novel ER α -associated interactions will be characterised in depth by integrating different technologies. For this, the interactome of the novel candidate will be assessed, using our new-developed quantitative pipeline to identify its partners and understand better its association with ER α . The functional role of this candidate interactor will be further assessed by combining siRNA with proliferation assays mainly using different breast cancer cell lines as models. We will also perform RNA-sequencing and whole proteome analysis to study the global effect in gene and protein expression following the knockdown of this factor. The expression data will be integrated with ChIP-sequencing experiments for profiling the genome binding of this factor in promoters or enhancers that are close to genes that are regulated upon knockdown. ChIP-sequencing experiments will also be performed for other important factors such as ER α and known ER α interactors to evaluate the effect on their chromatin binding upon knockdown of the candidate factor. Building on the acquired knowledge from the above objectives, the

ultimate goal of the project is to investigate the dynamics of the ER α complex composition in cell models and clinical samples using proteomic approaches and to discover novel ER α -associated proteins that may affect ER α function and its transcriptional activity.

Chapter 2

2. Materials and Methods

2.1 Cell lines and cell treatments

ER α -positive breast cancer cell lines MCF7, T47D and ZR751, the Human embryonic kidney 293 cell line (HEK293) and the AR-positive prostate cancer cell line LNCaP, were obtained from ATCC. MCF7 and HEK293 cells were grown in DMEM medium (Gibco, #41966-029), T47D, LNCaP, MDA-MB-231 and ZR751 cells in RPMI-1640 medium (Gibco, #21875-034). Both media were supplemented with 10% foetal bovine serum (Gibco, #A3160402), 50U/ml penicillin, 50 μ g/ml streptomycin (Gibco, #15070-063) and 2mM L-glutamine (Gibco, #25030). All cell lines were genotyped by short-tandem repeat (STR) genetic profiling using the PowerPlex 16HS Cell Line panel and analysed using the Applied Biosystems Gene Mapper ID v3.2.1 software by the external provider Genetica DNA Laboratories (LabCorp Specialty Testing Group). Cell cultures were tested routinely for mycoplasma contamination using the MycoProbe Mycoplasma detection kit (R&D Systems, #CUL001B) (Both were performed by the Research Instrumentation and Cell Services Core-CRUK-CI). For the cell treatments, 4-Hydroxytamoxifen (Sigma-Aldrich, #HG278) or Fulvestrant (Selleckchem, #S1191) were prepared in ethanol and used at final concentration 100nM.

2.1.1 Whole cell lysate preparation and western blot analysis

Cells reaching about 70-80% confluency were collected and washed twice in ice-cold Phosphate Buffered Saline (PBS) and scraped in 500-1ml PBS contained protease inhibitors (Roche, #505648900). Cells were centrifuged at 8000g for 3min at 4°C and the supernatant was discarded. Cell pellets were reconstituted in 50-200 μ l (based on the pellet size) RIPA buffer (Thermo Scientific, #89901), that was supplemented with protease inhibitors (Roche, #505648900), followed by sonication using the Bioruptor Plus (Diagenode) for 2 cycles (30 seconds on/30 seconds off) to degrade the DNA. The cellular debris was removed by centrifugation at 20,000rpm for 10min, at 4°C and supernatant was transferred to new tubes. A protein quantification assay was performed using the Direct Detect™ infrared (IR)-based quantification system (Millipore) to calculate protein concentration of the whole lysate. 25-40 μ g from each sample was loaded on the gel (Invitrogen Bolt™ 4-12%, #NW0412BOX) and the Precision Plus Protein™ Dual Colour molecular weight marker (BIO-RAD, #1610374) was

used for the determination of protein sizes. The proteins were transferred onto a nitrocellulose membrane (Invitrogen, #IB23001/02) using the iBlot 2 Dry Blotting System (Invitrogen), followed by one hour blocking at room temperature using Odyssey® Blocking Buffer (Li-Cor, #927-60001). The membrane was immunoblotted with ER α antibody (Novocastra, #6045332, 1:100) or ZNF207 antibody (Invitrogen, #PA5-30641, 1:1000) and rabbit (Cell signalling, #4970, 1:1000) or mouse beta-actin (Sigma-Aldrich, #A5441, 1:5000). Detection of the ER α and ZNF207 proteins was achieved using the IRDye 800CW Goat Anti-mouse (Li-Cor, #926-32210) diluted to 1:5000 or the IRDye 800CW Goat Anti-rabbit (Li-Cor, #926-32211) diluted to 1:5000 retrospectively, while the loading control was detected using the IRDye 680RD Goat Anti-rabbit (Li-Cor, #926-68071,) diluted to 1:15000 or the IRDye 680RD Goat Anti-mouse (Li-Cor, #926-68070) diluted to 1:15000. All antibodies were diluted in Odyssey Buffer contained 0.1% Tween. The proteins were visualised using the Odyssey CLx Imaging System (Li-Cor) and images were taken with the automated capture option of the Image studio Version 4.0 software.

2.1.2 Small interfering RNA (siRNA) assay

Cells were seeded in 100mm or 150mm plates and allowed to grow in complete media for 24h. The next day, cells were transfected with siNT (Horizon Discovery, #D-001810-10) or siZNF207 SMARTpool (Horizon Discovery, #L-019557-00) using Lipofectamine RNAiMax transfection reagent (Invitrogen, #13778150) diluted in Opti-MEM (Gibco, #31985047). The siRNAs were prepared in nuclease free water and aliquots of final concentration 20 μ M were stored at -80°C. For cell transfection, siRNAs were diluted in Opti-MEM and incubated for 5min, before added to the transfection reagent/Opti-MEM mix and incubated for 20min at room temperature. The mix was added to the plates at a final concentration of 10nM and the cells were incubated with the respective siRNA for 48h. For DNA, flow cytometry, protein and RNA experiments, the media was changed 24h after the transfection. For proliferation assays, cells were left in transfection medium for the duration of the assay.

2.1.3 Cell proliferation assays

Cell proliferation assays were performed using the IncuCyte® ZOOM Live Cell Analysis System (Essen BioScience). Different cell lines were seeded in 48 or 96-well plates and after 24h were transfected with siNT or siZNF207 in at least six technical replicates per condition. For each cell line the assay was conducted in three biological replicates, performing three

independent cultures and transfection treatments. Following the addition of the transfection mix to each well, plates were immediately placed in the IncuCyte® ZOOM Live Cell Analysis System (37°C with 5% CO₂) and cell proliferation was monitored for 4-5 days via phase-contrast images taken every 3 hours. Confluence was assessed using the default settings of the IncuCyte® ZOOM software.

2.2 RNA-sequencing analysis

MCF7 cells were washed twice with cold PBS and harvested in 350µl of lysis buffer (RLT, Qiagen #74106). For the RNA extraction the RNeasy® kit (Qiagen, #74106) was used according to the manufacturer's instructions and for the quantification of the extracted RNA the NanoDrop® ND-1000 Spectrophotometer (Thermo Scientific) was used. For the library preparation the TruSeq Stranded mRNA Library Prep Kit High Throughput (Illumina) was used according to the manufacturer's instructions and 2 lanes of 50 bp single-end reads were run on HiSeq 4000 to reach around 30M reads per sample (Library preparation and sequencing was conducted by the Genomics Core Facility-CRUK-CI). RNA-seq data processing and bioinformatic analysis was performed by the CRUK Cambridge Institute Bioinformatics core. Reads were aligned to the human genome version GRCh37.75 or GRCh38 using the aligner TopHat v2.1.0 (Trapnell et al., 2009) or STAR v2.5.3a (Dobin et al., 2013). Read counts were obtained using feature Counts function in Subread v1.5.2 (Liao et al., 2013) and read counts were normalized and tested for differential gene expression using the DESeq2 workflow (Love et al., 2014). Multiple testing correction was applied using the Benjamini–Hochberg method. Sashimi plots were generated from aligned bams in the Integrative Genomics Viewer (IGV) version 2.8.2 (<http://software.broadinstitute.org/software/igv/>). The minimum junction coverage was set to the IGV default, which is 1, when Sashimi plots were generated.

2.3 Proximity ligation assay

Ice-cold methanol (-20°C) was added (3min) for cell fixation and permeabilization, followed by three washing steps with cold PBS. PLA was carried out according to manufacturer's instructions (Sigma-Aldrich, #DUO92007) and the following primary antibodies were used: ERα (Santa Cruz, sc-543 or sc-8002, 1:250), HP1γ (Santa Cruz, sc-365085, 1:400), NIPBL (Santa Cruz, sc-374625, 1:200), FOXK1 (Santa Cruz, sc-373810, 1:200), GFP (Abcam, ab1218, 1:200), NCOA3 (Bethyl Laboratories, A300-347A, 1:200), CBP (Bethyl Laboratories, A300-363A, 1:200), BAF170 (Santa Cruz, sc-17838, 1:200), HDAC1 (Santa Cruz, sc-81598,

1:200). The incubation with the primary antibodies and the secondary proximity probes (Sigma-Aldrich, Rabbit-PLUS, #DUO92002 and Mouse-MINUS, #DUO92004) was performed for 1h at 37°C. For the single PLA recognition experiment two ER α antibodies (Santa Cruz, sc-543, 1:800 and Invitrogen, MA5-13191, 1:1200) were used in combination. The Leica DFC340FX microscope was used for visualisation and images were captured at high resolution for a total of 8 separate observation fields with Leica Imaging software. Cell numeration and PLA labelling were performed using Image J software and cells along with red PLA dots were counted using the 'Analyse Particles' function. For each condition at least 200 cells were imaged and analysed and the average value of number of spots per nucleus was calculated. For the statistical analysis student's t-test was performed.

2.4 Immunofluorescence

Cells were fixed and permeabilised using ice-cold (-20°C) methanol for 3min followed by blocking with PBS-5% (w/v) Bovine Serum Albumin (BSA) (Sigma-Aldrich, #A9418) for 30min at room temperature. The primary ER α antibody (Santa Cruz, sc-543, 1:250) was prepared in blocking solution (PBS-5 % (w/v) BSA) and added on coverslips for an incubation of 1h at 37 °C. The coverslips were washed 4 times in washing buffer (PBS-0.5% Tween) followed by incubation with the secondary antibody conjugated to Alexa Fluor 488 (Invitrogen, #A-21206, 1:500) for 1 hour at 37°C in the dark. After the completion of the incubation with the secondary antibody, coverslips were washed three times in washing buffer and once in PBS. For the α -Tubulin and CREST immunofluorescence staining, cells were fixed with 4% paraformaldehyde at room temperature for 10min, permeabilized with 0.5% Tween in PBS for 10min and blocked for 10min at room temperature (RT) in blocking solution (PBS-5 % (w/v) BSA). The primary CREST (Antibodies Incorporated, #15-234, 1:2000) and α -Tubulin antibodies (Sigma-Aldrich, #T9026, 1:500) were diluted in blocking solution (PBS-5 % (w/v) BSA) and incubated on coverslips for 1h at RT. The coverslips were washed 3 times in washing buffer (PBS-0.2% Tween) followed by incubation with the secondary antibodies conjugated to Alexa Fluor 488 or 555 (Invitrogen #A-21206 and #A-31570, 1:500) for 1h at RT in the dark. Coverslips were then washed again three times in washing buffer and Hoechst 33342 (Sigma-Aldrich, #B2261, 1:500 in PBS) was used for DNA staining. For cell scoring the Zeiss widefield system was used at 63x magnification (oil immersion objectives).

2.5 PDX propagation and tissue collection

PDX propagation and tissue collection was conducted by our collaborators in Australia (A/Professor Alex Swarbrick and A/Professor Elgene Lim labs). Viably frozen PDX tumour tissue was propagated in immune-compromised mice. Briefly, 1mm³ tumour pieces were implanted into the 4th mammary pad of NSG mice. All mice were supplemented with estrogen, using E2 pellets (made in-house) inserted into the dorsal scruff. A standard monitoring and measurements of tumour size were performed twice a week. When tumours reached appropriate size (~1000mm³), mice were sacrificed by cervical dislocation under deep, isoflurane induced anaesthesia. Tumours were resected, diced and processed by either snap freezing in liquid nitrogen, fixing in 10% neutral buffered formalin solution for subsequent paraffin embedding, embedding in optimal cutting temperature compound (OCT), or viably freezing in foetal calf serum (FCS) supplemented with 5% DMSO. This method section was provided by our collaborators in Australia (A/Professor Alex Swarbrick and A/Professor Elgene Lim labs).

2.5.1 Sample preparation of clinical tumour material

Human clinical tumours were collected from our collaborators in Australia (Professor Alex Swarbrick and Professor Elgene Lim labs) and frozen tissues were cryosectioned in 30µm slices by the Histopathology Core (CRUK-CI) using the Leica CM 3050 S cryostat. Tissue sections were fixed in a two-step procedure by adding 2mM DSG for 25min first. In the same suspension of tissue sections, 1% formaldehyde was added for another 20min without removing the DSG, followed by crosslinking quenching using glycine at 0.25M final concentration. Then samples were centrifuged for 3min at 2,500g, the supernatant was discarded and tissue pellets were washed twice with cold PBS. For pellet resuspension, 6ml LB3 buffer (10 mM Tris-HCl [pH 8], 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, and 0.5% N-lauroylsarcosine) was added to each tumour, followed by tip sonication for 6min. Each extracted chromatin amount was split for ERα and IgG RIME pull-down assays and added to the bead-bound antibody for overnight incubation at 4°C. The next day, the beads were washed 10 times with ice-cold RIPA buffer (150mM NaCl, 10mM Tris, pH 7.2, 0.1% SDS, 1% Triton X-100, 1% NaDeoxycholate) and twice with ice-cold (100nM) ammonium bicarbonate (AMBIC) prior to mass spectrometry analysis. Patient, and patient derived tissues used in this work were collected under protocol X13-0133, HREC/13/RPAH/187. HREC approval was obtained through the SLHD (Sydney Local Health

District) Ethics Committee [RPAH (Royal Prince Alfred Hospital) zone)] and site-specific approvals were obtained for all additional sites. Written consent was obtained from all patients prior to collection of tissue and clinical data stored in a de-identified manner, following pre-approved protocols. All animal procedures were carried out in accordance to the relevant national and international guidelines and animal protocols approved by the Garvan/St Vincent's Animal Ethics Committee (Animal ethics number 15/10).

2.6 Immunohistochemistry

FFPE blocks from PDX tumours were sectioned at 4µm onto Superfrost Plus slides. Immunohistochemistry was carried out by our collaborators in Australia (A/Professor Alex Swarbrick and A/Professor Elgene Lim labs) using the DAKO Bond Autostainer. Sections underwent dewaxing, heat induced antigen retrieval (DAKO reagent ER2, 30mins) and primary and secondary antibody incubations, using ERα antibody (Abcam, # ab108398, 1:500) and the EnVision+ Rabbit secondary system, respectively. Sections were counterstained with haematoxylin. This method section was provided by our collaborators in Australia (A/Professor Alex Swarbrick and Professor A/Elgene Lim labs).

2.7 Flow cytometry

MCF7 cells were grown in complete media for 24h before they were treated with siNT or siZNF207 for 48h at final concentration 10nM. Asynchronous MCF7 cells were fixed in 70% ethanol for 30min on ice and washed twice with ice-cold PBS. Each sample was resuspended in 1ml Flow stain solution [PBS+0.1%(v/v) Triton-X-100, Propidium Iodide solution (20µg/ml, Sigma-Aldrich, #P4864), Ribonuclease A (200µg/ml, Sigma-Aldrich, #R6513)] and the cell suspension was transferred to 5ml round bottom polystyrene tubes (Corning, #352235). Flow cytometry was performed using a Becton Dickson flow cytometer and analysed using the FlowJo® software package (Tree Star). The experiment was repeated twice and all statistical analyses were carried out by performing student's t-test.

2.8 ChIP-Seq and RIME assays

Cells (2×10^6) were grown in complete media and for the cell crosslinking 1% formaldehyde (Thermo, #28908) prepared in PBS was added, followed by an incubation for 10min at RT. For the double crosslinking cells were incubated in PBS containing 2mM DSG (disuccinimidyl glutarate-Santa Cruz Biotechnology, #sc-285455A) for 20min at RT followed by incubation in

1% formaldehyde for 10min at RT. The cell crosslinking was quenched by adding glycine to a final concentration of 0.1M and the cells were washed twice in ice-cold PBS and scraped in 1ml PBS contained protease (Roche, #5056489001) and phosphatase inhibitors (Thermo Scientific, #78427). Cells were centrifuged at 8000g for 3min at 4°C and the supernatant was discarded. For both ChIP-seq and RIME experiments, cell pellets were resuspended in Lysis Buffer 1 (50mM Hepes–KOH, pH 7.5, 140mM NaCl, 1mM EDTA, 10% Glycerol, 0.5% NP-40/Igepal CA-630, 0.25% Triton X-100) and rotated for 10min, at 4°C for nuclear extraction. Then nuclei were pelleted, resuspended in Lysis buffer 2 (10mM Tris–HCL, pH8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA) and rotated at 4°C for 5min. Samples were resuspended in 300µl Lysis buffer 3 (10mM Tris–HCl, pH 8, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na–Deoxycholate) and sonicated using the Bioruptor Plus (Diagenode) for 10-20 cycles (30 seconds on, 30 seconds off). After sonication the samples were centrifuged at 20,000g for 10 minutes at 4°C and a small aliquot of the supernatant was kept as input for ChIP-seq. For each sample, 50-100ul of Dynabeads™ Protein A (Invitrogen, #0002D) and 5-10ug of specific antibody were used. **Table 1** lists all the antibodies used for RIME and ChIP-seq assays. The bead-bound antibody and chromatin were incubated overnight at 4°C with rotation. For the spike-in ChIP-seq experiment, 5ug drosophila antibody (Active motif, #61686) and 50ng drosophila chromatin (Active motif, #53083) were added to each sample prior to overnight incubation. The next day, the beads for RIME assays were washed 10 times with ice-cold RIPA buffer (150mM NaCl, 10mM Tris, pH 7.2, 0.1% SDS, 1% Triton X-100, 1% NaDeoxycholate) and twice with 100mM ammonium bicarbonate (AMBIC, Fisher Scientific, #10207183) prior to mass spectrometry analysis.

The beads for ChIP-seq were washed six times with RIPA buffer, followed by one wash with TE (Tris EDTA, pH 7.4). Then, both ChIP samples and inputs were de-crosslinked by adding 200µl elution buffer (1% SDS, 0.1 M NaHCO₃), followed by overnight incubation at 65°C. After reversing crosslinking, DNA was purified using the phenol-chloroform-isoamyl DNA extraction method. Briefly, samples were treated with RNase A (Thermo Scientific, #AM2271) for 45min at 37°C, followed by Proteinase K treatment (Thermo Scientific, #25530049) for 2 hours at 55°C, before DNA was purified by phenol-chloroform extraction. A volume of 800µl 100% ethanol was added and samples were left overnight at -20°C for DNA precipitation. The next day, pellets were washed with 70% ethanol and reconstituted in 13µl Tris-HCl. ChIP-seq and input libraries were prepared using the ThruPlex Sample Prep Kit (Illumina, #R400676). The sample pool was subjected to next generation sequencing using HiSeq 4000 or Novaseq

(Illumina) to reach approximately 30M reads per sample (The sequencing was performed by the Genomics core-CRUK-CI).

Table 1. Antibodies used for RIME and ChIP-seq assays.

Bait protein	Antibody	Application
ZNF207	Santa Cruz sc-271942	ChIP, RIME
Pol II (phospho S5)	Abcam ab5131	ChIP, RIME
ER α	Abcam ab3575 Millipore 06-935	ChIP, RIME
MED1	Bethyl A300-793A	ChIP, RIME
MED12	Bethyl A300-774A	RIME
MED26	Cell signalling 14950S	RIME
MED14	Abcam ab72141	RIME
MED4	Santa Cruz sc-398179	RIME
MED15	Abcam ab176593	RIME
SMC1A	Bethyl A300-055A	ChIP
TAF3	Abcam ab188332	ChIP
FOXA1	Abcam ab5089	ChIP
CBP	Diagenode, C15410224	RIME
AR	Millipore 06-680	RIME
NCOA3	Bethyl, A300-347A	RIME
IgG rabbit	Abcam ab171870	RIME
IgG mouse	Santa Cruz sc-2025	RIME

2.8.1 ChIP-seq data analysis

The ChIP-seq materials and methods section has been provided from the CRUK-CI Bioinformatics core (Dr Ashley Sawle) and Dr Igor Chernukhin. Raw paired end reads in fastq files were aligned against the hg38 reference genome with the bwa-mem algorithm in bwa version 0.7.17 (Li and Durbin, 2009) by applying the default parameters. Read quality was assessed using FASTQC; the base calling qualities and the contamination with adapters was checked. The alignment rates should exceed 95% alignment and the duplication rates for ChIP samples should not exceed 20%. For inputs this should not be greater than 5%. Prior to peak calling, reads were filtered in four ways: (a) Unaligned reads were removed (b) Only reads aligned to canonical chromosomes were considered (c) Reads lying in regions of anomalous

enrichment in the inputs were filtered out using the R package GreyListChIP (DOI: 10.18129/B9.bioc.GreyListChIP) (d) Reads with a Mapping Quality of less than 15 were filtered out. Replicate input samples from each condition (siCtrl or siZNF207) were merged and separate inputs were made for each ChIP sample by down-sampling the merged input file to the same depth as the ChIP sample. Peaks were called using MACS2 version 2.1.2 (Zhang et al., 2008) with the default parameters. Differential binding analysis was carried out using the Bioconductor package DiffBind version 2.12.0 (Ross-Innes et al., 2012) in R version 3.6 with the default parameters. DESeq2 version 1.24.0 (Love et al., 2014) was used to test statistical significance of differential binding in consensus regions. The Benjamini-Hochberg multiple testing correction was used to adjust p-values. Binding regions were called as statistically significant at $FDR < 0.05$. Binding profiles were generated using Deeptools version 3.3.0 (Ramirez et al., 2016). For the enrichment analysis, the ratio of the number of differentially bound regions for each factor in the promoters and enhancers of differentially expressed genes to the number of differentially bound regions in the promoters/enhancers of non-changing genes was calculated. To identify promoters or enhancers two methods were applied; (a) Proximity: Promoters: -1000/+250 bases of TSS; Enhancers: -5e4/+5e4 bases of TSS excluding the promoter regions (b) Promoter capture Hi-C: Promoters are the bait regions of the Hi-C loops and enhancers are the distal regions of the Hi-C loops. For the Hi-C approach, the RNAseq data were filtered for genes that Hi-C baits were detected and then the top 200 downregulated genes were selected at $FDR < 0.01$. In parallel, 200 constitutive (non-changing genes) were selected by filtering for genes with average expression in the same range as the average expression of the top 200 downregulated genes and then 200 genes with fold change closest to one were used for further analysis. For each of the repressed/constitutive genes the number of lost/gained/non-changing binding sites was counted in promoters and enhancers and for each of the lost/gained/non-changing sites the ratio of binding sites for repressed genes to constitutive genes was calculated.

For the spike-in experiment, a combined genome reference was generated by merging the hg38 human and the dm6 fruit fly genomes and the raw paired end reads in fastq files were aligned against the combined reference genome with the bwa-mem algorithm in bwa version 0.7.17 (Li and Durbin, 2009) using default parameters. Then, the resulting bam files were split using SAMtools version 1.9 (Li et al., 2009) to generate separate files for reads aligned to human sequences and fruit fly sequences. Peaks were called using MACS2 version 2.1.2 (Zhang et al., 2008). Differential binding analysis was carried out using the Bioconductor package DiffBind version 2.12.0 (Ross-Innes et al., 2012) in R version 3.6 with the default parameters and

normalisation of the human read counts was achieved by replacing the library sizes in the DiffBind object with the library sizes of the corresponding fruit fly bam. DiffBind calculated the scaling factors by dividing the library size for each sample by minimum library size. DESeq2 version 1.24.0 (Love et al., 2014) was used to test statistical significance of differential binding in consensus regions and the Benjamini-Hochberg multiple testing correction for adjusting p-values. Binding regions were called as statistically significant at $FDR < 0.05$. For the Upset plots, DiffBind was used to generate a consensus set of binding regions that were present in at least two replicates for each factor. Using the consensus set from all factors, an overall consensus set was created that was the union of all the regions. This gave the total number of regions in the Upset plot. Each of these regions was deemed to have been detected for a particular factor, if there was an overlapping region in that factor's specific consensus set. To map a structural relation between binding factors and active transcriptome we measured the Binding Site Density (BSD) relative to the genomic regions with detectable gene expression. Genes that had a meaningful expression level detected in the MCF7 RNA-seq analysis were selected as active genes. BSD was calculated as $sO/nG*corf$ where sO is the cumulative frequency, nG is the number of tested features (genes) and corf is the density mean-correction factor. The cumulative frequency of chip locations was calculated as a sum of ChIP-seq sites overlapping 1kb bin size within +/-500kb window of genomic region with TSS of tested genes as a midpoint. Denovo motif analysis was performed using Meme version 4.9.1 (Bailey et al., 2009) to detect known and discover novel binding motifs amongst tag-enriched sequences.

2.8.2 ChIP-qPCR

After the DNA purification, 1-2µl per sample was diluted (1:10 dilution) for ChIP-qPCR analysis. The reaction mixture contained Power SYBR® Green PCR Master Mix (Applied Biosystems, #4309155), forward and reverse primers (final concentration 10µM each), 2µl diluted DNA template and nuclease-free water to a final volume of 15µl. Each reaction was performed in triplicates and for the analysis the BioRad CFX Connect Real Time System was used (10min at 95°C; 45 cycles of 15 seconds at 95°C and 30 seconds at 60°C). The sequences of primers used for ChIP-qPCR are shown in **Table 2**. Relative enrichment was determined as percentage input or ER3.

Table 2. Primer sequences for ChIP-qPCR at ER α binding sites.

Target	Forward primer	Reverse primer
XBP1	ATACTTGGCAGCCTGTGACC	GGTCCACAAAGCAGGAAAAA
GREB1	GAAGGGCAGAGCTGATAACG	GACCCAGTTGCCACACTTTT
RARA	GCTGGGTCCTCTGGCTGTTC	CCGGGATAAAGCCACTCCAA
MYC	GCTCTGGGCACACACATTGG	GGCTCACCTTGCTGATGCT
ESR1	GAAACAGCCCCAAATCTCAA	TTGTAGCCAGCAAGCAAATG
CA12	GGAGGCGTAACCCCTGTGTG	ACGGCAAGGGACTTGCTGAC
ER3 Control	GCCACCAGCCTGCTTCTGT	CGTGGATGGGTCCGAGAAAC

2.9 Protein digestion and TMT labelling

Digestion of bead-bound proteins was done by adding 10 μ L trypsin solution (15ng/ μ L) (Pierce) in 100mM AMBIC, followed by overnight incubation at 37°C. A second digestion step was performed the next day for 4h at 37°C and the supernatant solution was collected by placing the tubes on a magnet. The resultant peptides were acidified by the addition of 3 μ L 5% formic acid and purified using the Ultra-Micro C18 Spin Columns (Harvard Apparatus) according to manufacturer's instructions. For qPLEX-RIME, the peptide samples were reconstituted in 100 μ L 0.1M triethylammonium bicarbonate (TEAB) and labelled with the TMT reagents (Thermo Fisher). The peptide mixture was fractionated with Reversed-Phase cartridges at high pH (Pierce, #84868). Nine fractions were collected using different elution solutions in a range from 5% ACN to 50% ACN. For the whole proteome analysis, 200 μ L of 0.1 M TEAB, 0.1% SDS buffer was added to each cell pellet followed by probe sonication and boiling at 95°C. To estimate protein concentration, Bradford assay (BIO-RAD-Quick start) was performed according to manufacturer's instructions. For protein reduction, 2 μ L 50mM tris-2-carboxymethyl phosphine (TCEP, Sigma-Aldrich, #C4706) was added to each sample for 1h incubation at 60°C, followed by cysteine blocking for 10min at room temperature with the addition of 1 μ L 200mM methyl methanethiosulfonate (MMTS, Sigma-Aldrich, #64306). For proteolysis, trypsin (Pierce, #90058) solution was added at ratio protein/trypsin ~30:1 for overnight digestion at 37°C. The next day peptides were labelled with the TMT10plex reagents (Thermo Scientific) and the reaction was quenched with 8 μ L of 5% hydroxylamine (Thermo

Scientific, #90115). Following the completion of the chemical reaction with the TMT tags, the samples were mixed and dried with speedvac concentrator. The TMT mix was fractionated into 30 fractions on a Dionex Ultimate 3000 system at high pH using the X-Bridge C18 column (3.5 μ m 2.1x150mm, Waters). Fractions were dried with speedvac and stored at -80°C prior to mass spectrometry analysis.

2.9.1 Fe-NTA Phosphopeptide enrichment

After the completion of the high pH reversed phase fractionation, twelve fractions were collected and subjected to phosphoenrichment using the Fe-NTA phosphopeptide enrichment kit (Thermo Scientific, #A32992) according to manufacturer's instructions. Briefly, the peptide fractions were loaded to the spin columns in a low pH buffer followed by incubation with the resin material for 30min. At the end of the incubation the columns washed three times to remove non-phospho peptides that have been bound to the resin. After the washing step, the phosphopeptides were eluted from the column with a high pH buffer and collected to microcentrifuge tubes. The eluents were dried with speedvac concentrator and reconstituted in 10ul loading solution (2 % acetonitrile, 0.1 % formic acid) for LC-MS analysis.

2.9.2 LC-MS analysis

For the analysis of the peptide fractions, the Dionex Ultimate 3000 UHPLC system coupled with the nano-ESI Fusion Lumos or Q-Exactive or Q-Exactive HF mass spectrometer (Thermo Scientific) was used. Each peptide fraction was reconstituted in solution contained 2% acetonitrile/0.1% formic acid and was loaded on the Acclaim PepMap 100, 100 μ m \times 2cm C18, 5 μ m, 100 Å trapping column using the ulPickUp injection method at 5 μ L/min flow rate for 10min. For the peptide separation the EASY-Spray analytical column 75 μ m \times 25cm, C18, 2 μ m, 100 Å column was used for multi-step gradient elution. Mobile phase (A) was 2% acetonitrile, 0.1% formic acid and mobile phase (B) was 80% acetonitrile, 0.1% formic acid. The gradient elution method at flow rate 300nL/min was as follows: for 95min gradient up to 45% (B), for 5min gradient up to 95% (B), for 8min isocratic 95% (B), for 2min down to 5% (B), for 10min isocratic equilibration 5% (B) at 40°C. For the clinical sample analysis, a longer gradient separation was used as follows: for 160min gradient up to 40% (B), for 10min gradient up to 95% (B), for 8min isocratic 95% (B), for 2min down to 5% (B), for 10min isocratic equilibration 5% (B). The Lumos was operated in a data-dependent mode for both MS2 and SPS-MS3 methods. The full scans were performed in the Orbitrap in the range of 380-1500m/z

at 120K resolution. For MS2, peptides were selected in the quadrupole with MS isolation window 0.7Th and the MS2 scans were performed in the ion trap with collision energy 35%. The top 10 most intense fragments were selected for Synchronous Precursor Selection (SPS) HCD-MS3 analysis with MS2 isolation window 2.0Th. The HCD collision energy was set at 55% and the detection was performed in the Orbitrap in scan range 100-400 m/z, at 60K resolution. For the analysis of phospho-fractions, an MS2 approach was used on the Fusion and peptides were selected for fragmentation in the Orbitrap with MS isolation window 1.2Th and collision energy 38% at 50K resolution. For the Q-Exactive or Q-Exactive HF analysis, the full scan was performed in the Orbitrap in the range of 400-1600m/z at 60K and 70K resolution respectively. For MS2, the ten most intense precursors were selected at 17,5K and 30K resolution respectively. A 2.0Th isolation window was used and the HCD collision energy was 28%.

2.9.3 Data processing of mass spectrometry raw data

The collected HCD or CID tandem mass spectra were processed with the SequestHT search engine on the Proteome Discoverer 1.4 or 2.1 software. All spectra were searched against a UniProtKB/Swiss-Prot fasta file containing 20,365 reviewed human entries. The node for SequestHT included the following parameters: Precursor Mass Tolerance 20ppm, Fragment Mass Tolerance 0.5Da for CID spectra or 0.02Da for HCD spectra, Dynamic Modifications were Oxidation of Methionine (M) (+15.995 Da), Deamidation of Asparagine (N) and Glutamine (Q) (+0.984 Da) and Static Modifications were TMT6plex at any N-Terminus and Lysine (K) (+229.163 Da) for the quantitative data. Methylthio at Cysteine (C) (+45.988) was included for the total proteome and phosphoproteome data. Also, for the phosphorylation data phosphorylation at S, T, Y (+79.96) was added as dynamic modification and for the confidence of localization of phosphorylation sites the IMP-ptmRS mode was used. The Reporter Ion Quantifier node included a custom TMT 6plex (Thermo Scientific Instruments) Quantification Method, for MS3 scan events, HCD activation type, integration window tolerance 20ppm and integration method Most Confident Centroid. The consensus workflow included S/N (Signal-to-noise ratio) calculation for TMT intensities as previously described by McAlister et al. (McAlister et al., 2014) and the level of confidence for peptide identifications was estimated using the Percolator node with decoy database search. Strict FDR was set at q-value<0.01. For the label-free quantification, the HCD tandem mass spectra were processed with the SequestHT search engine on Proteome Discoverer 2.2 software. The node for SequestHT included the

following parameters: Precursor Mass Tolerance 20ppm, Maximum Missed Cleavages sites 2, Fragment Mass Tolerance 0.02Da and Dynamic Modifications were Oxidation of M (+15.995Da) and Deamidation of N, Q (+0.984Da). The Minora Feature Detector node was used for label-free quantification and the consensus workflow included the Feature Mapper and the Precursor Ion Quantifier nodes using intensity for the precursor quantification.

2.9.4 Analysis of proteomics data

The Bioinformatics core at the CRUK Cambridge Institute developed an R package (qPLEXanalyzer) to perform downstream analysis of quantitative proteomics data. The analysis for the different datasets was conducted using only unique peptides identified with high confidence (peptide FDR<1%). Firstly, various statistics methods were used to check the quality of the different datasets and any outlier samples were excluded at this step from downstream analysis. For qPLEX-RIME experiments the efficiency of the pull-down was evaluated by generating a plot with the peptide sequence of the bait protein based on high confidence identified peptides (FDR<1%). Then, peptide-level S/N TMT values were corrected for equal loading across samples using different normalisation approaches depending on the experiment type (median scaling or median scaling within the group). The normalised peptides intensities were aggregated by calculating the sum to protein intensities and any further analysis was conducted at the protein level. For the regression-based correction, unique peptides were aggregated and proteins detected in all the three TMT experiments were kept for analysis. The normalisation on the bait protein level was carried out at protein level using \log_2 row-mean scaled values. To filter-out non-specific proteins, a limma-based differential analysis was performed comparing ER α and IgG control samples. In the regression analysis, the ER α profile was used as the independent variable (x) and the profile of any other protein as the dependent variable (y) excluding the IgG controls. The residuals of the $y=ax+b$ linear model represent the protein quantification profiles that are not driven by ER α amount in the pull-down. For the detection of differentially regulated or bound proteins a limma-based analysis was carried out and a multiple testing correction was applied on p-value using the Benjamini-Hochberg method to control the False Discovery Rate (FDR).

2.10 Promoter Capture Hi-C sample preparation

Three biological replicates from MCF7 cells treated with siNT or siZNF207 were crosslinked with 2% formaldehyde prepared in PBS and incubated for 10min at room temperature. The

crosslinking reaction was quenched by adding 1M glycine to a final concentration of 0.125M for 5min at room temperature, followed by 15min incubation on ice. Crosslinked cells were washed twice in cold PBS, the supernatant discarded and the pellets were flash-frozen in liquid nitrogen. Cells were resuspended in 50ml freshly prepared ice-cold lysis buffer (10mM Tris-HCl pH 8, 10mM NaCl, 0.2% Igepal CA-630) contained protease inhibitors, followed by incubation for 30min on ice with occasional mixing. Following lysis, the chromatin was pelleted and washed with 1.25x NEB Buffer 2. Then, samples were resuspended in 1.25x NEB Buffer 2 and aliquots of $5-6 \times 10^6$ cells per sample were prepared. For the digestion, the HindIII (NEB, #R0104M) restriction enzyme was added to each aliquot and incubated overnight at 37°C. The next day, digested chromatin ends were filled with biotin-14-dATP (Life Technologies, #19524-016) in a Klenow end-filling reaction (NEB, #M0210L). For the ligation, T4 DNA ligase (Invitrogen, 15224-025) was added to each sample followed by incubation for 4 hours at 16°C. DNA was de-crosslinked by adding proteinase K (Roche, 03115879001) and incubated overnight at 65°C. DNA was purified by performing phenol (Sigma-Aldrich, #P4557) and phenol pH 8.0: chloroform (Sigma-Aldrich, #P3803) extractions. To monitor the library integrity, 2ul and 6ul aliquots of diluted Hi-C libraries (1:10 dilution in TE buffer) were loaded and run on an 2% E-gel (Invitrogen, #G402002) using a DNA ladder (Thermo Fisher, #1048809). The ligation efficiency was checked by conducting a PCR assay using the primers listed in the **Table 3**. An amount of 200ng of each library was purified using a PCR purification kit (Qiagen, #28106) and split into four samples; undigested, digested with HindIII-HF (NEB, #R3104S, High Fidelity), digested with NheI-HF and digested with both HindIII-HF and NheI-HF (NEB, #R3131S) enzymes. Digested samples for each primer pair were run on separate 2% E-gels. Biotin-14-dATP at non-ligated DNA ends was removed with T4 DNA polymerase (NEB, M0203L) and DNA was sheared to an average size of 400bp, using the manufacturer's instructions (Covaris E220). The sheared DNA was end repaired and fragments between 200–650 base pairs were selected using AMPure XP beads (Beckman Coulter, #A63881). Biotin-tagged DNA was pulled-down with streptavidin beads and ligated with paired-end adapter primers at final concentration 100μM. Eight cycles of PCR were performed to amplify libraries before capture (PE adapter 1: 5'-P-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3', PE adapter 2: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3, Integrated DNA Technologies). The quality and quantity of the libraries was evaluated by Bioanalyzer (Agilent) and quantitative PCR using the KAPA Library Quantification kit (Illumina, #KK4824). Hi-C libraries corresponding to 750ng were dried with a speedvac concentrator and re-suspended in 4μl water.

Hybridisation of Hi-C libraries was carried out using the SureSelectXT custom DNA bait library (Agilent, #5190-4831) and the SureSelect SSEL TE Reagent ILM PE full adaptor kit (Agilent, #931108). Hi-C libraries along with the hybridisation buffers and RNA bait was incubated for 24 hours at 65°C. Then, the SureSelect adaptor kit was used to isolate promoter fragment-containing ligation products with Streptavidin beads (Life technologies, #65601). Capture Hi-C library was amplified using 4 PCR cycles (1 cycle of 30sec at 98°C, 30sec at 65°C and 30sec at 72°C; 2 cycles of 10sec at 98°C, 30sec at 65°C and 30sec at 72°C; 1 cycle of 10sec at 98°C, 30sec at 65°C and 7min at 72°C) and purified twice using Ampure XP beads (Beckman Coulter, #A63881). Promoter capture Hi-C libraries were sequenced on the Novaseq platform to reach approximately one billion reads per sample.

Table 3. Primer sequences for quality control of human Hi-C libraries.

Primer	Sequence
hs AHF64 Dekker	GCATGCATTAGCCTCTGCTGTTCTCTGAAATC
hs AHF66 Dekker	CTGTCCAAGTACATTCCTGTTTACAAACCC
hs MYC -538	TGCCTGATGGATAGTGCTTTC
hs MYC +1820	AAAATGCCCATTTCCTTCTCC
hs HIST1 F	AAGCAGGAAAAGGCATAGCA
hs HIST1 R	TCTTGGGTTGTGGGACTTTC

2.10.1 PCHi-C data processing and interaction calling

The analysis of the PCHi-C data was performed by the CRUK-CI Bioinformatics core and this method section was provided by Kamal Kishore; Hi-C paired end raw sequencing reads were truncated through HiCUP v0.7.2 (Wingett et al., 2015). The truncated FASTQ files were mapped to the human reference genome (hg38) using bowtie (v1.2.1.1). The forward reads bam files were merged with reverse reads bam files (using only mapped reads from reverse reads bam file). After alignment, experimental artefacts and duplicate reads were removed. PCHi-C libraries contained three types of valid di-tags: i) promoter-promoter (reads between baits), ii) promoter-genome (reads between bait and non-promoter HindIII fragment) iii) genome-genome (reads where neither is the capture bait). We determined how many of the valid di-tags are between promoter-promoter and promoter-genome to calculate the capture efficiency. Di-tags consisting of only genome-genome reads were discarded prior to downstream analysis.

Significantly interacting regions were called using the CHiCAGO (Cairns et al., 2016) package. Interactions with a score ≥ 5 were considered to be statistically significant. To detect the significant differential interactions in Capture Hi-C data, the Chicdiff (Cairns et al., 2019) pipeline was used. The output files produced by the CHiCAGO pipeline, that contain interactions with CHiCAGO scores above a predefined cutoff (5) in at least one replicate, were used for the differential analysis. The analysis resulted in p-values for interactions adjusted based on the distance-dependent weights, followed by Benjamini-Hochberg (BH) multiple testing correction.

Chapter 3

3. Results

3.1 Development of a quantitative tool

3.1.1 The qPLEX-RIME method

To study the architecture and dynamics of chromatin-associated protein complexes, we have developed and optimised a quantitative pipeline which combines the previously described RIME method (Mohammed et al., 2013; Mohammed et al., 2016) with multiplex TMT chemical isobaric labelling (McAlister et al., 2012; Roumeliotis et al., 2017) (qPLEX-RIME). The development of our qPLEX-RIME quantitative pipeline along with various applications have been published (Papachristou et al., 2018). To demonstrate the feasibility of the method, we applied the developed workflow to study the Estrogen Receptor alpha interactome (ER α), a transcription factor involved in the development of breast cancer. The qPLEX-RIME workflow starts with a two-step fixation approach using disuccinimidyl glutarate (DSG) followed by formaldehyde (FA) that has been previously combined with ChIP assays to capture transient interactions with increased efficiency (Engelen et al., 2015; Nowak et al., 2005). Cross-linked protein complexes bound to the target protein are extracted from a nuclear-enriched lysate by immunoprecipitation, using a specific antibody, followed by proteolysis and peptide labelling with different TMT tags. For reduction of sample complexity and purification, peptides are fractionated at high pH on a reversed phase spin column prior to mass spectrometry analysis. Peptide fractions are collected at different compositions of the elution buffer and analysed on a tribrid Orbitrap Lumos mass spectrometer with the MultiNotch MS3 method as previously described by McAlister et. al. (McAlister et al., 2014). The main steps of the qPLEX-RIME method are shown in **Figure 7**.

The qPLEX-RIME method can be used for the discovery of proteins that are significantly and specifically associated with the bait protein directly or indirectly in a quantitative fashion. Negative control samples such as IgG pull-downs can be incorporated in the same experiment to discriminate unspecific binding based on the TMT signal intensity across specific (bait pull-downs) and non-specific (IgG-pull downs) conditions. The main utility of the qPLEX-RIME method is to identify changes in the composition of protein complexes in response to cell perturbation and/or in variable genomic backgrounds (e.g. different cell lines or mutated

conditions), using multiple technical and biological replicates in a single TMT experiment for enhanced statistical power. The multiplexity can increase the quantitative sensitivity and accuracy in capturing the interaction interplay between different factors. For the downstream data analysis, the Bioinformatics core of the CRUK Cambridge Institute has developed a comprehensive bioinformatics tool termed qPLEXanalyzer, that allows the user to perform data processing, visualisation, normalisation and statistical analysis steps (DOI:10.18129/B9.bioc.qPLEXanalyzer).

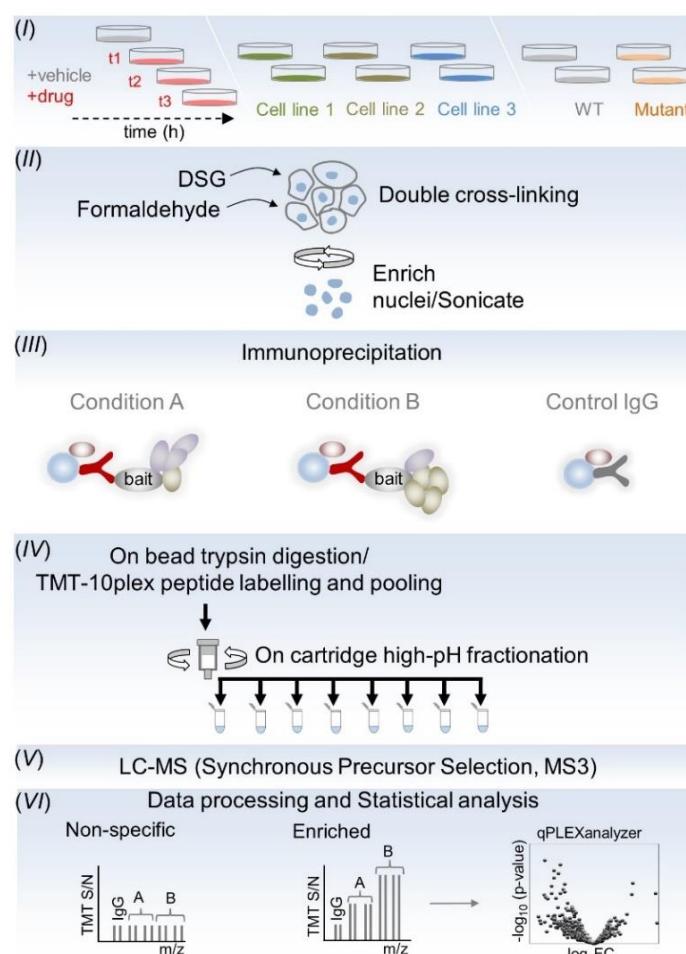


Figure 7. The qPLEX-RIME workflow.

Cells in different conditions or variable genomic backgrounds are double-crosslinked followed by nuclei isolation and sonication (I and II). Target protein complexes are immunoprecipitated and digested on the beads with trypsin (III, IV). The peptides are labelled using the TMT tags and the pooled sample is fractionated using Reversed-Phase spin columns at high pH (IV). Peptide fractions are analysed with the MultiNotch MS3 approach (V), followed by data processing and statistical analysis using the qPLEXanalyzer tool (VI).

3.1.2 Characterisation of the ER α interactome in MCF7 cells

Firstly, we applied qPLEX-RIME in asynchronous MCF7 breast cancer cells to assess whether we could successfully recover the ER α interactome. To this end, ER α qPLEX-RIME pull-downs were performed (N=5) in independently grown cells at different passage numbers. To discriminate non-specific binding, an equal number (N=5) of matched IgG control samples was included in the analysis. In this experiment, single crosslinking with formaldehyde was used, to permit a comparison with previously published data (Mohammed et al., 2013).

Data analysis with Proteome Discoverer 2.1 software identified 23,610 peptides (FDR<1%) of which 19,106 were quantified and matched to 2,955 proteins across the multiplexed set of all positive and negative samples. To examine the efficiency of the method in capturing and quantifying well-described ER α -associated proteins, we generated a list of known ER α interactors from BioGRID (Chatr-Aryamontri et al., 2017) (Biological General Repository for Interaction Datasets) and STRING (Szklarczyk et al., 2017) (Search Tool for the Retrieval of Interacting Genes/Proteins) databases. For BioGRID, we used only a subset of 386 proteins identified by high-throughput assays that were similar to the approach used here and for STRING we used only experimental associations (383 proteins, score>200). Noteworthy, only 37 proteins were common between the two reference subsets. The qPLEX-RIME method identified 295 (76%) and 171 (45%) of the known ER α -associated proteins from BioGRID and STRING respectively of which 225 (58%) and 154 (40%) showed positive enrichment at adj. p-value<0.1 (**Figure 8a**). Specifically, we found known co-regulators (e.g. P300, NCOA3, CBP, NRIP1, TRIM24, GREB1, RAR α , NCOR2 and HDACs), ER α -associated pioneer factors (e.g. FOXA1, AP-2 γ) and putative pioneer factors (e.g. GATA-3) significantly enriched in ER α samples compared to control samples (mean log₂Fold-Change of 2.4) (**Figure 8b**) (Anzick et al., 1997; Carroll et al., 2005; Jepsen et al., 2000; Liu and Bagchi, 2004; Mohammed et al., 2013; Shang et al., 2000; Tan et al., 2011; Theodorou et al., 2013; Tsai et al., 2010). ER α was one of the most significantly enriched proteins (adj. p-value<0.1, log₂Fold-Change=3.47) identified with 19 unique peptides (34.45% peptide coverage) (**Figure 8c**), consistent with previously published ER α RIME datasets (Mohammed et al., 2016). Taken together, our data demonstrate the ability of the qPLEX-RIME method to comprehensively characterise protein interactome networks.

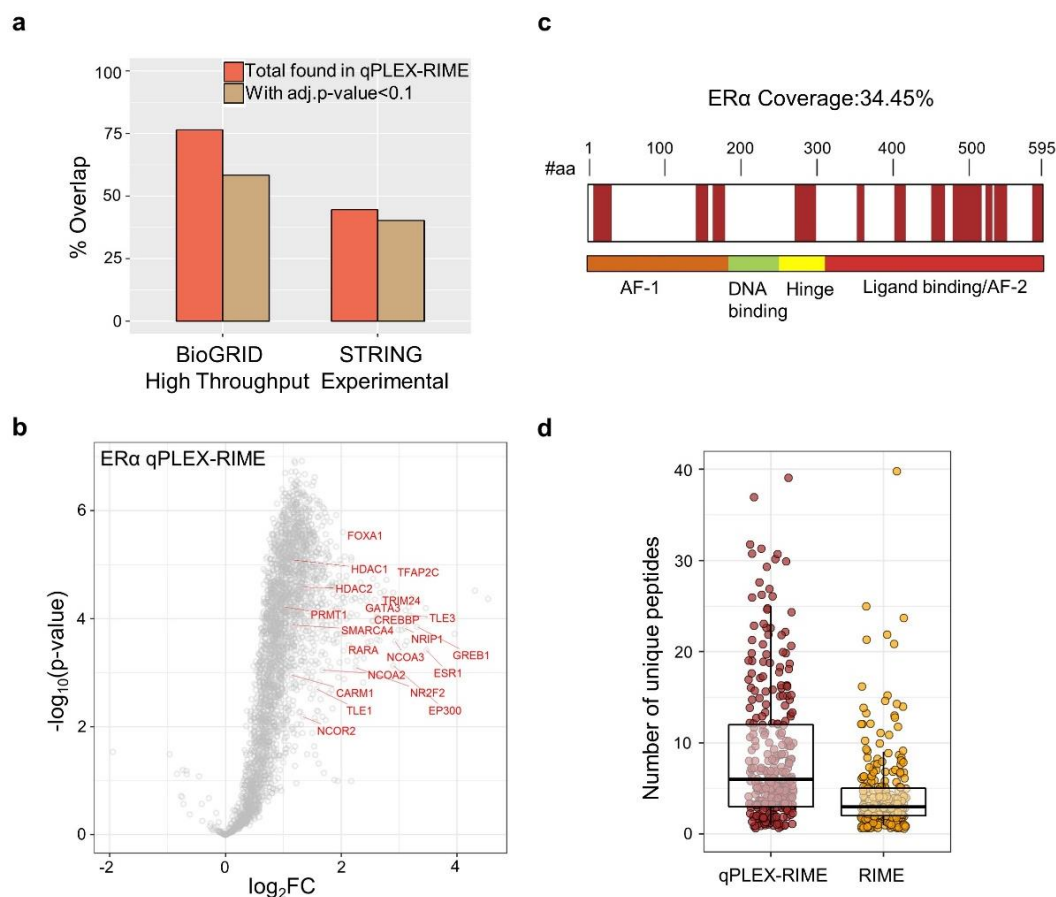


Figure 8. Application of qPLEX-RIME for the identification of ERα-associated proteins.

a) Bar plots illustrating the overlap of the qPLEX-RIME data with known ERα-associated proteins from BioGRID and STRING databases. b) Volcano plot summarising the quantitative results of the ERα qPLEX-RIME. ERα and several of its known interactors with significant enrichment are labelled. c) Sequence coverage of the ERα protein in the qPLEX-RIME analysis. d) Boxplots of the number of unique peptides for the overlapping ERα-associated proteins between the non-quantitative RIME and the qPLEX-RIME method. The RIME and qPLEX-RIME have been performed in MCF7 cells.

3.1.3 Comparison between qPLEX-RIME and non-quantitative RIME

We next performed a non-quantitative ERα RIME experiment with matched IgG controls to carry out a comparison between RIME and qPLEX-RIME methods in terms of recovery of known ERα interactors. The RIME was performed in two biological replicates and the comparison was limited to specific interactors after removing proteins found in both IgG samples. For more accurate comparison, we followed the same sample preparation workflow and samples were analysed on the same mass spectrometer for both approaches.

The data analysis showed that 302 of the 323 (93%) proteins identified as ERα-specific in the non-quantitative ERα-RIME pull-downs, were also identified by qPLEX-RIME with significant enrichment over the IgG controls (mean log₂Fold-Change of 2.5). Interestingly, we

observed an overall better peptide coverage for the overlapping ER α -associated proteins by qPLEX-RIME compared to the non-quantitative RIME method (**Figure 8d**), indicating higher sensitivity. The higher number of identified unique peptides can increase the reliability of the detected interactions and improve the accuracy of quantification.

Furthermore, qPLEX-RIME recovered 124 additional known BioGRID and STRING interactors compared to the non-quantitative RIME analysis (175 proteins >1 -log₂Fold-Change and adj. p-value <0.01 in qPLEX-RIME versus 51 proteins in non-quantitative RIME). Importantly, the higher sensitivity of the qPLEX-RIME method led to the identification of novel candidate ER α -associated proteins not previously found in RIME experiments. To validate selected findings as a means to assess the performance of the qPLEX-RIME approach, we conducted Proximity Ligation Assay (PLA) that allows the in situ detection of protein interactions with high specificity and sensitivity (Soderberg et al., 2006). The PLA confirmed the novel interactions of CBX3 (HP1 γ), NIPBL and FOXK1 with ER α (**Figure 9a**). The GFP (green fluorescent protein) protein was included in the analysis as negative control to discriminate non-specific interactions (**Figure 9b**). Notably, treatment of the MCF7 cells with the Selective ER α Degradar (SERD) Fulvestrant (Wardell et al., 2011) disrupted the above interactions three hours post treatment, demonstrating the specificity of the PLA assay and validating the new ER α interactors detected by qPLEX-RIME method (**Figure 9c**). The efficiency of the drug treatment and the effect on ER α protein levels was validated with immunofluorescence staining (**Figure 10**). Overall, these findings demonstrate a gain in sensitivity using the qPLEX-RIME method that can promote the identification of known and novel interacting proteins with statistical robustness.

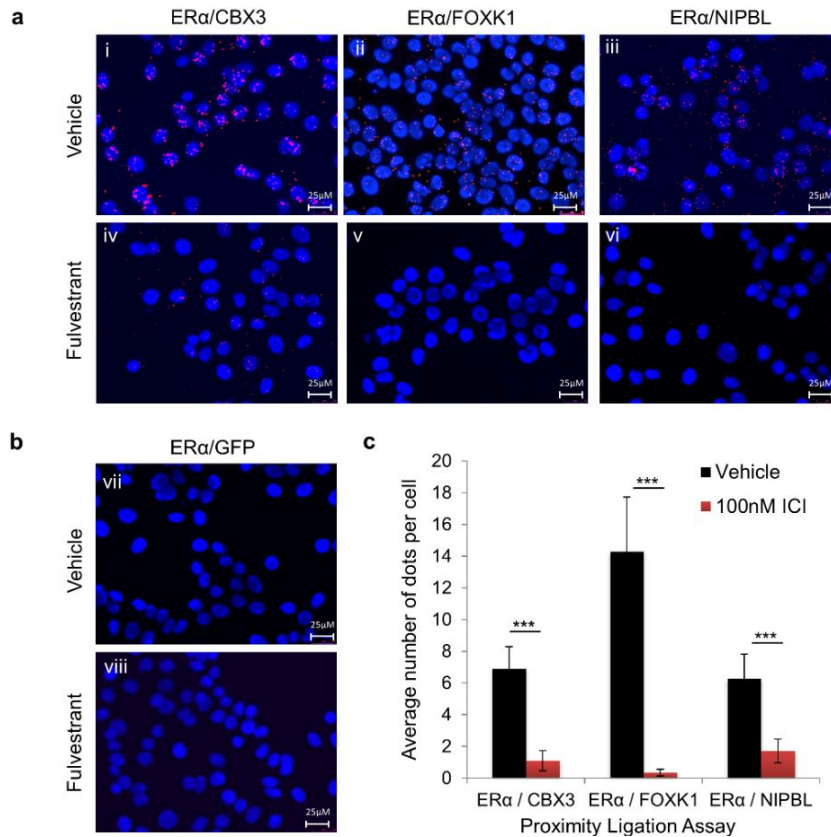


Figure 9. Validation of novel ERα-associated cofactors using Proximity Ligation Assay (PLA).

a) MCF7 cells were treated for 3h with vehicle (ethanol) (panels i–iii) or with 100nM Fulvestrant (ICI) (panels iv–vi) and were analysed by PLA to validate protein interactions between ERα and CBX3 (panels i and iv), FOXK1 (panels ii and v) and NIPBL (panels iii and vi). The experiment was performed in duplicate and the graphs are representative of one of the experiments. b) PLA assay using antibodies against ERα and GFP was used as a negative control (panels vii and viii). c) Quantification of the number of red PLA dots per cell was performed using Image J software (***)Student's *t*-test *p*-value<0.001). The error bars indicate standard deviation (SD).

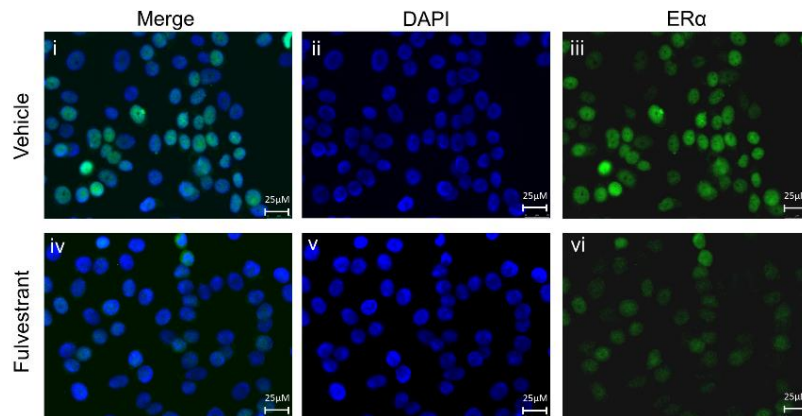


Figure 10. Decrease in ERα protein levels upon Fulvestrant treatment.

Immunofluorescence staining for ERα in MCF7 cells treated with vehicle (ethanol) (panels i–iii) or with 100nM Fulvestrant (panels iv–vi) for 3h.

3.1.4 Optimisation of crosslinking, starting material and antibody concentration

We next carried out a comparison between single and double crosslinking to examine whether the addition of DSG followed by FA can improve the crosslinking efficiency of nuclear-bound proteins to the bait, compared to standard FA crosslinking. To this end, we conducted ER α qPLEX-RIME pull-downs (4 biological replicates for each condition) using single or double-crosslinked MCF7 cells. One IgG pull-down was performed by mixing chromatin of all replicates, for each crosslinking method separately. These were included in the 10plex-TMT experiment to filter for specific interactors for further comparisons between the two approaches. The comparison revealed that double crosslinking resulted in a stronger enrichment of well-known and previously validated ER α interactors related to transcription and chromatin organisation such as FOXA1 (Mohammed et al., 2013), NR2F2 (Erdos and Balint, 2020) and NCOR2 (Jepsen et al., 2000) (**Figure 11a**). The enrichment obtained by double crosslinking can further improve the accuracy and robustness of the quantitative workflow.

Based on the successful preliminary experiments using the qPLEX-RIME method for the study of ER α -associated factors, we next performed titration experiments to assess the lower limit of starting material required for our method without compromising the recovery of the target complex. To this end, we applied a two-step fixation followed by LC-MS analysis to define the minimum number of cells required for an efficient ER α non-quantitative RIME pull-down in MCF7 cells. Reducing the number of cell culture plates will increase the method's throughput and facilitate the preparation of multiple replicates that can be accommodated in a single TMT experiment. The number of plates (150mm) was scaled down from four to one and the amount of beads and the antibody concentration for ER α or negative control IgG, was scaled proportionally. We started with four plates to match the number of cells used in previous published RIME experiments (Mohammed et al., 2013; Mohammed et al., 2015). Interestingly, the number of ER α peptides had small variation across the different numbers of plates as shown in **Figure 11b**. Regarding the coverage in known ER α interactors, an optimum number of peptides was obtained using two plates, above which no significant improvement was observed (**Figure 11c**). These results indicate that the starting material can be scaled down to one or two plates without significant loss in sensitivity of the assay. Therefore, for all the subsequent qPLEX-RIME, non-quantitative RIME and ChIP-seq experiments, we adopted the double crosslinking approach, using two plates (150mm) for each pull-down.

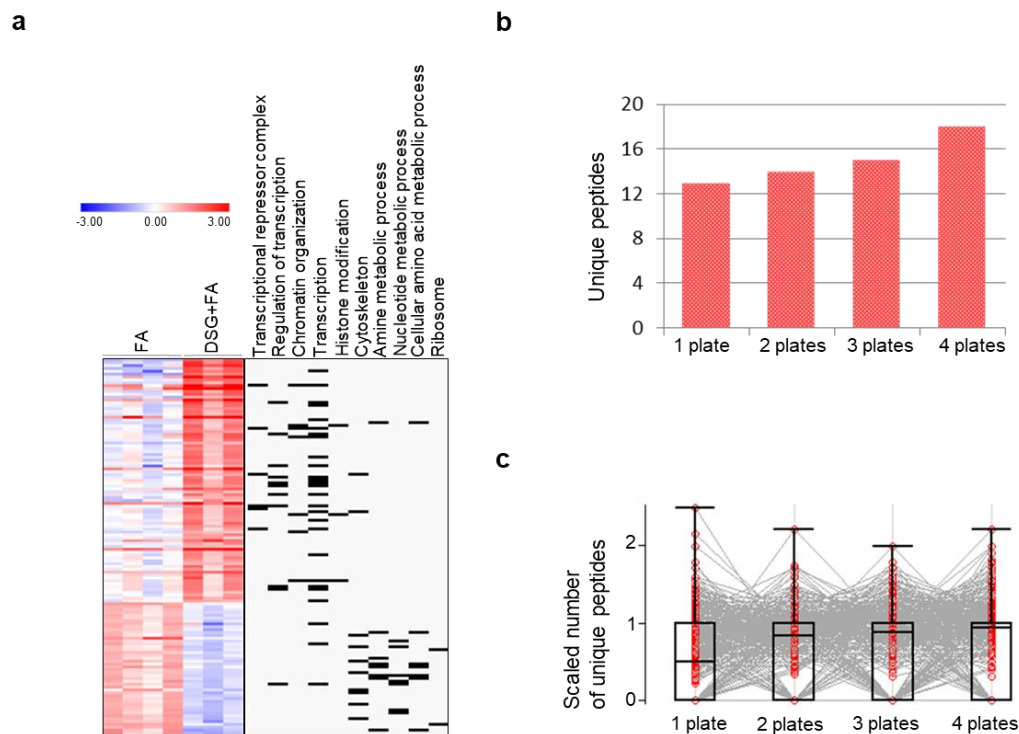


Figure 11. Comparison of crosslinking approaches and titration of starting material.

a) Comparison of double versus single crosslinking using qPLEX-RIME analysis. The scale bar represents log₂-scaled protein abundance values. b) Number of unique peptides of ERα across different number of plates. c) Scaled number of unique peptides for well-known ERα interactors across different number of plates in an ERα RIME experiment.

3.1.5 Evaluation of ChIP-seq and RIME grade antibodies for ERα

An important factor for the performance of the pull-down approaches is the specificity and efficiency of the antibody that is used to target the bait protein. For the newly-developed qPLEX-RIME method and the comparison with previous non-quantitative RIME data, we used the ERα antibody sc-543 from Santa Cruz Biotechnology. Most published ChIP-seq and RIME experiments have also been conducted using the sc-543 antibody (Lei et al., 2018; Mohammed et al., 2015; Mohammed et al., 2016; Ross-Innes et al., 2012), yielding high quality data. Recently, the sc-543 antibody was discontinued, impacting the ability to study ERα function using established ChIP and other pull-down assays.

To overcome this limitation, we compared the sc-543 (Santa Cruz Biotechnology) with other commercially available antibodies using MCF7 cells as a model. For this project I worked with my colleague Silvia E. Glont from the Carroll lab and our work on the evaluation of ChIP-seq and RIME grade antibodies was recently published (Glont et al., 2019b); Silvia performed the ChIP-qPCR and ChIP-seq experiments and I conducted the RIME experiments. Firstly, we

compared the commonly used sc-543 (Santa Cruz Biotechnology) antibody with the ab80922 (Abcam), ab3575 (Abcam), sc-514857 (C-3) (Santa Cruz Biotechnology), C15100066 (Diagenode) and 06-935 (Millipore) ER α available antibodies. We performed a ChIP-qPCR experiment in two biological replicates to assess ER α binding at well-known target regions (Carroll et al., 2006; Glont et al., 2019a) (XBP1, GREB1, RARA, MYC, ESR1, CA12). A known non-ER α binding site was included as negative control (ER3) (Carroll et al., 2006; Glont et al., 2019a). Interestingly, the ChIP-qPCR data suggested that 06-935 and ab3575 antibodies could successfully enrich for ER α -bound chromatin at selected loci compared to sc-543 antibody (**Figure 12a**).

In addition to ChIP-qPCR, we performed ChIP-seq experiments to further compare the performance of 06-935 and ab3575 with sc-543 in a global scale, including IgG as negative control. Each ER α ChIP-seq was performed in at least duplicates and the ER α negative MDA-MB-231 cell line was used to assess non-specific binding. We observed 6,031 ER α binding sites for sc-543 (Santa Cruz) antibody, 6,192 peaks for ab3575 (Abcam) and 6,552 for 06-935 (Millipore) with significant overlap between the three antibodies (**Figure 12b**). All antibodies showed robust enrichment at binding sites compared to background and motif analysis identified a high enrichment for the ER α response element (ERE) motif in the considered sequences (**Figure 12c**). Notably, neither of the ab3575 and 06-935 antibodies showed significant enrichment in the MDA-MB-231 cell line, highlighting their specificity (**Figure 12c**). **Figure 12d** illustrates the ER α binding profile in two well-known ER α target genes, (GREB1 and RARA) in both MCF7 and MDA-MB-231 cell lines in the different ER α ChIP-seq or IgG samples.

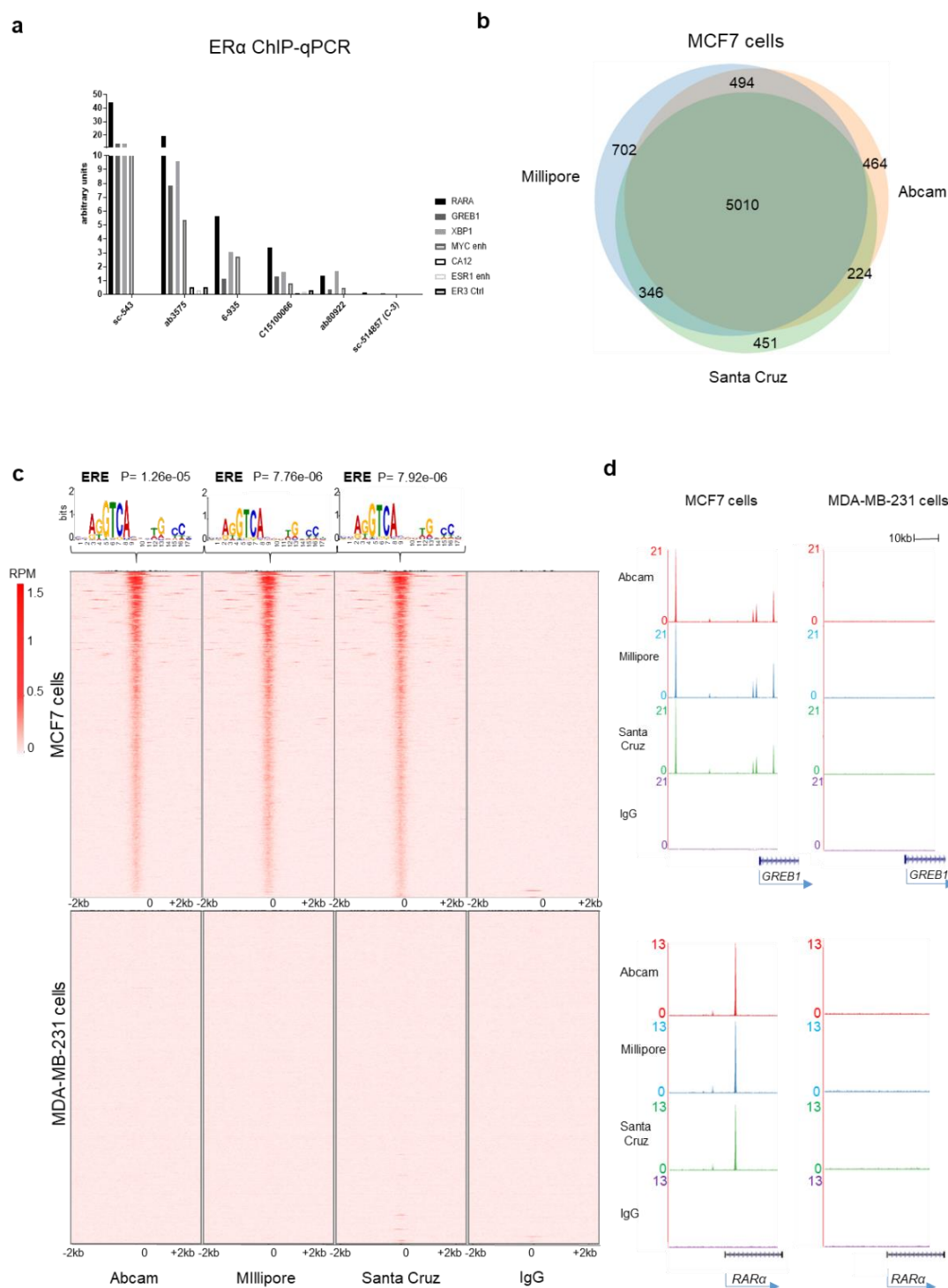


Figure 12. A comparison between different ERα antibodies by ChIP-qPCR and ChIP-seq.

a) ChIP-qPCR analysis for known ERα binding sites in MCF7 cells. Results are shown as arbitrary units. Antibodies used: sc-543 (Santa Cruz Biotechnology), ab80922 (Abcam), ab3575 (Abcam), sc-514857 (C-3) (Santa Cruz Biotechnology), C15100066 (Diagenode) and 06-935 (EMD). b) Venn diagram showing the overlap between ERα binding sites for Santa Cruz (sc-543), Millipore (06-935) and Abcam (ab3575) antibodies in MCF7 cells. c) Top: De novo motif analysis of ERα binding sites using MEME. Bottom: Heatmap of total number of ERα binding sites identified in both technical replicates of MCF7 and in all three biological replicates for MDA-MB-231, respectively. d) ERα binding profile in two well-known ERα target genes in MCF7 and MDA-MB-231 cells.

We next aimed to evaluate the performance of ab3575 and 06-935 in RIME experiments compared to sc-543 antibody. For this analysis, we tested the 06-935, ab3575 and sc-543 antibodies in two technical replicates using MCF7 cells. We included IgGs as negative controls to discriminate specific interactors and assess the antibody specificity. Firstly, we evaluated the pull-down efficiencies by comparing the sequence coverage of the bait protein achieved from the different antibodies. Interestingly, a similar number of unique ER α peptides was obtained (**Figure 13a**) across the different pull-downs, confirming that all three antibodies achieved efficient immunoprecipitation of the bait protein.

Secondly, we evaluated the efficiency of the different antibodies in detecting ER α -associated proteins. For data interpretation, we used a label-free quantification method based on the Minora algorithm implemented in Proteome Discoverer 2.2 software. Principal component analysis (PCA) using scaled intensities of known ER α -associated proteins, from BIOGRID and STRING databases (n=319) displayed a strong separation between ER α RIME samples and IgG controls, demonstrating the high specificity of the antibodies (**Figure 13b**). Importantly, we found only small variation between the three antibodies, suggesting that they all efficiently pull-down known ER α -associated proteins. Hierarchical clustering showed an overall consistency between the antibodies (**Figure 13c**) and revealed an enrichment of well-known ER α interactors such as FOXA1, GATA3, NCOA3 and EP300 by all three antibodies compared to IgG samples (**Figure 13d**). Collectively, the data support that all three ER α antibodies perform similarly in ChIP-seq and RIME analyses and both ab3575 and 06-935 can replace the sc-543 antibody for characterising ER α interactors or profiling ER α binding on chromatin. Interestingly, the ab3575 and 06-935 antibodies recognise different protein epitopes (ab3575: amino-acids 21-32; 06-935: C-terminus). To this end, the Carroll lab tested the combination of 06-935 and ab3575 antibodies compared to the individual use of these antibodies (ab3575, 06-935, sc-543) by ChIP-qPCR. The ChIP-qPCR was performed for two well-known ER α binding sites adjacent to GREB1 and RARA genes. Additionally, the ER α negative control binding region (ER3) was included (experiment conducted by Dr Sanjeev Kumar). Notably, ChIP-qPCR results showed a higher enrichment of ER α -bound chromatin following the addition of both antibodies compared to the application of individual antibodies (**Figure 14**). Taken together, our data show that the combination of two ER α antibodies can perform very well for the enrichment of ER α -bound chromatin and is suitable to substitute the sc-543 without affecting the sensitivity and specificity of the ChIP-experiments. For any new RIME or ChIP-seq experiment performed after the sc-543 was discontinued, the addition of both 06-935 and ab3575 antibodies was used.

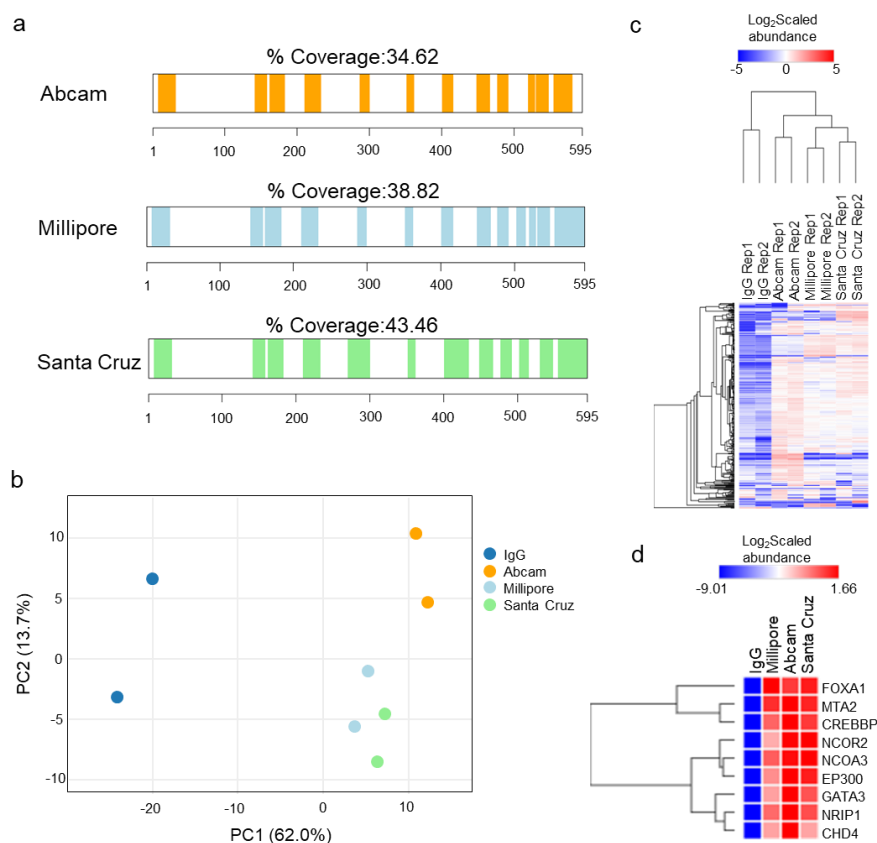


Figure 13. Comparison of RIME data between Santa Cruz, Millipore and Abcam antibodies.

a) Protein sequence coverage of ERα achieved by the use of Abcam (ab3575), Millipore (06-935) and Santa Cruz (sc-543) antibodies in RIME. b) PCA plot using a subset of known ERα interactors (n = 319, BIOGRID and STRING databases) for the four different RIME pull-downs. c) Hierarchical clustering of the scaled intensities of known ERα interactors from BIOGRID and STRING databases (n=319). d) Hierarchical clustering of well-characterised ERα interactors.

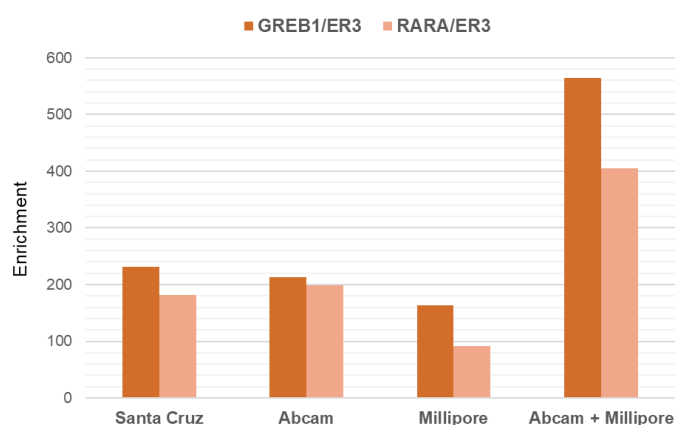


Figure 14. A comparison between different ERα antibodies by ChIP-qPCR.

ChIP-qPCR analysis for ERα known binding sites in MCF7 cells. Results are shown as enrichment over negative control site (ER3). Antibodies used: sc-543 (Santa Cruz Biotechnology), ab3575 (Abcam), 06-935 (Millipore), combined ab3575 and 06-935. The experiment was performed including one replicate for each antibody.

3.2 Different applications of the qPLEX-RIME method

3.2.1 Characterisation of different protein interactomes

To investigate whether our quantitative pipeline can be used to study interactors of bait proteins other than ER α , we performed qPLEX-RIME pull-downs on three additional factors. Five biological replicates were collected for each bait and an equal number (N=5) of matched IgG control samples was included to discriminate the specific interactors for each experiment.

Firstly, the qPLEX-RIME method was applied to explore the interactome of two well-characterised co-activators of nuclear receptors (Lonard and O'Malley B, 2007); CBP (CREB-binding protein) and NCOA3 (SRC-3). We quantified 1,437 and 1,135 proteins for CBP and NCOA3 respectively in the TMT 10plex sets of bait and IgG pull-downs (FDR<1%). Importantly, the target proteins CBP and NCOA3 were highly enriched in the bait pull-downs compared to the IgG controls (CBP: log₂Fold-Change=3.2, adj. p-value<0.01; NCOA3: log₂Fold-Change=3.39, adj. p-value<0.01), identified with 44 and 36 unique peptides respectively (**Figures 15a & b**). Well-described interactors of both bait proteins were identified including EP300, p160 coactivators, arginine methyltransferases and components of the ER α complex (Lonard and O'Malley B, 2007) (**Figures 15a & b**). Additionally, several subunits of the SWI/SNF chromatin remodelling complex such as SMARCA4 (BRG1), SMARCE1 (BAF57), SMARCB1 (BAF47) and SMARCC2 (BAF170) were identified. We also revealed a strong enrichment of corepressors such NCORs and HDACs in both datasets, suggesting that both co-activators and co-repressors are part of the same complex. This finding is consistent with previously published data showing an extensive co-localisation of corepressors and coactivators using ChIP-seq (Siersbaek et al., 2017).

Secondly, we explored the interactome of the phosphorylated-RNA polymerase II (Pol II) at serine-5 in the carboxy-terminal domain (CTD), to examine the efficiency of qPLEX-RIME on capturing phospho-interactomes. CDK7 in human TFIIF phosphorylates the serine 5 (S5) of the Pol II carboxy-terminal domain (CTD), after the assembly of the pre-initiation complex (Mosley et al., 2009) and studies have shown that it is implicated in the regulation of transcription initiation, elongation, termination and mRNA processing (Phatnani and Greenleaf, 2006). We quantified 1,442 proteins across all multiplexed samples at peptide FDR<1%; the bait protein was one of the top enriched proteins (adj. p-value<0.01, log₂Fold-Change =4.2), identified with 96 unique peptides. The **Figure 15c** highlights the sequence coverage of the bait protein and the identification of known Pol II interactors, including

subunits of the SWI/SNF complex, proteins of the mediator complex, initiation and elongation factors (Hahn, 2004; Orphanides and Reinberg, 2000).

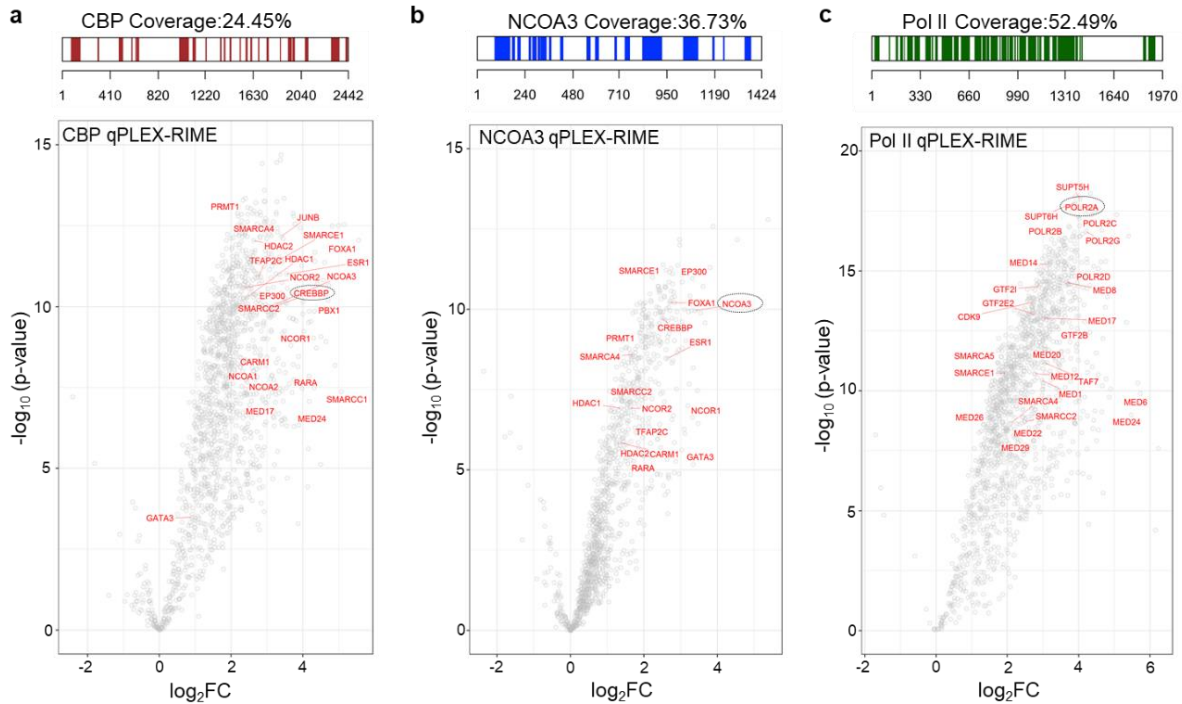


Figure 15. Application of qPLEX-RIME in CBP, NCOA3 and Pol II.

a) Sequence coverage of the CBP protein in the qPLEX-RIME analysis (top panel). The volcano plot shows the quantitative results of the CBP qPLEX-RIME (bottom panel). b) NCOA3 sequence coverage in the qPLEX-RIME analysis (top panel). The volcano plot displays the quantitative results of the NCOA3 qPLEX-RIME (bottom panel). c) Pol II sequence coverage in the qPLEX-RIME analysis (top panel). The volcano plot illustrates the quantitative results of the Pol II qPLEX-RIME (bottom panel). In all volcano plots, several known interactors of the bait proteins are shown in red font.

Next, we performed a comparison of all four interactomes (ER α , CBP, NCOA3, Pol II), that were obtained by qPLEX-RIME. This revealed the identification of a significant number of uniquely identified interactors as well as a partial overlap across the four datasets (**Figure 16a**). We further examined, whether the shared proteins are more likely due to the common underlying biology of the four baits rather than an intrinsic technical bias that could be introduced by the method. To this end, we made a Venn diagram comparing random selections of proteins identified in the four qPLEX-RIME experiments, without filtering for specificity (**Figure 16b**). This analysis identified a smaller portion of overlapping proteins between the four interactomes, suggesting small contribution of technical factors to the observed overlap.

Overall, our data demonstrate that qPLEX-RIME can be widely used for the characterisation of different interactomes with specificity and statistical robustness.

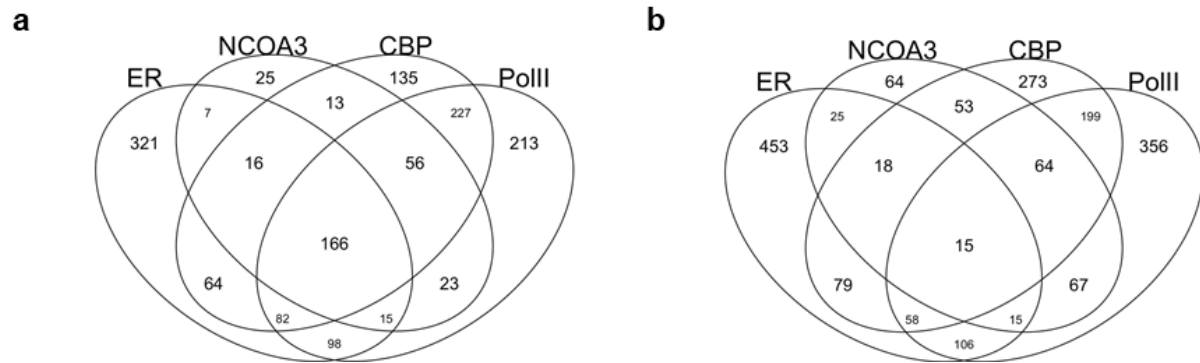


Figure 16. Comparison of the four interactomes (ER α , CBP, NCOA3, Pol II).

a) The Venn diagram illustrates the comparison of the enriched interactors ($\log_2\text{Fold-Change} > 1$, adj. $p\text{-value} < 0.05$, > 1 unique peptide) of the four bait proteins (ER α , CBP, NCOA3, Pol II). b) Venn diagram highlighting the overlap of the four bait proteins using random selection of proteins identified in the four qPLEX-RIME experiments.

3.2.2 Study of ER α complex dynamics upon OHT treatment

As part of our quantitative method optimisation, we performed three qPLEX-RIME experiments ($3 \times 10\text{plex}$), to explore the dynamics of the ER α complex assembly upon treatment with the Selective Estrogen Receptor Modulator (SERM) 4-hydroxytamoxifen (OHT) (Dutertre and Smith, 2000). To this end, double cross-linked MCF7 cells were treated with 100nM OHT for 2h, 6h and 24h or with ethanol for 24h (vehicle) in two biological replicates each condition, in each 10-plex-experiment to obtain a total of six replicates per time point. Additionally, control IgG pull-downs samples treated with OHT or ethanol for 24h were included in the same analysis to help eliminate non-specific binding in data analysis.

In parallel, RNA-seq analysis was performed in six biological replicates using matched OHT treated samples to evaluate the efficiency of the drug treatment. The response to the drug treatment was evidenced by the transcriptional repression of a number of known ER α target genes at 6h and 24h, as OHT exhibits ER α antagonist activity in breast and blocks estrogen action. **Figure 17a** illustrates the suppression of ER α target genes such as *PGR*, *PDZK1*, *TFF1*, *AREG*, *PKIB*, *SIAH2*, *MYB*, *HEY2*, *FOS*, *GREB1* and *TFF3* (Caliceti et al., 2013; Frasor et al., 2005; Frasor et al., 2003; Frasor et al., 2004; Ghosh et al., 2000; May and Westley, 1997), after 24h OHT treatment compared to the vehicle ($\log_2\text{Fold-Change} < -0.5$, adj. $p\text{-value} < 0.05$).

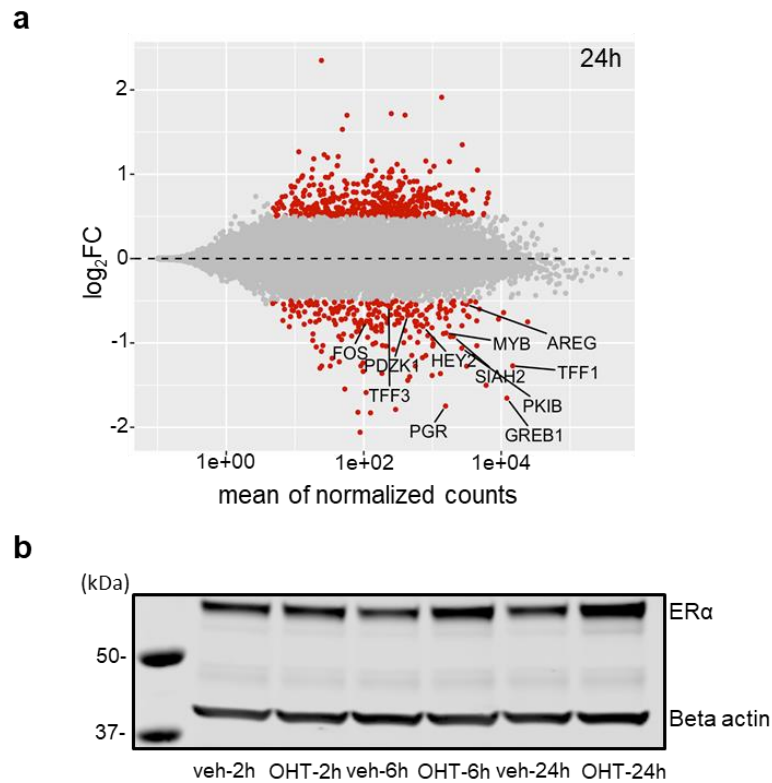


Figure 17. RNA-seq analysis and profile of ERα total protein levels.

a) Scatter plot displaying the RNA-seq quantification results at 24h time point. Significantly regulated genes are highlighted ($|\log_2\text{Fold-Change}| > 0.5$, adj. $p\text{-value} < 0.05$) and known ERα target genes are labelled. b) Western blot showing the temporal changes of ERα upon OHT treatment in whole cell lysate (MCF7). Beta-actin had been used as a loading control.

qPLEX-RIME analysis of the samples using the MultiNotch MS3 approach quantified 1,105 proteins (FDR<1%) across all three replicate TMT experiments combined. Of these, 412 proteins were significantly enriched in ERα pull-downs compared to IgG samples ($\log_2\text{Fold-Change} > 1$, adj. $p\text{-value} < 0.01$). During the data interpretation of the qPLEX-RIME dataset, we observed a change in ERα levels upon OHT treatment, a finding that was validated by Western Blot (**Figure 17b**). A change in the levels of the bait protein due to biological mechanisms underlying the drug treatment, may influence the amount of purified proteins and result in systematic quantification biases. Indeed, our data showed a significant correlation of the quantified proteins on the amount of ERα pulled-down (**Figure 18a**). To correct for this effect, the Bioinformatics core applied a linear regression approach (McCarthy et al., 2017; Roumeliotis et al., 2017); the main advantage of this approach is that proteins with strong dependency on the target protein are subjected to more significant correction compared to those with small dependency, which are only slightly adjusted. **Figure 18b** illustrates the quantitative

profile of two well-known ER α interactors (GATA3 and HDAC2) before and after correction. After correcting the quantitative data for the abundance of ER α using the linear regression approach, we identified 249 ER α -enriched proteins with altered profile in the interactome in at least one time point ($|\log_2\text{Fold-Change}|>0.5$, adj. p-value<0.05), allowing for a comprehensive mapping of the dynamic organisation of the ER α complex in response to OHT treatment as discussed in the next section.

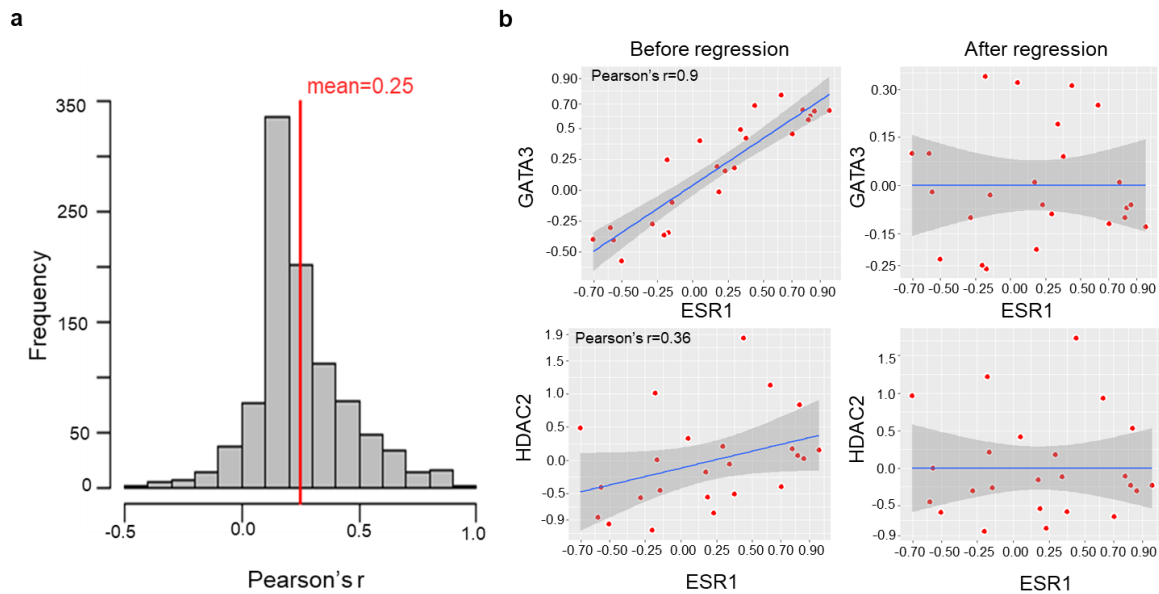


Figure 18. Correction of quantitative results for the dependency on ER α target protein by linear regression.
a) Histogram of Pearson's correlation coefficients between ER α protein profile and all ER α -enriched proteins.
b) Scatter plots of ER α profile versus the raw (left panel) and corrected (right panel) quantitative values of GATA3 (with strong dependency on ER α) and HDAC2 (with low dependency on ER α).

3.2.2.1 Dissociation and recruitment of co-factors upon OHT treatment

To better understand the cascade of molecular changes following OHT treatment of MCF7 cells and evaluate whether the changes captured by qPLEX-RIME are consistent with previous reports and provide additional findings, we interrogated the significant differences observed in the ER α interactome at each time point (2h, 6h, 24h). Firstly, at 2h treatment with OHT, we observed a loss of 12 proteins including known ER α co-activators such as NCOA3 and CBP (**Figure 19a**). These proteins have been linked to histone acetylation and modulation of chromatin organisation and structure leading to activation of gene transcription (Metivier et al., 2003; Shang et al., 2000). Their loss in the ER α complex is in line with previously published data, showing that the access of co-activators is blocked by the binding of antagonists (Shiau

et al., 1998). Additionally, a significant loss of the interaction between ER α and NRIP1 (RIP140) protein was detected. NRIP1 can act as a corepressor or as a coactivator (Rosell et al., 2014) with previous evidence suggesting that NRIP1 is required for ER α -complex formation as well as for gene expression mediated by ER α (Rosell et al., 2014). **Figure 19b** depicts the quantification profile of these key ER α co-factors across all biological replicates. Furthermore, OHT treatment resulted in the loss of GREB1 and its paralog gene product GREB1L. Loss of GREB1 upon OHT treatment has been previously reported (Mohammed et al., 2013), but for the first time we observed the loss of both proteins simultaneously.

Next, we examined the changes detected in the ER α interactome 6h post OHT treatment. The analysis of the data using the qPLEXanalyzer tool revealed 237 significantly enriched interactors (\log_2 Fold-Change>0.5, adj. p-value<0.05) compared to the vehicle treatment. Notably, NCOA3, NRIP1, GREB1 and GREB1L remained at decreased levels in the interactome (\log_2 Fold-Change<-0.5, adj. p-value<0.05) (**Figure 19a**). Amongst the enriched proteins we found recruitment of several components of the NuRD (Nucleosome Remodelling and Deacetylase) complex, e.g. HDAC1/2 and the signature components MTA1/2 (Lai and Wade, 2011; Liu and Bagchi, 2004), as well as an enrichment of the co-repressor NCOR2 (SMRT) (Chen and Evans, 1995; Jepsen et al., 2000) (**Figure 19b**). Consistent with our data, NURD complex and NCOR2 have been previously shown by ChIP to be recruited to promoter regions of ER α target genes in a sequential manner, following OHT treatment, to create a repressive chromatin conformation (Liu and Bagchi, 2004; Shang et al., 2000). Additionally, we detected an enrichment for subunits of the ATP-dependent chromatin remodelling complex SWI/SNF, including SMARCC2 (BAF170), SMARCE1 (BAF57) and SMARCA4 (BRG1) (**Figure 19b**), which is known to regulate both gene activation and gene repression (Belandia et al., 2002; Zhang et al., 2007). The SMARCA4 protein, which was previously shown to be required for repression of ER-mediated transcription (Zhang et al., 2007), was one of the top enriched SWI/SNF proteins.

A restoration of the ER α complex after 24 treatment with OHT was observed, with the exception of the NCOA3, NRIP1 and GREB1 proteins, which were still decreased (\log_2 Fold-Change<-0.5, adj. p-value<0.05) (**Figure 19a**). The complex restoration is likely associated with the half-life of the drug as the cells were treated once with OHT at the beginning of the assay. Also, the fact that co-activator levels remain decreased in the complex, highlights the important role of the elimination of co-activators in the antagonistic activity of OHT. To validate key findings from the qPLEX-RIME analysis, we performed PLA assays for factors

that change with OHT treatment after 2h or 6h. PLA confirmed the loss of NCOA3 and CBP at 2h and the enrichment of SMARCC2 (BAF170) and HDAC1 at 6h (**Figure 20a & b**).

Next, we compared our findings at 2h treatment with OHT with previously published ER α RIME data (Mohammed et al., 2013), where MCF7 cells were labelled with SILAC reagents followed by 3 hours treatment with tamoxifen, to detect changes in the ER α complex based on SILAC ratios. This revealed the high correlation between the two different datasets, confirming the accurate quantification obtained by the MultiNotch MS3 level mass spectrometry analysis (**Figure 21**). Taken together, our unbiased analysis captured the dynamic interplay of distinct coregulatory complexes with high quantification accuracy and revealed that the inhibitory effect of OHT peaks at 6h, where ATP-dependent remodelling and corepressor complexes may coordinate to create a transcriptionally inactive chromatin environment.

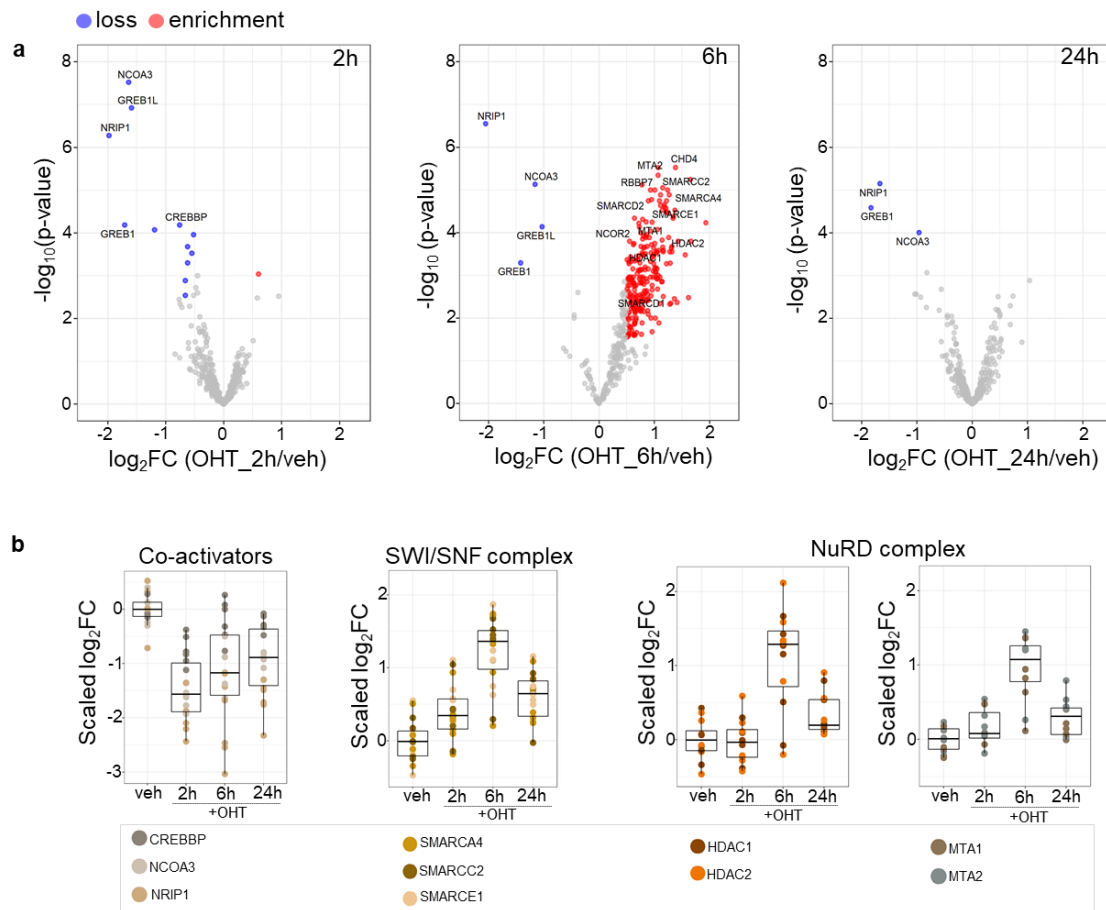


Figure 19. Temporal profiling of the ER α interactome following treatment of MCF7 cells with OHT.

a) Volcano plots showing the enrichment or loss of proteins in the ER α interactome upon OHT treatment for 2h, 6h and 24h. b) Boxplots illustrating the loss of ER α co-activator proteins CBP, NCOA3 and NRIP1 at 2h and the enrichment of SWI/SNF and NuRD complexes subunits at 6h (left to right). Quantitative values are normalized so that the median of the vehicle treated samples is zero (centred at the median of vehicle).

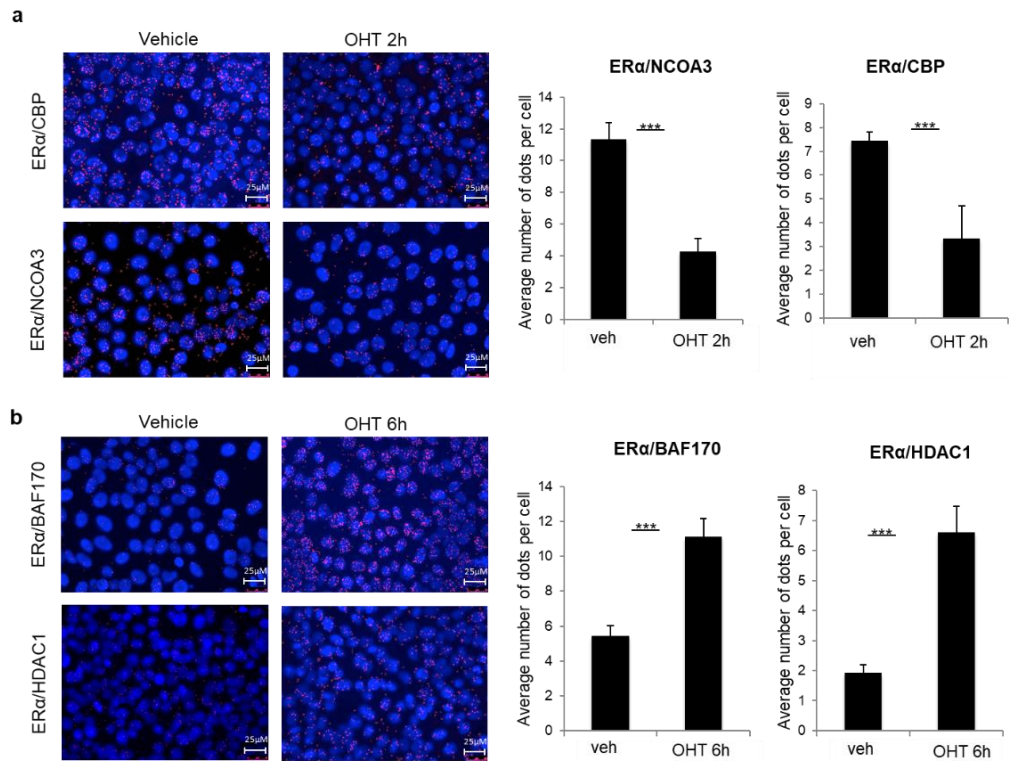


Figure 20. Validation of temporal changes in the ERα complex upon OHT treatment.

a) PLA assay showing the loss of NCOA3 and CBP from the ERα complex at 2h treatment with OHT compared to vehicle treatments (left panel). b) PLA assay showing the enrichment of BAF170 and HDAC1 in the ERα complex at 6h treatment with OHT (left panel). Both experiments were performed in duplicate and the graphs are representative of one of the experiments. Bar plots showing the quantification of the number of red PLA dots per cell normalized to ERα counts (single recognition PLA assay) using Image J software (***) Student's t-test p-value <0.001) (right panels). The error bars indicate standard deviation (SD).

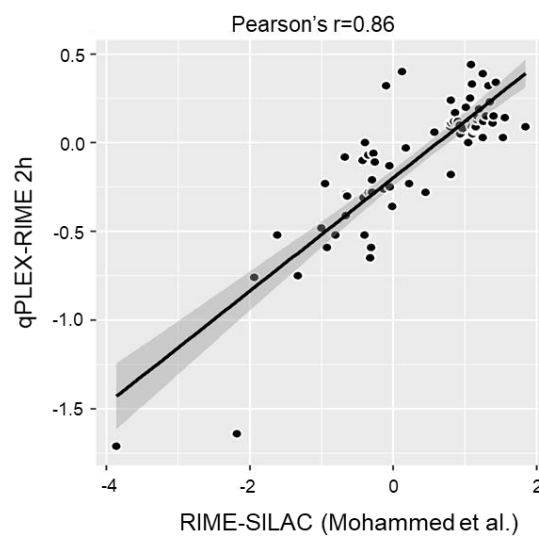


Figure 21. Comparison between qPLEX-RIME and RIME-SILAC data.

Comparison of the qPLEX-RIME data at 2h OHT treatment with published RIME-SILAC data, where MCF7 cells were treated with OHT for 3h, displayed a good correlation between the two datasets.

3.2.3 Identification of net changes in the ER α complex

The qPLEX-RIME data indicated that OHT treatment in MCF7 cells caused significant changes in the composition of ER α complex. To evaluate whether these changes are specific for the interactome or result from changes in total protein levels, we conducted a time-course whole proteome quantification in matched samples under the same conditions (vehicle, 2h, 6h and 24h, four biological replicates each, 2 \times 8plex). The TMT labelled fractions were analysed with the MutliNotch MS3 approach, which enabled the quantification of 8,916 and 8,342 proteins in the first and second experiments respectively (FDR<1%).

The analysis of the whole proteome data confirmed the up-regulation of ER α protein levels (log₂Fold-Change: 2h 0.21, 6h 0.5 and 24h 1) upon OHT treatment, which was not due to an increase in gene transcription. This finding is in line with previous reports showing an increased ER α protein stability in the presence of OHT (Wijayaratne and McDonnell, 2001). Most importantly, the comparison between the qPLEX-RIME results and the whole proteome data showed that the changes detected in the ER α complex upon OHT treatment represent changes in protein recruitment as the respective total protein and mRNA levels were unchanged (**Figure 22a**). GREB1 was the only ER α interactor with decreased mRNA and total protein levels at 24h treatment. It is known that GREB1 is an ER α target gene (Ghosh et al., 2000; Mohammed et al., 2013) and this explains the decreased association between ER α and GREB1 at the later time point.

Further k-means clustering of the significantly regulated proteins (adj. p-value<0.05) across the three time points in the whole proteome, identified clusters of up- and down- regulated proteins (**Figure 22b**). Notably, Gene Set Enrichment Analysis (GSEA) of the regulated proteins, revealed an enrichment of genes linked to estrogen response and tamoxifen resistance (**Figure 22c**). We also observed a down-regulation of proteins involved in cell cycle (Whitfield et al., 2002) (**Figure 23a**), consistent with the antiproliferative effects of OHT (Liu and Bagchi, 2004). Taken together, the RNA-seq data demonstrated an early effect on gene expression, following OHT treatment (6h), coinciding with the strong stoichiometric changes in the ER α complex. Additionally, the whole proteome results revealed protein expression changes at the later time point (24h) and confirmed that the changes of ER α -associated proteins are specific for the complex and not due to global changes in protein levels. **Figure 23b** illustrates the low mRNA-to-protein correlation at 2h and 6h and the respective strong correlation at 24h. We conclude that our qPLEX-RIME data in combination with the total proteome measurements

outline both changes in the ER interactome and the associated downstream global effects using an unbiased and accurate approach.

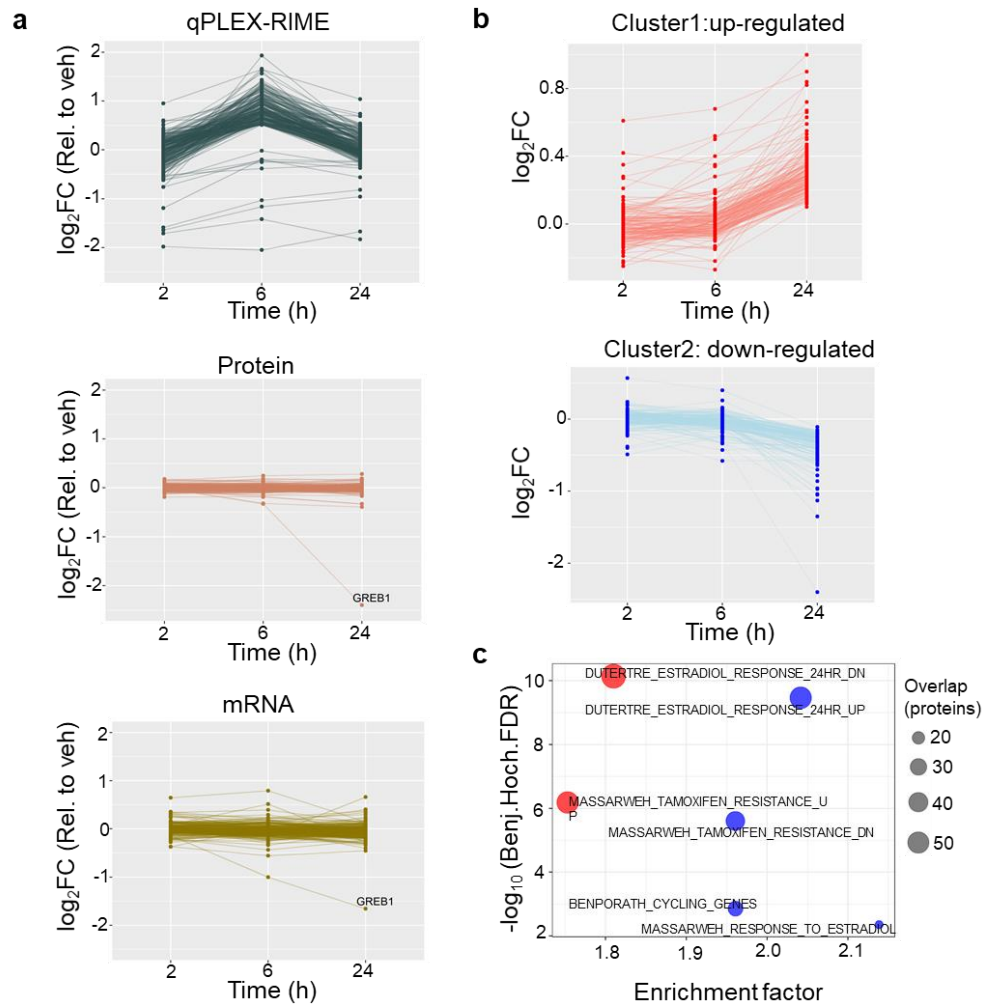


Figure 22. Comparison of qPLEX-RIME data with whole proteome and RNA-seq data.

a) Line plots of the significantly enriched or lost proteins in the qPLEX-RIME data (top panel) and their respective profiles in the whole proteome (middle panel) and RNA-seq analysis (bottom panel). b) Line plots representing two k-means clusters of down- and up-regulated proteins identified in total proteome analysis. c) Gene set enrichment analysis for the down- and up-regulated protein clusters.

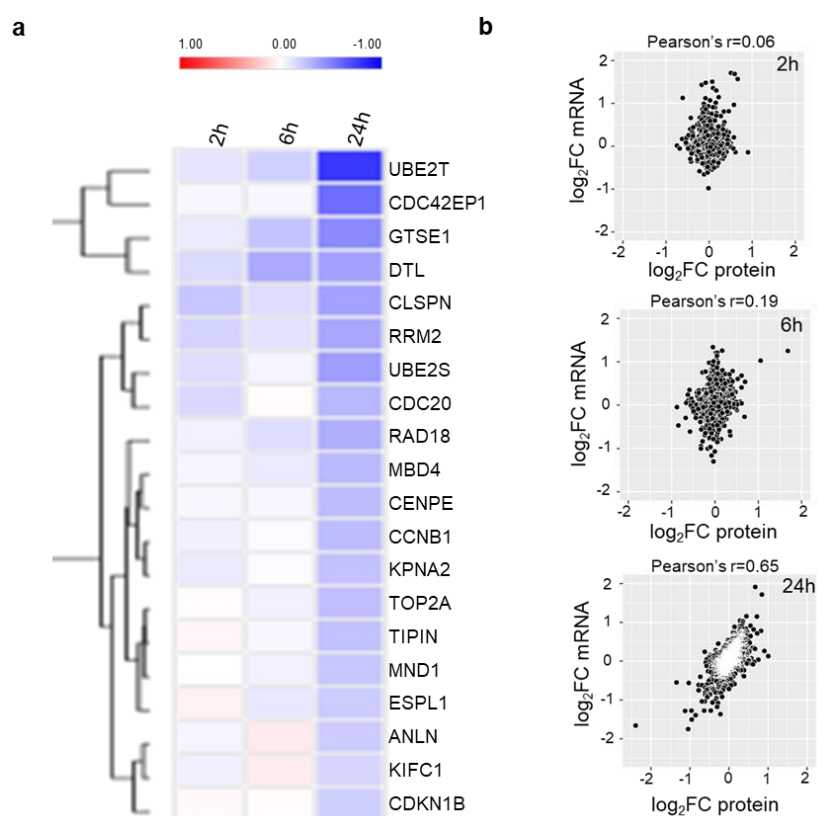


Figure 23. Differentially regulated proteins and correlation with gene expression.

a) Hierarchical clustering of down-regulated proteins involved in cell cycle at 24h using total proteome measurements. The scale bars represent log₂ ratios versus the vehicle treatment. b) Scatter plots of mRNA versus total protein at 2h, 6h and 24h treatment with OHT.

3.2.4 Application of qPLEX-RIME in clinical tumour material

3.2.4.1 Characterisation of ER α interactome *in vivo*

To assess whether our method can be used to study chromatin associated protein-protein interactions in tumours, we conducted an ER α qPLEX-RIME experiment using three independent ER α positive human Patient Derived Xenograft (PDX) tumours (HCI-003, HCI-005, HCI-006), that have been previously described (DeRose et al., 2011). To increase the efficiency of the crosslinking, we adjusted our pipeline by including a cryosectioning step to dissociate the tumours. To this end, cryosections (30 μ m) of each tumour were double-crosslinked and each tumour was split into two parts to be used for ER α and IgG pull-downs. The MS3 analysis identified 2,319 proteins (FDR<1%) across all multiplexed samples with highly reproducible profiles (**Figure 24a**). We successfully detected ER α (log₂Fold-Change=1.72, adj. p-value=0.026, unique peptides=3) and known ER α interactors, such as

CBP (Shang et al., 2000; Zwart et al., 2011), NCOA2 (Zwart et al., 2011), HDAC1 (Liu and Bagchi, 2004), GREB1 (Mohammed et al., 2013), SMARCE1 (BAF57) (Belandia et al., 2002), SMARCA4 (BRG1) (Metivier et al., 2003) and NCOA5 (CIA) (Sauve et al., 2001) (**Figure 24b**). We further analysed the sequences of the identified interactors and their respective quantified peptides to examine possible contamination from mouse stroma. The analysis revealed that 60% of the significant interactors were identified with at least one unique human peptide (i.e. a peptide that does not align to the mouse proteome), highlighting that the identified proteins were primarily derived from human cancer cells. Consistently, the tumour samples showed high cellularity and positive staining for human ER α exclusively in the cancer cells and not in the stroma (**Figure 24c**).

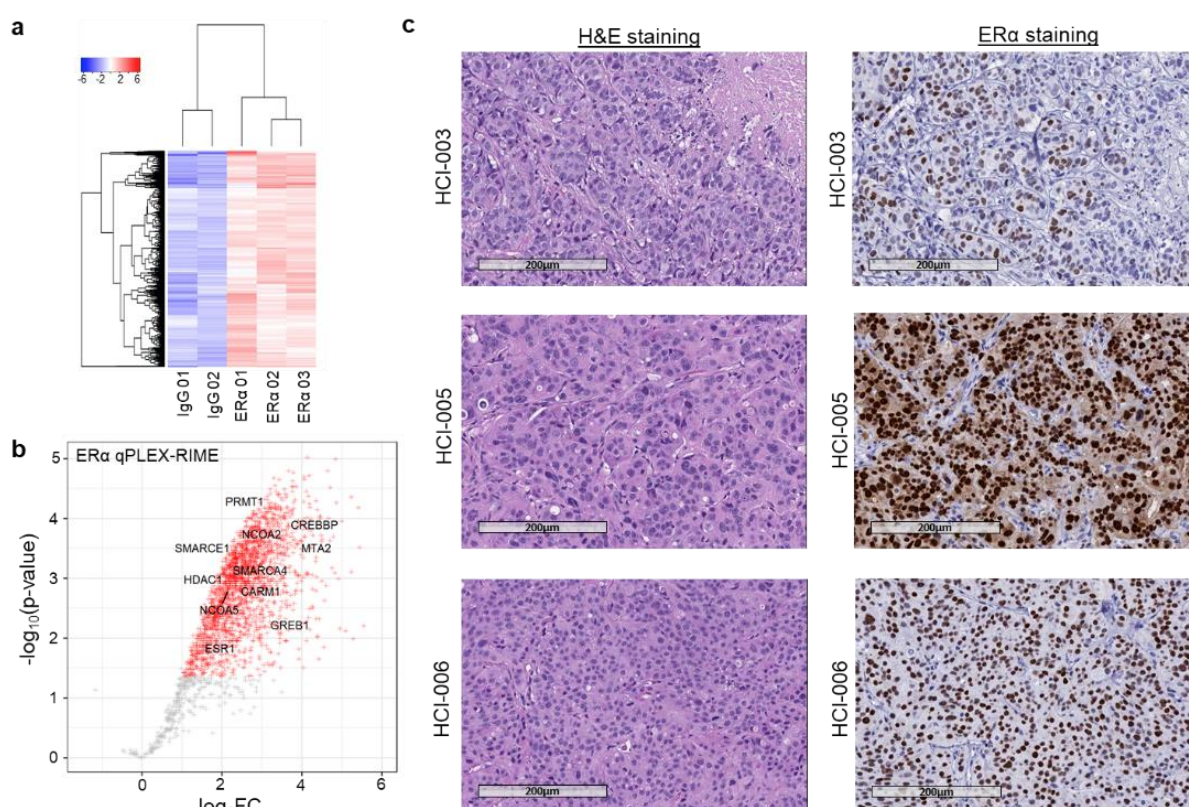


Figure 24. ER α qPLEX-RIME application in PDX tissues.

a) Hierarchical clustering of the qPLEX-RIME quantified ER α -enriched proteins in PDX tissues. The scale bar represents row-mean scaled \log_2 values. b) Volcano plot illustrating the quantitative results of qPLEX-RIME application in PDX tissues. Statistically significant proteins ($\log_2\text{Fold-Change} > 1$, $\text{adj. } p\text{-value} < 0.05$) are shown in red colour and several well-known ER α interactors are labelled. c) H&E (Haemotoxylin and Eosin) staining and ER α staining for the three PDX tumours that were used in the qPLEX-RIME experiment (magnification=20x, scale bar=200 μm).

Next, we aimed to apply the same workflow in human cancer clinical tissues, using five independent human breast cancer tumours [ER α positive, PR (Progesterone Receptor) positive, HER2 (Human epidermal growth factor receptor 2) negative, Grade2/Grade3]. Approximately 60 sections (30 μ m) were obtained per tumour and were double-crosslinked before they were split for ER α and matched IgG pull-downs (**Figure 25a**). The analysis of the samples yielded the quantification of 2,191 proteins (FDR<1%) across the multiplexed set. Notably ER α was recovered with excellent sequence coverage (17 unique peptides), similar to the coverage we previously achieved from the ER α qPLEX-RIME in MCF7 cells. Additionally, well-described ER α interactors such as FOXA1, GATA3, GREB1, EP300, CBP, HDACs, NCORs, NCOA2 and subunits of the SWI/SNF complex were identified (**Figure 25b**). **Figure 25c** displays the high enrichment of several key ER α interactors in the bait samples compare to IgG controls. Integrating all the datasets of the ER α interactome in MCF7 cells, we identified a compendium of 253 novel ER α -associated proteins with consistent presence in all MCF7 datasets (**Figure 26**); the vast majority of these factors, approximately 83%, can now be studied either in PDX or in human clinical tissues validating the relevance of these factors *in vivo*. To conclude, our data highlights the method's sensitivity and ability to identify endogenous protein networks from heterogeneous PDX and human tumour material, facilitating the study of known and novel associations in a more clinically relevant environment.

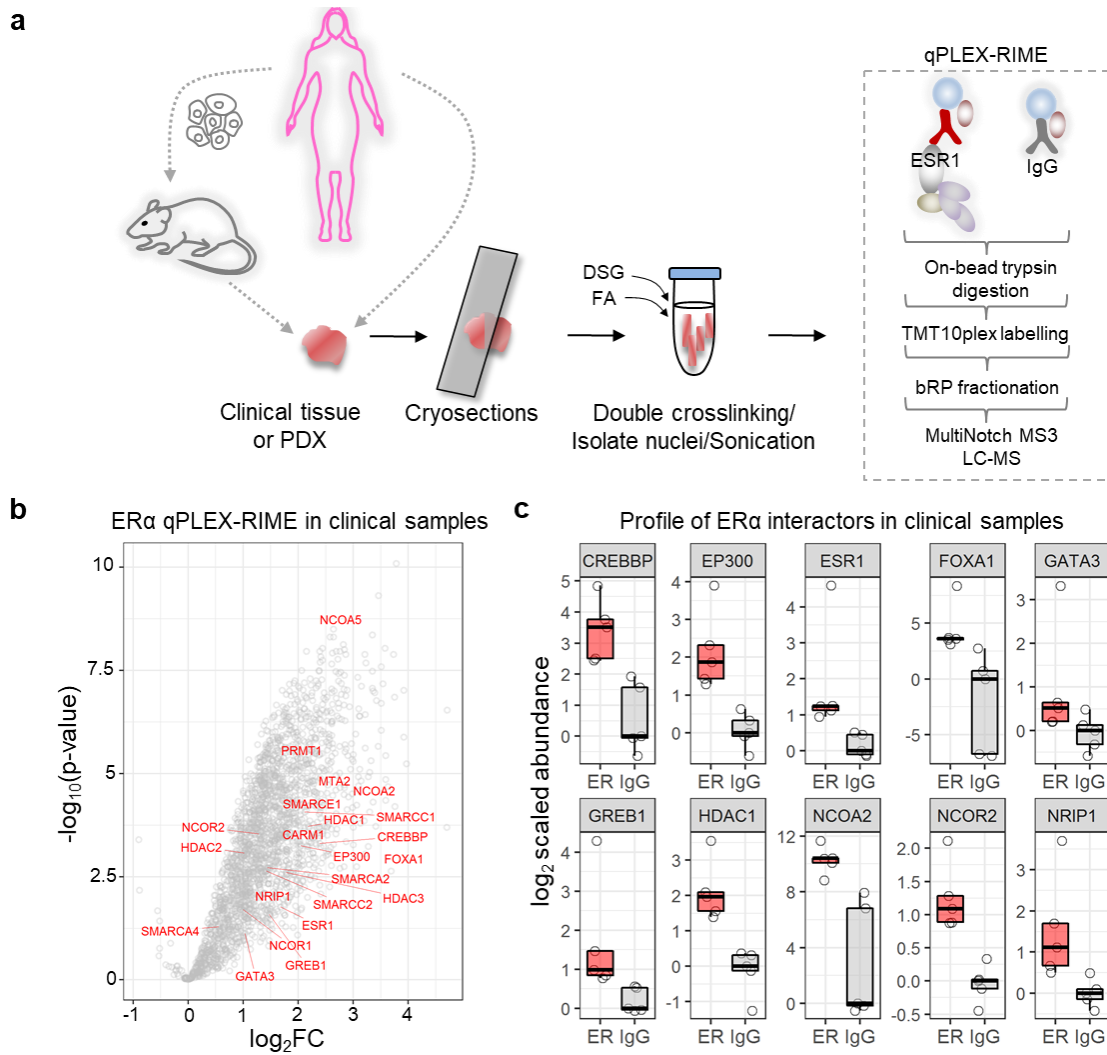


Figure 25. Characterisation of ERα interactors from *in vivo* samples.

a) ERα qPLEX-RIME workflow in human xenograft tissues or in human breast cancer tumours. b) Volcano plot summarising the quantitative results of the ERα interactome in human breast cancer tumours. Several well-known ERα interactors are labelled. c) Box plots illustrating the enrichment of selected known ERα interactors in the ERα samples compared to IgG controls in human breast cancer tissues. The \log_2 values are normalised so that the median of IgGs is zero.

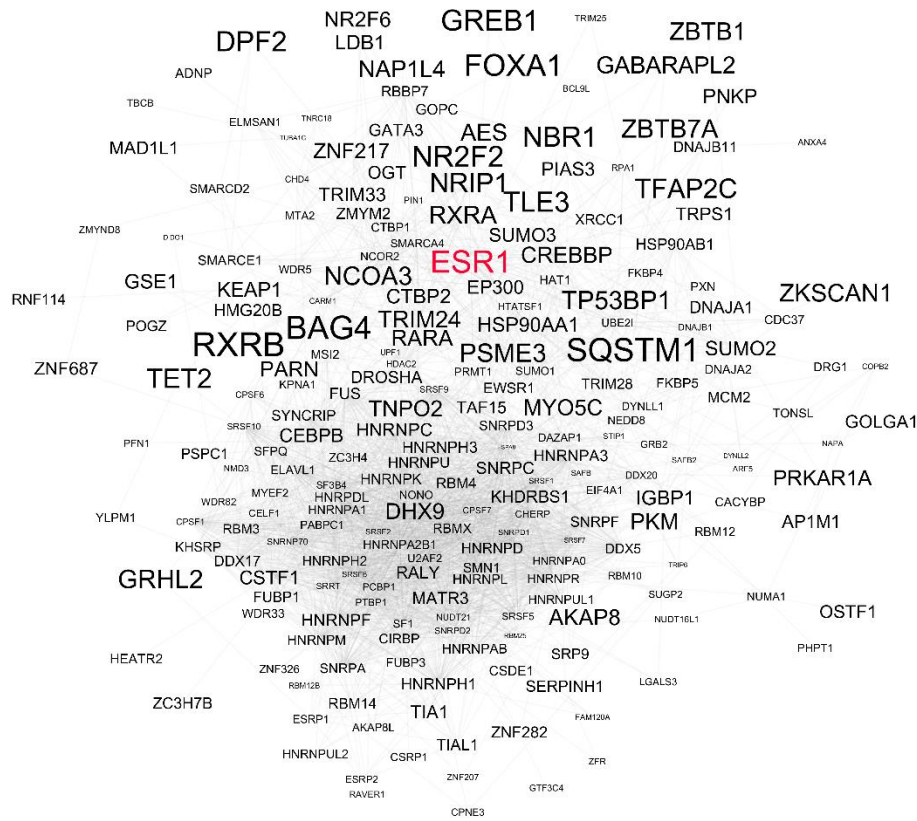


Figure 26. Most frequently enriched ER α interactors.

STRING network of 253 ER α interactors identified consistently across all the qPLEX-RIME analyses performed in MCF7 cells. The font size increases proportionally to the average fold change enrichment of these proteins across all the ER α samples compared to IgG controls.

3.2.4.2 ER α interactome in clinical material with variable expression profile of PR

As we showed in section 3.2.4.1, we detected proteins with high sensitivity, significantly enriched in the bait samples compared to IgGs in clinical samples. Next, we carried out a qPLEX-RIME-based analysis to study the dynamics of specific ER α -associated proteins in human clinical material with variable expression profile of progesterone receptor (PR), that has been correlated with clinical outcome (Blows et al., 2010; Purdie et al., 2014). PR expression is used as biomarker of ER α function and high expression of PR has been correlated with good prognosis in breast cancer. It is a known ER α target gene and it has been shown to interact directly with ER α and to regulate ER α chromatin binding and transcriptional activity (Mohammed et al., 2015). Our collaborators from Netherlands Cancer Institute (Wilbert Zwart lab) collected ten independent human breast cancer tumours, five ER α positive/PR positive (ER α + PR+) and five ER α positive/PR negative tumours (ER α + PR-). Importantly, 30 sections (30 μ m) were obtained per tumour, which is less starting material compared to our previous

application of qPLEX-RIME on the ER α positive clinical tumours for the characterisation of ER α interactome *in vivo* (section 3.2.4.1). This will evaluate the sensitivity of the pipeline and assess whether it can be used to study dynamics of interactomes, where the starting material is limited. To filter for non-specific binding, we used the list of specific interactors that was obtained in the section 3.2.4.1 describing the specificity experiment in human tumours.

The MS3 analysis identified 1,041 specific ER α -associated proteins (FDR<1%) across all multiplexed samples, including the bait protein and known ER α interactors such as GREB1, CARM1 and PRTM1. The identification of ER α and key ER α interactors from small amounts of primary tumour material, highlights the improved sensitivity of the pipeline. Additionally, PR was detected, in line with previous findings showing a physical interaction between these two nuclear receptors (Mohammed et al., 2015). Interestingly, the association between PR and ER α was the most enriched interaction in the ER α + PR+ compared to the ER α + PR- tumours (\log_2 Fold-Change=1.51, adj. p-value<0.1), a change that correlates with the low levels of PR in the latter patient cohort. Other significantly enriched interactions in ER α + PR+ compared to the ER α + PR- tumours were between ER α and TRPS1 protein (\log_2 Fold-Change=1.06, adj. p-value<0.1) and ER α and GREB1. These interactions have been previously detected in ER α -RIME experiments (Mohammed et al., 2013; Serandour et al., 2018). TRPS1 expression has been positively correlated with ER α and PR expression status (Wu et al., 2014) that may explain the loss of interaction in the ER α +PR- patients. GREB1 has been defined as a novel progesterone-responsive gene (Camden et al., 2017), revealing that loss of the interaction may reflect changes in gene expression. Also, it has been described that breast tumours that express both GREB1 and PR have the best prognosis (Mohammed et al., 2013). These findings indicate that the application of qPLEX-RIME can reveal the identification and characterisation of possible prognostic markers for breast cancer. Surprisingly, only a few changes were observed on the ER α interactome (**Figure 27**), suggesting that variable expression of PR may not affect the assembly of ER α complex and its association with different cofactors. Also, further functional assays may give a better insight in role of progesterone expression in the complex dynamics, but this is beyond the scope of this experiment. Taken together, the high sensitivity of the qPLEX-RIME method enabled the detection and quantitative study of ER α interactors in patients with variable expression of PR, revealing important findings.

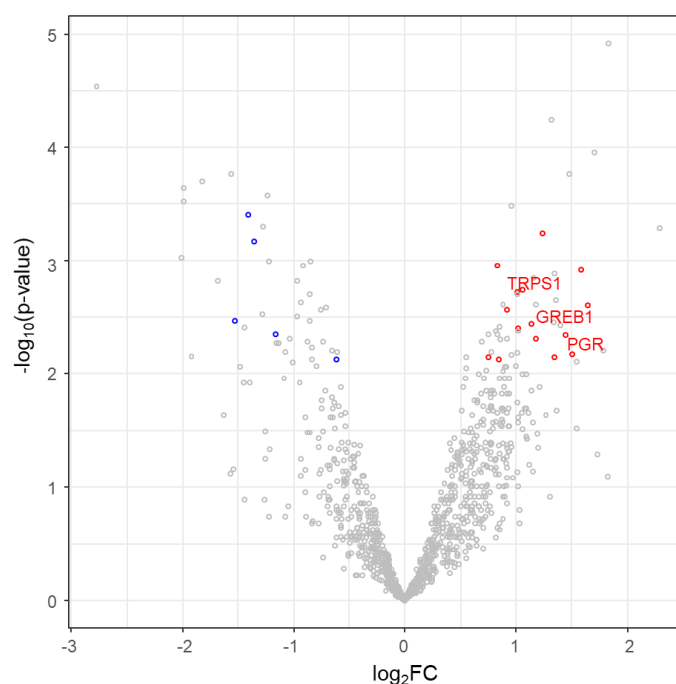


Figure 27. Analysis of tumours with variable expression profile of PR.

Volcano plot summarising the quantitative results from the qPLEX-RIME analysis of ER α + PR+ and ER α + PR- tumours. We have compared ER α + PR+ versus ER α + PR- tumours. The components of the ER α complex that mainly change are highlighted in red.

3.3 Characterisation of a novel ER α -associated transcription factor

3.3.1 Identification of ZNF207 protein by qPLEX-RIME

The application of the qPLEX-RIME method across different samples led to the generation of various interactome datasets that provided in-depth characterisation of the transcriptional regulation in ER α positive breast cancer cells and tumours. Integration and interpretation of the *in vitro* and *in vivo* interactome data revealed a very frequent identification of the ZNF207 protein, a transcription factor that belongs to the zinc finger protein family (Fang et al., 2018). The **Figure 28** summarizes the detection of ZNF207 protein in different interactome experimental datasets. Transcription factors can regulate ER α function, can influence its binding to chromatin and can affect response to therapeutic treatments (Bi et al., 2020; Hua et al., 2018). Because they hold such important roles, we decided to investigate further the role of ZNF207, as it could represent a promising novel co-regulatory transcription factor of ER α . Zinc finger-containing proteins (ZNFs) comprise the largest transcription factor family in the human genome and work together with other transcription factors or co-factors to regulate

development, differentiation and metabolism (Fang et al., 2018; Jen and Wang, 2016). Studies have described an important role of ZNF207 in mitosis in HeLa cells and Glioblastoma multiforme stem cells (GSCs) (Jiang et al., 2014) and more recently in self-renewal and pluripotency of human embryonic stem cells (Fang et al., 2018).

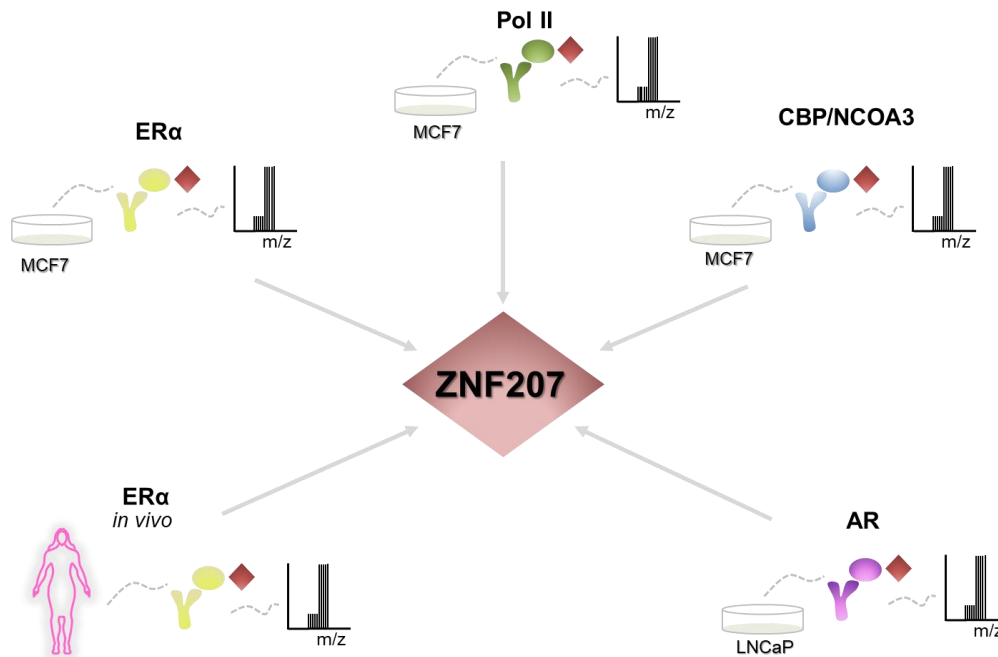


Figure 28. Overview of ZNF207 discovery.

Identification of ZNF207 factor in different datasets of qPLEX-RIME (ER α , Pol II, CBP, NCOA3) or non-quantitative RIME (AR) experiments using as a model cell lines (MCF7 or LNCaP cells) or human clinical samples.

In our data, the identification of ZNF207 coincided with the presence of nuclear receptors, essential coactivators as well as with components of the basal transcriptional machinery. Specifically, ZNF207 was significantly enriched in ER α compared to IgG controls *in vitro* (MCF7: adj. p-value<0.01, log₂Fold-Change=1.32) and *in vivo* (Clinical samples: adj. p-value<0.01, log₂Fold-Change=2.86) (**Figure 29a**). Additionally, the application of label-free quantification using the Minora algorithm on AR (Androgen Receptor) and IgG RIME data in LNCaP cells, confirmed the association of ZNF207 with nuclear receptors. Each pull-down was performed in two biological replicates and ZNF207 protein had a significant enrichment in AR-RIME compared to IgG samples (adj. p-value<0.01, log₂Fold-Change=7.5) (**Figure 29b**).

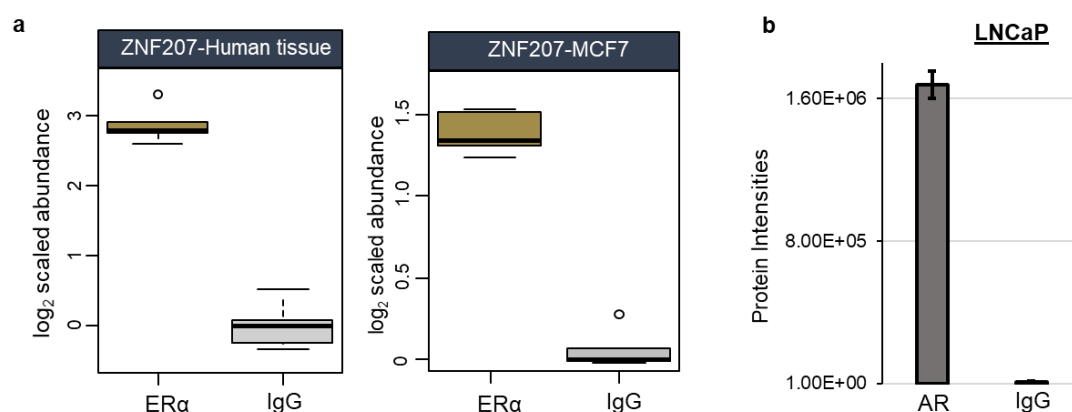


Figure 29. Association of ZNF207 with nuclear receptors.

a) Box plots illustrating the enrichment of ZNF207 in ERα pull-downs (human tissue and cell lines) compared to IgG controls *in vitro* and *in vivo*. The log₂ values are normalized so that the median of IgGs is zero. b) Bar plot displaying the enrichment of ZNF207 in AR pull-downs compared to IgG controls. The protein intensities represent the average of two biological replicates.

Moreover, ZNF207 was significantly enriched in the qPLEX-RIME interaction data of two well-known nuclear receptor co-activators; NCOA3 and CBP, (CBP-qPLEX-RIME: adj. p-value<0.01, log₂Fold-Change=2.67; NCOA3-qPLEX-RIME: adj. p-value<0.01, log₂Fold-Change=2.59), confirming its association with active nuclear receptor complexes (**Figure 30a**). Interestingly, ZNF207 was also identified as one of the top enriched proteins in the RNA polymerase II (Pol II) qPLEX-RIME experiment compared to IgG samples (adj. p-value<0.01, log₂Fold-Change=4.6) (**Figure 30b**). For the Pol II pull-down an antibody that recognized the phosphorylated polymerase II at serine-5 was used, suggesting a potential involvement of ZNF207 in the transcription initiation complex. Taken together, the data support that ZNF207 associates with nuclear receptors and coactivators and may have a distinct and vital role in ERα positive breast cancer but also in prostate cancer, via association with important components of gene transcription. Based on these results we decided to investigate further the functional role of this transcription factor.

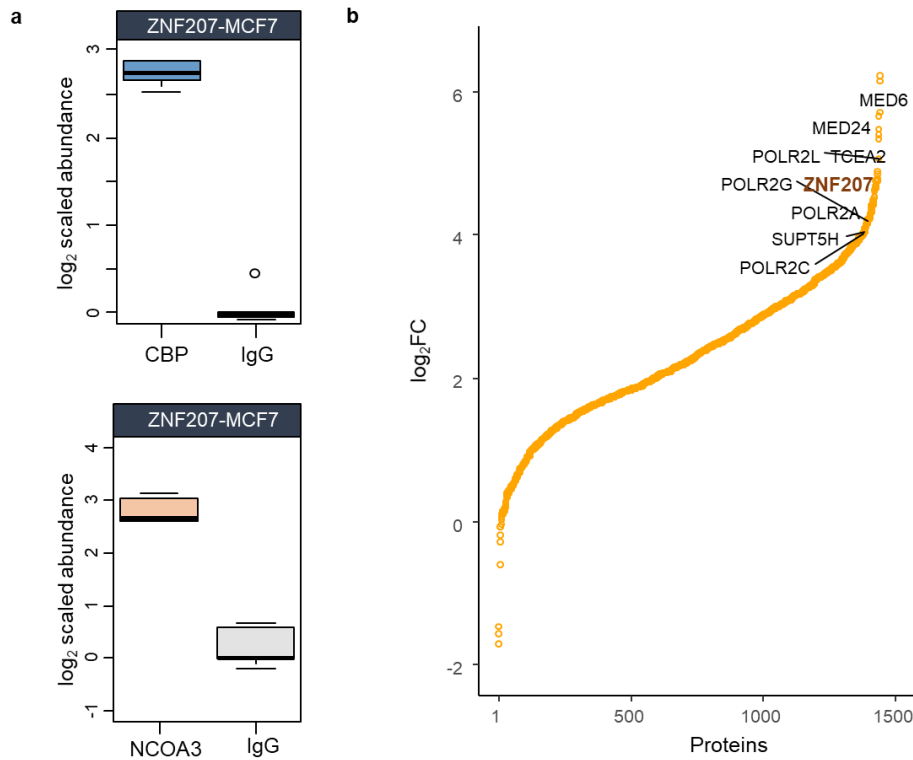


Figure 30. Association of ZNF207 with co-activators and Pol II in MCF7 cells.

a) Box plot illustrating the enrichment of ZNF207 in CBP pull-downs compared to IgGs. b) Box plot displaying the significant association between ZNF207 and NCOA3 compare to IgGs. For all the box plots, the \log_2 values are normalized so that the median of IgGs is zero. c) Scatter plot displaying the significant enrichment of ZNF207 and known key interactors in Pol II pull-downs compared to IgGs.

3.3.2 Functional importance of ZNF207 for cell proliferation

To start delineating the role of this factor in cellular behavior, we looked up published data from a CRISPR high-throughput screening assay (Nagarajan et al., 2020) that was conducted in our lab to identify genes that are important in the proliferation of MCF7 cells. This analysis revealed the depletion of all guide RNAs targeting ZNF207 (**Figure 31a**), indicating an essential role of ZNF207 in cell proliferation ($\text{FDR} < 0.05$). To follow up further on the CRISPR screening findings in different contexts, we developed an assay using small interfering RNA (siRNA) targeting ZNF207. We examined whether the siRNA-mediated ZNF207 knockdown affects the proliferation of cancer and non-cancer cell lines. To this end, different cell lines were transfected with a SMARTpool siRNA targeting-ZNF207 or a non-targeting SMARTpool siRNA control (siNT). Firstly, the efficiency of the siRNA was tested in different concentrations (10nM, 20nM and 30nM) (**Figure 31b**) and at different time points (24h, 48h, 72h and 96h) (**Figure 31c**) in MCF7 cells by Western blot. For further experiments the lowest

concentration of siZNF207 or siNT (10nM) at 48h time point was used, based on the optimal efficiency in decreasing ZNF207 protein expression with lowest cell toxicity.

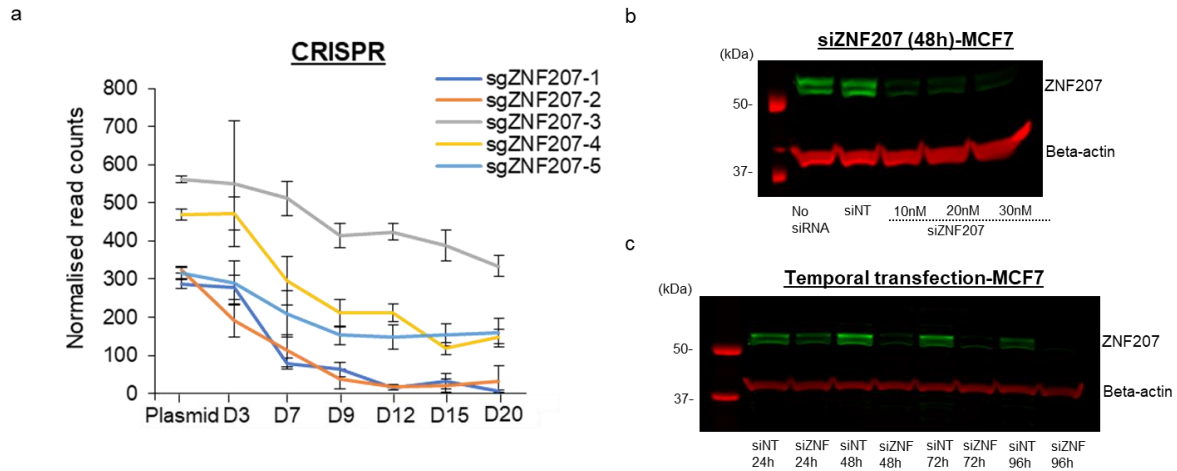


Figure 31. CRISPR screening in MCF7 cells and siRNA transfection conditions.

a) CRISPR screening revealed the temporal depletion of all five guide RNAs targeting ZNF207. Plasmid represents the gRNA counts from control library before infecting to cells. b) Western blot showing the effect of the siRNA-targeting ZNF207 in different concentrations (10nM, 20nM and 30nM) using whole lysate of MCF7 cells. Cells treated with no siRNA or siNT were used as controls. c) Western blot showing the temporal decrease of ZNF207 levels upon siRNA treatment at 10nM. In both Western blot assays, beta-actin was used as loading control. The figure (a) was provided by my colleague Dr Sankari Nagarajan.

Next, using the above optimised transfection conditions, we studied the effect of ZNF207 knockdown on three different ER α positive breast cancer cell lines; MCF7, T47Ds and ZR751. Data showed that ZNF207 knockdown effectively decreased cell proliferation in all three cell lines compared to siNT (**Figure 32**), with a more pronounced effect on the proliferation of MCF7 cells. This independent phenotypic validation of the published CRISPR screen results confirmed the essentiality of ZNF207 in breast cancer cell lines. To test whether this effect is limited to breast cancer cell lines, we also performed a cell proliferation assay in the LNCaP prostate cancer cell line, following ZNF207 knockdown. The results showed a strong decrease in cell proliferation of LNCaP cells (**Figure 33a**), indicating that ZNF207 might be a general essential factor in hormone-dependent cancer cell lines. Interestingly, the profile of cell proliferation following knockdown of ZNF207 is similar between MCF7 and LNCaP cells, with cell growth decreasing significantly after 48h of the transfection, suggesting a strong dependency of these two different cell line models on ZNF207 expression.

To investigate if ZNF207 is essential for cancer cell proliferation in a broader range of cancer cell lines beyond hormone-dependent cancers, we analysed the publicly available data from the Cancer Dependency Map (DepMap) (Meyers et al., 2017), that uncovers gene dependency across hundreds of cancer cell lines using CRISPR and RNAi (RNA interference) screening. Both technologies from the DepMap data revealed that ZNF207 is an essential gene for cell lines across the majority of cancer types. The **Figure 33b** shows the data from the CRISPR screening. Specifically, in the CRISPR screening, 722 out of 739 cell lines that were included in the analysis, showed decrease proliferation and therefore, high dependency on ZNF207 expression. Similarly, in the RNAi assay, the growth of 648 out of 710 cell lines was strongly affected by ZNF207 knockdown, classifying ZNF207 as a common essential gene. Additionally, we assessed the specificity of DepMap results by analysing the screening data for ESR1 gene. Notably, the results confirmed the unique role and essentiality of ER α gene in the proliferation of multiple breast cancer cell lines, highlighting the specificity and accuracy of these public datasets (**Figure 33b**).

Next, we tested the essentiality of ZNF207 in HEK293 cells as a model representing a non-cancer cell line. The efficiency of the knockdown HEK293 was validated with western blot and the proliferation assay showed a significant inhibition of HEK293 growth after 72h of transfection with siRNA-targeting ZNF207 (**Figure 33c & d**). Altogether, the data show that ZNF207 knockdown negatively affects cell fitness of both cancer and non-cancer cells. These results in combination with public RNA expression data from the Human Protein Atlas (<https://www.proteinatlas.org/>), where ZNF207 gene is found constitutively expressed across multiple different cell lines (**Figure 34**), suggest that the functional role of this factor is important in the majority of cellular contexts.

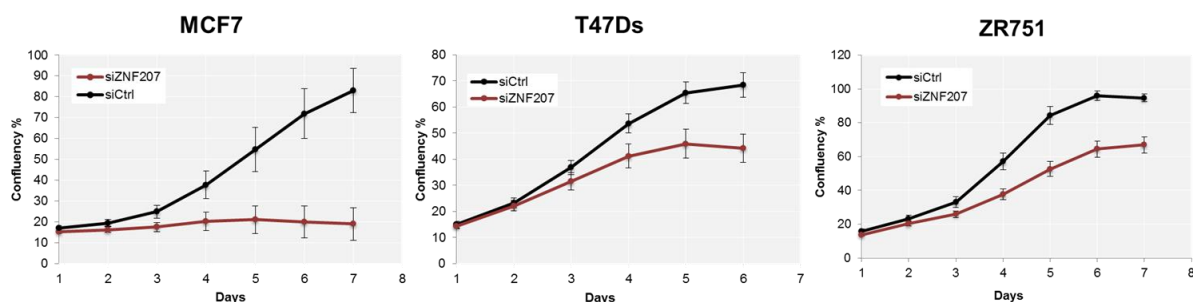


Figure 32. Proliferation assays.

Growth assays displaying the temporal effect on proliferation of three different breast cancer cell lines (MCF7, T47Ds and ZR751) upon ZNF207 knockdown. The assays have been performed in triplicates and the graphs are representative of one of the experiments. The error bars indicate standard deviation (SD).

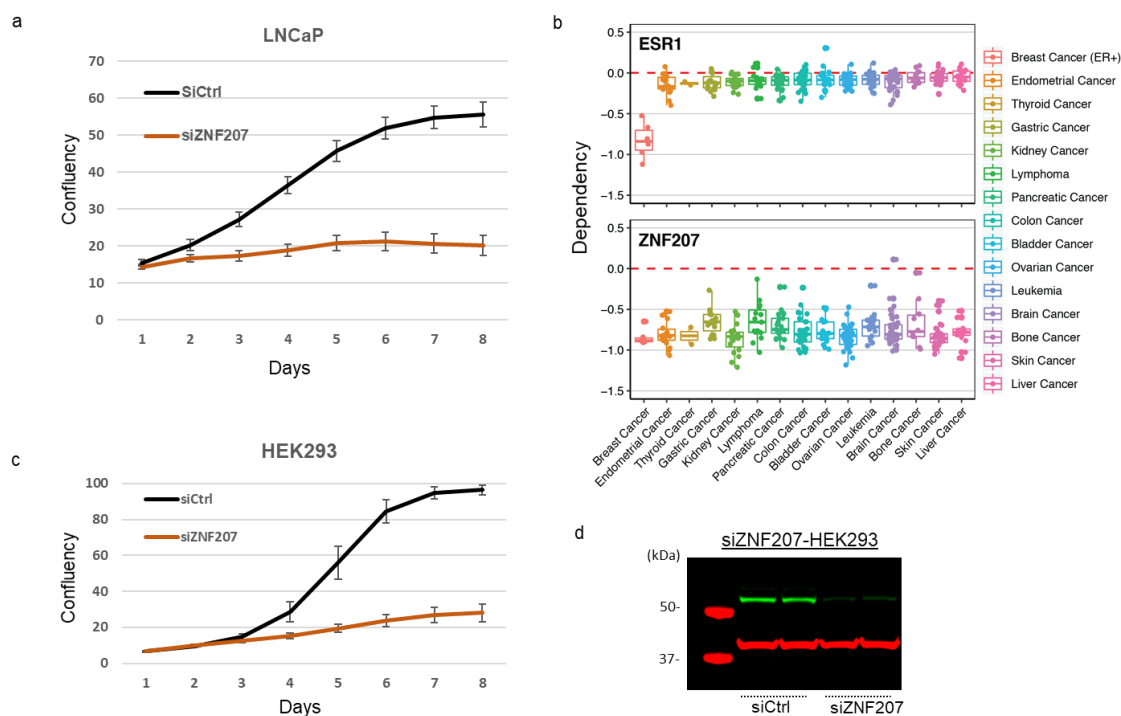


Figure 33. Proliferation assays and analysis of DepMap data.

a) Proliferation assay displaying the temporal effect on proliferation of LNCaP cells upon ZNF207 knockdown. b) CRISPR DepMap data displaying the effect of ERα or ZNF207 knock-out in a broad spectrum of cancer cell lines. c) Proliferation assay displaying the temporal effect on proliferation of HEK293 cells upon ZNF207 knockdown. d) Western blot showing the efficiency of the siRNA assay in HEK293 cells (10nM, 48h). The proliferation assays have been performed in triplicates and the graphs are representative of one of the experiments. The error bars indicate standard deviation (SD).

seq data from vehicle or OHT-treated MCF7 cells, which confirmed the presence of all three isoforms in the MCF7 cells (**Figure 36**). This analysis of two different RNA-seq datasets supports the presence of all ZNF207 isoforms in our system, displaying no isoform specificity. The isoform analysis was performed by the CRUK-CI Bioinformatics core.

We next investigated the association between ZNF207 and ER α by examining whether the transcription of ER α target genes was modulated by ZNF207 knockdown. To this end, genes significantly downregulated by ZNF207 knockdown (adj. p-value<0.05, log₂Fold-Change<-0.5) were compared with genes significantly repressed (adj. p-value<0.05, log₂Fold-Change<-0.5) after 24h OHT treatment using our RNA-seq data from OHT-treated MCF7 cells. We observed a small overlap between the two datasets as only 2% of the down-regulated genes by ZNF207 knockdown (total 41 genes), were overlapped with genes that were repressed by OHT treatment (**Figure 37**), highlighting an alternative effect of the ZNF207 knockdown. Further analysis of the common genes revealed known ER α targets but not an enrichment of specific pathways. Also, the expression levels of ZNF207 were unchanged upon OHT treatment, indicating that ZNF207 is not an ER α target gene.

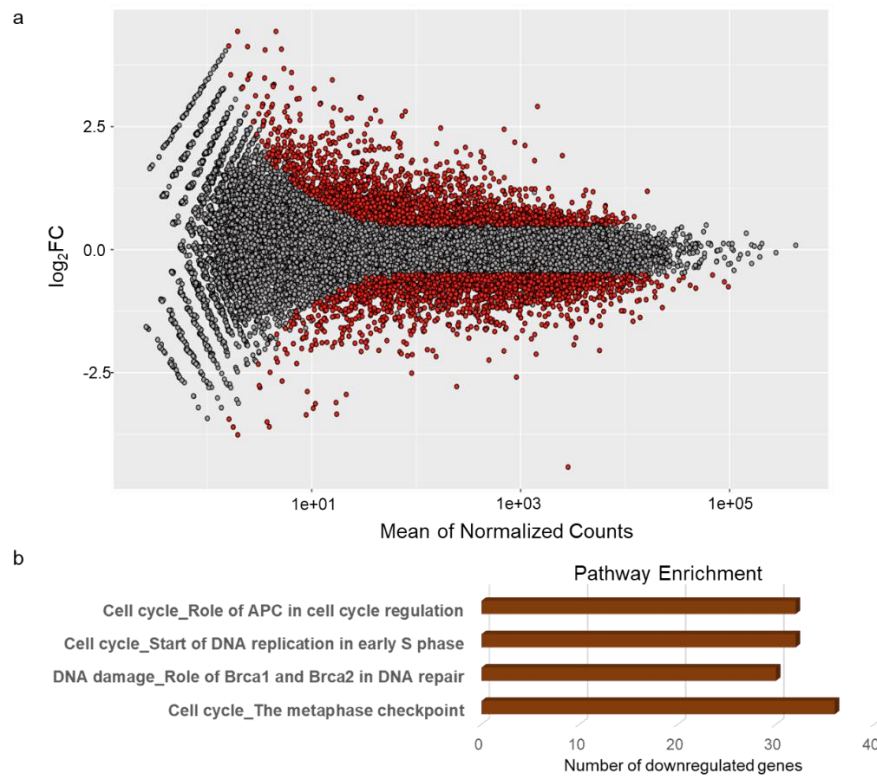


Figure 35. RNA-seq analysis following ZNF207 knockdown.

a) Scatter plot summarising the RNA-seq quantification results. Regulated genes are highlighted in red ($|\log_2\text{Fold-Change}| > 0.5$, adj. $p\text{-value} < 0.05$). b) Pathway enrichment for the downregulated genes identified in the RNA-seq analysis ($p\text{-value} < 0.01$). The pathway enrichment was provided by the Bioinformatics core.

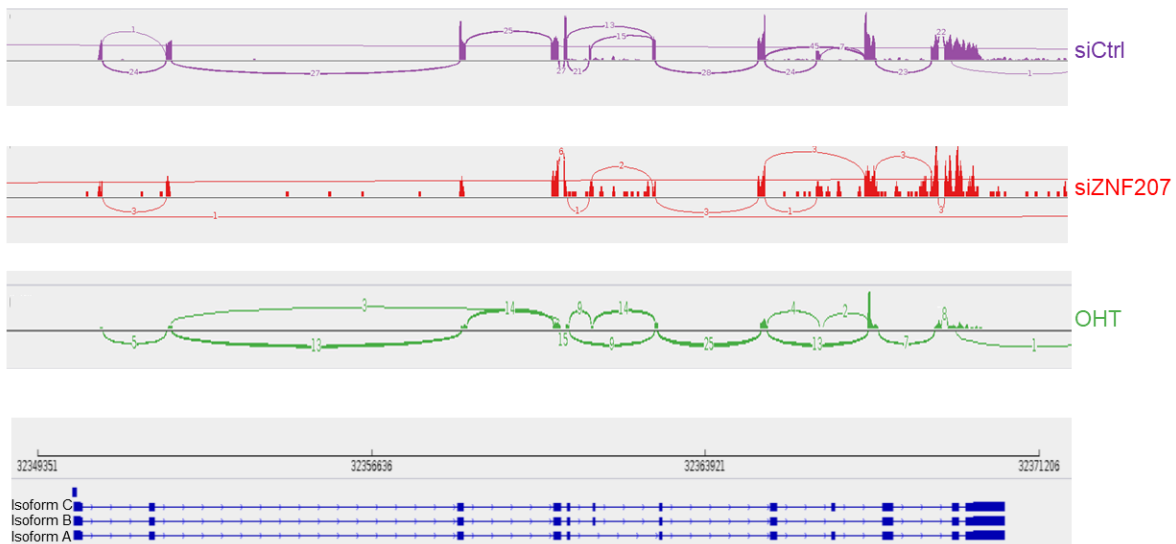


Figure 36. Sashimi plots.

Sashimi plot of ZNF207 gene in MCF7 cells treated with siNT (siCtrl sample-top panel) or siZNF207 (middle panel) or MCF7 treated with OHT for 6h (bottom panel) from RNA-seq data. The frequency of exon skipping is indicated by the numbers. The figures are representative of one of the replicates for each condition. The figure was provided by the Bioinformatics core.

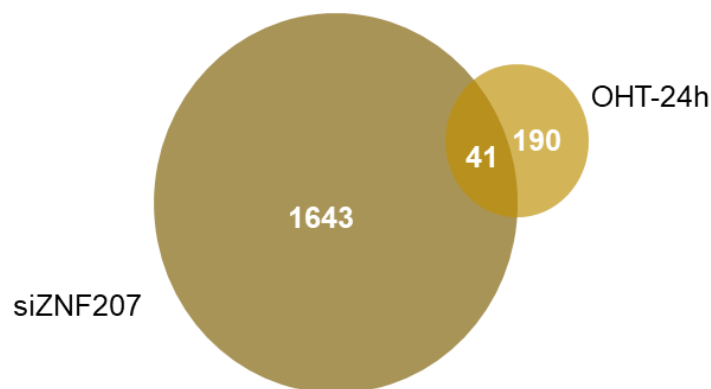


Figure 37. Comparison of two RNA-seq datasets.

A Venn diagram comparison between the significantly downregulated genes identified in the RNA-seq analysis following ZNF207 knockdown or OHT treatment for 24 hours.

To assess whether the changes in gene expression correspond to altered protein levels, a TMT-quantitative whole proteome experiment was conducted in MCF7 cells using the same transfection conditions with the RNA-seq experiments, in independently collected cells. The analysis included five biological replicates of ZNF207 knockdown in MCF7 cells using matched non-targeting siRNA treated cells as control samples. We quantified a total of 8,520 proteins at FDR<1% and detected 308 significantly regulated proteins (adj. p-value<0.05, $|\log_2\text{Fold-Change}|>0.3$) (**Figure 38**), including ZNF207 as the most significantly downregulated protein in the whole proteome dataset (adj. p-value<0.01, $\log_2\text{Fold-Change} = -1.71$). The analysis of the proteomics data did not indicate if any of the ZNF207 isoforms is more prominent, as only shared peptides between the three isoforms were identified. Although, a smaller number of changes were detected at the protein level compared to the transcript level, pathway enrichment analysis confirmed the downregulation of proteins significantly linked to cell cycle (**Figure 38**). Additionally, there was a significant correlation between mRNA and protein suggesting that changes at protein levels are largely regulated at the transcriptional level (**Figure 39**). Importantly, the largest proportion of changes was representing down-regulated proteins/genes, as a shift was observed in the density plot, including cell cycle related factors at both RNA and protein level.

Next, we conducted a phosphoproteome analysis to study if the loss of ZNF207 affects the phosphorylation status of proteins through downstream pathways. We used matched cells pellets with the whole proteome analysis to permit any comparison between the two datasets for the detection of specific changes at phosphorylation status. An additional phosphoenrichment step, using IMAC, was applied in our pipeline to increase the sensitivity

and the discovery of low-abundant phosphopeptides. We quantified approximately 14,000 unique phosphopeptides across the TMT multiplexed set with FDR<1%. Interestingly we did not observe many significant changes in the phosphoproteome or enrichment of specific pathways, suggesting that the negative effect on cell cycle gene transcription most likely is not regulated by prominent phosphorylation mechanisms.

The findings from both the RNA-seq and whole proteome are in line with the essentiality of the ZNF207 and suggest that ZNF207 may have a role in the transcriptional regulation of genes related to cell cycle and cell proliferation. In addition to these observations, the analysis of the phosphoproteome is suggestive of a mechanism not dependent on phosphorylation of central signalling proteins.

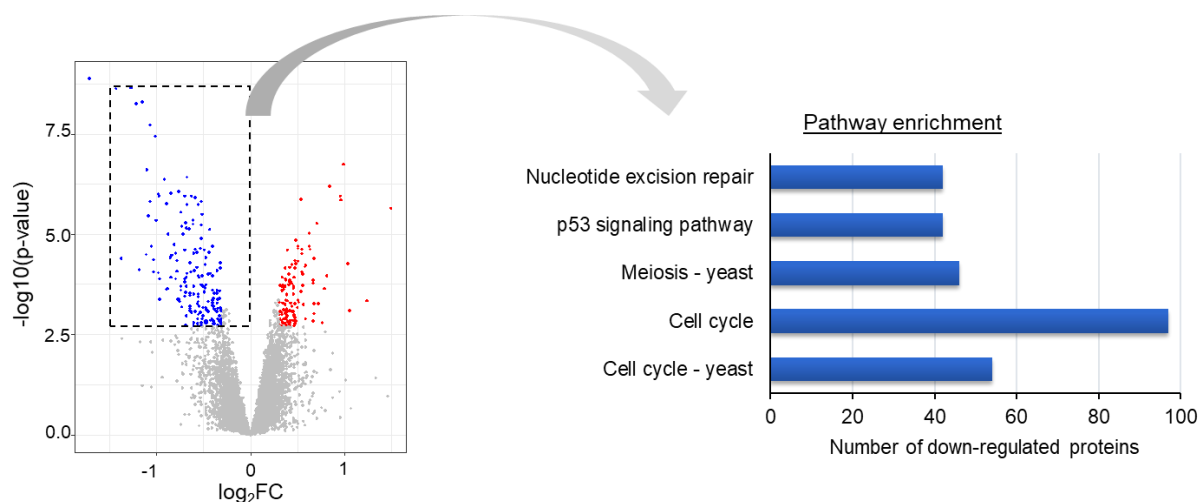


Figure 38. Whole proteome analysis.

Volcano plot (left panel) summarising the quantitative results from the whole proteome analysis in MCF7 cells following ZNF207 knockdown. Significantly regulated proteins are highlighted in red ($|\log_2\text{Fold-Change}| > 0.3$, adj. $p\text{-value} < 0.05$). Pathway enrichment analysis (right panel) for the downregulated proteins showing strong association with cell cycle ($p\text{-value} < 0.01$).

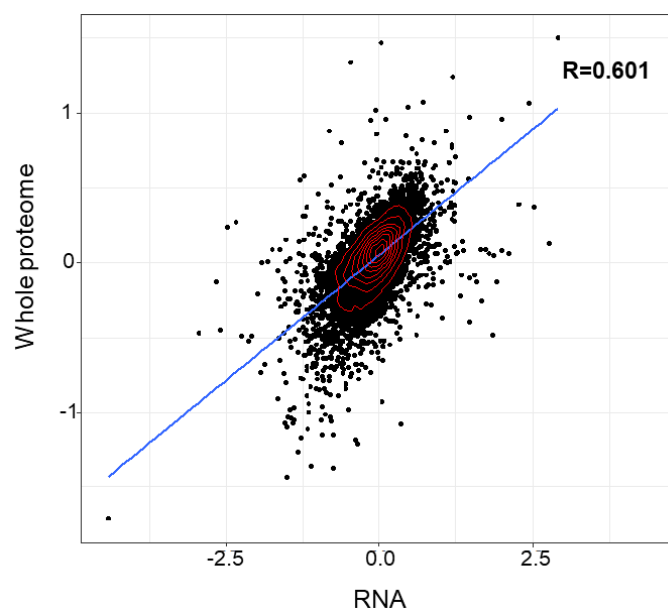


Figure 39. Correlation between protein and gene expression data.

Scatter plot illustrating the correlation of all proteins versus mRNA data following ZNF207 knockdown.

3.3.4 Effect of ZNF207 knockdown on cell cycle progression

ZNF207 was originally discovered as a regulator of mitotic chromosome alignment in HeLa cells and Glioblastoma multiforme stem cells (GSCs) (Jiang et al., 2014). Whether ZNF207 has a role in proper chromosome alignment and in the metaphase to anaphase progression in breast cancer cell lines is unknown. Given that abnormal mitosis can lead to impaired cell proliferation, it was important to assess the potential involvement of ZNF207 in mitosis in MCF7 cells and whether ZNF207 knockdown affects cell cycle progression. To this end, we performed a flow cytometry experiment to determine the relative proportion of cells in various stages of the cell cycle. The experiment was performed in two biological replicates by collecting and treating independently grown MCF7 cells. However, a third replicate will be included to increase the confidence of the findings. The flow cytometry analysis showed that the percentage of cells in G1, S and G2/M cell cycle phases remained unaffected after 48h treatment with siZNF207 at a concentration of 10nM (p-value cut-off <0.01) (**Figure 40a**). This observation suggests that knockdown of ZNF207 for 48h does not significantly affect cell cycle progression in MCF7 cells. Additionally, the absence of an increased G2/M cell population after siZNF207 treatment, indicates that ZNF207 knockdown is unlikely to have a profound effect upon metaphase chromosome alignment in MCF7 cells, since this would be anticipated to increase the time spent in mitosis and result in an increased mitotic cell fraction.

Abnormal chromosome alignment can result in incorrect attachment of chromosomes to the mitotic spindle (Jiang et al., 2014) and lead to chromosome segregation errors. Elevated rates of chromosome segregation errors and subsequent increased aneuploidy may affect cell viability (Janssen et al., 2009). We therefore assessed whether ZNF207 knockdown leads to an increase in chromosomal segregation errors. Asynchronous MCF7 cells grown on coverslips were fixed and stained by immunofluorescence (kinetochores were stained with CREST (centromere) antibodies, microtubules were stained with an anti- α -tubulin antibody and chromosomes were stained with Hoechst 33342). We scored 50 anaphase cells per sample and we did not observe any significant increase in the number of cells making chromosomal segregation errors between the control and ZNF207 knockdown cells (**Figure 40b**). The high number of chromosomal segregation errors in control MCF7 cells is consistent with previously published data quantifying the well documented chromosomal instability of this cell line (Bakhoum et al., 2014; Thompson and Compton, 2008). For this experiment I worked together with Rebecca Burrell from the Carroll group.

To conclude, the changes we detected in cell proliferation assays for various cell models, following ZNF207 knockdown, cannot be explained by altered cell cycle progression or mitotic abnormalities generating highly aneuploid cells. Despite excellent knockdown of ZNF207 at 48h, no effect on cell cycle profile or chromosome segregation was seen. Given the strong effect on cell viability after 72h of ZNF207 knockdown, we would expect a significant mitotic phenotype at 48h if this was the explanation for reduced viability. While we cannot exclude that a longer siRNA treatment might reveal a mitotic phenotype, our observations after 48h treatment with siZNF207 indicate an alternative mechanism for the impact of siZNF207 on cell proliferation. Notably, recent published data in human embryonic stem cells described that ZNF207 regulates pluripotency and developmental gene expression without observing any significant role in mitotic chromosome alignment (Fang et al., 2018). These previous findings in combination with our data suggest that ZNF207 may have additional important functions besides a possible role in mitosis.

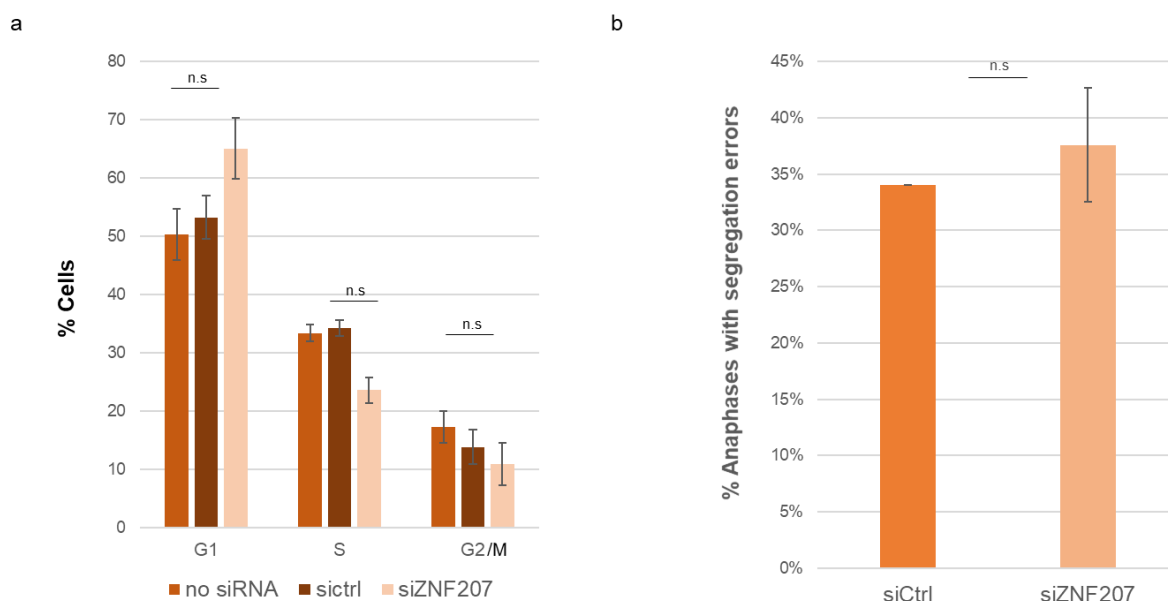


Figure 40. FACS and microscopy experiments.

a) Bar plot summarising the cell population at different cycle stages upon treatment with siCtrl, siZNF207 or without siRNA. Phases of the cell cycle: G1 phase, S phase (synthesis) and G2/M (interphase/mitosis). b) Bar plot showing the percentage of anaphase cells with chromosomal segregation errors in cells treated with siCtrl or siZNF207. Both experiments were performed in two biological replicates and the error bars indicate standard deviation (SD). Both bar plots show no significant (n.s) difference between the conditions. The cell scoring was performed by Rebecca Burrell. Student's t-test p-value cut-off < 0.01).

3.3.5 ZNF207 antibody testing using RIME and ChIP-seq

As our data suggest a different role for ZNF207 compared to what has been described in the literature so far, we next sought to study the ZNF207 interactome in depth and characterise the proteins associated with this transcription factor in MCF7 breast cancer cells, by performing non-quantitative RIME and qPLEX-RIME experiments. This may help to better understand its function and whether ZNF207 interacts with specific proteins or protein complexes. Firstly, we conducted non-quantitative RIME experiments on untreated, asynchronous MCF7 cells to test the performance of six different antibodies; Abcam (ab123322), Atlas (HPA017013), Santa Cruz D12 (sc-271943), Genetex (GTX116214), Life technologies (LF) (PA530641), Santa Cruz E2 (sc-271942) against ZNF207, as there are no available ZNF207 interactome data in the literature. We included species-matched IgG negative controls (IgG-rabbit and IgG-mouse) to filter for non-specific interactors. To evaluate the pull-down efficiencies of the different antibodies, we compared the sequence coverage of the bait protein and the detection of specific ZNF207-associated proteins, by subtracting proteins identified in both IgG samples.

All antibodies achieved a good sequence coverage for ZNF207 except the D12 antibody (**Figure 41a**), where the bait protein was detected with one unique peptide. The Atlas and Genetex antibodies gave the highest peptide coverage for the bait protein, 7 and 6 peptides respectively (**Figure 41a**), that represents 85-100% of the theoretical number of tryptic peptides that can be detectable for ZNF207 (**Figure 41b**). The sc-E2 antibody achieved a good coverage for the bait protein (3 peptides, 40% of theoretical peptide coverage) and gave the highest number of unique peptides for ZNF207-associated proteins that are known components of the ER α complex and the transcription machinery, including co-activators, Pol II initiation and elongation factors (**Figure 41c**). A further comparison of the detected specific associated proteins in each of the three best-performing antibodies (Atlas, Genetex, sc-E2), showed a good overlap and interestingly revealed a high number of interactors uniquely identified with the sc-E2 antibody (**Figure 41d**).

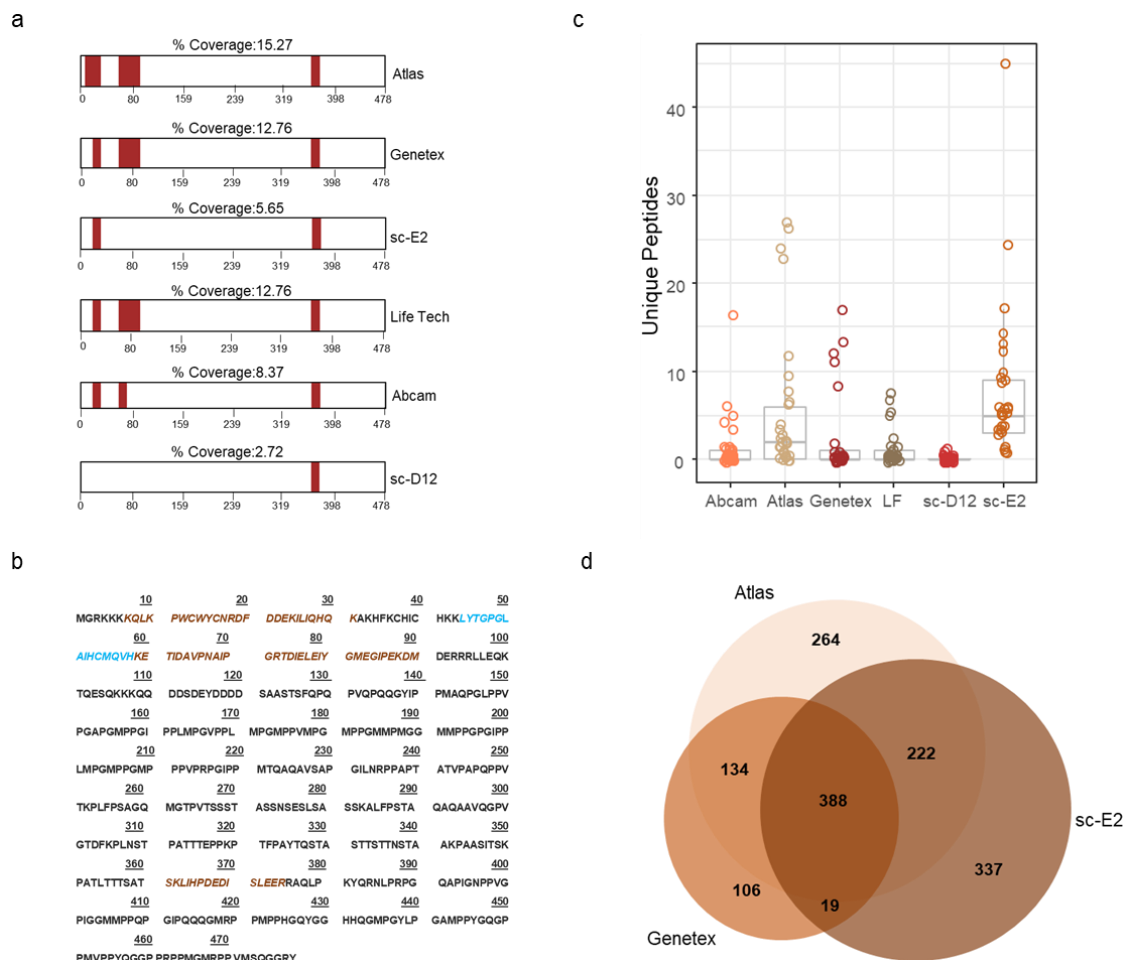


Figure 41. Comparison of 6 antibodies specific for ZNF207 with RIME.

Protein sequence coverage of ZNF207 achieved by the use of different antibodies. b) Sequence of ZNF207 protein. Letters in italics indicate tryptic peptides theoretically detectable from the mass spectrometer. Orange font indicates tryptic peptides that were consistently identified in ZNF207-RIME experiments and blue font indicates amino acid sequence that was not detectable in our assays. c) Boxplot of the number of identified unique peptides for components of the ER α complex and the transcription machinery across the 6 antibodies. d) Venn diagram comparison of the specific interactors identified using Atlas, Genetex and sc-E2 antibodies.

Despite the similarities between the RIME and ChIP-seq protocols, we have observed that antibodies performing robustly in RIME, do not always perform comparably in ChIP-seq. For this reason, we tested in parallel the performance of five out of the six antibodies (we excluded sc-D12) in a ChIP-seq experiment in untreated, asynchronous MCF7 cells, including ER α as a positive control. The ER α control ChIP-seq detected a high number of binding events (60,868) (**Figure 42a**), demonstrating the good quality of the assay. Surprisingly all antibodies performed poorly in terms of the number of binding events with the exception of sc-E2 antibody, which achieved a large number of total peaks (45,643) (**Figure 42a**). Additionally,

the sc-E2 antibody has been used before for ZNF207 ChIP-seq experiments (Fang et al., 2018), in line with the conclusion from our evaluation experiment. **Figure 42b** displays snapshots of ChIP-seq tracks at GREB1 gene locus, a key ER α -regulated gene, demonstrating the high signal and the low background for the positive control. Additionally, the quality and signal intensity of the peaks confirmed the good enrichment achieved using the sc-E2 antibody and the poor binding profiling obtained using the other four ZNF207 antibodies. Interestingly, this pilot experiment revealed a strong overlap between ER α and ZNF207 as 70% of the ZNF207 binding sites was overlapping with ER α binding, a common feature of co-localisation of chromatin-associated factors (**Figure 42c**). This finding is in an agreement with the qPLEX-RIME data showing an association between ZNF207 and the ER α complex. As the sc-E2 antibody performed well in both RIME and ChIP-seq assays, we used it for all the subsequent experiments that were conducted to study the function of ZNF207.

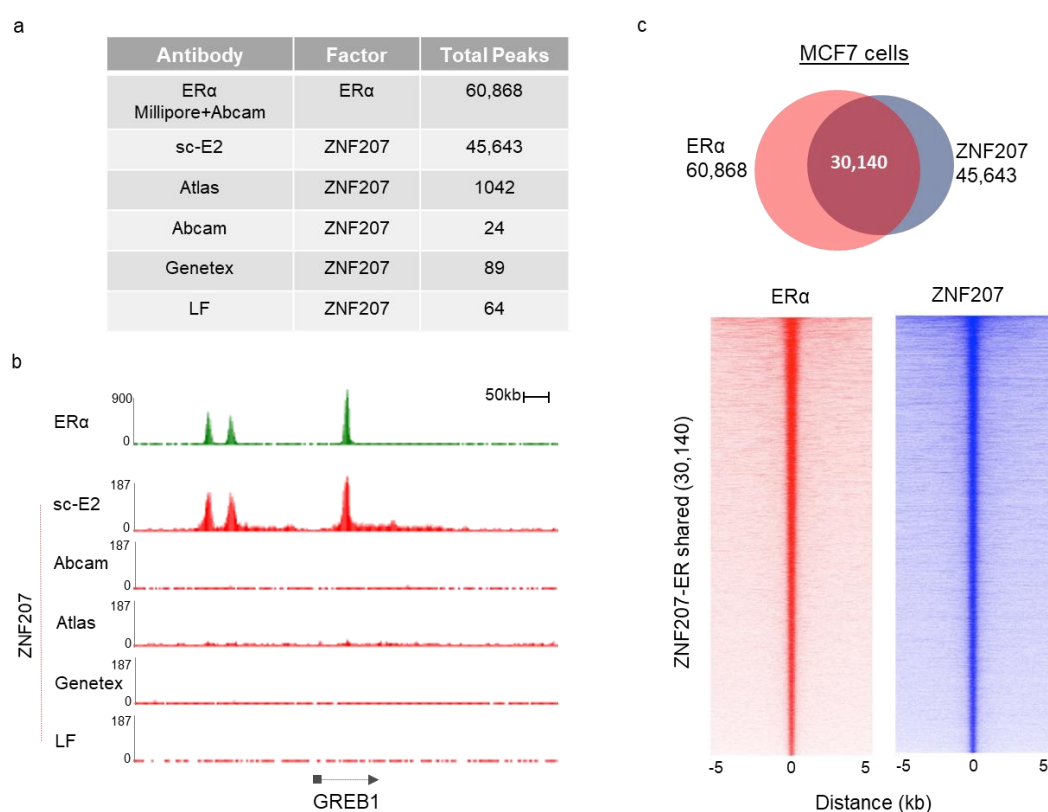


Figure 42. ZNF207 antibody testing using ChIP-seq in MCF7 cells.

a) The number of binding sites for 5 different ZNF207 antibodies (sc-E2, Atlas, Abcam, Genetex, LF) along with the number of binding sites detected for ER α . b) UCSC genome browser tracks of ChIP-seq signal near a key ER α target gene (GREB1) for the five ZNF207 antibodies and ER α . c) Venn diagram showing the overlap of binding sites between ZNF207 (using sc-E2 antibody) and ER α (top). Heatmap of the ZNF207-ER α shared binding sites (30,140) (bottom).

3.3.5.1 Characterisation of ZNF207 interactome

Having optimised the pull-down conditions, we next studied the ZNF207 interactome by conducting ZNF207 qPLEX-RIME analysis in five independent biological replicates using the sc-E2 antibody. An equal number of IgG control samples was included to filter for non-specific binding. Data analysis using the qPLEXanalyzer tool confirmed the strong association of ZNF207 with ER α and well-characterised ER α interactors such as FOXA1, GATA3, NCOA3 and CBP (adj. p-value<0.01, log₂Fold-Change>1), in line with the overlap that we observed in the binding sites between ZNF207 and ER α by ChIP-seq. Interestingly, we identified six out of the twelve Pol II subunits in the pull down experiment (RPB1-4, RPB7, RPB9); all subunits were significantly enriched in the ZNF207 pull-downs compared to IgG samples (adj. p-value<0.01, mean log₂Fold-Change=2.13). Additionally, transcription initiation and elongation factors were detected, including TAF3 (Louder et al., 2016), SPT6 and SPT5 (Wada et al., 1998) (adj. p-value<0.01, log₂Fold-Change>1), highlighting a strong association of ZNF207 with the Pol II complex. Importantly, we found that ZNF207 interacts with several subunits of the mediator complex, such as MED1, MED12, MED4 and MED15 (adj. p-value<0.01, log₂Fold-Change>1). Notably, Mediator subunits are rarely detected in our pull-down experiments of ER α and associated proteins, highlighting the importance and specificity of the association between ZNF207 and the Mediator complex. **Figure 43a** summarises the quantitative findings from the ZNF207 qPLEX-RIME analysis.

To validate the interaction between ZNF207 and Mediator complex, we executed non-quantitative RIME experiments on different critical Mediator subunits (MED1, MED12, MED4, MED14 and MED26), covering the three out of the four distinct modules of the complex termed the middle, tail and CDK8 kinase module (Soutourina, 2018). We successfully identified the bait proteins with high peptide sequence coverage in the respective RIME samples along with most of the Mediator subunits and other well-known interactors, including Pol II, validating the efficiency of the pull-downs (**Figure 43b**). Notably, the RIME data confirmed the association between ZNF207, different subunits of the Mediator complex and components of the transcription initiation complex. Several coactivator proteins were also detected, validating the association between Mediator subunits and transcription coactivators, which is essential for the recruitment of Pol II and components of the general transcription machinery to promoters (Soutourina, 2018). Taken together, the interactome data demonstrate that ZNF207 has a strong association with ER α complex but also with subunits of the Mediator

complex, suggestive of a role of ZNF207 in Mediator function and regulation of gene transcription initiation.

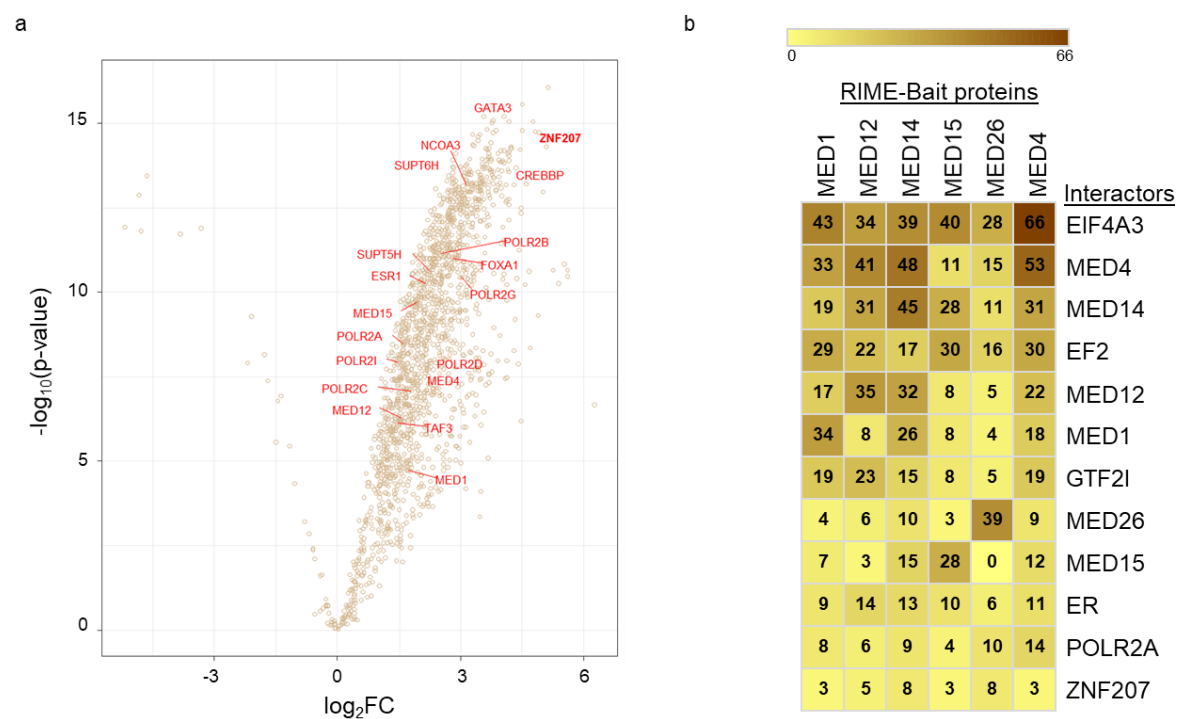


Figure 43. ZNF207 interactome and RIME for Mediator subunits.

a) Volcano plot summarising the quantitative data from the ZNF207 qPLEX-RIME. The bait protein along with components of ERα complex and the transcription machinery complexes are highlighted in red. b) Heatmap displaying the sequence coverage (%) for bait proteins (MED1, MED12, MED14, MED15, MED26, MED4) and different Mediator-associated proteins identified in non-quantitative RIME experiments.

3.3.6 ZNF207 knockdown does not affect the assembly of the ERα complex

Our interactome data have revealed the association of ZNF207 with ERα and well-described ERα interactors (such as FOXA1, NCOA3 and CBP). To determine whether ZNF207 knockdown has any effect on the assembly of ERα complex, we conducted an ERα qPLEX-RIME experiment on MCF7 cells, that have been treated with siZNF207 or siNT (5 biological replicates for siZNF207 or siNT-treated cells). We quantified 2,197 proteins with FDR<1% across the multiplexed TMT set; 122 proteins were significantly regulated (adj. p-value<0.1) (**Supplementary Data 1**). Interestingly, we observed a significant enrichment in the complex for transcription corepressors following ZNF207 knockdown; (1) CBX8, a component of the Polycomb PRC1 complex (Tang et al., 2019), (2) BAZ2A, a component of the NoRC

(nucleolar remodeling complex) that leads to heterochromatin formation and transcriptional silencing (Gu et al., 2015), (3) the MTG16 corepressor (Kumar et al., 2015) and the (4) RCOR3 which is a member of the REST (repressor element-1-silencing transcription factor) corepressor (CoREST) family (Xue et al., 2011). This could be partly due to differential expression of these cofactors, as they were detected to be upregulated in our RNA-seq and whole proteome data, suggesting that knockdown of ZNF207 causes an increase on the gene expression of transcription co-repressors. Similarly, other significantly regulated ER α -associated proteins (adj. p-value<0.1) detected in this dataset were mainly linked to cell cycle and DNA damage and their expression levels in both our RNA-seq and whole proteome data were changed.

For further analysis and filtering for specific interactors, we used the ER α qPLEX-RIME experiment (5-ER α qPLEX-RIME pull-downs compared to 5-IgG controls), that we conducted in MCF7s to characterise the ER α interactome. We defined the significant specific ER α -associated factors from this dataset (adj. p-value<0.1) and compared this against a representative well-known STRING interactome of proteins directly associated with ER α (**Figure 44a**). Next, we examined if any of these factors were significantly changed as ER α interactors upon ZNF207 knockdown compared to siNT. Notably, the knockdown of ZNF207 did not affect the assembly of the ER α complex, as most of the well-known interactors and ER α itself remained unaffected (**Figure 44b**).

Overall, our data demonstrate that ZNF207 associates significantly with ER α and components of the ER α complex, but is not required for the assembly of the ER α complex, as the well-known ER α interactors were unchanged following ZNF207 knockdown. Interestingly, we observed an enrichment in the ER α complex of specific co-repressors proteins, a likely consequence of an up-regulation of transcription of these factors. This upregulation may contribute to the significant number of down-regulated genes that were detected in the gene expression analysis. Notably, well-known ER α -target genes were only a small portion of the total number of decreased genes following ZNF207 knockdown, validating the qPLEX-RIME findings of minimal impact on the ER α complex in the absence of ZNF207. We speculate that ZNF207 was detected in ER α pull-down experiments given that nuclear receptors comprise a large family of transcription factors interacting with Mediator subunits as well as with members of the transcription machinery via chromatin loops that bridge distinct regulatory complexes (Kang et al., 2002). To conclude, our data suggest a more generic role for ZNF207, likely associated not only with ER α but with other factors to regulate gene expression and proliferation.

qPLEX-RIME, 158 proteins were significantly regulated (adj. p-value<0.1) compared to siNT samples (**Supplementary Data 2**). Gene Ontology enrichment analysis of all quantified proteins revealed a loss of interaction between MED1 and proteins that were related to transcription initiation and elongation. Specifically, we detected a significant loss on the interaction between MED1 and Pol II (adj. p-value<0.1) as well as with general transcription factors that are required for the assembly of the Pol II pre-initiation complex and the phosphorylation of the Pol II C-terminal domain, including TATA box binding protein-associated factors (TAFs) (Tora, 2002). Additionally, the enrichment analysis showed a significant loss of the Mediator complex; the ZNF207 knockdown led to loss of interaction between MED1 and other Mediator subunits. In this experiment, all 33 subunits of the Mediator complex were detected and interestingly all showed an overall negative enrichment in the complex following ZNF207 knockdown, suggesting a global negative effect on the Mediator complex assembly in the absence of ZNF207.

Previously published studies have shown that Mediator complex affects Pol II elongation through interactions with the super elongation complex (SEC) (Donner et al., 2010; Takahashi et al., 2011). Interestingly, in our interactome data we observed a decrease in the interaction between MED1 and members of the super elongation complex. Specifically, ZNF207 knockdown decreased the association between MED1 and the positive elongation factor b (P-TEFb), which comprises CDK9 as a catalytic subunit and cyclin T1 (CCNT1) or T2 as a regulatory subunit (Luo et al., 2012). We detected CDK9 and CCNT1 subunits in our data and both showed a decreased association with MED1 (adj. p-value<0.1). Additionally, we identified other members of the SEC that had the same profile as P-TEFb, including AFF4, MLLT3 (AF9) and MLLT1 (ENL) (Luo et al., 2012). Notably, AFF4 and AF9 were the top depleted proteins in our data (AF9: adj. p-value<0.01, log₂Fold-Change=-0.89; AFF4: adj. p-value<0.01, log₂Fold-Change=-0.72), suggesting a significant effect in the elongation process when ZNF207 was silenced.

Analysis of the MED12 qPLEX-RIME data showed 211 regulated proteins (adj. p-value<0.1) in the MED12 siZNF207 pull-downs compared to siNT samples (**Supplementary Data 3**). Pathway enrichment analysis of the data, reproducibly identified the effect on the Mediator complex and the loss of interaction between the different subunits, which was also observed in the MED1 qPLEX-RIME. Interestingly, both quantitative experiments showed a loss of interaction with the MED14 subunit, the architectural and functional backbone of the Mediator complex (Cevher et al., 2014), suggesting a strong effect on the complex assembly and stabilisation. Structural and functional studies have indicated that the association between

subunits of the head or middle subcomplexes and MED14 is critical to reconstitute a functional and structural core that strongly associates with Pol II (Cevher et al., 2014). This finding may explain the loss of interaction between MED1 and Pol II in our data as the knockdown of ZNF207 had a significant effect on the Mediator subunits, including MED14.

The findings from the above interactome studies were reproducibly confirmed by the MED4 qPLEX-RIME data. Specifically, we identified 158 significantly regulated proteins (adj. p-value<0.1) in the siZNF207 compared to siCtrl samples (**Supplementary Data 4**). The CORUM analysis confirmed the negative enrichment for Mediator complex (score=-0.72), indicative of a more general effect on the Mediator assembly upon ZNF207 knockdown as observations are consistent when different Mediator subunits were used as bait proteins. **Figure 45** illustrates the effect on Mediator complex, Pol II core complex, elongation and initiation factors in MED1, MED4 and MED12 qPLEX-RIME experiments in MCF7 cells, following ZNF207 knockdown.

Next, we evaluated whether the disruption of the Mediator complex, following ZNF207 knockdown, is specific for breast cancer models or a more general effect. To this end, we conducted two qPLEX-RIME experiments on MED1 and MED14 subunits in HEK293 cells, a non-cancer cell line, which as we showed above, its proliferative rate was decreased upon ZNF207 knockdown. Five independent biological replicates of MED1 or MED14 qPLEX-RIME samples were collected from HEK293 cells treated with either siRNA specific for ZNF207 or with siNT. We quantified 2,638 and 2,274 proteins for MED1 and MED14 qPLEX-RIME respectively across knockdown and control samples at peptide FDR<1%. The analysis of the MED1 and MED14 qPLEX-RIME data, showed a significant loss of 403 and 98 proteins (adj. p-value<0.1) respectively in the siZNF207 pull-downs compared to siNT samples (**Supplementary Data 5 & 6**). Gene Ontology enrichment analysis of both datasets revealed the effect on Mediator complex, RNA polymerase II core complex and general transcription factors (**Figure 45**), in agreement with the findings from the qPLEX-RIME experiments in MCF7 cells. Overall, the data suggest a disruption of the Mediator complex and the implication of ZNF207 on its assembly. We also observed the loss of interaction between the bait proteins and members of the super elongation complex such as AFF4 (MED1 qPLEX-RIME/AFF4: adj. p-value<0.01, log₂Fold-Change=-0.64, MED14 qPLEX-RIME/AFF4: adj. p-value<0.01, log₂Fold-Change=-0.84), validating the effect on the elongation process upon the decrease on ZNF207 expression levels.

Knockdown of ZNF207 in cancer and non-cancer cell line models resulted in a less efficient assembly of the Mediator subunits along with a decreased interaction between Mediator

complex and important factors that control transcription initiation and elongation. These results suggest an important role of ZNF207 in the Mediator complex assembly that affects the recruitment of factors that play a crucial role in the initiation and elongation steps of gene expression by facilitating PIC establishment and function.

3.3.8 ZNF207 knockdown affects the assembly of the Pol II complex

In order to follow up on the results from the qPLEX-RIME experiments for different Mediator subunits and to assess the effect on the general transcription machinery, given that the Mediator complex has been showed to be involved in most stages of Pol II activation (Allen and Taatjes, 2015; El Khattabi et al., 2019), we carried out a qPLEX-RIME experiment on the phosphorylated form of Pol II at serine-5. As before, five independent biological replicates were collected from MCF7 cells transfected with an siRNA-pool targeting-ZNF207 or siNT. We quantified 1,725 proteins (FDR<1%) across the multiplexed set of all different conditions. Of these, 275 proteins were significantly regulated compared to siNT control samples (adj. p-value<0.1) (**Supplementary Data 7**). Notably, in all qPLEX-RIME experiments (Pol II, MED1, MED12, MED4 and MED14) where cells were treated with siRNA targeting-ZNF207, the most significant loss was observed in the interaction between the different bait proteins and the ZNF207 (Average log₂Fold-Change=-1.5, adj. p-value<0.01), validating the efficacy of the siRNA assay.

Gene set enrichment analysis (GSEA) showed decreased interaction between Pol II and general transcription factors (GTFs), such as TFIIA and TFIIH that are required for basal transcription (Johnson et al., 2002). Additionally, enrichment analysis using the CORUM database of protein complexes (Giurgiu et al., 2019) showed a significant decrease in the interaction with TAF factors that belong to the TFIID complex such as TAF3, TAF1 and TAF6, that have a vital role in facilitating the assembly of the pre-initiation complex (PIC) at the core promoters (Johnson et al., 2002; Louder et al., 2016), in line with the findings of the MED1 qPLEX-RIME experiment. Furthermore, a reduced interaction between Pol II and the complex formed by TFIID and TFIIA (DA complex) and the DAB complex that contains TFIID, TFIIA and TFIIB transcription factors was observed.

Additionally, a negative enrichment for members of the SEC complex was detected, including AFF4, ENL and P-TEFb factors (adj. p-value<0.1, mean log₂Fold-Change =-0.4), further supporting the MED1 data and the effect of ZNF207 knockdown on both transcription and elongation processes. We also observed an enrichment for corepressors, a finding consistent

with the ER α interactome analysis and the RNA-seq and whole proteome analysis. The enrichment of known corepressor proteins, following ZNF207 knockdown, may suggest a potential mechanism of preventing gene transcription. Dot plot (**Figure 45**) summarising the findings from the Pol II qPLEX-RIME in MCF7 cells upon ZNF207 knockdown.

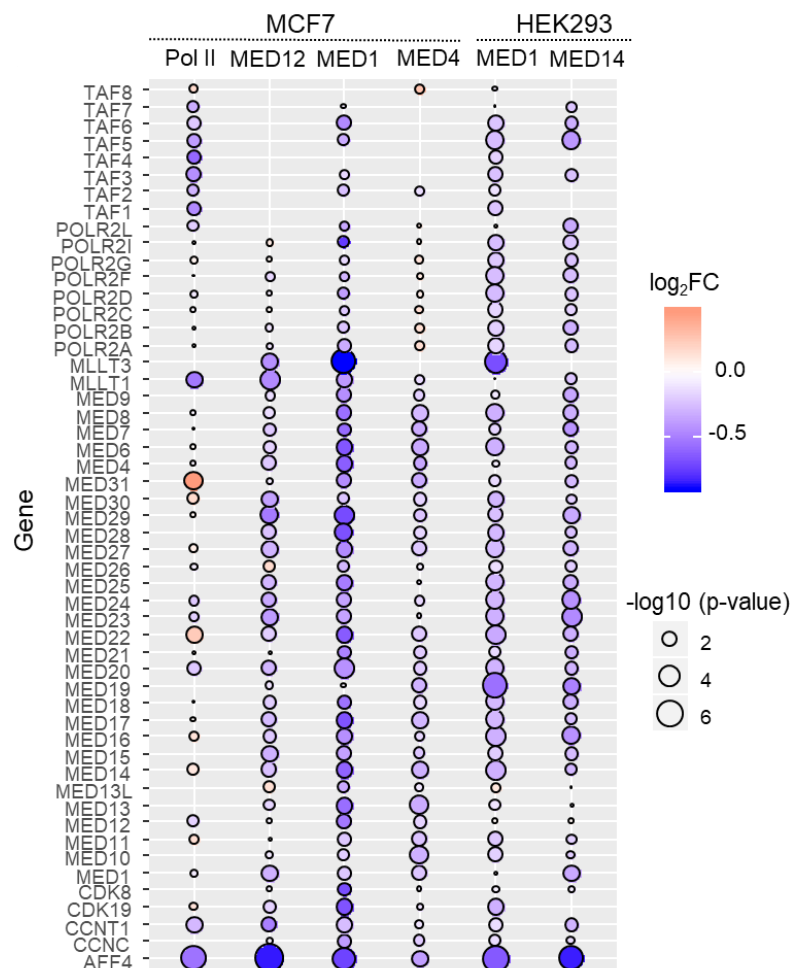


Figure 45. Overview of qPLEX-RIME experiments after ZNF207 knockdown.

a) Dot plot summarising the effect on initiation factors, elongation factors, Mediator subunits and Pol II core complex detected in MED1, MED12, MED4 and Pol II qPLEX-RIME experiments in MCF7 cells and MED1 and MED14 qPLEX-RIME experiments in HEK293 cells following ZNF207 knockdown. The colour of the dots illustrates log₂FC calculated by comparing siZNF07 versus sictrl samples and the size of the dot highlights the significance of the findings in -log₁₀(p-value).

To verify whether the changes identified in core factors of the transcription machinery, by the qPLEX-RIME analysis, are specific for the respective complexes or result from changes in

total protein levels, we compared the findings of all the qPLEX-RIME experiments with the whole proteome analysis. We only compared the qPLEX-RIME data from MCF7 cells as the full proteome was performed using the same cell model. The comparison confirmed that the changes in the Pol II machinery and Mediator complex were specific to changes in recruitment of the complexes, since total protein levels were stable (**Figure 46**). Notably, many of the gene expression changes that we observed were probably linked to the disassembly on these complexes upon decrease of the ZNF207 levels, as these factors have a vital role in the regulation of gene transcription. The results indicate an important role of ZNF207 in the regulation of gene transcription, as knockdown of this factor is globally affecting the assembly of the Mediator complex and the interaction with transcription/elongation factors and Pol II, that is required for the formation of the PIC and the transcription elongation.

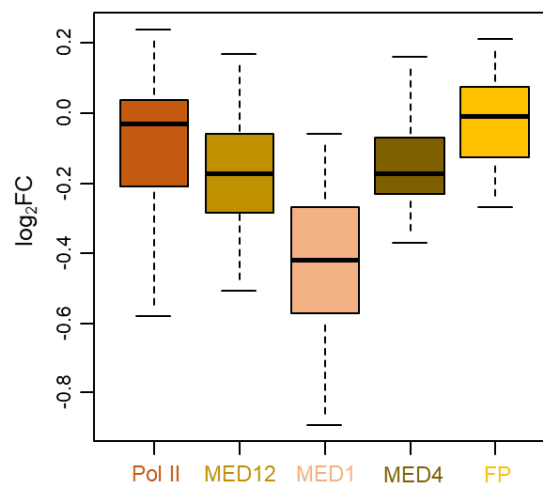


Figure 46. Comparison of qPLEX-RIME changes with full proteome.

Bar plots illustrating \log_2FC for the most significantly depleted proteins detected in the different qPLEX-RIME experiments (Pol II, MED12, MED1, MED4) and their respective \log_2FC detected in the full proteome (FP) following ZNF207 knockdown in MCF7 cells.

3.3.9 siZNF207 does not affect chromatin loops

It has been suggested that the Mediator complex has an important role in gene transcription by being involved in the formation and stabilisation of DNA chromatin loops that connect promoters with distal elements (Kagey et al., 2010). To investigate whether the changes in the Mediator complex, following ZNF207 knockdown, affect the chromatin loops tethering promoters to enhancers, we performed a promoter capture Hi-C experiment (PCHi-C) targeting

21,000 promoters (Schoenfelder et al., 2018). Three biological replicates were collected for each condition (siZNF207 or siNT) using MCF7 cells. A total number of 30×10^6 cells per sample was fixed by adding formaldehyde at a final concentration of 2% and the chromatin was digested by the restriction enzyme HindIII (Mifsud et al., 2015; Schoenfelder et al., 2018). Although this process generates small DNA fragments, the basic three-dimensional architecture of chromatin is maintained by the DNA-protein crosslinks introduced during the fixation step. For the PCHi-C experiment, we conducted the 18-days protocol that was publicly available (Schoenfelder et al., 2018) and we monitored the quality of the libraries at several steps.

At the early steps of the protocol, we carried out two quality controls to assess the PCHi-C library quality. Firstly, we evaluated the quality and quantity of the libraries by checking the size of the different libraries into an E-Gel (agarose Gel). All samples ran in the gel as a narrow band over 10kb size (**Figure 47a**), indicating the good quality and integrity of the library. Next, we tested the ligation efficiency by performing a PCR digest assay using primer sequences (AHF64+AHF66, myc locus and HIST1) for quality control of human Hi-C libraries, that have been published previously (Schoenfelder et al., 2018). A successful ligation would result in the disappearance of the original HindIII restriction site and the formation of a new NheI recognition site (**Figure 47b**). Indeed, we achieved a complete NheI digestion of ligation products, highlighting the high-quality of the library (**Figure 47b**). After the successful assessment of all quality steps and the completion of the protocol, the libraries were sequenced on the Illumina NovaSeq (Zhou et al., 2019) to an average depth of ~1 billion reads per sample.

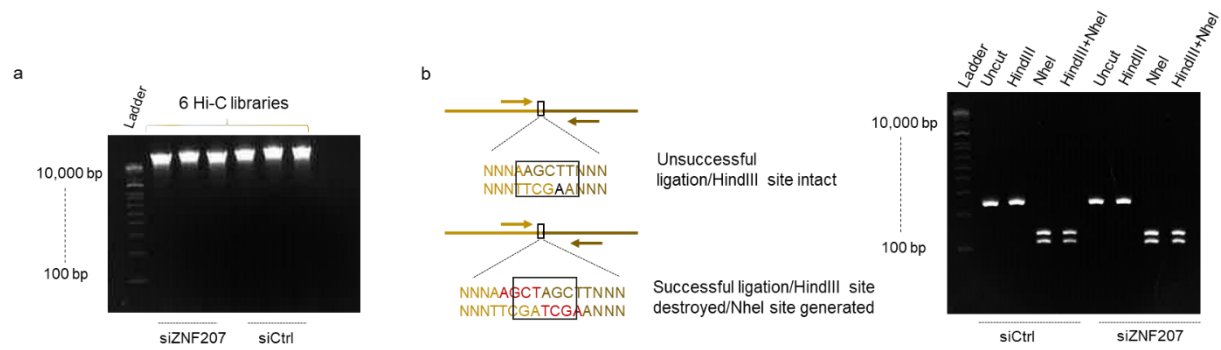


Figure 47. PCHi-C pre-sequencing quality controls.

a) Monitoring the library integrity and ligation by running 200ng of each library on a 2% E-gel. Each DNA library (total six) run as a band over 10kb. b) Left, schematic of DNA sequence after Hi-C ligation following unsuccessful (top) or successful (bottom) dNTP Klenow fill-in of restriction junctions and subsequent ligation. The figure adapted from Schoenfelder et al., 2018. Right, representative example of HindIII, NheI and HindIII/NheI restriction digests of our Hi-C ligation products using the AHF64+AHF66 Dekker primer pair. The figure is representative of two Hi-C libraries from the two different conditions (siCtrl+siZNF207). An undigested library that contained only water was included as control. For both gels a DNA Ladder was used to monitor the DNA size.

After sequencing, the quality of the Capture Hi-C data was assessed using HiCUP, a pipeline that was designed for mapping, filtering and processing Hi-C data (Wingett et al., 2015). The downstream analysis of the PCHi-C data was performed by the Bioinformatics core. Firstly, for improving the mapping efficiency, Hi-C reads that contain sequences that align to two separate regions for the genome, were truncated at the putative Hi-C junction ligation. The paired-end read data were mapped against the hg38 reference genome and filtered for experimental Hi-C artefacts, invalid di-tags and identical di-tags to avoid incorrect interpretations regarding the genomic structure. The valid pairs and unique di-tags in our sequences were more than 70% and 60% respectively across all different samples. Notably, the percentage of unique di-tags was lower compared to published Hi-C data (Wingett et al., 2015), which can be explained by the very deep sequencing, generating 7.0 to 10.0 million di-tags with both ends uniquely mapped to the human reference genome. Additionally, the capture efficiency was between 67-73% (Total-Captured/Di-tags-Processed*100), validating the high-quality PCHi-C libraries.

Next, the input of the HiCUP pipeline was imported into the CHiCAGO (Capture Hi-C Analysis of Genomic Organization) algorithm (Cairns et al., 2016) to filter for robust promoters-distal elements interactions based on the number of reads and distance between the two elements. Interactions with a CHiCAGO score ≥ 5 were considered as high-confidence

interactions and used for downstream statistical analysis. Specifically, we detected 289,202 and 274,009 significant interactions for the siZNF207 and siCtrl samples respectively between promoters and promoter-interacting regions. Analysis of CHiCAGO other ends, which are regulatory regions that interact with the bait (promoter), showed the enrichment over expected values for active histone marks (H3K4me1, H3K4me3, H3K27ac) in both samples (siZNF207 and siCtrl), in line with the assumption that DNA loops are formed between promoters and distal regulatory regions such as enhancers. **Figure 48a** summarise the chromatin features of promoter-interacting fragments detected using CHiCAGO for the ZNF207 knockdown samples.

Specifically, more than 90% of the significant promoter-genome contacts were between promoters and distal elements (**Figure 48b**). To assess further these interactions, we integrated our PCHi-C data with published epigenome datasets. We examined different histone modifications (Active enhancers: Distal H3K4me1 and H3K27ac, Inactive enhancers: Distal H3K4me3 and no H3K27ac) and we found that ~ 10% of the significant interactions were between promoters and active enhancers and ~ 18% were between promoters and inactive enhancers. These percentages of enhancer-promoter contacts are in line with previously published data (Schoenfelder et al., 2015), supporting that other types of elements may have a role in promoter regulation. Approximately 5-8% of the interactions (23,291 for siZNF207 and 14,067 for siCtrl samples) were between two promoters (**Figure 48b**), suggesting that promoters can act as regulatory elements for distal genes. Further data interpretation highlighted a correlation between high enrichment for active histone marks (H3K4me1, H3K4me3, H3K27ac) and high levels of promoter expression (**Figure 48c**). In contrast, low promoter expression was correlated with low levels of active histone marks and an enrichment of the repressive histone mark H3K27me3 (**Figure 48c**). These results indicate that promoters of highly expressed genes interact with distal regions that are enriched for active histone markers, whereas promoters of low expressed or silenced genes interact with regions enriched for repressive histone marks that are associated with heterochromatin formation. This pattern of the epigenetic modifications at distal interacting sites has been observed before (Schoenfelder et al., 2015), indicating the finding of functional and relevant loops in our dataset.

Confirming the high quality of the libraries and the consistency in our data with previously published Hi-C data, we next performed a differential analysis of the PCHi-C data using the R package Chicdiff (Cairns et al., 2019). Notably, the statistical analysis revealed no significant difference in promoters-distal regulatory elements interactions between ZNF207-knockdown

and control samples (**Figure 48d**). Despite, the significant effect on the Mediator complex assembly, the analysis showed no global effect on the contact frequency of the long-range promoter interactions. These findings are consistent with recently published work arguing that Mediator and Pol II are not required to bridge regulatory DNA regions (El Khattabi et al., 2019). The authors from El Khattabi et al, showed that the Mediator complex does not create a stable topological bridge but regulates promoter-enhancer contacts indirectly, via recruitment of transcription factors and co-activators (El Khattabi et al., 2019). In line with these findings a more recent study confirmed the minimal effect on genome architecture following the degradation of the MED14 subunit, which acts as a central Mediator scaffold (Jaeger et al., 2020). Overall, our data support that ZNF207 knockdown affects the assembly of the Mediator complex, the recruitment of transcriptional regulatory proteins, including Pol II, general initiation and transcription elongation factors and the regulation of the initiation and elongation stages of transcription, without disrupting any of the DNA chromatin loops.

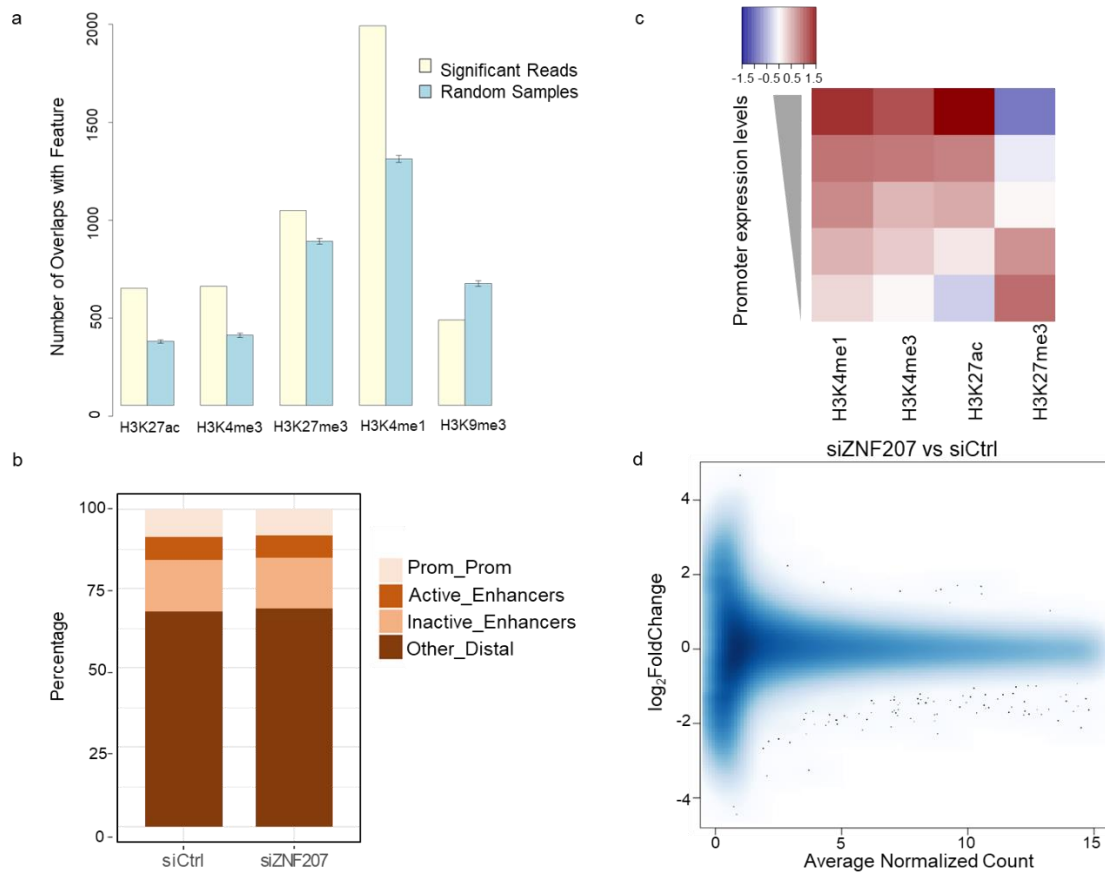


Figure 48. Promoter capture Hi-C data.

a) Chromatin features of promoter-interacting fragments detected using CHiCAGO. Chromatin features were obtained from the ENCODE project. b) Stack bar plot showing the distribution of different significant promoter-genome interactions. c) Heat map showing the enrichment/depletion for histone modifications and the correlation with promoter expression levels (promoter expression levels have been divided into 5 classes). d) Scatter plot summarising the differential analysis results using the Chicdiff package. The figures have been provided by Kamal Kishore.

3.3.10 Genomic characterisation of ZNF207 function

To study the potential genomic interplay between ZNF207, general transcription factors, Mediator complex, ER α , FOXA1 and Pol II, we performed chromatin immunoprecipitation sequencing for ZNF207, MED1, Pol II, TAF3, ER α and FOXA1 in asynchronous MCF7 cells treated with siZNF207 or siNT, using five independent biological replicates. The efficiency of the siRNA assay was validated for all five replicates with western blot before the ChIP-seq experiment (**Figure 49a**). It has been shown that cohesin depletion affects the promoter-promoter and promoter-enhancer pairs, supporting the role of cohesin to tether regulatory elements (El Khattabi et al., 2019). We therefore conducted ChIP-seq for SMC1A (a subunit of the cohesin complex) to investigate whether the chromatin binding of cohesin is affected or

not following ZNF207 knockdown and if the presence of cohesin is enough to stabilise the chromatin loops even if Mediator complex is affected. Matched inputs from both conditions were included and peaks were called using MACS2 (Zhang et al., 2008). Peaks occurring in at least two out of the five independent replicates were considered for downstream analysis. This resulted in the identification of 63,086, 42,422, 38,250, 65,059, 55,050 73,017 and 77,994 binding sites for ZNF207, MED1, Pol II, TAF3, ER α , FOXA1 and SMC1A respectively. The analysis of the ChIP-seq data was performed by the CRUK Cambridge Institute Bioinformatics core and Igor Chernukhin.

Firstly, we investigated the distribution of binding sites for ZNF207 and we observed that the majority of ZNF207 binding sites were detected in intergenic and other gene distal regions (**Figure 49b**). Interestingly, although only a small fraction (13%) of ZNF207 binding events were detected in promoters, the strongest binding density for ZNF207 binding sites was clustered around the transcription start sites (TSS), in line with previously published data (Fang et al., 2018) (**Figure 49c**). The binding sites for siZNF207 samples on TSS sites had lower binding density compared to the siCtrl samples, validating the siRNA efficiency (**Figure 49c**). To investigate further the binding intensity profile of ZNF207 in our cell model system, we subdivided all ZNF207 binding sites into promoter and non-promoter sites (peaks 1250 bp downstream and 250 bp upstream of the TSS were defined as promoters and peaks outside of this region but within 50,000 bp of the TSS were defined as enhancers). In the same analysis we included all the factors to study their binding intensity profile and compare it with the ZNF207 binding. We observed binding of ZNF207 to both promoters and enhancers with the strongest occupancy in promoters (**Figure 49d**). The profiles were similar for the other factors, with TAF3 and Pol II showing a very strong binding to promoters. It has been reported before that 67% of RNA Pol II sites (Carroll et al., 2006) and 87% of TFIID-binding sites (Kim et al., 2005) map to promoter-proximal regions, validating our findings.

Next, we compared the ChIP-seq binding events between ZNF207, Pol II, MED1 and TAF3, confirming a significant overlap, which indicates colocalisation of ZNF207 with these master regulators of transcription (**Figure 50a**). This finding is consistent with the strong association of these factors that was observed in the qPLEX-RIME and non-quantitative RIME experiments. Interestingly, on the TSS sites, we detected a very strong overlap between ZNF207, MED1, Pol II and TAF3 (**Figure 50b**), highlighting the strong association between these factors. We also observed a colocalisation between ZNF207, ER α and the pioneer factor FOXA1 with 44% of the ZNF207 binding sites (28,162 sites) overlapping with both ER α and FOXA1 (**Figure 51a**). This finding is in line with the interactome data, where ZNF207 was

observed to significantly associate with ER α and components of the ER α complex. The higher number of unique binding events for FOXA1 may be explained from the fact that we had a higher number of detected binding sites for the FOXA1 ChIP-seq compared to the other two factors. Motif analysis of ZNF207 binding sites outside of promoters confirmed the association between ZNF207, ER α and FOXA1, with forkhead and AP-1 motifs being the most significant sequences within ZNF207 peaks (**Figure 51b**). The motif analysis for ZNF207 binding sites in promoters showed an enrichment for zinc finger transcription factor motifs (SP1, KLF5 and EGR1) that all belong to the C2H2 zinc finger class as ZNF207 (**Figure 51b**).

The association between binding of ZNF207 and the different factors was further validated with Pearson correlation coefficient analysis, which showed that the binding profiles of MED1, TAF3 and ZNF207 were highly correlated mainly in promoters (0.96 positive linear correlation) (**Figure 51c**). As such, ZNF207 appears to have two distinct binding patterns, with co-occupancy at promoters with MED1 and TAF3 and co-binding at distal enhancers with ER α and FOXA1.

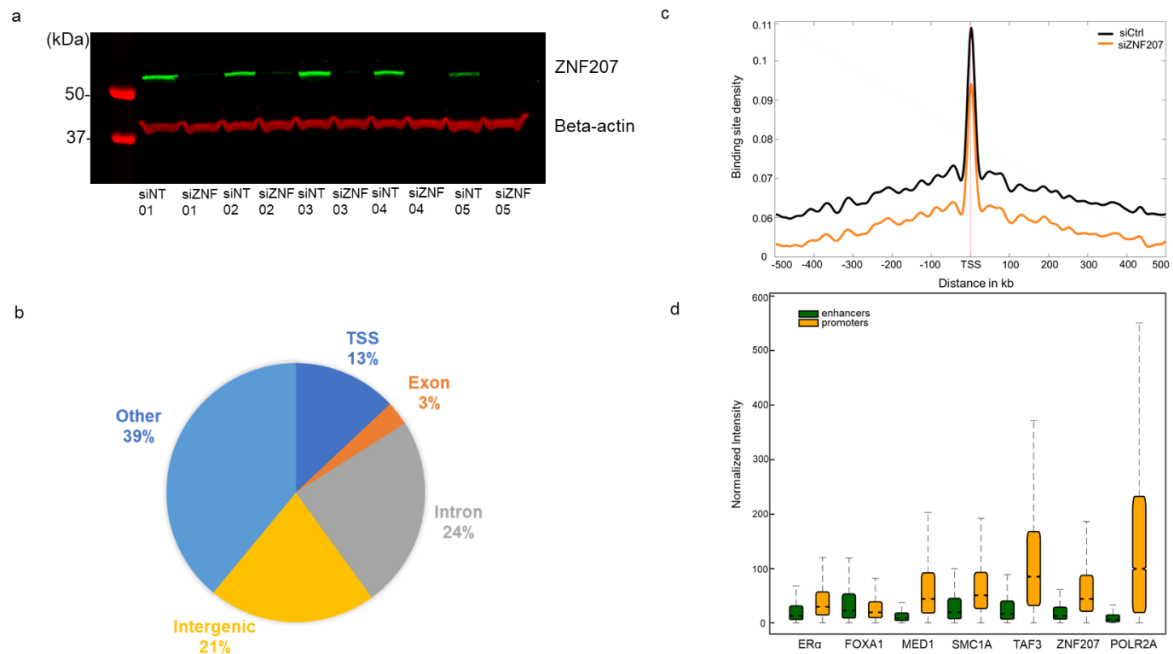


Figure 49. ChIP-seq data.

a) Western blot showing the decrease of ZNF207 levels upon siRNA treatment in all five biological replicates that were used for the ChIP-seq analysis. b) Pie chart showing the distribution of ZNF207 binding sites. c) Density plot illustrating the enrichment of ZNF207 binding sites close to TSS sites for the siCtrl and siZNF207 samples. d) Boxplots illustrating the protein-DNA binding intensity profile of the different factors (ER α , FOXA1, MED1, SMC1A, TAF3, ZNF207, POLR2A) in promoters and enhancers separately. The figures b, c and d have been provided by Igor Chernukhin.

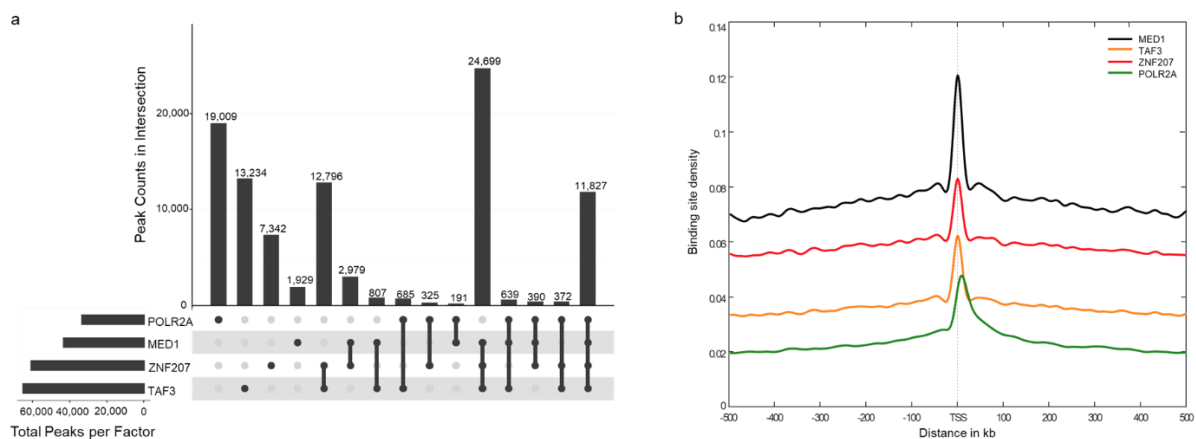


Figure 50. Overlap between ZNF207, MED1, TAF3 and Pol II binding sites.

a) UpSet plot summarises the comparison of total binding sites between ZNF207, TAF3, Pol II (POLR2A) and MED1. b) Density plot illustrating the enrichment of ZNF207, TAF3, Pol II (POLR2A) and MED1 binding sites close to TSS sites in the control samples and the overlap between the different factors. Figures have been provided by the Bioinformatics core and Igor Chernukhin.

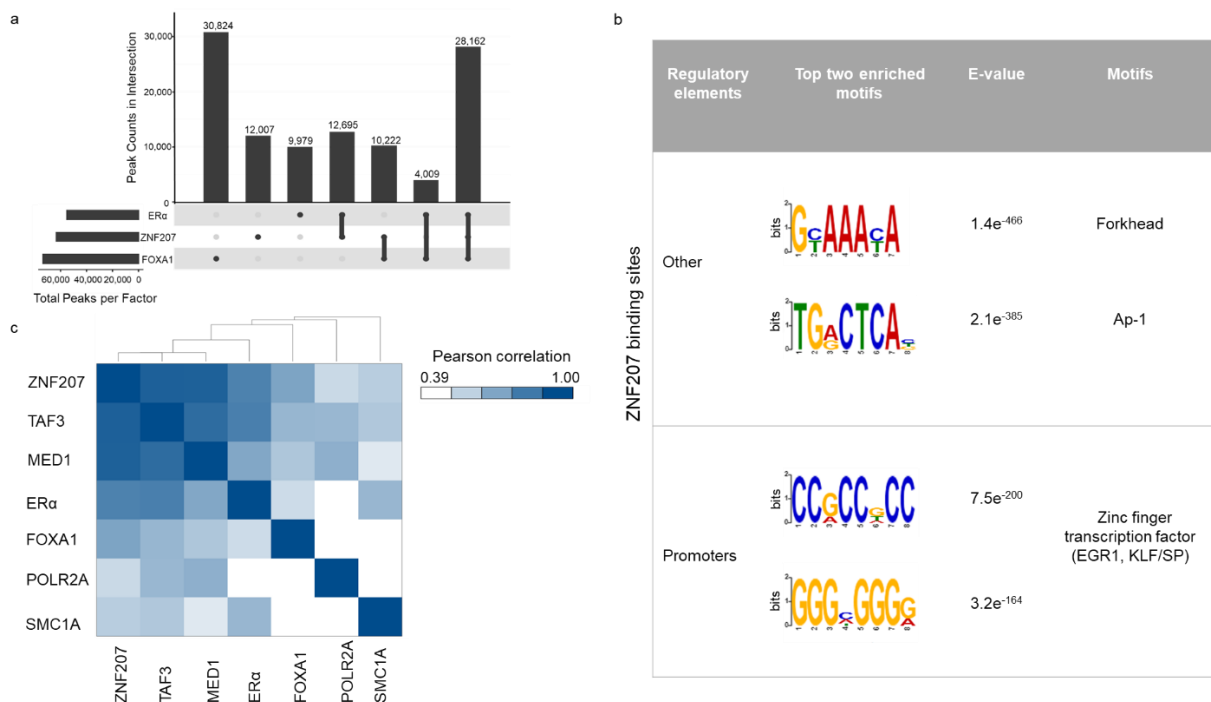


Figure 51. Overlap of ZNF207, ERα and FOXA1 binding sites and de novo motif analysis.

a) UpSet plot summarises the comparison of total binding sites between ZNF207, ERα and FOXA1. b) De novo motif analysis of ZNF207 peak regions reveals enrichment of Forkhead and Ap-1 motifs in non-promoter regulatory elements (other) and enrichment of zinc finger transcription factor motifs in promoter regions. c) Heatmap illustrating the correlation in the promoter binding sites of the different factors (ZNF207, MED1, TAF3, SMC1A, POLR2A, ERα, FOXA1). Figures have been provided by the Bioinformatics core and Dr Igor Chernukhin.

Analysis using Diffbind (Ross-Innes et al., 2012) revealed the loss of 4,094 and 11,090 peaks for MED1 and TAF3 respectively in the siZNF207 compared to siNT samples (adj. p-value<0.05). The effect on MED1 and TAF3 chromatin binding is consistent with depletion of these factors in the Pol II and Mediator complex in the qPLEX-RIME data, following ZNF207 knockdown. Additionally, 2,223 and 3,433 peaks for MED1 and TAF3 respectively had a positive enrichment (adj. p-value<0.05). The knockdown of ZNF207 had a modest effect on ER α and FOXA1 binding as 296 and 1091 binding sites were significantly lost (adj. p-value<0.05) and 1,876 and 2,205 binding sites were detected to be significantly gained respectively (adj. p-value<0.05). This finding is in line with the very mild effect that was detected on ER α complex following ZNF207 knockdown. The Diffbind analysis also revealed a mild effect on Pol II binding; a loss of 881 sites and a gain of 2,834 sites were detected following ZNF207 knockdown (adj. p-value<0.05). This result is suggestive of a more specific effect on the Mediator and general transcription factors recruitment, following ZNF207 knockdown, with Pol II possible still be in the promoters but not be able to initiate a proper assembly of the pre-initiation complex. The loss of MED1 binding following ZNF207 knockdown was confirmed by ChIP-qPCR in two well-known ER α binding sites (GREB1, XBP1), that were found with strong binding of MED1 in our data. The ER3 control site was included and each reaction was conducted in three technical replicates (**Figure 52**).

To better understand the effect of ZNF207 knockdown and its association with MED1 and TAF3 in regulating gene expression, we mapped all genes that were close to the significantly regulated binding sites (gained or lost) for all three factors and performed a functional enrichment analysis using all human genes as a background/control. Following this approach, we tested whether there is any overrepresentation of genes that are linked to particular biological pathways. We also divided the regulated binding sites into promoters and distal regions and mapped them to genes that were close to these regulatory regions to assess any specific effect of regulation at promoters or enhancers. For this analysis, peaks 1250 bp downstream and 250 bp upstream of the TSS were defined as promoters and peaks outside of this region but within 50,000 bp of the TSS were defined as enhancers. The functional enrichment of the MED1 binding sites revealed a significant correlation between lost sites and genes linked to cell cycle. Importantly, the highest odds ratio (quantifies the strength of the association) was observed on the lost sites within promoters (FDR<0.05; average odds ratio>4) (**Figure 53**), indicating an effect on the binding of MED1 on promoters close to cell cycle genes following ZNF207 knockdown. Additionally, the functional enrichment for genes near TAF3 regulated sites correlated well with the observations from the MED1 ChIP-seq, as an

enrichment for lost TAF3 sites was detected near promoters of genes related to cell cycle (FDR<0.05; average odds ratio>5) (**Figure 53**). The enrichment analysis did not indicate any significant enrichment for the detected gained binding sites for any of these two factors, suggesting that the effect following ZNF207 knockdown is related to loss of chromatin binding of Mediator subunits and transcription-related components. As ZNF207 mainly binds strongly to promoters, as we showed above, depletion of ZNF207 is likely to have a substantial impact on promoter biology and activity. Notably, it is the first time that ZNF207 knockdown is associated with an effect on chromatin binding of factors that have a vital role in the initiation of transcription. The enrichment analysis for the detected gained/lost binding sites for any of the other factors (ER α , FOXA1, Pol II), did not reveal any significant enrichment, highlighting the specific effect on TAF3 and MED1.

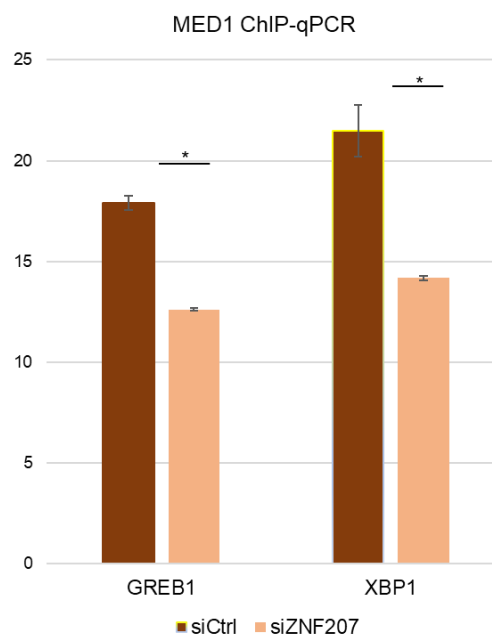


Figure 52. MED1 ChIP-qPCR.

*ChIP-qPCR analysis for MED1 in two well-known ER α binding sites (GREB1, XBP1) in MCF7 cells. Results are shown as enrichment over input (*Student's *t*-test *p*-value<0.005). The experiment was performed in three replicates and the replicates have been averaged. The error bars represent the standard deviation (SD).*

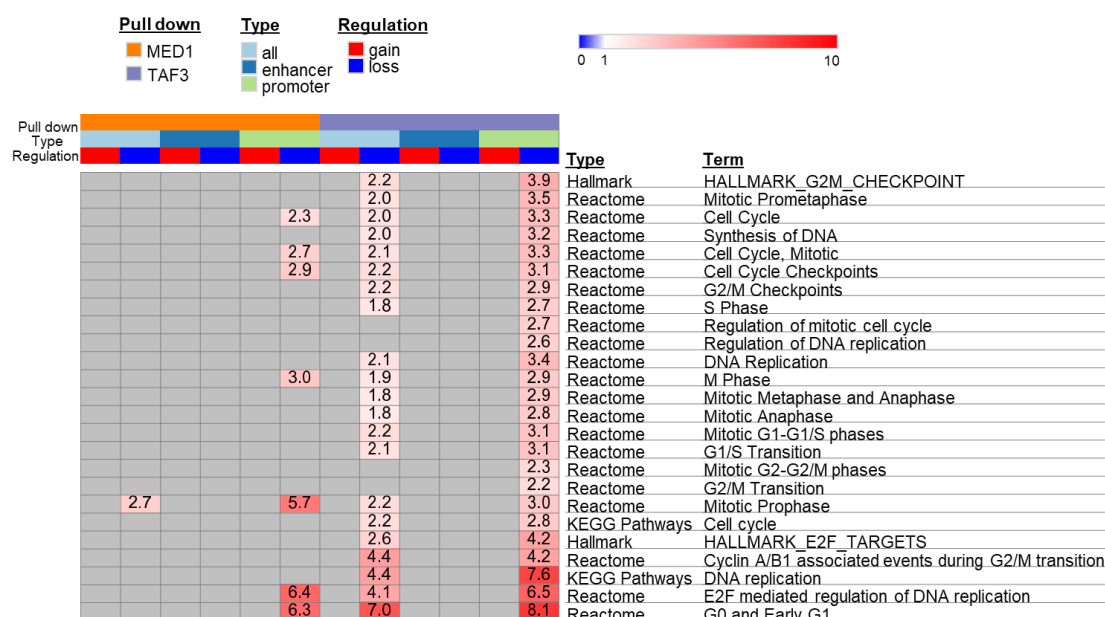


Figure 53. Functional enrichment analysis.

Heatmap summarising the results from the functional enrichment analysis for the lost or gained sites identified from the DiffBind analysis for MED1 and TAF3 ChIP-seq ($FDR < 0.05$). The sites have been defined as sites close to promoters or enhancers for the functional enrichment analysis. The results highlight the significant enrichment ($FDR < 0.05$) for the lost sites of MED1 and TAF3 close to promoters of genes that linked to cell cycle. Colour scale shows odds ratio and different colours at the top of the heatmap define the bait protein, the regulatory element (promoter or enhancer) and the regulation of binding sites upon ZNF207 knockdown (gain or lost sites).

Next, we integrated the results from the functional enrichment analysis for TAF3 and MED1 with our RNA-seq differential gene expression data. This comparison revealed that the genes that were linked to cell cycle and were mapped close to MED1 and TAF3 promoter binding sites were downregulated in our RNA-seq, indicating a direct effect of the MED1 and TAF3 promoter loss binding following ZNF207 knockdown on gene expression (**Figure 54**). It is known from the literature that recruitment of Mediator complex facilitates the recruitment and stabilisation of Pol II, TAF factors and other PIC components (Soutourina, 2018). Thus, we hypothesise that ZNF207 knockdown affects the recruitment of Mediator complex and especially MED1 that cause a decreased recruitment of components of the pre-initiation complex, resulting in downregulation of a set of genes that are associated with cell cycle.

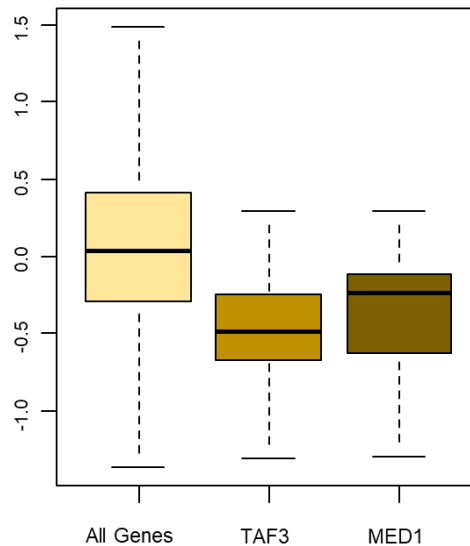


Figure 54. Correlation between ChIP-seq and RNA-seq data.

Box plots summarising the quantitative values (siZNF207 compared to siCtrl) of all genes detected in RNAseq and the cell cycle genes detected close to MED1 or TAF3 depleted promoter binding sites in the ChIP-seq.

We also conducted SMC1A (a subunit of the cohesin complex) ChIP-seq and Diffbind analysis to assess the effect of ZNF207 knockdown in cohesin recruitment. Here, we observed 1,091 sites to be significantly gained and 867 binding sites to be significantly lost (adj. p-value<0.05). The functional enrichment analysis did not show any significant enrichment of gained or lost sites of SMC1A close to any gene groups (odds ratio<1). This finding supports the intact promoter-enhancer contacts that we observed in our PChi-C experiment. It is possible that the presence of cohesin complex is sufficient to stabilise these architectural bridges even in the absence of specific subunits of the Mediator complex, following ZNF207 knockdown.

In parallel, we integrated our RNA-seq and PChi-C data from MCF7 cells with the ChIP-seq data to define more accurately promoters and enhancers, as the PChi-C data provide great detail on the interaction between promoters (bait) and enhancers (distal HindIII fragments). We focused on regulatory elements that are connected to genes that were found significantly down-regulated in the RNA-seq data (adj. p-value<0.05, |log₂Fold-Change|>0.5). The analysis of all lost, gained and not changing sites for MED1, ER α , FOXA1, TAF3, SMC1A and ZNF207, confirmed the significant enrichment for MED1 lost sites on enhancers and promoters of the top 200 repressed genes identified from the RNA-seq analysis (adj. p-value<0.05) (**Figure 55**). Notably the enrichment was significantly higher in the promoters (enrichment score>7, adj. p-value<0.05) validating the data from the initial functional analysis. We also observed a significant enrichment for TAF3 (adj. p-value<0.05) and ZNF207 lost sites (adj. p-value<0.05)

on promoters of the top 200 repressed genes, validating the effect on TAF3 binding near to promoters of repressed genes following loss of ZNF207 binding. Although, there were fewer genes for the analysis, as we lack regulatory element contacts for all the regulated genes, we gained confidence for our findings by having the same observations for MED1 and TAF3 and validating the strong association between MED1, TAF3 and ZNF207 on using two different approaches for the functional analysis of the ChIP-seq data.

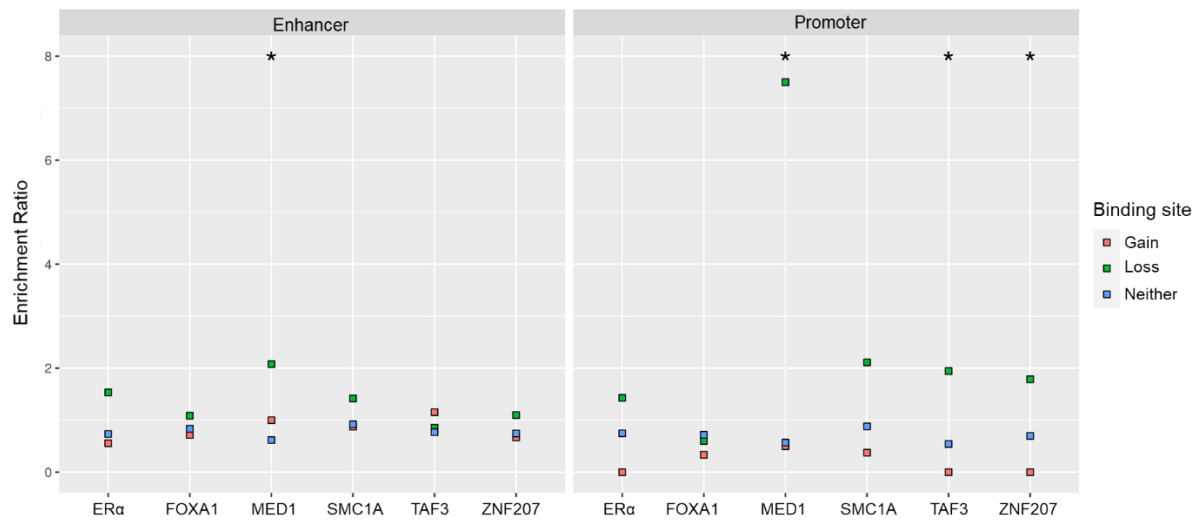


Figure 55. Correlation between ChIP-seq, RNA-seq and PChi-C.

Enrichment analysis for all gained, lost and unchanged binding sites for ERα, FOXA1, MED1, SMC1A, TAF3 and ZNF207 on promoters (right panel) or enhancers (left panel) of the significantly downregulated genes from the RNA-seq analysis. For the analysis the top 200 downregulated genes were selected that had loops detected in the PChi-C analysis. The enrichment ratio is the ratio of the number of binding sites associated with top 200 repressed genes to the number of binding sites associated with 200 constitutively expressed genes. The 200 constitutive genes were selected by filtering for genes having mean expression levels in the same range as the top 200 repressed genes and then selecting the 200 with fold changes closest to one. The asterisks indicate Fisher's test adj. p -value < 0.05. The figure has been provided by the Bioinformatics core.

To allow more accurate data interpretation, a spike-in control can be included in the ChIP-seq analysis for normalisation (Chen et al., 2015), an important point given recent findings that global changes in ChIP-seq data can be masked by lack of normalisation between samples. To confirm whether ZNF207 knockdown affects the binding of MED1 on genes related to cell cycle, we repeated the ChIP-seq experiment for ZNF207 and MED1 by adding *Drosophila melanogaster* chromatin in an amount proportional to the number of cells for subsequent normalisation of the data, prior to chromatin-antibody incubation. Notably both analyses were

highly correlated, giving similar enrichment ratios for MED1 and ZNF207 lost or gained sites. **Figure 56** illustrates the comparison between the two datasets for the MED1 lost and gained sites. This observation increased the confidence in our previous findings and confirmed the absence of any inherent normalisation issues. As discussed above, the ZNF207 knockdown selectively affects the functionality of cell-cycle transcriptional networks via disrupting MED1 recruitment and subsequent assembly of the Mediator complex. A similar selective role of Mediator subunits disruption has been recently described, where the authors observed a selective and pronounced disruption of cell-type specific transcriptional circuits following mediator degradation (Jaeger et al., 2020).

The integration of the different genomic approaches provided a clearer picture of how ZNF207 impacts chromatin binding of Mediator complex and components of the transcription machinery mainly in promoters, leading to changes in transcriptional networks that regulate cell cycle. The depletion of important transcriptional networks, following ZNF207 knockdown, may explain the essentially of ZNF207 across multiple cell types. To conclude, knockdown of ZNF207 affected Mediator complex assembly and MED1 recruitment to promoters of cell cycle genes resulting in inhibited recruitment of initiation factors such as TAF3 in the promoters of these genes, displaying a negative effect on their expression profile.

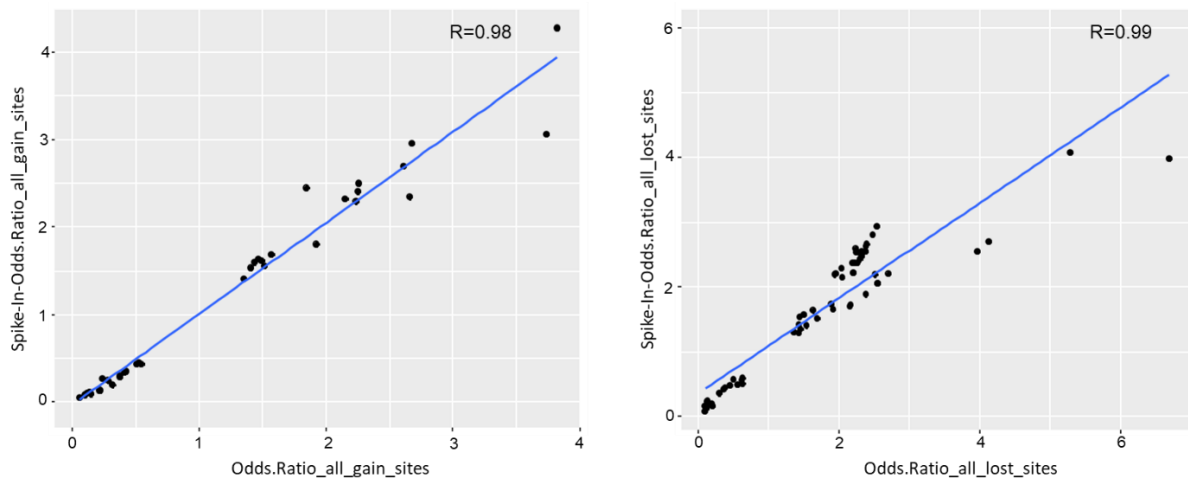


Figure 56. Correlation between ChIP-seq experiments with or without spike-in control (*Drosophila melanogaster* chromatin).

Scatter plots showing the linear correlation of odds ratio for the MED1 significantly lost (right panel-Pearson's $R=0.99$) or gained binding sites (left panel-Pearson's $R=0.98$) (adj. p -value<0.05) that were detected in both experiments.

Chapter 4

4. Discussion

Cellular functions and signalling pathways are regulated by molecular events that involve dynamic protein-protein networks, as well as protein abundance and post-translational modification changes. In depth and detailed study of the dynamics of these spatiotemporal changes can provide a significant insight into the physiological cellular processes and abnormal protein networks linked to diseases. Breast cancer is the most common form of cancer among women in the western world and 75% of cases is estrogen dependent. The association between estrogen signalling and breast cancer has been the main focus of our lab and has also been studied by other groups mainly by genomics and gene expression methods (Rueda et al., 2019; Sorlie et al., 2001; Sorlie et al., 2003), as proteomic characterisation remains less studied. However, changes at protein level cannot be predicted accurately by genomic profiling methods alone, as it has been shown that the correlation between gene expression and relative changes at the protein levels in cancer tissues is moderate (Mertins et al., 2016). Advances in mass spectrometry together with innovations in sample preparation, data software and quantitative approaches have permitted the development of multiple strategies to characterise whole proteomes and protein interactomes with high efficiency and accuracy, providing a better understanding of the events that occur at the proteomic level. Although, recent studies (Lawrence et al., 2015; Mertins et al., 2016; Tyanova et al., 2016a) have accomplished the proteomic profiling of breast cancer clinical samples, offering a better understanding of key players and driving mechanisms, the dynamic crosstalk between factors in protein complexes and the mapping of tumour cell protein networks *in vivo* has been less extensively studied.

To address this, we developed a quantitative method, called qPLEX-RIME, which enables the study of endogenous protein interactomes with high sensitivity and statistical robustness, giving insight into the functional interplay between proteins. The qPLEX-RIME approach integrates the well-established RIME immunoprecipitation method with advanced high-resolution quantitative multiplexed mass spectrometry analysis. It quantifies cross-linked protein interactomes with multiplex isobaric labelling and MS3 analysis, overcoming the effect of interfering ions and improving the dynamic range of quantification. The multiplexed analysis of our pipeline eliminates the need to compare multiple data obtained by individual LC-MS runs, thereby increasing the quantification coverage in very low abundant protein interactors, that are stochastically captured between independent replicate runs. Our

quantitative pipeline can be further empowered with the recent advances in the TMT quantification approach and the availability of reagents with higher multiplexing capacity, termed TMTpro, that enable the analysis of up to sixteen samples in one experiment. This can offer an increased number of technical or biological replicates within one experiment and better coverage of proteomes and interactomes, as the samples are not distributed across separate multiplexed sets, reducing the missing values. The increased power of our qPLEX-RIME can delineate the dynamics of chromatin associated protein complexes in different biological contexts and experimental settings, offering in-depth protein detection, reproducible quantification and increased sensitivity and sample throughput. Preliminary data obtained in our lab using the 16plex qPLEX-RIME method have demonstrated the additional gain of accuracy, statistical power and depth (data not shown here), minimising the sample preparation and mass spectrometry time. Additionally, the combination of the qPLEX-RIME pipeline with mass spectrometry acquisitions, designed to identify cross-linked peptides (XL-MS), can offer a powerful approach for the detailed understanding of the structure and dynamics of heterogeneous protein networks. The cross-linking analysis identifies proximal residues between subunits of protein complexes, uncovering spatial orientation and protein connectivity. XL-MS studies have been performed for the structural characterisation of various protein systems by purifying specific complexes (Yu and Huang, 2018). The combination of the qPLEX-RIME approach with XL-MS can provide a tool for *in vivo* characterisation of physical interactions, increasing the information that can be obtained from the study of protein interactomes. Preliminary data from our lab have shown the feasibility of this approach and further optimisations can improve our strategy and provide important information for protein complexes in different disease states, increasing the variety of studies that qPLEX-RIME pipeline can be applied.

In the present thesis, we focused on ER α , the major driving transcription factor in luminal breast cancer, which can be targeted by different drugs such as tamoxifen, that has been used for the treatment of ER⁺ breast cancer (Ali and Coombes, 2002; Rangel and Huang, 2013). ER α functions as part of a large transcriptional complex involving multiple transcriptional factors that can modulate the ER α transcriptional activity. While several ER α interactors involved in ER α -mediated gene expression have been studied, our knowledge about their importance at the tissue level and the impact of drugs such as tamoxifen on their global association with ER α remains incomplete. An efficient pull-down of ER α interactors was achieved by *in vivo* double cross-linking using a combination of cross-linkers; the crosslinker DSG was added, followed by formaldehyde, that has been used routinely for the detection and

quantification of protein-DNA interactions and interactions between chromatin proteins that are in close proximity (Orlando, 2000). The application of double-crosslinking increased the efficiency of our qPLEX-RIME and ChIP-seq experiments, without increasing the background and the unspecific protein or DNA binding. The combination of the two crosslinkers has been successfully used in our lab to characterise the interactome and genome binding of several transcription factors (Nagarajan et al., 2020; Siersbaek et al., 2020).

The quantitative data obtained from the ER α qPLEX-RIME experiments and rigorous statistical analysis enabled the identification of significantly enriched key ER α specific interactors and novel ER α -associated proteins compared to the negative controls. These can include transient, direct, indirect or weak interactions, as it is known that ER α associates with a number of different cofactors rapidly in a cyclic fashion (Metivier et al., 2003; Shang et al., 2000). As such the final result of our crosslinking-based pipeline represents the sum of all these interactions and the different complexes that are formed under particular conditions in the cell. Among these, we validated the novel interactions between ER α and three proteins, namely CBX3, NIPBL and FOXK1 using PLA assays. The PLA protocol does not require modification or tagging of the proteins for which their physical interaction is investigated, permitting the study of endogenous interactions with high sensitivity. This approach has been employed before to detect interactions between ER α and signalling proteins (Poulard et al., 2012). CBX3 protein belongs to HP1 protein family, that has been implicated in gene regulation, DNA replication and nuclear architecture (Eissenberg and Elgin, 2000); NIPBL is a core subunit of the highly conserved protein complex cohesin that has an important role in chromatin structure, gene expression and DNA repair (Hill et al., 2016). The transcription factor FOXK1 is a member of the forkhead family and has an vital role in tumourigenesis (Katoh and Katoh, 2004; Wu et al., 2016). These findings display that the increased sensitivity gained by qPLEX-RIME can reveal novel ER α interactors that can be validated by alternative methods. Furthermore, our qPLEX-RIME data on three additional factors; CREBBP, NCOA3 and Pol II highlighted the wide applicability of our pipeline and displayed a comprehensive characterisation of the interactome of known nuclear receptors co-activators as well as the interactome of the phosphorylated form of Pol II at serine 5, that has a fundamental role in the initiation of transcription (Phatnani and Greenleaf, 2006).

The qPLEX-RIME experiment targeting ER α after OHT treatment at three different time points, demonstrated dynamic changes in ER α co-regulators, following drug treatment, recapitulating but also broadening the existing knowledge of OHT mechanism. At 2h treatment with OHT, we detected a loss of important transcriptional co-activators such as NCOA3 and

CBP, whereas at 6h an enrichment in the recruitment of two well-conserved chromatin remodelling complexes, namely the NuRD and the SWI/SNF complex was observed. It is the first time that the SWI/SNF complex has been described to be recruited upon tamoxifen treatment, a finding that highlights the important role of this complex and its various functions as coactivator or corepressor. We also observed an enrichment of the basal corepressor NCOR2, which assists in the recruitment of HDAC proteins (Underhill et al., 2000). At 24h treatment with OHT, we observed a restoration of the ER α complex, which may be linked to the half-life of OHT, as we treated our cell line once with the drug prior to double-crosslinking. The exceptions were NRIP1, GREB1 and NCOA3; notably, NCOA3 is amplified in breast cancer (Anzick et al., 1997) and its expression levels have been linked to the effectiveness of tamoxifen treatment (Osborne et al., 2003). Previous ChIP-seq analysis has enabled the discovery of a number of binding sites of NCOA3 that are associated with genes that have predictive value for breast cancer patient outcome (Zwart et al., 2011), supporting an important role of this co-regulator in tamoxifen response.

Our time course data suggest a switch between activation and repression of transcription following OHT treatment. This transition engages a two-step process with the quick loss of co-activators, followed by the recruitment of co-repressors and ATP-chromatin remodelling complexes that may act together or in a sequential manner to achieve transcriptional repression. In parallel we integrated our qPLEX-RIME data with whole protein and mRNA analysis, displaying an extensive view of the activity of a transcription-associated complex over time. This strategy provides a better understanding of the drug mechanism and the discovery of protein changes that are associated with the complex assembly or gene and protein expression. Additionally, changes in the recruitment of these factors on chromatin or on their expression levels may be linked to endocrine resistance, highlighting the importance of studying drug function and the mechanisms that result in drug resistance, which remain elusive. A proposed model of OHT mechanism is depicted in **Figure 57**.

In our qPLEX-RIME pipeline, the ability to combine the labelled peptides derived from multiple samples increased the sensitivity of the method and enabled the in-depth *in vivo* characterisation of the ER α interactome in PDX and clinical tumours. Key ER α interactors have previously been detected from clinical material, however this required a targeted mass spectrometry-based approach and has not been done in an unbiased manner (Mohammed et al., 2013). The application of qPLEX-RIME in primary tumour material or PDX models can permit the study of protein network regulation across different types of cancers, leading to the discovery of protein signatures and potential therapeutic markers in heterogenous and more

physiologically relevant systems, where starting material is limited. In addition to increased sensitivity, the application of isobaric labelling resolves the experimental difficulties with cell lines that are not compatible with stable-isotope labelled culture media and provides a tool for the quantitative analysis of clinical samples that are not amenable to *in vivo* isotopic labelling techniques. Notably, our isobaric-labelling data showed high correlation with previously published SILAC data (Mohammed et al., 2013), confirming the highly accuracy of the quantification obtained by the MultiNotch MS3 level mass spectrometry analysis.

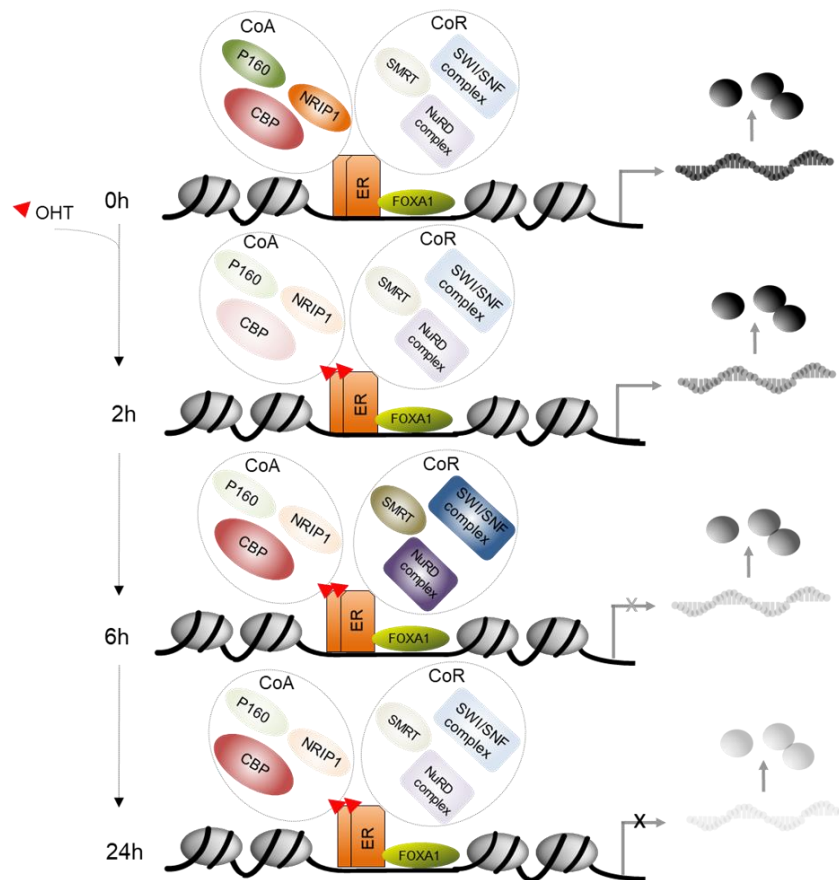


Figure 57. A proposed model of OHT mechanism.

In untreated conditions (0h), both co-activators and co-repressors are interacting with ERα in the chromatin. The pioneer factor FOXA1 is on chromatin as well. After 2h treatment, a loss of co-activators such as NCOA3 and CBP takes place. At 6h, an enhanced recruitment of the NuRD and SWI/SNF complexes is observed coinciding with the enrichment of the NCOR2 corepressor. The proteins NCOA3 and NRIP1 remain at low levels in the interactome. At 24h, we observed a full restoration of the ERα complex with the exception of the NCOA3 and NRIP1 proteins, which were maintained at decreased amounts in the complex. After 24h treatment, a downregulation of ERα target genes at transcript and protein level is observed. In the cartoon, intense and faint colours indicate enrichment or loss respectively.

CRISPR and RNAi screens can allow the systematic identification of core and context-specific essential/fitness genes governing cell proliferation across all cell types. Integrating the qPLEX-RIME datasets from various cell line models and clinical tumours with previously published CRISPR and RNAi data, we identified a novel ER α -associated factor, namely ZNF207, which is a member of the zinc finger family, the largest transcription family in human genome (Jen and Wang, 2016). Zinc finger proteins (ZNFs) are implicated in development, differentiation, metabolism, autophagy and more recently have been linked to cancer progression, highlighting the importance to elucidate the functional role of this protein family (Jen and Wang, 2016). Specifically, amplification and overexpression of ZNF306 have been reported in invasive colorectal cancers, multiple myeloma and prostate cancer (Yang et al., 2008) and ZNF388 has been identified as an oncogene (Jen et al., 2017). Other members such as ZNF191, ZNF668, ZNF348 have been implicated in angiogenesis, proliferation and metastasis of breast cancer (Jen and Wang, 2016). The ZNF703 gene has been mapped in the 8p12-p11 region, that is commonly amplified in breast cancer and has been associated with poor prognosis in luminal B tumours (Reynisdottir et al., 2013) and high levels of expression of ZNF217 mRNA is associated with poor prognosis and the development of metastasis in breast cancer (Vendrell et al., 2012). Recently, a ZNF factor termed ZNF143 has been described as a regulator of chromatin loops (Ye et al., 2020). These studies highlight the diverse roles of ZNF factors in various cancer types and stimuli as well as their involvement in chromatin organisation and gene expression, illustrating the complexity and the different layers of regulation. Importantly, these transcription factors not only contain zinc finger domains for DNA binding, but also functional domains that may regulate subcellular localisation and binding with different interacting partners, affecting gene regulation (Jen and Wang, 2016). Our qPLEX-RIME data displayed a wide variety of ZNF207 interacting proteins, including nuclear receptors, co-activators, Pol II and Mediator subunits.

ZNF207 was initially discovered as a kinetochore-binding protein, that promotes mitosis in HeLa cells and Glioblastoma multiforme stem cells (GSCs) (Jiang et al., 2014) and more recently as a transcription factor that controls self-renewal and pluripotency of human embryonic cells (Fang et al., 2018). Here, our findings suggest that ZNF207 is a pan-essential factor in different cellular contexts and functions as a key regulator of the transcription machinery. We hypothesise that ZNF207 is required for the recruitment of the Mediator complex and members of the pre-initiation complex in the promoters of cell-cycle related genes, highlighting the diverse role of zinc finger proteins and the importance of studying further this protein family. The association between members of the zinc finger family and the

Mediator complex has not been described before, revealing a potential mechanism controlling Mediator complex recruitment and function. Although, the Mediator complex has emerged as the most critical coactivator, having a role in transferring signals from DNA-binding transcription factors to Pol II, the mechanistic understanding of its functions in human cells remains incomplete. This is because of experimental challenges related to its large size, conformational flexibility and the generation of Mediator subcomplexes, that display various functionalities, highlighting the importance of understanding the processes prior or after the recruitment of Mediator and the initiation of gene transcription (Soutourina, 2018).

Our functional characterisation of ZNF207 using proteomic techniques indicated that ZNF207 may act as a transcription co-activator affecting the assembly of the Mediator complex, especially for the subcomplex containing the MED1 subunit. In addition to MED1, we investigated the effect on MED4 and MED12 subunits to test whether the effect is general for the Mediator complex assembly or specific for the MED1 subcomplex, which only exists in a subpopulation (less than 20% of the total) (Zhang et al., 2005). We observed the disruption on Mediator complex in all three different qPLEX-RIME, validating the general effect on the Mediator complex following ZNF207 knockdown. Interestingly the MED1 subcomplex is tightly associated with Pol II (Zhang et al., 2005), that may support the effect on the PIC complex following ZNF207 knockdown. Importantly, this MED1-containing holoenzyme is involved in both basal- and activator-dependent transcription (Zhang et al., 2005), which is in line with the effect on gene expression and especially to genes that are associated with cell cycle signalling. Zinc finger proteins can also act either as transcription activators or as repressors (Jen and Wang, 2016), that may explain the equal numbers of up- and downregulated genes detected in the RNA-seq analysis, following ZNF207 knockdown. Recruitment of the Mediator complex results in the subsequent recruitment of the transcription machinery components and initiation of RNA polymerase II-dependent transcription. We hypothesise that knockdown of ZNF207 affects the assembly of the Mediator complex and leads to impaired stabilisation of the PIC complex in promoters of genes linked to cell cycle. The Mediator complex and its construct with the components of the pre-imitation complex are highly conserved from yeast to humans (Soutourina, 2018). The fact that ZNF207 knockdown affects such essential components of the basal transcription, indicates a more generic transcriptional role of ZNF207, in line with its essentiality across different cell lines. The effect on the Mediator complex in MCF7 cells was confirmed in additional experiments on a non-cancer cell line, where qPLEX-RIME experiments against MED1 and MED14 subunits in HEK293 cells revealed the disruption of the Mediator complex, evidenced by reduced assembly.

Specifically, the effect on MED14 interactome was intriguing as it is known that it is the central backbone of the complex that has a vital role in bridging all the main modules of the Mediator complex (Cevher et al., 2014), highlighting the importance of ZNF207 protein in the regulation of the complex assembly.

Characterisation of the genome binding profile of various factors, following ZNF207 knockdown, validated the findings from the gene expression and interactome analyses and uncovered the specific effect on binding of MED1 and TAF3 close to promoters of genes linked to cell cycle. The genome-wide binding sites of ZNF207 have been previously profiled in human embryonic stem cells (Fang et al., 2018), however to our knowledge, our data is the deepest and highest quality ChIP-seq experiment, for ZNF207, that has been performed in a breast cancer model. An experiment that can be considered is to investigate the DNA binding profile of this factor in various cell systems and assess its regulatory role in different cancers or different stages of development. Additionally, application of ChIP-seq experiments on other Mediator subunits will be important to assess the effect of ZNF207 on subunits from different domains of the Mediator complex. It appears that ZNF207 knockdown mainly affects the association of Mediator with core promoters, but not the association with enhancers. Here, we chose to study the genome binding profile of MED1, which is part of the middle module, however one could consider to study the effect on subunits that belong to the tail module, which mainly associates with enhancers (Petrenko et al., 2016). Notably, we tested several antibodies targeting different Mediator subunits for ChIP-seq experiments, however most antibodies did not provide good binding profile coverage, except the one for the MED1 subunit. For future experiments, the development of customised antibodies for additional Mediator subunits or chromatin fractionation analysis (Raab et al., 2019), following ZNF207 knockdown, could provide a more comprehensive study of the effect of ZNF207 loss on the chromatin binding patterns of the Mediator complex.

To our knowledge, this is the first time that a member of the zinc finger family is associated with the Mediator complex, the disassembly of the pre-initiation complex and the inhibition of the expression of cell cycle related genes. Overall, our findings support that knockdown of ZNF207 has an effect on the classic function of Mediator complex and the transmission of signals between transcription factors and pre-initiation complex affecting mainly the recruitment of general transcription factors. The Mediator complex is also involved in many other aspects of transcription, including elongation, termination, RNA processing, as well as its epigenetic regulation. In the qPLEX-RIME experiments for Mediator subunits and Pol II, we observed an effect on the recruitment of elongation factors and its association with Mediator

complex and Pol II, following ZNF207 knockdown. It would be interesting to further investigate this effect and discover if ZNF207 regulates other mechanisms than those, where the Mediator complex is involved. Another perspective would be to study the effect of ZNF207 knockdown on Pol II pausing. In the human genome, Pol II pauses at promoter-proximal sites for several minutes, before starting a productive elongation, which serves as a quality-control checkpoint (Luo et al., 2012). Recently, a compensatory feedback loop has been described with Pol II pausing being sufficient to maintain the gene transcription, compensating for defects in Pol II initiation, following MED14 degradation (Jaeger et al., 2020). Based on these findings, the study of the Pol II pause state and whether is affected by ZNF207 knockdown can be important to illuminate further the function of ZNF207. Furthermore, the establishment of an XL-MS approach can provide structural resolution to these different complexes and delineate how ZNF207 interacts with the various components. It can provide important answers about the physical interactions between the different molecules and potentially identify other factors that may mediate these different interactions.

The observation of an upregulation of corepressors in the RNA-seq analysis may indicate a mechanism via which ZNF207 affects negatively the gene expression. It is possible that the disassembly of the Mediator complex affects the expression of these corepressors, as the Mediator complex has been linked with activation and repression of transcription (Balciunas et al., 1999). On the other hand, an interplay between the different corepressors may occur to enhance the effect of ZNF207 knockdown. BAZ2A is the largest subunit of the nucleolar remodelling complex NoRC, that is known to establish epigenetic silencing and transcriptional repression through association with DNA methyltransferases and histone modifier complexes. It has also been shown that BAZ2A regulates numerous protein-coding genes and directly interacts with EZH2, the catalytic subunit of the Polycomb repressive complex 2 (PRC2) to maintain epigenetic silencing at genes repressed in metastasis (Gu et al., 2015). CBX8 is a transcriptional repressor, component of the Polycomb repressive complex 1 (PRC1), that has been described to have a critical role in the pathogenesis of cancer (Tang et al., 2019), whereas MTG corepressors can interact with other corepressor proteins such as Swi-independent 3A (SIN3A), NCOR1 and silencing mediator for retinoid and thyroid receptors (SMRT) to recruit histone deacetylases (HDACs), leading to gene repression by lysine deacetylation of histones (Kumar et al., 2015). All these corepressors that have been identified in our datasets play an important role in gene expression and it will be interesting to study further how they associate with ZNF207 and whether they affect the assembly of the Mediator complex and the expression of genes linked to cell cycle.

The combination of global proteome and phosphoproteome, complementing the study of endogenous interactions in cell lines and tissues in an unbiased manner, can lead to the discovery of novel protein function and uncover the links between functional alterations and phenotypes. Our phosphoproteome study revealed that knockdown of ZNF207 does not affect the phosphorylation status of members of the transcription machinery or proteins linked to cell cycle, at least for the phosphosites covered in this study. It has been reported that phosphorylation on serine or threonine residues of ZNFs regulate DNA binding abilities and recruitment of interactors, whereas acetylation facilitates protein interactions (Jen and Wang, 2016). To this end, further study of phosphorylation or other post-translational modifications on ZNF207 and its interactors may reveal additional information of how the interaction between ZNF207 and the Mediator complex is regulated, with implications in gene activation or repression. This will be of high importance as the Mediator and components of the pre-initiation complex are regulated by post-translational modifications especially phosphorylation. Two main phosphosites in Pol II, serine 5 and serine 2, have crucial roles in initiation of transcription and the release of Pol II into a productive transcription elongation (Soutourina, 2018).

ZNF207 mRNA is the target of the splicing factor SFRS11 and during somatic reprogramming, ZNF207 changes from isoform B to A and C (Fang et al., 2018). It has been reported that the canonical protein is the dominant isoform in the embryonic stem cells and represent the functional form that interacts with different co-factors. This alternative isoform switches the function of ZNF207, indicating a distinct role for each isoform during development (Fang et al., 2018). Thus, we investigated whether there is a dominant ZNF207 isoform in MCF7 cells, that is linked to the changes in the gene expression and the effect on Mediator recruitment. Our analysis showed the existence of the canonical isoform and a significant number of reads for the exon 6 and 9 skipping, indicating the presence of all three isoforms. Further functional analysis may provide additional evidence for a dominant role of a specific isoform in our breast cancer system and whether there is a mechanism of switching isoforms that may play distinct roles in various cell types and systems.

To study further the functional role of ZNF207, we performed Promoter Capture Hi-C to specifically enrich and capture interactions involving promoters and distal elements across the whole genome. We were interested to study whether ZNF207 has any role on chromatin loops formation, as we observed a significant effect on Mediator complex, that has been previously reported to regulate chromatin loops (Kagey et al., 2010). Additionally, it has been reported that other members of the zinc finger family can regulate DNA loops (Wen et al., 2018).

Analysis of the statistically significant interactions detected by the CHiCAGO pipeline revealed that interactions between promoters and annotated enhancers represent only a small fraction (~30%) of the total promoter-distal interactions, supporting that other regions that do not harbour classical enhancer markers can have regulatory function. To this end, studies have identified regulatory elements that control gene expression, but are devoid of enhancer marks (Mifsud et al., 2015). This is an interesting observation that can be studied further to gain more knowledge about the interactions between regulatory elements in different systems and understand better the transcriptional regulation of genes. Regarding the role of ZNF207 in regulating chromatin looping, we did not detect any significant change in the chromatin loops, following ZNF207 knockdown. Our data indicate, that the negative effect on the Mediator complex does not translate into changes in the formation and stabilisation of chromatin loops, a finding that was quite surprising. However, our observation is in line with a recent published study (El Khattabi et al., 2019) and suggests that the Mediator complex is not required for the DNA loop formation between promoters and enhancers as no significant effect was observed on promoter-enhancer contacts, following ZNF207 knockdown. Despite the effect on Mediator chromatin binding and Mediator complex assembly, the chromatin loops remain unaffected. For this experiment, each condition was represented by three biological replicates that were sequenced and aligned, which provides increased statistical power and confidence for our findings compared to previously published three-dimensional studies (El Khattabi et al., 2019; Schoenfelder et al., 2015), where no more than two biological replicates were included. These findings, taken together, are inconsistent with initial models that proposed that Mediator complex along with transcription factors, coactivators and Pol II creates an architectural bridge between promoters and distal elements (Kagey et al., 2010). This apparent discrepancy with the initial studies may be explained by experimental challenges or the study of particular Mediator subunits that may not be expected to participate in the looping formation. In our data, the observation of no effect in promoter-enhancer contacts may be explained and supported by the fact the cohesin complex, which is the primary factor in facilitating long-range interactions (El Khattabi et al., 2019), remains unaffected in the genomic regions close to genes linked to cell cycle following ZNF207 knockdown. It is possible that cohesin is the main factor for establishing and stabilising the chromatin loops and the Mediator complex or the specific subunits that have been studied in the present study may not be involved on this mechanism. Overall, studying the three-dimensional chromatin interactions can be critical and provide important information for different biological systems. It has been shown that three-dimensional chromatin landscape remodelling is associated with endocrine resistance in ER α

positive breast cancer (Achinger-Kawecka et al., 2020) and the establishment of the method in our lab can facilitate the study of these underlying mechanisms and lead to the detection of changes that can be targeted for therapeutic purposes. The development and establishment of chromosome conformation capture studies can provide genome-wide evidence and establish a better understanding of the interactions between regulatory elements.

Our conclusions resulted from the integration of different high-throughput approaches including interactome characterisation, genome binding profiling and three-dimensional structure, highlighting the importance of following different strategies to unravel the functional role of transcription factors. Two different approaches were also applied for the analysis of the ChIP-seq data, increasing the accuracy of our findings. Even if the two different experimental strategies (spike-in or without spike-in control) did not show significant differences that could affect the outcome of the analysis, the spike-in strategy will be applied to ensure an accurate comparison between experimental conditions for future experiments. Collectively, the results from the different functional approaches lead us to the model proposed for ZNF207 function in **Figure 58**.

The findings from the various functional assays attribute an important and novel role for a factor, that belongs to zinc finger family. Future work is needed to further investigate the consequences of ZNF207 loss on the Mediator assembly and PIC formation. One of the perspectives is to investigate, whether ZNF207 knockdown directly affects the Mediator assembly or other factors may be implicated in the Mediator recruitment and initiation of transcription. It seems that ZNF207 knockdown mainly influences gene promoters and the assembly of the mediator, which as our data suggest, is important for recruitment of the pre-initiation complex and RNA pol II. It will be important to consider performing an experiment, where we remove Mediator function by knocking down the central MED14 subunit or other essential subunits and assess whether it has the same effect on recruitment of TAFs and other initiation and elongation factors as ZNF207 knockdown. This would show that removing ZNF207 has the same effect as removing Mediator activity, validating the vital role of this transcription factor. It would also confirm that the effect we see on the initiation and elongation factors, as well as in the Pol II core complex, is likely linked to the effect on the Mediator. It would also be important to understand why this factor has various roles in different biological systems and to consider that the expression of other factors may result in this multifunctionality. Additional ChIP-seq or ChIP-qPCR experiments on various Mediator subunits can be important for the in-depth study of the effect on the Mediator complex and to test whether ZNF207 has a specific role on the assembly of a particular domain. One more

experiment that can be considered, is the establishment of stable cells models overexpressing ZNF207, followed by interactome and genome profiling studies to test the effect on the recruitment of the Mediator complex or the components of the PIC complex. It will be interesting to see whether the overexpression affects other important pathways that may be associated or not with the Mediator assembly and the expression of genes linked to cell cycle. It may be of interest to overexpress different isoforms of ZNF207 to better understand their various roles in the different biological systems. We can also consider overexpressing ZNF207 in cellular models that express ZNF207 in much lower levels compared to the cell models that we used in the present study. A recent proteomics study by Gygi's lab has provided quantification profiling of the proteome of 375 cell lines from diverse lineages (Nusimov et al., 2020). This large-scale profiling could help identify cell lines that have the lowest protein levels of ZNF207 transcription factor to be used as cellular models for the perspective experiments proposed above.

To conclude, our qPLEX-RIME method can be used to monitor the dynamic changes of the composition of protein complexes and importantly can be applied to clinical samples to study interactome regulation in tumours with variable genomic backgrounds or numerous other biological and clinical questions. It provides a robust tool for the quantitative analysis of complexes that can be applied to generate comprehensive endogenous protein-protein interaction maps. The application of qPLEX-RIME led to the identification of the ZNF207 transcription factor as an important player in the assembly of the Mediator complex and subsequently in gene expression regulation. Future work is warranted to assess the role of this factor in different biological systems. Given the ubiquitous expression of ZNF207 and its essential function, further investigation of its role and its possible association with other co-factors or members of the zinc finger protein family to control important cellular signalling processes is particularly promising. We suggest ZNF207 as a transcription factor with a vital role in the initiation of transcription and particular in the recruitment of the Mediator complex and components of the PIC complex.

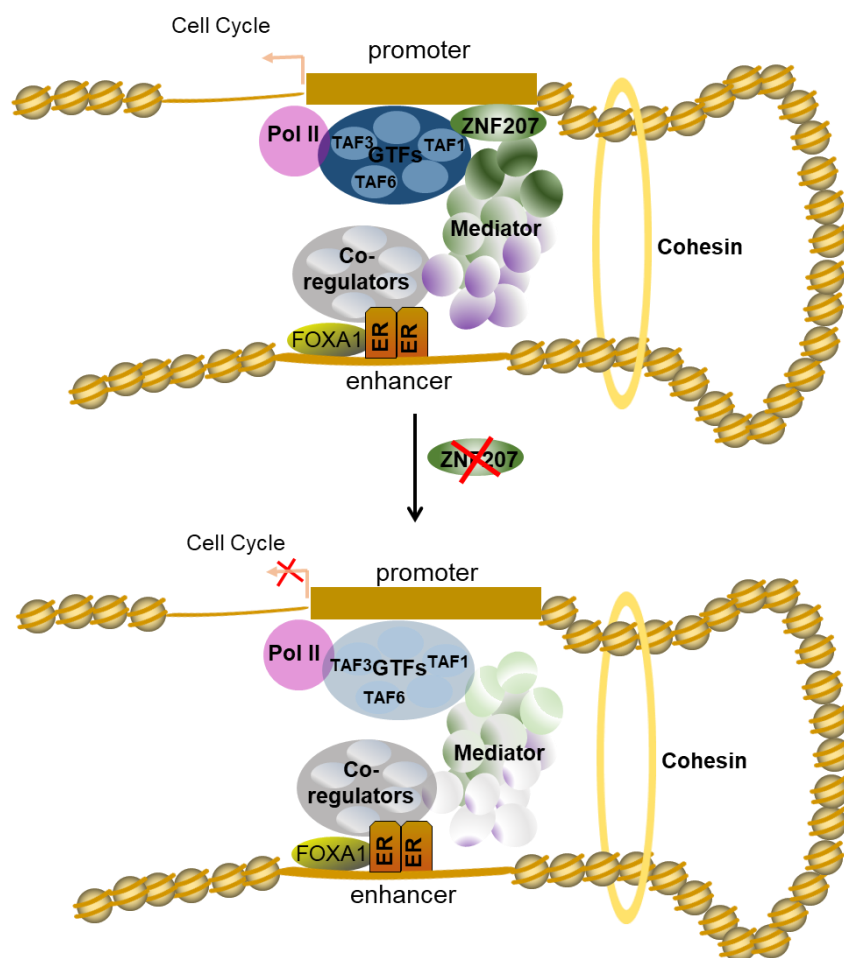


Figure 58. A proposed model of ZNF207 functional role.

ZNF207 binds strongly to promoters and interacts with the Mediator complex and components of the pre-initiation complex, as well as with the ER α complex. Following ZNF207 knockdown, we observed changes in the assembly of the Mediator complex and decreases in MED1 recruitment to promoters of cell cycle genes, resulting in inhibited recruitment of initiation factors such as TAF3 in the promoters of these genes, displaying a negative effect on their expression profile. No effect was observed on cohesin, ER α complex and the formation of enhancer-promoters chromatin loops, following ZNF207 knockdown. In the cartoon, faint colours indicate loss following ZNF207 knockdown.

4.1 Supplementary Data

Supplementary Data 1: *Regulated proteins identified in ERα-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.*

Gene Symbol	log ₂ FC	adj.P.Val	Gene Symbol	log ₂ FC	adj.P.Val	Gene Symbol	log ₂ FC	adj.P.Val
SLTM	-0.24	0.063	CBX8	1.28	0.011	ZYX	-0.44	0.011
SRRT	-0.24	0.063	FAM199X	0.79	0.014	SPECC1	-0.44	0.018
RBM17	-0.24	0.063	CBFA2T3	0.76	0.011	DHRS4	-0.44	0.069
EIF3E	-0.25	0.048	LMX1B	0.71	0.003	TTK	-0.45	0.072
MAGED1	-0.25	0.063	PFN2	0.58	0.0032	SEC24B	-0.46	0.094
HSPH1	-0.25	0.074	ARID3A	0.55	0.0021	ECT2	-0.47	0.094
SF3B2	-0.26	0.061	LSM8	0.53	0.048	QKI	-0.48	0.011
DDX46	-0.26	0.063	OTX1	0.52	0.038	FYCO1	-0.48	0.074
SRSF2	-0.27	0.061	NUCKS1	0.5	0.0057	USO1	-0.49	0.063
APC	-0.27	0.087	LDB1	0.47	0.074	MAEA	-0.52	0.067
SF3B1	-0.28	0.036	H1FO	0.44	0.03	TOP2A	-0.52	0.091
UHRF1	-0.28	0.044	ERF	0.43	0.074	MPHOSPH8	-0.54	0.0053
JUNB	-0.28	0.061	RCOR3	0.42	0.082	CIZ1	-0.57	0.0065
PIK3C2A	-0.29	0.027	AKAP8	0.39	0.061	KCTD20	-0.58	0.00063
ELAVL1	-0.29	0.055	DDX47	0.38	0.027	RUNX2	-0.58	0.066
FBRSL1	-0.29	0.069	RXRB	0.37	0.035	MOB4	-0.6	0.09
PNN	-0.3	0.029	ARHGDI A	0.37	0.072	TFAP4	-0.65	0.00063
RPL36A	-0.3	0.074	RBM4B	0.37	0.086	SQSTM1	-0.65	0.0031
CFAP20	-0.3	0.078	RREB1	0.36	0.063	CCND1	-0.72	0.024
PDLIM7	-0.3	0.094	GTF2I	0.35	0.0031	BTBD10	-0.92	5.50E-05
VEZF1	-0.31	0.082	RXRA	0.35	0.069	BUB3	-0.97	6.00E-07
CDK4	-0.32	0.027	USP15	0.34	0.066	TDG	-1.06	0.00018
GBE1	-0.32	0.089	EWSR1	0.34	0.069	ZNF207	-1.72	4.10E-06
NMD3	-0.33	0.04	RARA	0.34	0.096			
CDK11B	-0.33	0.074	EP300	0.33	0.085			
TET2	-0.33	0.081	CANX	0.32	0.027			
CTCF	-0.34	0.072	BAZ2A	0.31	0.033			
SDCCAG3	-0.34	0.072	CRIP2	0.3	0.063			
RIF1	-0.34	0.089	TARS	0.3	0.091			
TRIP6	-0.35	0.0057	NFATC2IP	0.29	0.08			
ZC3H4	-0.35	0.01	NR2F6	0.28	0.082			
SRSF11	-0.35	0.014	RERE	0.28	0.094			
CDK5RAP2	-0.35	0.066	PRDX6	0.27	0.04			
SEC16A	-0.35	0.082	HMCES	0.27	0.063			
CBFB	-0.36	0.042	NACC1	0.26	0.074			
PAWR	-0.36	0.048	DCP1A	0.26	0.077			
KRT80	-0.37	0.038	CBX5	0.25	0.063			
TAPT1	-0.37	0.063	SRP14	0.23	0.09			
YBX3	-0.38	0.0057	CLK3	0.22	0.061			
SEC24A	-0.39	0.077	KMT2D	0.22	0.072			
PDXDC1	-0.39	0.088	SMU1	0.21	0.066			
PPP1R12A	-0.4	0.01	WRNIP1	0.19	0.082			
SRRM2	-0.41	0.0031	SF1	-0.21	0.087			
H3F3A	-0.42	0.011	RSRC1	-0.21	0.09			
HNRNPUL1	-0.42	0.027	FAU	-0.22	0.063			
L3MBTL2	-0.42	0.061	PPHLN1	-0.23	0.062			
UPF2	-0.42	0.072	TRIM25	-0.23	0.067			
DNAJC8	-0.43	0.0018	CHERP	-0.23	0.082			
TJP1	-0.43	0.0018	FAM120A	-0.24	0.027			
FAM208A	-0.43	0.0021						

Supplementary Data 2: Regulated proteins identified in MED1-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.

GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val
BUB3	-0.75	2.90E-05	SRP14	0.33	0.037	CLMP	0.37	0.074
TDG	-0.88	0.00029	UBN2	0.4	0.038	MED13	-0.56	0.074
NSD3	-0.6	0.0019	POLDIP3	0.29	0.04	HADHA	0.63	0.076
TET2	-0.57	0.002	TRIM33	-0.27	0.041	BCLAF1	0.36	0.076
MLLT3	-0.89	0.0028	NUDCD1	-0.33	0.041	AKAP8	0.25	0.077
AFF4	-0.72	0.0034	ANP32E	-0.4	0.041	SS18	-0.32	0.078
LUC7L	0.44	0.0034	TBL1XR1	-0.27	0.041	EWSR1	0.23	0.078
TRIM27	-0.42	0.0034	UBXN7	-0.31	0.042	ARL6IP5	0.4	0.08
USP11	-0.64	0.0034	PSIP1	-0.39	0.042	EP300	-0.34	0.08
L3MBTL2	-0.43	0.0034	RBM12B	0.42	0.043	FSBP	-0.67	0.081
RCOR1	-0.42	0.0034	CLCN5	0.4	0.044	CCNT1	-0.25	0.081
DNAJC8	-0.53	0.0055	PTMA	-0.38	0.044	NOP56	0.46	0.081
LUC7L2	0.36	0.0066	DCAF7	-0.31	0.047	DLD	0.45	0.081
NFIC	-0.59	0.0066	CCND1	-0.55	0.05	BICRA	-0.52	0.081
YAP1	-0.88	0.0066	CREBBP	-0.44	0.05	ARF5	0.26	0.081
HMG20A	-0.42	0.0074	NPAT	-0.34	0.051	MCM3	-0.31	0.081
PFN2	0.59	0.0081	QKI	-0.41	0.052	MORF4L1	-0.3	0.081
DDR1	0.48	0.0083	HCFC1	-0.22	0.052	PKP3	0.33	0.083
VANGL1	-0.44	0.0083	HADHB	0.5	0.052	FXR1	0.47	0.084
SMC1A	-0.37	0.0083	HSPD1	0.43	0.054	GBE1	-0.5	0.092
SUMO3	-0.44	0.0083	ZC3H4	-0.36	0.056	LUC7L3	0.22	0.092
IRF2BP2	-0.68	0.0091	MED6	-0.65	0.083	UBAP2L	0.24	0.092
STAG2	-0.7	0.0091	RBM6	0.73	0.06	NCOA5	0.34	0.092
GFPT1	-0.5	0.012	RREB1	0.3	0.06	TKFC	0.46	0.092
SMC3	-0.31	0.015	MED22	-0.63	0.076	ZNF638	0.39	0.092
CLK3	0.36	0.016	UHRF1	-0.45	0.061	RPL36A	-0.28	0.092
SQSTM1	-0.58	0.019	KRR1	0.47	0.061	MGMT	0.26	0.092
TLK2	0.31	0.02	RIF1	-0.43	0.061	PATZ1	-0.34	0.092
EDC4	0.32	0.02	DDX27	0.41	0.062	EMG1	0.32	0.092
NMD3	-0.4	0.02	CRIP2	0.26	0.062	MED25	-0.5	0.06
MED12L	-0.85	0.06	ANKRD11	0.82	0.064	MAGI1	0.51	0.094
MED29	-0.69	0.02	SSBP1	0.38	0.065	DNAJA2	0.23	0.094
IDH2	0.38	0.02	SPINT2	0.37	0.067	H1FO	0.45	0.096
HNRNPUL1	-0.43	0.02	SUGP2	0.32	0.067	CHMP5	0.35	0.096
JUNB	-0.47	0.023	ELAVL1	-0.25	0.067	TOR1A	0.41	0.098
SMARCA5	-0.41	0.023	EHD1	0.55	0.067	MED27	-0.45	0.093
SRSF5	0.36	0.024	USP10	-0.38	0.067	MED16	-0.43	0.074
CKS1B	-0.59	0.024	KRT9	-0.85	0.068	MED20	-0.42	0.028
CBX3	-0.31	0.024	MFAP1	0.27	0.068	TMED10	0.35	0.02
AGO2	0.53	0.024	CCT7	-0.3	0.068			
GRHL2	-0.27	0.025	OGT	-0.33	0.068			
KDM1A	-0.27	0.025	MED4	-0.62	0.068			
PLK1	-0.56	0.025	HK1	0.26	0.068			
FAM199X	0.33	0.025	SNRPD1	0.29	0.068			
FUBP3	0.34	0.025	MCM7	-0.32	0.068			
RCOR3	0.5	0.028	CDK19	-0.66	0.068			
TJP1	-0.37	0.028	MLH1	-0.63	0.068			
MED28	-0.67	0.033	ADSS	-0.55	0.068			
NCOA3	-0.37	0.028	EMD	0.22	0.071			
SERPINB12	-0.47	0.028	VEZF1	-0.46	0.073			
DDX47	0.29	0.028	RBM23	0.32	0.073			
EIF4E2	0.45	0.031	LSM2	0.3	0.074			
PIAS1	-0.31	0.031	CIT	-0.59	0.074			
DIS3	-0.37	0.031	MED14	-0.6	0.078			
MED17	-0.66	0.081	MLLT1	-0.39	0.074			
SLC5A2	-0.62	0.035	JADE3	-0.37	0.074			
ALDH18A1	-0.34	0.035	KRT17	-0.3	0.074			
TONSL	-0.48	0.037	HMGB1	-0.26	0.074			
NUP50	-0.45	0.037	TRPS1	-0.35	0.074			
			HIRA	0.23	0.074			

Supplementary Data 3: Regulated proteins identified in MED12-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.

GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val
ZNF207	-1.4	1.20E-06	TRIP6	-0.33	0.036	DSTN	0.34	0.062
BUB3	-1.29	9.50E-07	MCM3	-0.33	0.037	LUC7L2	0.35	0.027
BMP2K	-0.99	0.064	TJP1	-0.33	0.062	SRSF5	0.35	0.041
SETD2	-0.91	0.0016	MED24	-0.33	0.085	CELF1	0.35	0.044
SCOC	-0.91	0.018	SF3B5	-0.31	0.064	UBE2M	0.35	0.049
ANAPC4	-0.87	0.064	DICER1	-0.31	0.083	ANXA7	0.35	0.056
AFF4	-0.85	6.10E-05	MED1	-0.3	0.055	C7orf50	0.35	0.061
ANAPC1	-0.84	0.00097	MED15	-0.3	0.062	PRKRIP1	0.35	0.062
CDC16	-0.78	0.0024	CSE1L	-0.3	0.096	NONO	0.35	0.071
CDC23	-0.76	4.00E-04	NUP160	-0.3	0.096	MPG	0.35	0.073
KPNA3	-0.76	0.0016	CSDE1	-0.29	0.039	NFIB	0.35	0.075
PIK3C2A	-0.73	0.018	FAM120A	-0.29	0.047	GIGYF2	0.36	0.041
CDC26	-0.68	0.0091	SRRM2	-0.29	0.056	ALDH16A1	0.36	0.056
PICALM	-0.67	0.062	MED27	-0.29	0.063	HNRNPLL	0.37	0.03
FCHO2	-0.65	0.02	KPNA2	-0.28	0.036	HNRNPF	0.37	0.059
FOSL2	-0.64	0.038	NCBP1	-0.28	0.069	HNRNPA2B1	0.37	0.062
GLUL	-0.63	0.04	MED25	-0.28	0.096	NOP16	0.37	0.066
BRIP1	-0.63	0.049	SF3B1	-0.28	0.099	ATP6V0A1	0.38	0.018
CDC27	-0.62	4.00E-04	MCM5	-0.25	0.085	PABPN1	0.38	0.03
USP10	-0.61	0.085	G3BP1	-0.24	0.065	SH3BP5	0.38	0.038
TET2	-0.58	0.0028	PRKDC	-0.24	0.096	SFPQ	0.38	0.062
CDK5RAP3	-0.57	0.0031	MED14	-0.24	0.097	EMG1	0.38	0.062
YAP1	-0.56	0.058	OGT	-0.23	0.082	RAVER1	0.39	0.026
DIS3	-0.56	0.067	SMC1A	-0.22	0.062	HMGAI	0.39	0.031
TRAFD1	-0.55	0.061	SMC3	-0.21	0.094	HNRNPUL2	0.39	0.035
CLTB	-0.55	0.093	PRKCD	0.2	0.092	HNRNPDL	0.39	0.058
ITSN1	-0.54	0.027	NACC1	0.21	0.094	GAR1	0.39	0.064
TFRC	-0.54	0.044	DHX38	0.22	0.095	PPP4C	0.39	0.067
CBFB	-0.53	0.0088	TRIR	0.25	0.085	SEMA4C	0.4	0.018
AP2M1	-0.53	0.07	RNGTT	0.26	0.062	HNRNPR	0.4	0.041
SMARCC1	-0.52	0.045	RCC1	0.26	0.07	HINT1	0.4	0.068
CDK9	-0.52	0.048	RBM25	0.26	0.085	ZCCHC17	0.41	0.049
MED29	-0.51	0.03	EIF4H	0.27	0.087	SRP14	0.42	0.02
RFC5	-0.51	0.038	H1FO	0.28	0.062	ARHGDI	0.42	0.044
HLTF	-0.49	0.03	TIAL1	0.28	0.073	DAZAP1	0.42	0.059
CCNT1	-0.49	0.082	ARMT1	0.28	0.075	FUBP3	0.43	0.024
NUP62	-0.48	0.024	MMTAG2	0.28	0.096	MBNL1	0.43	0.068
NUP107	-0.47	0.058	TALDO1	0.29	0.038	CTSG	0.43	0.071
ANAPC16	-0.46	0.0068	HMGB2	0.29	0.058	LMX1B	0.43	0.077
ANAPC7	-0.45	0.0053	HNRNPU	0.29	0.064	EBNA1BP2	0.43	0.082
MLLT1	-0.45	0.018	SUGP2	0.29	0.064	CRIP2	0.44	0.0068
XPO6	-0.43	0.061	GTF2I	0.29	0.073	HNRNPA3	0.44	0.038
MLLT3	-0.43	0.062	HMCES	0.3	0.03	PFN2	0.44	0.044
MAD1L1	-0.42	0.037	SSB	0.3	0.047	LUC7L	0.45	0.018
CKS1B	-0.42	0.062	HMG2	0.3	0.058	KHDC4	0.45	0.04
TPR	-0.41	0.024	REPIN1	0.3	0.068	INTS4	0.45	0.065
AP2A1	-0.41	0.071	TRA2B	0.3	0.094	LANCL1	0.46	0.0048
CIZ1	-0.4	0.064	APEX1	0.31	0.037	ESRR	0.46	0.047
AP2B1	-0.4	0.096	PCBP1	0.31	0.058	CIRBP	0.46	0.049
CDK4	-0.39	0.03	STRBP	0.31	0.061	VPS33A	0.46	0.058
WDHD1	-0.39	0.037	PDCD6	0.31	0.062	PSMB3	0.46	0.071
NUP93	-0.39	0.058	PCBP2	0.31	0.062	NUCKS1	0.48	0.0097
FAM98A	-0.38	0.03	TARDBP	0.31	0.062	AKAP8	0.48	0.018
DCAF7	-0.38	0.03	C11orf98	0.31	0.062	PRR3	0.48	0.064
MED23	-0.38	0.059	EZR	0.31	0.083	CLK3	0.49	0.018
SEC23IP	-0.38	0.064	RBM6	0.31	0.089	ARGLU1	0.5	0.037
NUP58	-0.38	0.068	RBMX	0.31	0.097	HNRNPA0	0.51	0.024
QKI	-0.37	0.03	HMG2	0.32	0.051	KHDRBS1	0.51	0.085
SUPT20H	-0.37	0.031	MYEF2	0.32	0.051	HSPD1	0.53	0.068
NUP50	-0.37	0.045	KHSRP	0.32	0.064	TSR1	0.54	0.077
DNAJC8	-0.36	0.026	PYGO2	0.32	0.069	PHF6	0.55	0.0028
FNBP4	-0.36	0.031	DHX9	0.32	0.071	RBM23	0.55	0.032
USP39	-0.36	0.044	ZRANB2	0.33	0.037	ZMAT5	0.55	0.097
NUP205	-0.36	0.058	DBR1	0.33	0.039	EWSR1	0.56	0.019
MED30	-0.36	0.059	FUBP1	0.33	0.062	EIF6	0.6	0.058
CAPRIN1	-0.36	0.064	RBM45	0.33	0.078	ZNF664	0.61	0.0068
CTCF	-0.35	0.037	RBMXL1	0.33	0.096	GCA	0.61	0.0088
NUP133	-0.35	0.058	HNRNPL	0.33	0.096	CXXC5	0.61	0.037
TSEN34	-0.34	0.056	HDGF	0.34	0.022	MEAF6	0.62	0.024
NUP98	-0.34	0.063	THUMP1	0.34	0.033	RBM4B	0.74	0.0016
			VIM	0.34	0.058			

Supplementary Data 4: Regulated proteins identified in MED4-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.

GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val
ZNF207	-1.69	3.00E-07	NMD3	-0.32	0.02	CTNNBL1	0.25	0.043
LONRF2	-1.24	0.0012	HBS1L	-0.32	0.032	FUBP3	0.25	0.071
BUB3	-0.85	3.00E-05	RPL21	-0.32	0.032	CRIP2	0.26	0.034
BRIP1	-0.77	0.0095	RPS21	-0.32	0.04	AGO2	0.26	0.099
TET2	-0.71	0.0017	RPL18A	-0.32	0.041	GLYR1	0.27	0.046
SPTLC1	-0.68	0.00043	MED13	-0.32	0.041	GTF2B	0.28	0.041
MED12L	-0.67	0.00043	RPS17	-0.32	0.071	PNISR	0.28	0.041
SMTN	-0.67	0.011	RPL35A	-0.31	0.028	PFN2	0.28	0.041
RABGAP1	-0.64	0.0095	ABCE1	-0.31	0.039	ALB	0.28	0.06
SRPK2	-0.64	0.012	HNRNPH3	-0.31	0.041	TCERG1	0.29	0.016
CCND1	-0.62	0.033	KDM1A	-0.31	0.049	WDR18	0.29	0.089
SYNE4	-0.62	0.043	LCMT1	-0.31	0.082	RBM23	0.3	0.045
CCNL2	-0.56	0.061	RPRD1A	-0.3	0.025	HDGF	0.3	0.079
TNS3	-0.56	0.092	MED10	-0.3	0.047	PELP1	0.33	0.052
RAB6A	-0.53	0.04	RPL37A	-0.3	0.058	PURB	0.34	0.032
BIRC5	-0.52	0.058	CCNA2	-0.3	0.06	IDH2	0.34	0.071
CDCA8	-0.52	0.066	MED14	-0.3	0.076	PHB2	0.34	0.091
SLC27A3	-0.52	0.076	RPS12	-0.29	0.032	NKTR	0.35	0.032
KANK2	-0.51	0.071	RPL28	-0.29	0.041	DCP1A	0.36	0.0078
CKS1B	-0.5	0.0062	EMC2	-0.29	0.071	LAS1L	0.36	0.049
RNF216	-0.5	0.025	ANAPC7	-0.29	0.073	LANCL1	0.37	0.0068
SMCHD1	-0.5	0.028	NQO1	-0.29	0.091	TRPM4	0.37	0.041
GPRC5A	-0.5	0.061	RPS23	-0.28	0.076	RBM4B	0.38	0.032
PLD1	-0.5	0.071	MED17	-0.28	0.088	STOML2	0.38	0.068
PLK1	-0.49	0.0085	RPL23	-0.27	0.041	ZFC3H1	0.47	0.0017
SPAG5	-0.49	0.058	MED8	-0.27	0.079	EIF4E2	0.47	0.024
CASP8AP2	-0.48	0.062	NAPG	-0.26	0.032	ATP5F1A	0.48	0.093
IQSEC1	-0.48	0.063	RNF114	-0.26	0.038	ATP5F1B	0.49	0.038
ASAP3	-0.46	0.023	PPP6C	-0.26	0.049	DOCK6	0.49	0.071
RSBN1L	-0.45	0.011	RPS2	-0.26	0.076	ATP5PO	0.52	0.071
GSE1	-0.44	0.046	RAB14	-0.26	0.082	TBKBP1	0.55	0.046
NUDT16L1	-0.43	0.028	RPL3	-0.26	0.091	PRDM11	0.56	0.011
TSEN34	-0.42	0.021	LIMS1	-0.26	0.094	NUCKS1	0.64	0.0021
RHOBTB3	-0.42	0.055	ESRP1	-0.25	0.032	NOS1AP	0.64	0.045
HLTF	-0.41	0.032	TRIP6	-0.25	0.041	GIGYF1	0.73	0.00047
RRM2	-0.41	0.041	TRIM25	-0.25	0.046	ATP5F1C	0.77	0.047
KRT80	-0.41	0.058	PMM2	-0.25	0.064	ZBTB6	0.94	0.036
DYNLL2	-0.4	0.036	RPS15A	-0.23	0.041	ERVFC1	1	0.068
RPL37	-0.4	0.04	GRB2	-0.23	0.067			
JUNB	-0.4	0.088	RPL30	-0.23	0.082			
SRRM2	-0.4	0.091	TPP2	-0.23	0.089			
TOP3A	-0.4	0.096	RPL11	-0.23	0.09			
RPL34	-0.39	0.033	TRIM27	-0.23	0.091			
RPL12	-0.39	0.036	GTF2H3	-0.22	0.049			
PRMT3	-0.38	0.022	PABPC1	-0.21	0.043			
INCENP	-0.38	0.046	UHRF1	-0.21	0.071			
RCOR1	-0.37	0.0017	COL6A6	-0.21	0.08			
AFF4	-0.37	0.083	SUMO3	-0.21	0.09			
RPL15	-0.36	0.053	EIF5	-0.2	0.071			
RPL32	-0.35	0.041	DCAF7	-0.19	0.071			
RPL29	-0.35	0.062	AKAP8	0.18	0.071			
RPL8	-0.35	0.076	PRCC	0.19	0.088			
RPL36	-0.34	0.033	ZMAT2	0.2	0.077			
RPL36A	-0.34	0.041	EDC3	0.21	0.091			
RPL13A	-0.34	0.057	RAVER1	0.22	0.063			
RPS28	-0.33	0.032	HNRNPR	0.22	0.065			
PPP6R3	-0.33	0.049	MFAP1	0.22	0.08			
MED6	-0.33	0.071	GSTK1	0.23	0.086			
ZNF277	-0.33	0.084	IK	0.24	0.036			
			WBP11	0.24	0.05			
			HNRNPUL2	0.25	0.043			

Supplementary Data 5: Regulated proteins identified in MED1-qPLEX-RIME following ZNF207 knockdown in HEK293 cells.

GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val
BUB3	-0.96	5.00E-06	FBR5	-0.37	0.0091	TRRAP	-0.18	0.022	GATA6	-0.22	0.046	HIC2	0.38	0.083
CREB5	-1.17	1.50E-05	AFF1	0.46	0.0097	PPP2R1B	-0.51	0.022	POLR2I	-0.25	0.046	AKAP17A	0.21	0.084
MORC3	-0.95	3.60E-05	ATRX	0.33	0.0098	FBRS1	-0.25	0.022	CETN2	0.25	0.047	LDHA	-0.16	0.084
LUC7L	0.74	3.80E-05	DNAJC9	0.26	0.0099	PRRC2B	0.33	0.022	TMA7	0.24	0.047	CASP2	0.24	0.085
SPEN	-0.54	0.00012	MED22	-0.33	0.01	POLR2F	-0.25	0.022	KHSRP	0.16	0.047	HMG83	0.14	0.085
RPRD1A	-0.64	0.00015	NUP58	0.33	0.01	NSMCE1	-0.3	0.022	OPA1	0.46	0.048	MED15	-0.19	0.085
ZNF207	-2.01	0.00015	WRNIP1	-0.31	0.01	SCAF8	0.21	0.022	RNF2	-0.23	0.047	CELF1	0.3	0.087
DCAF7	-0.58	0.00034	TRIM28	0.22	0.011	CREBBP	-0.24	0.022	PARK7	0.21	0.048	GAPDH	0.19	0.088
AFF4	-0.64	0.00034	TAF5L	-0.27	0.011	RNF20	-0.19	0.023	FAM199X	0.57	0.048	ARPC2	-0.26	0.088
RNF216	-1.48	0.00034	ARNT	-0.41	0.011	TMEM214	0.38	0.023	THUMPD1	0.32	0.048	MFAP1	0.17	0.088
TRAFD1	-0.66	0.00034	CTBP2	-0.37	0.011	MED25	-0.28	0.023	ARHGDI	0.28	0.049	SF1	-0.14	0.088
MYC	-0.76	0.00034	NUP35	0.37	0.011	MNAT1	-0.34	0.023	CSNK2A2	-0.23	0.049	P3H4	-0.88	0.089
MED19	-0.55	5.00E-04	GCFC2	-0.25	0.012	RUVBL1	-0.17	0.023	PCGF3	-0.3	0.049	NUDT16L1	-0.26	0.09
TBL3	0.56	5.00E-04	MED16	-0.3	0.012	BIRC6	-0.41	0.023	WAC	-0.18	0.049	TNPO3	0.23	0.09
ABCE1	-0.79	5.00E-04	HMG82	0.22	0.012	CHMP4B	0.22	0.023	GON4L	-0.25	0.049	KCTD15	-0.27	0.091
HOGF	0.43	0.00058	PPP4R3A	0.32	0.012	DDX24	-0.24	0.023	NVL	-0.23	0.049	IVD	-0.25	0.091
RCOR1	-0.6	0.00058	MPHOSPH8	-0.27	0.012	NUP205	0.29	0.024	CBX4	-0.3	0.049	ASNS	-0.22	0.092
FEN1	0.36	0.00058	HNRNPR	0.29	0.012	HSPH1	-0.22	0.025	POLR2A	-0.16	0.049	AAAS	0.21	0.092
LUC7L3	0.42	0.00058	NUP98	0.38	0.013	GTFL2	0.19	0.026	SMC6	-0.21	0.049	DNAH2	0.24	0.092
WDR3	0.49	0.00058	MAT2A	-0.37	0.013	MED27	-0.26	0.026	UBA2	0.25	0.05	CENPE	-0.26	0.092
FUBP1	0.45	6.00E-04	SALL2	-0.4	0.013	DBR1	0.35	0.026	CDK19	-0.3	0.05	PPP2R1A	-0.16	0.094
NUP62	0.46	0.00074	MED14	-0.3	0.013	GNL3	-0.26	0.027	MCM6	0.17	0.051	VEZF1	-0.25	0.094
MED12L	-0.63	0.00085	SRSF5	0.28	0.013	ATN1	-0.3	0.027	SRRT	-0.14	0.051	BCAP31	-0.28	0.094
RCOR3	0.5	0.00085	RAD21	-0.27	0.014	ZMYND11	-0.33	0.027	RAVER1	0.21	0.051	TCERG1	0.12	0.094
HU1F	-0.44	0.0011	SRRM2	-0.27	0.014	UBE2M	0.22	0.027	RADS1AP1	0.26	0.051	CARHSP1	0.22	0.094
RBPJ	-0.48	0.0011	GLE1	0.35	0.014	POLR2K	-0.21	0.027	UBXN1	0.22	0.051	YTHDF2	0.34	0.095
LUC7L2	0.41	0.0013	ELAVL1	-0.26	0.014	NCL	0.23	0.027	PNKP	0.26	0.051	ZNF281	-0.17	0.095
GLUL	-0.9	0.0014	CDK9	-0.25	0.014	CARM1	-0.24	0.027	RPL37	0.27	0.052	HNRNPA2B1	0.14	0.098
SUPT20H	-0.42	0.0014	METAP1	0.4	0.014	DDX23	0.17	0.027	HMGN1	0.25	0.052	UTS2	-0.31	0.098
GFPT1	-0.53	0.0014	NACC1	0.36	0.015	MED18	-0.27	0.027	SRSF6	0.19	0.052	TAF3	-0.24	0.098
NUCKS1	0.46	0.0014	UNG	0.35	0.015	CDC27	-0.28	0.028	SAE1	0.21	0.052	STRAP	0.14	0.098
DIS3	-0.33	0.0016	SS18	-0.36	0.015	YWHAQ	0.17	0.029	NAN5	0.21	0.052	HMG20A	-0.18	0.098
GTf2H1	-0.63	0.0016	KDM18	-0.31	0.015	C9orf78	0.26	0.029	CENPB	0.44	0.052	ZNF830	0.19	0.098
GTf2H4	-0.35	0.0016	GID8	-0.34	0.015	ANAPC16	-0.81	0.029	ANP32A	-0.27	0.053	LARP7	-0.13	0.099
UBXN7	-0.44	0.0017	RBM3	-0.28	0.015	MED30	-0.28	0.029	ARGLU1	0.27	0.053	PBRM1	0.19	0.099
ERCC2	-0.38	0.0017	TAF6L	-0.25	0.015	POLR1A	-0.33	0.03	BCOR	-0.23	0.053	PES1	-0.23	0.099
MLLT3	-0.68	0.0018	RNF138	-0.51	0.015	CTDP1	-0.26	0.03	ARPC3	-0.27	0.053	RFC1	0.15	0.099
TXNDC12	-0.44	0.0019	ERCC3	-0.3	0.015	BMS1	-0.36	0.03	SGF29	-0.22	0.053	TAF1	-0.22	0.099
ZNF703	-0.49	0.0025	KCNH1	-0.33	0.015	SUPT17L	-0.25	0.03	FOXC1	-0.16	0.053	INTS11	-0.18	0.099
ANAPC7	-0.39	0.0026	MED6	-0.3	0.015	MED28	-0.27	0.03	CSNK1A1	0.21	0.054	ING3	-0.16	0.099
PEF1	-0.32	0.0026	UBE2S	-0.23	0.015	POLR2G	-0.21	0.03	ALKBH5	0.19	0.055	ARID3A	0.29	0.099
HOXB9	-0.49	0.0026	MED23	-0.3	0.015	DGCR8	-0.24	0.031	LSM7	0.21	0.055	NCBP1	-0.2	0.099
PHF6	0.36	0.0027	NUP88	0.28	0.015	UFC1	0.28	0.031	TRIM27	-0.23	0.055	WBP11	0.14	0.099
SPTLCL	-0.64	0.0028	YWHAZ	0.25	0.015	INTS14	-0.29	0.031	KPNB1	0.16	0.056			
PLK1	-0.32	0.003	CDK7	-0.34	0.016	PRPF3	0.22	0.032	ZBTB10	-0.2	0.056			
SMPD4	0.36	0.0031	MMS22L	-0.43	0.016	FOXK1	-0.25	0.032	NUMA1	0.17	0.058			
SMARCA1	0.53	0.0031	PALLD	-0.33	0.016	ZMAT2	0.2	0.032	CTCF	-0.23	0.058			
TAF12	-0.43	0.0035	ESD	0.24	0.016	PARP1	0.17	0.033	ZC3HC1	-0.17	0.058			
PWP2	0.39	0.0035	MED8	-0.3	0.017	TALDO1	0.27	0.033	GSTP1	0.24	0.058			
ATF3	-1.16	0.0039	PGD	0.25	0.017	H3F3A	-0.25	0.033	NUDT21	-0.17	0.059			
HNRNPUL2	0.31	0.0038	MED24	-0.28	0.017	SMC2	0.2	0.034	PFN1	0.17	0.059			
NDC1	0.41	0.0039	DCUN1D5	-0.27	0.018	RBM25	0.17	0.034	KIAA1143	0.2	0.061			
NUP210	0.44	0.0039	TAF5	-0.25	0.018	DNAJB1	0.23	0.034	GNL3L	-0.3	0.061			
PCGF5	-0.4	0.0041	NUP188	0.3	0.018	TMEM209	0.36	0.034	AKAP8	0.21	0.062			
C15orf39	-0.48	0.0042	TP53	0.36	0.018	POLR2B	-0.18	0.035	MED11	-0.21	0.062			
HSPA1B	-0.22	0.0042	TSEN34	-0.41	0.018	TAF9B	-0.38	0.035	EDF1	0.26	0.062			
DIMT1	-0.51	0.0043	EWSR1	0.29	0.019	PRPF38A	0.24	0.035	ANXA7	0.21	0.063			
CCDC97	-0.62	0.0046	TXNDC17	0.43	0.019	EED	-0.31	0.035	SIX4	-0.56	0.063			
TP53BP1	0.37	0.0051	BAG3	-0.26	0.019	ZNF608	-0.31	0.035	DHX40	0.17	0.063			
SUGP2	0.47	0.0052	PPME1	0.21	0.019	CAVIN1	-0.36	0.037	GTF3C5	-0.2	0.064			
CPNE3	-0.34	0.0054	RBM12	-0.2	0.019	DCAF1	0.29	0.037	MED29	-0.24	0.065			
CHEK2	-0.28	0.0054	MGA	-0.28	0.019	UBAP2L	0.37	0.038	DDA1	0.22	0.066			
CIT	-0.35	0.0055	NAGK	0.31	0.019	MCM8	-0.28	0.038	XRCC5	0.12	0.066			
RAE1	0.44	0.0066	SIRT1	-0.29	0.019	CMAS	-0.3	0.038	PIN4	0.2	0.067			
NUP93	0.31	0.0067	MED20	-0.29	0.019	CRKL	-0.23	0.038	BAP18	0.26	0.067			
SMCHD1	-0.44	0.0067	PCF11	-0.3	0.019	TAF6	-0.22	0.038	MYEF2	0.19	0.069			
HOXD13	-0.38	0.0068	POLR2D	-0.26	0.019	CCNA2	-0.28	0.038	RSU1	-0.32	0.069			
NUP54	0.38	0.0074	WDR36	0.34	0.019	ZRANB2	0.22	0.038	SDE2	-0.24	0.069			
ATXN7L3	-0.36	0.0074	KIF11	0.32	0.019	E2F3	-0.3	0.038	SF3B3	-0.15	0.069			
ZNF503	-0.33	0.0075	L3MBTL2	-0.25	0.019	LRRK1	0.51	0.04	DHX9	0.13	0.07			
ATF2	-0.65	0.0075	NUP214	0.23	0.019	HMGN4	0.4	0.04	CUL5	-0.18	0.07			
TXLNA	0.58	0.0075	DNAJB6	-0.26	0.019	BANF1	0.38	0.04	POLR2C	-0.16	0.07			
HOXA11	-0.51	0.0075	MED17	-0.27	0.019	NSMCE3	-0.22	0.042	CUL2	-0.17	0.071			
NSRP1	-0.31	0.0075	ATF7	-0.45	0.019	RBBP7	-0.33	0.042	RECQL	0.16	0.072			
KDM1A	-0.3	0.0075	INTS13	-0.22	0.019	HNRNPA0	0.25	0.042	POLR2E	-0.18	0.072			
SRP14	0.25	0.0075	RIOX2	-0.23	0.02	DEK	0.18	0.042	KIF4A	0.2	0.073			
ZIC2	-0.44	0.0075	SMC5	-0.24	0.02	TRPS1	-0.21	0.042	ATXN7L2	-0.25	0.073			
YWHAE	0.23	0.0076	TAF10	-0.34	0.02	JUN	-0.5	0.042	PHF14	-0.28	0.073			
FUBP3	0.33	0.0084	RAI1	-0.31	0.02	EDC4	0.36	0.043	STMN1	0.2	0.074			
LANCL1	0.31	0.0084	SFPQ	0.18	0.02	RPP30	0.22	0.044	RUVBL2	-0.12	0.075			
EP300	-0.33	0.0084	TARDBP	0.19	0.02	ZC3H4	-0.18	0.044	RBM23	0.33	0.075			
HNRNP3	-0.32	0.0088	TASOR	-0.25	0.02	RCN2	-0.26	0.044	BAZ1A	-0.14	0.075			
CBX5	0.28	0.0088	EIF4H	0.24	0.021	TFRC	-0.38	0.044	PNISR	0.22	0.075			
CCAR2	0.29	0.0088	CSR2	-0.31	0.021	PSME3	0.17	0.044	API5	0.13	0.076			
STAG2	-0.42	0.0088	ACP1	0.25	0.021	TADA3	-0.24	0.044	DBT	0.2	0.077			
OGT	-0.25	0.0088	KDM4A	0.21	0.021	TAF15	-0.24	0.044	CUTC	-0.25	0.078			
PDCD6	-0.24	0.0089	SIN3B	0.28	0.021	YAP1	-0.33	0.045	TLE4	-0.15	0.079			
TADA1	-0.33	0.0089	CKS1B	-0.23	0.021	TRIP6	-0.24	0.045	EMC7	-0.25	0.079			
NUP155	0.32	0.0089	NFYA	-0.35	0.021	ACLY	-0.17	0.046	ZNF579	0.29	0.079			
			NKTR	0.33	0.022	NUTF2	0.24	0.046	TTL12	0.29	0.079			
									NONO	0.16	0.081			

Supplementary Data 6: *Regulated proteins identified in MED14-qPLEX-RIME following ZNF207 knockdown in HEK293 cells.*

GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val
PELO	1.11	0.00065	NUCKS1	0.37	0.052
MORC3	-0.94	0.0013	KRT16	0.78	0.052
ATF3	-1.06	0.0013	ATXN7L3	-0.5	0.052
FSCN1	-0.94	0.0013	ERCC2	-0.37	0.052
AFF4	-0.84	0.0013	C15orf39	-0.42	0.054
CREB5	-1.12	0.0018	MED24	-0.42	0.054
BUB3	-0.99	0.0042	CCDC9	-0.38	0.055
RPRD1A	-0.62	0.0042	NUP37	-0.34	0.056
LUC7L	0.63	0.0064	IGF2BP1	0.35	0.056
YAP1	-0.76	0.0064	TAF10	-0.46	0.058
SPEN	-0.69	0.0064	ABHD14B	0.54	0.063
ZNF207	-1.37	0.0064	WDR36	0.52	0.063
DIMT1	-0.61	0.0064	TFAP4	-0.42	0.063
TBL1X	-0.71	0.0068	UTS2	0.49	0.064
LANCL1	0.55	0.0079	FEN1	0.31	0.065
MED12L	-0.52	0.018	KRT14	0.64	0.065
RBPJ	-0.5	0.023	SIRT1	-0.44	0.065
DCAF7	-0.49	0.025	CYB5R3	-0.49	0.065
ABCE1	-0.44	0.027	SUGP2	0.34	0.065
PWP2	0.63	0.031	RCOR3	0.43	0.065
SERPINB12	0.64	0.044	KRT9	0.77	0.065
HLTF	-0.36	0.044	CLK1	-0.44	0.066
HNRNPUL2	0.44	0.044	C7orf50	0.39	0.066
FBRSL1	-0.45	0.044	NFYA	-0.4	0.066
KRT77	0.81	0.044	ZNF629	0.43	0.066
MROH2B	-0.5	0.044	GSTP1	0.44	0.066
MED23	-0.45	0.044	NSRP1	-0.33	0.066
MNAT1	-0.43	0.044	RBBP7	-0.36	0.066
RRP1B	-0.72	0.044	TAF5	-0.4	0.066
PHF6	0.35	0.044	KRT78	0.61	0.073
SUPT20H	-0.43	0.044	CDK7	-0.38	0.073
TBL3	0.51	0.044	KRT79	0.64	0.073
PCGF5	-0.51	0.044	EYA4	-0.44	0.073
STAG2	-0.53	0.044	MED19	-0.48	0.073
PUM2	0.52	0.047	WRNIP1	-0.36	0.082
JUN	-0.78	0.047	APEX1	0.28	0.082
DNAJC9	0.33	0.047	SS18L2	-0.57	0.083
WDR3	0.5	0.047	HDGF	0.35	0.083
PLK1	-0.35	0.047	KRT1	0.64	0.083
CTBP2	-0.38	0.047	HMGN4	0.38	0.085
MED16	-0.42	0.047	UBE2M	0.25	0.087
KRT17	1.19	0.047	PHPT1	0.45	0.087
TADA2B	-0.47	0.047	INTS13	-0.27	0.087
FBR5	-0.42	0.048	MYEF2	0.38	0.092
TAF5L	-0.49	0.048	WDR55	0.37	0.092
BANF1	0.36	0.048	TADA3	-0.32	0.098
UBXN7	-0.44	0.049	SKP2	-0.32	0.099
SPTLC1	-0.72	0.05	AFF1	0.32	0.099
TAF6L	-0.44	0.052	MEIS1	-0.47	0.099

Supplementary Data 7: Regulated proteins identified in Pol II-qPLEX-RIME following ZNF207 knockdown in MCF7 cells.

GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val	GeneSymbol	log ₂ FC	adj.P.Val
ZNF207	-1.38	2.10E-06	SMC1A	-0.29	0.033	MRFAP1	-0.41	0.069	FAM208B	0.18	0.096
BUB3	-1	3.30E-06	GOLGB1	0.48	0.033	H3F3A	-0.4	0.071	CHMP1A	0.33	0.096
CCDC97	-1.88	3.30E-06	ZFR	-0.23	0.033	STAG2	-0.82	0.071	KCNT2	0.4	0.096
QKI	-0.8	1.30E-05	ILF3	-0.25	0.033	PKHD1L1	0.62	0.072	MCM7	-0.29	0.096
SETD2	-0.8	1.40E-05	ZCCHC17	0.58	0.033	IKZF2	0.46	0.072	SREK1IP1	0.23	0.096
HNRNPUL1	-0.53	0.00011	ZNF8	0.6	0.035	KMT2A	0.28	0.072	HSPA4	0.24	0.096
MORC3	-0.98	0.00012	CCAR1	-0.26	0.037	THUMPD1	0.2	0.072	ABT1	0.49	0.096
CBX5	0.7	7.00E-04	RBM17	-0.25	0.041	SF3B5	-0.3	0.072	PTBP3	0.36	0.096
ELAVL1	-0.39	0.00081	NAN5	0.35	0.041	CXCC1	-0.25	0.073	PSMA7	-0.24	0.096
AFF4	-0.54	0.0013	PCYT1A	0.27	0.041	MYO18A	0.6	0.075	CBX3	0.22	0.096
DNAJC8	-0.55	0.0013	CCDC59	0.31	0.041	SRSF5	0.31	0.076	NAPA	0.29	0.096
VEZF1	-0.75	0.0013	TRIM24	0.64	0.042	USP3	0.39	0.076	PHC2	0.27	0.096
HNRNPH3	-0.41	0.0013	PEBP1	0.23	0.042	TONSL	-0.31	0.076	ECT2	1.02	0.096
PRDM11	0.66	0.0015	VPS13C	0.34	0.042	ZMYND11	-0.24	0.076	MORF4L1	-0.35	0.096
CBX4	-0.46	0.0024	WRNIP1	0.33	0.042	GTF2I	0.22	0.076	ALDH18A1	-0.36	0.096
SRBD1	-0.47	0.0033	SET	0.26	0.042	A2M	0.46	0.078	RBPMS	-0.24	0.096
PAPOLA	-0.55	0.005	ZNF768	0.44	0.042	BUD23	-0.21	0.078	RPL32	-0.45	0.096
RBAK	0.57	0.0056	PPP4R2	0.44	0.042	SMC3	-0.21	0.078	RTCA	0.44	0.096
HLTF	-0.57	0.0056	CHCHD3	0.37	0.043	HNRNPC	-0.18	0.078	TJP2	0.47	0.096
CBFB	-0.66	0.0061	CHD1	-0.35	0.043	ESRP2	-0.22	0.078	ZSWIM8	0.43	0.097
ZNF689	0.94	0.0081	SCAF1	0.39	0.043	SETD1B	0.28	0.078	ZNF707	0.69	0.097
SFSWAP	-0.46	0.0081	RPL21	-0.38	0.043	AFF1	0.33	0.078	SMTN	-0.3	0.098
NUCKS1	0.31	0.0092	EPS8L1	0.77	0.044	AKAP8	0.17	0.078	RPRD2	0.25	0.098
CLK3	0.41	0.01	SF3B2	-0.34	0.044	C1QC	0.44	0.078	RPL13A	-0.43	0.098
TRIM28	0.67	0.01	NPLC4	-0.67	0.045	MLLT6	0.48	0.078	SMN1	-0.41	0.098
HNRNPH1	-0.26	0.011	PHF6	0.37	0.045	FAM76B	0.39	0.078	NET1	-0.38	0.098
PSIP1	-0.54	0.011	PRKDC	-0.37	0.045	RPS28	-0.25	0.081	MSH2	-0.23	0.098
NUDT16L1	-0.38	0.011	DXO	0.34	0.045	SNX12	0.52	0.081	PHF2	-0.24	0.098
FAM103A1	-0.34	0.011	CLASRP	-0.24	0.045	SUGP1	-0.2	0.081	PDS5B	-0.23	0.098
CBX8	0.44	0.011	HSPD1	-0.33	0.046	NDRG1	0.59	0.081	DCAF7	-0.18	0.098
KPNA3	-0.67	0.011	DNAJB1	0.22	0.046	HNRNPD	-0.19	0.084	MFAP1	0.23	0.098
CASP8AP2	-0.44	0.011	IKZF5	0.81	0.046	SYF2	0.3	0.086	MEAF6	-0.36	0.098
CENPB	0.6	0.012	SF3A3	-0.36	0.046	KHDRBS1	-0.19	0.086	RNGTT	0.18	0.098
SLBP	-0.38	0.012	PTMA	-0.27	0.048	LENG1	0.2	0.086	ARHGDI1A	0.25	0.098
LUC7L2	0.43	0.012	CMTR1	0.31	0.048	BUD13	0.29	0.087	SUPT4H1	-0.22	0.098
PSPC1	-0.26	0.012	CROCC2	0.57	0.048	RBM12B	0.21	0.088	SF3B3	-0.19	0.099
LUC7L	0.5	0.015	CCNT1	-0.27	0.048	SMS	0.39	0.089			
INTS4	0.28	0.015	MLLT1	-0.52	0.05	RBM6	0.21	0.089			
SCAI	-0.41	0.015	CCNT2	-0.27	0.052	IRS1	0.22	0.089			
THOC2	0.48	0.015	RPL15	-0.36	0.054	RIF1	-0.33	0.089			
CHERP	-0.27	0.015	BLVRB	0.22	0.055	PIN4	0.23	0.089			
CTCF	-0.54	0.015	CWC25	0.37	0.055	GRHL2	-0.32	0.089			
RPL18A	-0.37	0.016	EWSR1	0.19	0.055	STK32C	0.42	0.089			
U2SURP	-0.27	0.016	RPS17	-0.53	0.055	PCIF1	0.18	0.089			
RBM23	0.5	0.016	H1FO	0.34	0.055	NSRP1	0.35	0.089			
CDK9	-0.37	0.017	XRN2	-0.17	0.055	ZMAT2	0.24	0.089			
SETX	0.39	0.018	DDX39A	0.25	0.055	JUNB	-0.56	0.089			
SNIP1	0.46	0.018	SNRPB2	-0.27	0.055	NFIB	0.29	0.089			
POGK	0.39	0.018	HP1BP3	-0.29	0.057	GTPBP4	0.17	0.089			
ZNF268	0.57	0.018	UHRF1	-0.42	0.058	TRNT1	0.22	0.089			
SF1	-0.23	0.019	RBM4B	0.3	0.058	GTF2H3	-0.26	0.089			
CHD2	-0.34	0.02	NMD3	-0.27	0.058	PNPO	0.33	0.09			
FUS	-0.26	0.02	IRF2BP2	-0.26	0.058	INPP5K	0.38	0.09			
FKBP15	0.74	0.02	CLK2	0.4	0.059	ZMYM2	-0.36	0.09			
TCEA2	0.38	0.021	ILF2	-0.27	0.059	ERCC6	-0.19	0.091			
ESRP1	-0.26	0.023	DDX47	0.27	0.06	TAF3	-0.43	0.091			
FAM199X	0.59	0.023	MED22	0.24	0.06	TRPS1	-0.24	0.091			
SF3B1	-0.31	0.023	TSEN34	-0.24	0.06	NXF1	0.18	0.091			
SRP14	0.26	0.023	SF3B6	-0.21	0.06	HSP90AA1	-0.38	0.091			
SINHCAF	-0.86	0.023	KDM1A	-0.38	0.06	LYZ	0.41	0.092			
PM20D2	0.56	0.023	UIMC1	0.52	0.06	SSU72	0.29	0.092			
USP39	-0.37	0.024	CRIP2	0.23	0.06	BRX1	0.33	0.092			
CTBP1	0.36	0.024	TRPC4AP	0.44	0.06	XPO5	-0.23	0.092			
KRR1	0.37	0.024	C3	0.37	0.062	PDCD7	0.51	0.092			
BAZ2A	0.49	0.024	SF3A1	-0.32	0.062	LAD1	0.44	0.092			
SYNCRIP	-0.3	0.027	ZNF140	0.44	0.063	DUSP23	0.31	0.092			
RPS23	-0.25	0.027	RPS6KA4	0.28	0.064	CWF19L2	0.27	0.092			
CHMP5	0.34	0.027	TSGA10	0.37	0.064	ARGLU1	0.34	0.092			
IBTK	0.5	0.027	EIF5A	0.25	0.065	WTAP	0.24	0.092			
SRP9	0.4	0.027	SF3A2	-0.37	0.065	RRP15	-0.22	0.092			
PFN2	0.24	0.03	IMP4	0.51	0.065	TAF15	-0.16	0.092			
TTLL12	0.41	0.03	PPP1CC	0.29	0.065	L1RE1	0.28	0.092			
ABCE1	-0.51	0.03	XPC	-0.26	0.065	RBM4	-0.17	0.092			
HMGB1	-0.31	0.031	TCEAL4	0.23	0.065	EPM2AIP1	-0.27	0.092			
RBPJ	0.29	0.032	PPIL2	0.2	0.066	TBPL1	0.29	0.092			
MED31	0.46	0.032	CWC27	0.26	0.066	BCLAF1	0.32	0.092			
HNRNPA1	-0.24	0.032	HSPH1	-0.26	0.069	LUC7L3	0.29	0.092			
C7orf50	0.25	0.032	CLP1	-0.21	0.069	RPL22L1	0.31	0.092			
SNRPA1	-0.32	0.032	CFL1	0.23	0.069	RPL34	-0.3	0.095			
			RBM25	0.24	0.069	PCNA	-0.26	0.095			

Chapter 5

5. References

- Acevedo, M.L., and Kraus, W.L. (2003). Mediator and p300/CBP-steroid receptor coactivator complexes have distinct roles, but function synergistically, during estrogen receptor alpha-dependent transcription with chromatin templates. *Molecular and cellular biology* 23, 335-348.
- Achinger-Kawecka, J., Valdes-Mora, F., Luu, P.L., Giles, K.A., Caldon, C.E., Qu, W., Nair, S., Soto, S., Locke, W.J., Yeo-Teh, N.S., *et al.* (2020). Epigenetic reprogramming at estrogen-receptor binding sites alters 3D chromatin landscape in endocrine-resistant breast cancer. *Nature communications* 11, 320.
- Ali, S., and Coombes, R.C. (2002). Endocrine-responsive breast cancer and strategies for combating resistance. *Nature reviews Cancer* 2, 101-112.
- Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. *Nature reviews Molecular cell biology* 16, 155-166.
- Ambrosino, C., Tarallo, R., Bamundo, A., Cuomo, D., Franci, G., Nassa, G., Paris, O., Ravo, M., Giovane, A., Zambrano, N., *et al.* (2010). Identification of a hormone-regulated dynamic nuclear actin network associated with estrogen receptor alpha in human breast cancer cell nuclei. *Molecular & cellular proteomics : MCP* 9, 1352-1367.
- Anzick, S.L., Kononen, J., Walker, R.L., Azorsa, D.O., Tanner, M.M., Guan, X.Y., Sauter, G., Kallioniemi, O.P., Trent, J.M., and Meltzer, P.S. (1997). AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science* 277, 965-968.
- Arnesen, S., Blanchard, Z., Williams, M.M., Berrett, K.C., Li, Z., Oesterreich, S., Richer, J.K., and Gertz, J. (2020). Estrogen receptor alpha mutations in breast cancer cells cause gene expression changes through constant activity and secondary effects. *Cancer research*.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* 37, W202-208.
- Bakhoun, S.F., Silkworth, W.T., Nardi, I.K., Nicholson, J.M., Compton, D.A., and Cimini, D. (2014). The mitotic origin of chromosomal instability. *Current biology : CB* 24, R148-149.
- Balciunas, D., Galman, C., Ronne, H., and Bjorklund, S. (1999). The Med1 subunit of the yeast mediator complex is involved in both transcriptional activation and repression. *Proceedings of the National Academy of Sciences of the United States of America* 96, 376-381.
- Bauer, A., and Kuster, B. (2003). Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes. *European journal of biochemistry / FEBS* 270, 570-578.
- Bauer, P.M., Fulton, D., Boo, Y.C., Sorescu, G.P., Kemp, B.E., Jo, H., and Sessa, W.C. (2003). Compensatory phosphorylation and protein-protein interactions revealed by loss of function and gain of function mutants of multiple serine phosphorylation sites in endothelial nitric-oxide synthase. *The Journal of biological chemistry* 278, 14841-14849.
- Belandia, B., Orford, R.L., Hurst, H.C., and Parker, M.G. (2002). Targeting of SWI/SNF chromatin remodelling complexes to estrogen-responsive genes. *The EMBO journal* 21, 4094-4103.
- Berggard, T., Linse, S., and James, P. (2007). Methods for the detection and analysis of protein-protein interactions. *Proteomics* 7, 2833-2842.
- Bi, M., Zhang, Z., Jiang, Y.Z., Xue, P., Wang, H., Lai, Z., Fu, X., De Angelis, C., Gong, Y., Gao, Z., *et al.* (2020). Enhancer reprogramming driven by high-order assemblies of transcription factors promotes phenotypic plasticity and breast cancer endocrine resistance. *Nature cell biology* 22, 701-715.

Bjornstrom, L., and Sjoberg, M. (2005). Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. *Molecular endocrinology* 19, 833-842.

Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., Foster, L.J., and Mann, M. (2003). A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nature biotechnology* 21, 315-318.

Blows, F.M., Driver, K.E., Schmidt, M.K., Brooks, A., van Leeuwen, F.E., Wesseling, J., Cheang, M.C., Gelmon, K., Nielsen, T.O., Blomqvist, C., *et al.* (2010). Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med* 7, e1000279.

Bludau, I., and Aebersold, R. (2020). Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nature reviews Molecular cell biology* 21, 327-340.

Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., *et al.* (2004). A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nature cell biology* 6, 97-105.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68, 394-424.

Browne, A.L., Charmsaz, S., Vareslija, D., Fagan, A., Cosgrove, N., Cocchiglia, S., Purcell, S., Ward, E., Bane, F., Hudson, L., *et al.* (2018). Network analysis of SRC-1 reveals a novel transcription factor hub which regulates endocrine resistant breast cancer. *Oncogene* 37, 2008-2021.

Bruna, A., Rueda, O.M., Greenwood, W., Batra, A.S., Callari, M., Batra, R.N., Pogrebniak, K., Sandoval, J., Cassidy, J.W., Tufegdizic-Vidakovic, A., *et al.* (2016). A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. *Cell* 167, 260-274 e222.

Burdall, S.E., Hanby, A.M., Lansdown, M.R., and Speirs, V. (2003). Breast cancer cell lines: friend or foe? *Breast cancer research : BCR* 5, 89-95.

Cairns, J., Freire-Pritchett, P., Wingett, S.W., Varnai, C., Dimond, A., Plagnol, V., Zerbino, D., Schoenfelder, S., Javierre, B.M., Osborne, C., *et al.* (2016). CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome biology* 17, 127.

Cairns, J., Orchard, W.R., Malysheva, V., and Spivakov, M. (2019). Chicdiff: a computational pipeline for detecting differential chromosomal interactions in Capture Hi-C data. *Bioinformatics* 35, 4764-4766.

Caliceti, C., Aquila, G., Pannella, M., Morelli, M.B., Fortini, C., Pinton, P., Bonora, M., Hrelia, S., Pannuti, A., Miele, L., *et al.* (2013). 17beta-estradiol enhances signalling mediated by VEGF-A-delta-like ligand 4-notch1 axis in human endothelial cells. *PloS one* 8, e71440.

Camden, A.J., Szwarc, M.M., Chadchan, S.B., DeMayo, F.J., O'Malley, B.W., Lydon, J.P., and Kommagani, R. (2017). Growth regulation by estrogen in breast cancer 1 (GREB1) is a novel progesterone-responsive gene required for human endometrial stromal decidualization. *Mol Hum Reprod* 23, 646-653.

Cancer Genome Atlas, N. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61-70.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., *et al.* (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122, 33-43.

Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoutte, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F., *et al.* (2006). Genome-wide analysis of estrogen receptor binding sites. *Nature genetics* 38, 1289-1297.

Cevher, M.A., Shi, Y., Li, D., Chait, B.T., Malik, S., and Roeder, R.G. (2014). Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nature structural & molecular biology* 21, 1028-1034.

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., *et al.* (2017). The BioGRID interaction database: 2017 update. *Nucleic acids research* 45, D369-D379.

Chen, J.D., and Evans, R.M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* 377, 454.

Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., and Tyler, J.K. (2015). The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Molecular and cellular biology* 36, 662-667.

Choi, C.H., Hiromura, M., and Usheva, A. (2003). Transcription factor IIB acetylates itself to regulate transcription. *Nature* 424, 965-969.

Chumsri, S., Howes, T., Bao, T., Sabnis, G., and Brodie, A. (2011). Aromatase, aromatase inhibitors, and breast cancer. *The Journal of steroid biochemistry and molecular biology* 125, 13-22.

Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., *et al.* (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 163, 506-519.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular cell* 9, 279-289.

Comsa, S., Cimpean, A.M., and Raica, M. (2015). The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research. *Anticancer research* 35, 3147-3154.

Couse, J.F., Lindzey, J., Grandien, K., Gustafsson, J.A., and Korach, K.S. (1997). Tissue distribution and quantitative analysis of estrogen receptor-alpha (ERalpha) and estrogen receptor-beta (ERbeta) messenger ribonucleic acid in the wild-type and ERalpha-knockout mouse. *Endocrinology* 138, 4613-4621.

Cox, J., Hein, M.Y., Lubner, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP* 13, 2513-2526.

Csizmok, V., and Forman-Kay, J.D. (2018). Complex regulatory mechanisms mediated by the interplay of multiple post-translational modifications. *Current opinion in structural biology* 48, 58-67.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., *et al.* (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486, 346-352.

Dai, X., Cheng, H., Bai, Z., and Li, J. (2017). Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J Cancer* 8, 3131-3141.

Danielsen, J.M., Sylvestersen, K.B., Bekker-Jensen, S., Szklarczyk, D., Poulsen, J.W., Horn, H., Jensen, L.J., Mailand, N., and Nielsen, M.L. (2011). Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Molecular & cellular proteomics : MCP* 10, M110 003590.

Das Gupta, K., Shakespear, M.R., Curson, J.E.B., Murthy, A.M.V., Iyer, A., Hodson, M.P., Ramnath, D., Tillu, V.A., von Pein, J.B., Reid, R.C., *et al.* (2020). Class IIa Histone Deacetylases Drive Toll-like Receptor-Inducible Glycolysis and Macrophage Inflammatory Responses via Pyruvate Kinase M2. *Cell reports* 30, 2712-2728 e2718.

Dauvois, S., Danielian, P.S., White, R., and Parker, M.G. (1992). Antiestrogen ICI 164,384 reduces cellular estrogen receptor content by increasing its turnover. *Proceedings of the National Academy of Sciences of the United States of America* 89, 4037-4041.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306-1311.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233-1244.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. *Genes & development* 30, 1357-1382.

Dephoure, N., and Gygi, S.P. (2012). Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. *Science signaling* 5, rs2.

Deroo, B.J., and Korach, K.S. (2006). Estrogen receptors and human disease. *The Journal of clinical investigation* 116, 561-570.

DeRose, Y.S., Wang, G., Lin, Y.C., Bernard, P.S., Buys, S.S., Ebbert, M.T., Factor, R., Matsen, C., Milash, B.A., Nelson, E., *et al.* (2011). Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nature medicine* 17, 1514-1520.

Dhiman, V.K., Bolt, M.J., and White, K.P. (2018). Nuclear receptors in cancer - uncovering new and evolving roles through genomic analysis. *Nature reviews Genetics* 19, 160-174.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21.

Donner, A.J., Ebmeier, C.C., Taatjes, D.J., and Espinosa, J.M. (2010). CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nature structural & molecular biology* 17, 194-201.

Dutertre, M., and Smith, C.L. (2000). Molecular mechanisms of selective estrogen receptor modulator (SERM) action. *The Journal of pharmacology and experimental therapeutics* 295, 431-437.

Eissenberg, J.C., and Elgin, S.C. (2000). The HP1 protein family: getting a grip on chromatin. *Current opinion in genetics & development* 10, 204-210.

El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.R., Park, S., Wang, X., *et al.* (2019). A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell* 178, 1145-1158 e1120.

Engelen, E., Brandsma, J.H., Moen, M.J., Signorile, L., Dekkers, D.H., Demmers, J., Kockx, C.E., Ozgur, Z., van, I.W.F., van den Berg, D.L., *et al.* (2015). Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nature communications* 6, 7155.

Erdos, E., and Balint, B.L. (2020). NR2F2 Orphan Nuclear Receptor is Involved in Estrogen Receptor Alpha-Mediated Transcriptional Regulation in Luminal A Breast Cancer Cells. *Int J Mol Sci* 21.

Everley, P.A., Bakalarski, C.E., Elias, J.E., Waghorne, C.G., Beausoleil, S.A., Gerber, S.A., Faherty, B.K., Zetter, B.R., and Gygi, S.P. (2006). Enhanced analysis of metastatic prostate cancer using stable isotopes and high mass accuracy instrumentation. *Journal of proteome research* 5, 1224-1231.

Evertts, A.G., Zee, B.M., Dimaggio, P.A., Gonzales-Cope, M., Collier, H.A., and Garcia, B.A. (2013). Quantitative dynamics of the link between cellular metabolism and histone acetylation. *The Journal of biological chemistry* 288, 12142-12151.

Fang, F., Xia, N., Angulo, B., Carey, J., Cady, Z., Durruthy-Durruthy, J., Bennett, T., Sebastiano, V., and Reijo Pera, R.A. (2018). A distinct isoform of ZNF207 controls self-renewal and pluripotency of human embryonic stem cells. *Nature communications* 9, 4384.

Feng, Q., Yi, P., Wong, J., and O'Malley, B.W. (2006). Signaling within a coactivator complex: methylation of SRC-3/AIB1 is a molecular switch for complex disassembly. *Molecular and cellular biology* 26, 7846-7857.

Forler, D., Kocher, T., Rode, M., Gentzel, M., Izaurralde, E., and Wilm, M. (2003). An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nature biotechnology* 21, 89-92.

Frasor, J., Danes, J.M., Funk, C.C., and Katzenellenbogen, B.S. (2005). Estrogen down-regulation of the corepressor N-CoR: mechanism and implications for estrogen derepression of N-CoR-regulated genes. *Proceedings of the National Academy of Sciences of the United States of America* 102, 13153-13157.

Frasor, J., Danes, J.M., Komm, B., Chang, K.C., Lyttle, C.R., and Katzenellenbogen, B.S. (2003). Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. *Endocrinology* 144, 4562-4574.

Frasor, J., Stossi, F., Danes, J.M., Komm, B., Lyttle, C.R., and Katzenellenbogen, B.S. (2004). Selective estrogen receptor modulators: discrimination of agonistic versus antagonistic activities by gene expression profiling in breast cancer cells. *Cancer research* 64, 1522-1533.

Fu, X., Pereira, R., De Angelis, C., Veeraraghavan, J., Nanda, S., Qin, L., Cataldo, M.L., Sethunath, V., Mehravaran, S., Gutierrez, C., *et al.* (2019). FOXA1 upregulation promotes enhancer and transcriptional reprogramming in endocrine-resistant breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., *et al.* (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462, 58-64.

Gao, H., Korn, J.M., Ferretti, S., Monahan, J.E., Wang, Y., Singh, M., Zhang, C., Schnell, C., Yang, G., Zhang, Y., *et al.* (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nature medicine* 21, 1318-1325.

Gao, W.W., Xiao, R.Q., Zhang, W.J., Hu, Y.R., Peng, B.L., Li, W.J., He, Y.H., Shen, H.F., Ding, J.C., Huang, Q.X., *et al.* (2018). JMJD6 Licenses ERalpha-Dependent Enhancer and Coding Gene Activation by Modulating the Recruitment of the CARM1/MED12 Co-activator Complex. *Molecular cell* 70, 340-357 e348.

Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R., and Mann, M. (2010). Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature methods* 7, 383-385.

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

Gigantino, V., Salvati, A., Giurato, G., Palumbo, D., Strianese, O., Rizzo, F., Tarallo, R., Nyman, T.A., Weisz, A., and Nassa, G. (2020). Identification of Antiestrogen-Bound Estrogen Receptor alpha Interactomes in Hormone-Responsive Human Breast Cancer Cell Nuclei. *Proteomics* 20, e2000135.

Gingras, A.C., Aebersold, R., and Raught, B. (2005). Advances in protein complex analysis using mass spectrometry. *The Journal of physiology* 563, 11-21.

Gingras, A.C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nature reviews Molecular cell biology* 8, 645-654.

Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic acids research* 47, D559-D563.

Glont, S.E., Chernukhin, I., and Carroll, J.S. (2019a). Comprehensive Genomic Analysis Reveals that the Pioneering Function of FOXA1 Is Independent of Hormonal Signaling. *Cell reports* 26, 2558-2565 e2553.

Glont, S.E., Papachristou, E.K., Sawle, A., Holmes, K.A., Carroll, J.S., and Siersbaek, R. (2019b). Identification of ChIP-seq and RIME grade antibodies for Estrogen Receptor alpha. *PLoS one* 14, e0215340.

Gu, L., Frommel, S.C., Oakes, C.C., Simon, R., Grupp, K., Gerig, C.Y., Bar, D., Robinson, M.D., Baer, C., Weiss, M., *et al.* (2015). BAZ2A (TIP5) is involved in epigenetic alterations in prostate cancer and its overexpression predicts disease recurrence. *Nature genetics* 47, 22-30.

Guan, J., Zhou, W., Hafner, M., Blake, R.A., Chalouni, C., Chen, I.P., De Bruyn, T., Giltneane, J.M., Hartman, S.J., Heidersbach, A., *et al.* (2019). Therapeutic Ligands Antagonize Estrogen Receptor Function by Impairing Its Mobility. *Cell* 178, 949-963 e918.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nature structural & molecular biology* 11, 394-403.

Hall, J.M., and McDonnell, D.P. (2005). Coregulators in nuclear estrogen receptor action: from concept to therapeutic targeting. *Molecular interventions* 5, 343-357.

Hanker, A.B., Sudhan, D.R., and Arteaga, C.L. (2020). Overcoming Endocrine Resistance in Breast Cancer. *Cancer cell* 37, 496-513.

Hasegawa, N., Sumitomo, A., Fujita, A., Aritome, N., Mizuta, S., Matsui, K., Ishino, R., Inoue, K., Urahama, N., Nose, J., *et al.* (2012). Mediator subunits MED1 and MED24 cooperatively contribute to pubertal mammary gland development and growth of breast carcinoma cells. *Molecular and cellular biology* 32, 1483-1495.

Heery, D.M., Kalkhoven, E., Hoare, S., and Parker, M.G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387, 733-736.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., *et al.* (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163, 712-723.

Hill, V.K., Kim, J.S., and Waldman, T. (2016). Cohesin mutations in human cancer. *Biochimica et biophysica acta* 1866, 1-11.

Holen, I., Speirs, V., Morrissey, B., and Blyth, K. (2017). In vivo models in breast cancer research: progress, challenges and future directions. *Dis Model Mech* 10, 359-371.

Hua, H., Zhang, H., Kong, Q., and Jiang, Y. (2018). Mechanisms for estrogen receptor expression in human cancer. *Exp Hematol Oncol* 7, 24.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature genetics* 43, 27-33.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., *et al.* (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162, 425-440.

Ignatiadis, M., and Sotiriou, C. (2013). Luminal breast cancer: from biology to treatment. *Nat Rev Clin Oncol* 10, 494-506.

Ingle, J.N., Suman, V.J., Rowland, K.M., Mirchandani, D., Bernath, A.M., Camoriano, J.K., Fishkin, P.A., Nikcevich, D.A., Perez, E.A., and North Central Cancer Treatment Group Trial, N. (2006). Fulvestrant in women with advanced breast cancer after progression on prior aromatase inhibitor therapy: North Central Cancer Treatment Group Trial N0032. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 24, 1052-1056.

Iwao, K., Miyoshi, Y., Egawa, C., Ikeda, N., and Noguchi, S. (2000). Quantitative analysis of estrogen receptor-beta mRNA and its variants in human breast cancers. *International journal of cancer Journal international du cancer* 88, 733-736.

Jaeger, M.G., Schwalb, B., Mackowiak, S.D., Velychko, T., Hanzl, A., Imrichova, H., Brand, M., Agerer, B., Chorn, S., Nabet, B., *et al.* (2020). Selective Mediator dependence of cell-type-specifying transcription. *Nature genetics*.

Janssen, A., Kops, G.J., and Medema, R.H. (2009). Elevating the frequency of chromosome mis-segregation as a strategy to kill tumor cells. *Proceedings of the National Academy of Sciences of the United States of America* 106, 19108-19113.

Jen, J., Lin, L.L., Lo, F.Y., Chen, H.T., Liao, S.Y., Tang, Y.A., Su, W.C., Salgia, R., Hsu, C.L., Huang, H.C., *et al.* (2017). Oncoprotein ZNF322A transcriptionally deregulates alpha-adducin, cyclin D1 and p53 to promote tumor growth and metastasis in lung cancer. *Oncogene* 36, 5219.

Jen, J., and Wang, Y.C. (2016). Zinc finger proteins in cancer progression. *J Biomed Sci* 23, 53.

Jepsen, K., Hermanson, O., Onami, T.M., Gleiberman, A.S., Lunyak, V., McEvilly, R.J., Kurokawa, R., Kumar, V., Liu, F., Seto, E., *et al.* (2000). Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell* 102, 753-763.

Jeselsohn, R., Bergholz, J.S., Pun, M., Cornwell, M., Liu, W., Nardone, A., Xiao, T., Li, W., Qiu, X., Buchwalter, G., *et al.* (2018). Allele-Specific Chromatin Recruitment and Therapeutic Vulnerabilities of ESR1 Activating Mutations. *Cancer cell* 33, 173-186 e175.

Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M., and Schiff, R. (2015). ESR1 mutations-a mechanism for acquired endocrine resistance in breast cancer. *Nat Rev Clin Oncol* 12, 573-583.

Jiang, H., He, X., Wang, S., Jia, J., Wan, Y., Wang, Y., Zeng, R., Yates, J., 3rd, Zhu, X., and Zheng, Y. (2014). A microtubule-associated zinc finger protein, BuGZ, regulates mitotic chromosome alignment by ensuring Bub3 stability and kinetochore targeting. *Dev Cell* 28, 268-281.

Johnson, K.M., Wang, J., Smallwood, A., Arayata, C., and Carey, M. (2002). TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes & development* 16, 1852-1863.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430-435.

Kang, Y.K., Guermah, M., Yuan, C.X., and Roeder, R.G. (2002). The TRAP/Mediator coactivator complex interacts directly with estrogen receptors alpha and beta through the TRAP220 subunit and directly enhances estrogen receptor function in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 99, 2642-2647.

Katoh, M., and Katoh, M. (2004). Identification and characterization of human FOXK1 gene in silico. *International journal of molecular medicine* 14, 127-132.

Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., Airolidi, E.M., and Burge, C.B. (2015). Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics* 31, 2400-2402.

Kersten, K., de Visser, K.E., van Miltenburg, M.H., and Jonkers, J. (2017). Genetically engineered mouse models in oncology research and cancer medicine. *EMBO molecular medicine* 9, 137-153.

Kettner, N.M., Vijayaraghavan, S., Durak, M.G., Bui, T., Kohansal, M., Ha, M.J., Liu, B., Rao, X., Wang, J., Yi, M., *et al.* (2019). Combined Inhibition of STAT3 and DNA Repair in Palbociclib-Resistant ER-Positive Breast Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 25, 3996-4013.

Key, T.J., Verkasalo, P.K., and Banks, E. (2001). Epidemiology of breast cancer. *The Lancet Oncology* 2, 133-140.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature* 436, 876-880.

Kim, W., Bennett, E.J., Huttlin, E.L., Guo, A., Li, J., Possemato, A., Sowa, M.E., Rad, R., Rush, J., Comb, M.J., *et al.* (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. *Molecular cell* 44, 325-340.

Kong, S.L., Li, G., Loh, S.L., Sung, W.K., and Liu, E.T. (2011). Cellular reprogramming by the conjoint action of ERalpha, FOXA1, and GATA3 to a ligand-inducible growth state. *Molecular systems biology* 7, 526.

Korde, L.A., Zujewski, J.A., Kamin, L., Giordano, S., Domchek, S., Anderson, W.F., Bartlett, J.M., Gelmon, K., Nahleh, Z., Bergh, J., *et al.* (2010). Multidisciplinary meeting on male breast cancer: summary and research recommendations. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28, 2114-2122.

Kouros-Mehr, H., Bechis, S.K., Slorach, E.M., Littlepage, L.E., Egeblad, M., Ewald, A.J., Pai, S.Y., Ho, I.C., and Werb, Z. (2008). GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer cell* 13, 141-152.

Kumar, P., Gullberg, U., Olsson, I., and Ajore, R. (2015). Myeloid translocation gene-16 co-repressor promotes degradation of hypoxia-inducible factor 1. *PloS one* 10, e0123725.

Lacroix, M., and Leclercq, G. (2004). About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. *Molecular and cellular endocrinology* 219, 1-7.

Lai, A.Y., and Wade, P.A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nature reviews Cancer* 11, 588-596.

Lam, S.W., Jimenez, C.R., and Boven, E. (2014). Breast cancer classification by proteomic technologies: current state of knowledge. *Cancer treatment reviews* 40, 129-138.

Lapek, J.D., Jr., Greninger, P., Morris, R., Amzallag, A., Pruteanu-Malinici, I., Benes, C.H., and Haas, W. (2017). Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities. *Nature biotechnology* 35, 983-989.

Lawrence, R.T., Perez, E.M., Hernandez, D., Miller, C.P., Haas, K.M., Irie, H.Y., Lee, S.I., Blau, C.A., and Villen, J. (2015). The Proteomic Landscape of Triple-Negative Breast Cancer. *Cell reports* 11, 990.

Lazennec, G., Bresson, D., Lucas, A., Chauveau, C., and Vignon, F. (2001). ER beta inhibits proliferation and invasion of breast cancer cells. *Endocrinology* 142, 4120-4130.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of clinical investigation* 121, 2750-2767.

Lei, J.T., Shao, J., Zhang, J., Iglesia, M., Chan, D.W., Cao, J., Anurag, M., Singh, P., He, X., Kosaka, Y., *et al.* (2018). Functional Annotation of ESR1 Gene Fusions in Estrogen Receptor-Positive Breast Cancer. *Cell reports* 24, 1434-1444 e1437.

Lemeer, S., Hahne, H., Pachi, F., and Kuster, B. (2012). Software tools for MS-based quantitative proteomics: a brief overview. *Methods in molecular biology* 893, 489-499.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.

Li, J., Van Vranken, J.G., Pontano Vaiteș, L., Schweppe, D.K., Huttlin, E.L., Etienne, C., Nandhikonda, P., Viner, R., Robitaille, A.M., Thompson, A.H., *et al.* (2020a). TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nature methods* *17*, 399-404.

Li, Y., Haarhuis, J.H.I., Sedenò Cacciatore, A., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D., *et al.* (2020b). The structural basis for cohesin-CTCF-anchored loops. *Nature* *578*, 472-476.

Li, Z., Adams, R.M., Chourey, K., Hurst, G.B., Hettich, R.L., and Pan, C. (2012). Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *Journal of proteome research* *11*, 1582-1590.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research* *41*, e108.

Lin, C.Y., Strom, A., Vega, V.B., Kong, S.L., Yeo, A.L., Thomsen, J.S., Chan, W.C., Doray, B., Bangarusamy, D.K., Ramasamy, A., *et al.* (2004). Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. *Genome biology* *5*, R66.

Liu, X.F., and Bagchi, M.K. (2004). Recruitment of distinct chromatin-modifying complexes by tamoxifen-complexed estrogen receptor at natural target gene promoters in vivo. *The Journal of biological chemistry* *279*, 15050-15058.

Liu, Y., Beyer, A., and Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* *165*, 535-550.

Lonard, D.M., and O'Malley B, W. (2007). Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation. *Molecular cell* *27*, 691-700.

Lonning, P.E., and Eikesdal, H.P. (2013). Aromatase inhibition 2013: clinical state of the art and questions that remain to be solved. *Endocrine-related cancer* *20*, R183-201.

Louder, R.K., He, Y., Lopez-Blanco, J.R., Fang, J., Chacon, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* *531*, 604-609.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* *15*, 550.

Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Molecular systems biology* *14*, e8126.

Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. *Nature reviews Molecular cell biology* *13*, 543-547.

Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., *et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* *161*, 1012-1025.

Magnani, L., Ballantyne, E.B., Zhang, X., and Lupien, M. (2011). PBX1 genomic pioneer function drives ERalpha signaling underlying progression in breast cancer. *PLoS genetics* *7*, e1002368.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature reviews Genetics* *11*, 761-772.

Mann, M., Ong, S.E., Gronborg, M., Steen, H., Jensen, O.N., and Pandey, A. (2002). Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends in biotechnology* *20*, 261-268.

Mavaddat, N., Antoniou, A.C., Easton, D.F., and Garcia-Closas, M. (2010). Genetic susceptibility to breast cancer. *Molecular oncology* *4*, 174-191.

May, F.E., and Westley, B.R. (1997). Expression of human intestinal trefoil factor in malignant cells and its regulation by oestrogen in breast cancer cells. *The Journal of pathology* *182*, 404-413.

McAlister, G.C., Huttlin, E.L., Haas, W., Ting, L., Jedrychowski, M.P., Rogers, J.C., Kuhn, K., Pike, I., Grothe, R.A., Blethrow, J.D., *et al.* (2012). Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Analytical chemistry* 84, 7469-7478.

McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wuhr, M., Huttlin, E.L., Erickson, B.K., Rad, R., Haas, W., and Gygi, S.P. (2014). MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Analytical chemistry* 86, 7150-7158.

McCarthy, D.J., Campbell, K.R., Lun, A.T., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33, 1179-1186.

McCord, R.P., Kaplan, N., and Giorgetti, L. (2020). Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Molecular cell* 77, 688-708.

McFarland, M.A., Ellis, C.E., Markey, S.P., and Nussbaum, R.L. (2008). Proteomics analysis identifies phosphorylation-dependent alpha-synuclein protein interactions. *Molecular & cellular proteomics : MCP* 7, 2123-2137.

Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., *et al.* (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 534, 55-62.

Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* 115, 751-763.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., *et al.* (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature genetics* 49, 1779-1784.

Michaloglou, C., Crafter, C., Siersbaek, R., Delpuech, O., Curwen, J.O., Carnevalli, L.S., Staniszevska, A.D., Polanska, U.M., Cheraghchi-Bashi, A., Lawson, M., *et al.* (2018). Combined Inhibition of mTOR and CDK4/6 Is Required for Optimal Blockade of E2F Function and Long-term Growth Inhibition in Estrogen Receptor-positive Breast Cancer. *Molecular cancer therapeutics* 17, 908-920.

Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., *et al.* (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics* 47, 598-606.

Mohammed, H., and Carroll, J.S. (2013). Approaches for assessing and discovering protein interactions in cancer. *Molecular cancer research : MCR* 11, 1295-1302.

Mohammed, H., D'Santos, C., Serandour, A.A., Ali, H.R., Brown, G.D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L., *et al.* (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell reports* 3, 342-349.

Mohammed, H., Russell, I.A., Stark, R., Rueda, O.M., Hickey, T.E., Tarulli, G.A., Serandour, A.A., Birrell, S.N., Bruna, A., Saadi, A., *et al.* (2015). Progesterone receptor modulates ERalpha action in breast cancer. *Nature* 523, 313-317.

Mohammed, H., Taylor, C., Brown, G.D., Papachristou, E.K., Carroll, J.S., and D'Santos, C.S. (2016). Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nature protocols* 11, 316-326.

Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., and Washburn, M.P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Molecular cell* 34, 168-178.

Myers, S.A., Klaeger, S., Satpathy, S., Viner, R., Choi, J., Rogers, J., Clauser, K., Udeshi, N.D., and Carr, S.A. (2019). Evaluation of Advanced Precursor Determination for Tandem Mass Tag (TMT)-Based Quantitative Proteomics across Instrument Platforms. *Journal of proteome research* 18, 542-547.

Nagalingam, A., Tighiouart, M., Ryden, L., Joseph, L., Landberg, G., Saxena, N.K., and Sharma, D. (2012). Med1 plays a critical role in the development of tamoxifen resistance. *Carcinogenesis* 33, 918-930.

Nagarajan, S., Rao, S.V., Sutton, J., Cheeseman, D., Dunn, S., Papachristou, E.K., Prada, J.G., Couturier, D.L., Kumar, S., Kishore, K., *et al.* (2020). ARID1A influences HDAC1/BRD4 activity, intrinsic proliferative capacity and breast cancer treatment response. *Nature genetics* 52, 187-197.

Nagpal, N., Sharma, S., Maji, S., Durante, G., Ferracin, M., Thakur, J.K., and Kulshreshtha, R. (2018). Essential role of MED1 in the transcriptional regulation of ER-dependent oncogenic miRNAs in breast cancer. *Scientific reports* 8, 11805.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930-944 e922.

Nowak, D.E., Tian, B., and Brasier, A.R. (2005). Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *BioTechniques* 39, 715-725.

Nusinow, D.P., Szpyt, J., Ghandi, M., Rose, C.M., McDonald, E.R., 3rd, Kalocsay, M., Jane-Valbuena, J., Gelfand, E., Schweppe, D.K., Jedrychowski, M., *et al.* (2020). Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* 180, 387-402 e316.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* 1, 376-386.

Orlando, V. (2000). Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends in biochemical sciences* 25, 99-104.

Orphanides, G., and Reinberg, D. (2000). RNA polymerase II elongation through chromatin. *Nature* 407, 471-475.

Osborne, C.K., Bardou, V., Hopp, T.A., Chamness, G.C., Hilsenbeck, S.G., Fuqua, S.A., Wong, J., Allred, D.C., Clark, G.M., and Schiff, R. (2003). Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *Journal of the National Cancer Institute* 95, 353-361.

Papachristou, E.K., Kishore, K., Holding, A.N., Harvey, K., Roumeliotis, T.I., Chilamakuri, C.S.R., Omarjee, S., Chia, K.M., Swarbrick, A., Lim, E., *et al.* (2018). A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. *Nature communications* 9, 2311.

Parker, C.E., Mocanu, V., Mocanu, M., Dicheva, N., and Warren, M.R. (2010). Mass Spectrometry for Post-Translational Modifications. In *Neuroproteomics*, O. Alzate, ed. (Boca Raton (FL)).

Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., *et al.* (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 27, 1160-1167.

Parrish, J.R., Gulyas, K.D., and Finley, R.L., Jr. (2006). Yeast two-hybrid contributions to interactome mapping. *Current opinion in biotechnology* 17, 387-393.

Paulo, J.A., McAllister, F.E., Everley, R.A., Beausoleil, S.A., Banks, A.S., and Gygi, S.P. (2015). Effects of MEK inhibitors GSK1120212 and PD0325901 in vivo using 10-plex quantitative proteomics and phosphoproteomics. *Proteomics* 15, 462-473.

Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J., *et al.* (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nature communications* 7, 11479.

Petrenko, N., Jin, Y., Wong, K.H., and Struhl, K. (2016). Mediator Undergoes a Compositional Change during Transcriptional Activation. *Molecular cell* 64, 443-454.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes & development* 20, 2922-2936.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. *Nature reviews Molecular cell biology* 16, 245-257.

Poulard, C., Treilleux, I., Lavergne, E., Bouchekioua-Bouzaghrou, K., Goddard-Leon, S., Chabaud, S., Tredan, O., Corbo, L., and Le Romancer, M. (2012). Activation of rapid oestrogen signalling in aggressive human breast cancers. *EMBO molecular medicine* 4, 1200-1213.

Purdie, C.A., Quinlan, P., Jordan, L.B., Ashfield, A., Ogston, S., Dewar, J.A., and Thompson, A.M. (2014). Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. *British journal of cancer* 110, 565-572.

Raab, J.R., Smith, K.N., Spear, C.C., Manner, C.J., Calabrese, J.M., and Magnuson, T. (2019). SWI/SNF remains localized to chromatin in the presence of SCHLAP1. *Nature genetics* 51, 26-29.

Rafiee, M.R., Girardot, C., Sigismondo, G., and Krijgsveld, J. (2016). Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins. *Molecular cell* 64, 624-635.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* 44, W160-165.

Rangel, L.B., and Huang, T.H. (2013). Estrogen response in luminal breast cancer. *Oncotarget* 4, 1548-1549.

Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., *et al.* (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305-320 e324.

Rauniyar, N., and Yates, J.R., 3rd (2014). Isobaric labeling-based relative quantification in shotgun proteomics. *Journal of proteome research* 13, 5293-5309.

Reynisdottir, I., Arason, A., Einarsson, B.O., Gunnarsson, H., Staaf, J., Vallon-Christersson, J., Jonsson, G., Ringner, M., Agnarsson, B.A., Olafsdottir, K., *et al.* (2013). High expression of ZNF703 independent of amplification indicates worse prognosis in patients with luminal B breast cancer. *Cancer Med* 2, 437-446.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nature biotechnology* 17, 1030-1032.

Robinson, P.J., Trnka, M.J., Pellarin, R., Greenberg, C.H., Bushnell, D.A., Davis, R., Burlingame, A.L., Sali, A., and Kornberg, R.D. (2015). Molecular architecture of the yeast Mediator complex. *eLife* 4.

Rolland, T., Tasan, M., Charlotiaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., *et al.* (2014). A proteome-scale map of the human interactome network. *Cell* 159, 1212-1226.

Rose, C.M., Isasa, M., Ordureau, A., Prado, M.A., Beausoleil, S.A., Jedrychowski, M.P., Finley, D.J., Harper, J.W., and Gygi, S.P. (2016). Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes. *Cell Syst* 3, 395-403 e394.

Rosell, M., Nevedomskaya, E., Stelloo, S., Nautiyal, J., Poliandri, A., Steel, J.H., Wessels, L.F., Carroll, J.S., Parker, M.G., and Zwart, W. (2014). Complex formation and function of estrogen receptor alpha in transcription requires RIP140. *Cancer research* 74, 5469-5479.

Ross-Innes, C.S., Brown, G.D., and Carroll, J.S. (2011). A co-ordinated interaction between CTCF and ER in breast cancer cells. *BMC genomics* 12, 593.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., *et al.* (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389-393.

Roumeliotis, T.I., Williams, S.P., Goncalves, E., Alsinet, C., Del Castillo Velasco-Herrera, M., Aben, N., Ghavidel, F.Z., Michaut, M., Schubert, M., Price, S., *et al.* (2017). Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. *Cell reports* 20, 2201-2214.

Roux, K.J., Kim, D.I., Burke, B., and May, D.G. (2018). BioID: A Screen for Protein-Protein Interactions. *Current protocols in protein science / editorial board, John E Coligan [et al]* 91, 19 23 11-19 23 15.

Rueda, O.M., Sammut, S.J., Seoane, J.A., Chin, S.F., Caswell-Jin, J.L., Callari, M., Batra, R., Pereira, B., Bruna, A., Ali, H.R., *et al.* (2019). Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature* 567, 399-404.

Sauve, F., McBroom, L.D., Gallant, J., Moraitis, A.N., Labrie, F., and Giguere, V. (2001). CIA, a novel estrogen receptor coactivator with a bifunctional nuclear receptor interacting determinant. *Molecular and cellular biology* 21, 343-353.

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., *et al.* (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome research* 25, 582-597.

Schoenfelder, S., Javierre, B.M., Furlan-Magaril, M., Wingett, S.W., and Fraser, P. (2018). Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. *J Vis Exp*.

Serandour, A.A., Mohammed, H., Miremedi, A., Mulder, K.W., and Carroll, J.S. (2018). TRPS1 regulates oestrogen receptor binding and histone acetylation at enhancers. *Oncogene* 37, 5281-5291.

Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A., and Brown, M. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* 103, 843-852.

Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., and Greene, G.L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95, 927-937.

Shiino, S., Kinoshita, T., Yoshida, M., Jimbo, K., Asaga, S., Takayama, S., and Tsuda, H. (2016). Prognostic Impact of Discordance in Hormone Receptor Status Between Primary and Recurrent Sites in Patients With Recurrent Breast Cancer. *Clin Breast Cancer* 16, e133-140.

Shu, S., Lin, C.Y., He, H.H., Witwicki, R.M., Tabassum, D.P., Roberts, J.M., Janiszewska, M., Huh, S.J., Liang, Y., Ryan, J., *et al.* (2016). Response and resistance to BET bromodomain inhibitors in triple-negative breast cancer. *Nature* 529, 413-417.

Siersbaek, R., Madsen, J.G.S., Javierre, B.M., Nielsen, R., Bagge, E.K., Cairns, J., Wingett, S.W., Traynor, S., Spivakov, M., Fraser, P., *et al.* (2017). Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. *Molecular cell* 66, 420-435 e425.

Siersbaek, R., Scabia, V., Nagarajan, S., Chernukhin, I., Papachristou, E.K., Broome, R., Johnston, S.J., Joosten, S.E.P., Green, A.R., Kumar, S., *et al.* (2020). IL6/STAT3 Signaling Hijacks Estrogen Receptor alpha Enhancers to Drive Breast Cancer Metastasis. *Cancer cell* 38, 412-423 e419.

Soderberg, O., Gullberg, M., Jarvius, M., Ridderstrale, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G., *et al.* (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nature methods* 3, 995-1000.

Sogaard, T.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *The Journal of biological chemistry* 282, 14113-14120.

Sommer, S., and Fuqua, S.A. (2001). Estrogen receptor and breast cancer. *Semin Cancer Biol* 11, 339-352.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* 98, 10869-10874.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., *et al.* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences of the United States of America* 100, 8418-8423.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nature reviews Molecular cell biology* 19, 262-274.

Spange, S., Wagner, T., Heinzl, T., and Kramer, O.H. (2009). Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *The international journal of biochemistry & cell biology* 41, 185-198.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., *et al.* (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122, 957-968.

Stender, J.D., Nwachukwu, J.C., Kastrati, I., Kim, Y., Strid, T., Yakir, M., Srinivasan, S., Nowak, J., Izard, T., Rangarajan, E.S., *et al.* (2017). Structural and Molecular Mechanisms of Cytokine-Mediated Endocrine Resistance in Human Breast Cancer Cells. *Molecular cell* 65, 1122-1135 e1125.

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

Sur, I.K., Hallikas, O., Vaharautio, A., Yan, J., Turunen, M., Enge, M., Taipale, M., Karhu, A., Aaltonen, L.A., and Taipale, J. (2012). Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* 338, 1360-1363.

Svinkina, T., Gu, H., Silva, J.C., Mertins, P., Qiao, J., Fereshetian, S., Jaffe, J.D., Kuhn, E., Udeshi, N.D., and Carr, S.A. (2015). Deep, Quantitative Coverage of the Lysine Acetylome Using Novel Anti-acetyl-lysine Antibodies and an Optimized Proteomic Workflow. *Molecular & cellular proteomics : MCP* 14, 2429-2440.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., *et al.* (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* 45, D362-D368.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., *et al.* (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* 146, 92-104.

Takaku, M., Grimm, S.A., De Kumar, B., Bennett, B.D., and Wade, P.A. (2020). Cancer-specific mutation of GATA3 disrupts the transcriptional regulatory network governed by Estrogen Receptor alpha, FOXA1 and GATA3. *Nucleic acids research*.

Tan, S.K., Lin, Z.H., Chang, C.W., Varang, V., Chng, K.R., Pan, Y.F., Yong, E.L., Sung, W.K., and Cheung, E. (2011). AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. *The EMBO journal* *30*, 2569-2581.

Tang, B., Tian, Y., Liao, Y., Li, Z., Yu, S., Su, H., Zhong, F., Yuan, G., Wang, Y., Yu, H., *et al.* (2019). CBX8 exhibits oncogenic properties and serves as a prognostic factor in hepatocellular carcinoma. *Cell death & disease* *10*, 52.

Theodorou, V., Stark, R., Menon, S., and Carroll, J.S. (2013). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome research* *23*, 12-22.

Thingholm, T.E., Jensen, O.N., and Larsen, M.R. (2009). Analytical strategies for phosphoproteomics. *Proteomics* *9*, 1451-1468.

Thomas, C., and Gustafsson, J.A. (2015). Estrogen receptor mutations and functional consequences for breast cancer. *Trends Endocrinol Metab* *26*, 467-476.

Thompson, A., Wolmer, N., Koncarevic, S., Selzer, S., Bohm, G., Legner, H., Schmid, P., Kienle, S., Penning, P., Hohle, C., *et al.* (2019). TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Analytical chemistry* *91*, 15941-15950.

Thompson, S.L., and Compton, D.A. (2008). Examining the link between chromosomal instability and aneuploidy in human cells. *The Journal of cell biology* *180*, 665-672.

Tibes, R., Qiu, Y., Lu, Y., Hennessy, B., Andreeff, M., Mills, G.B., and Kornblau, S.M. (2006). Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Molecular cancer therapeutics* *5*, 2512-2521.

Ting, L., Rad, R., Gygi, S.P., and Haas, W. (2011). MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature methods* *8*, 937-940.

Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes & development* *16*, 673-675.

Torchy, M.P., Hamiche, A., and Klaholz, B.P. (2015). Structure and function insights into the NuRD chromatin remodeling complex. *Cellular and molecular life sciences : CMLS* *72*, 2491-2507.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* *25*, 1105-1111.

Tsai, W.W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.Y., Shi, X., Schwarzer, D., Plunkett, W., *et al.* (2010). TRIM24 links a non-canonical histone signature to breast cancer. *Nature* *468*, 927-932.

Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghoussaini, M., Hines, S., Healey, C.S., *et al.* (2010). Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature genetics* *42*, 504-507.

Tyanova, S., Albrechtsen, R., Kronqvist, P., Cox, J., Mann, M., and Geiger, T. (2016a). Proteomic maps of breast cancer subtypes. *Nature communications* *7*, 10259.

Tyanova, S., Temu, T., and Cox, J. (2016b). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols* *11*, 2301-2319.

Underhill, C., Qutob, M.S., Yee, S.P., and Torchia, J. (2000). A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. *The Journal of biological chemistry* *275*, 40463-40470.

Vandemoortele, G., De Sutter, D., Moliere, A., Pauwels, J., Gevaert, K., and Eyckerman, S. (2019). A Well-Controlled BioID Design for Endogenous Bait Proteins. *Journal of proteome research* *18*, 95-106.

Vendrell, J.A., Thollet, A., Nguyen, N.T., Ghayad, S.E., Vinot, S., Bieche, I., Grisard, E., Josserand, V., Coll, J.L., Roux, P., *et al.* (2012). ZNF217 is a marker of poor prognosis in breast

cancer that drives epithelial-mesenchymal transition and invasion. *Cancer research* 72, 3593-3606.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., *et al.* (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & development* 12, 343-356.

Walsh, C.A., Qin, L., Tien, J.C., Young, L.S., and Xu, J. (2012). The function of steroid receptor coactivator-1 in normal tissues and cancer. *Int J Biol Sci* 8, 470-485.

Wang, Z., Yu, K., Tan, H., Wu, Z., Cho, J.H., Han, X., Sun, H., Beach, T.G., and Peng, J. (2020). 27-Plex Tandem Mass Tag Mass Spectrometry for Profiling Brain Proteome in Alzheimer's Disease. *Analytical chemistry*.

Wardell, S.E., Marks, J.R., and McDonnell, D.P. (2011). The turnover of estrogen receptor alpha by the selective estrogen receptor degrader (SERD) fulvestrant is a saturable process that is not required for antagonist efficacy. *Biochemical pharmacology* 82, 122-130.

Welboren, W.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. (2009). Genomic actions of estrogen receptor alpha: what are the targets and how are they regulated? *Endocrine-related cancer* 16, 1073-1089.

Wen, Z., Huang, Z.T., Zhang, R., and Peng, C. (2018). ZNF143 is a regulator of chromatin loop. *Cell biology and toxicology* 34, 471-478.

Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., *et al.* (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular biology of the cell* 13, 1977-2000.

Wierer, M., and Mann, M. (2016). Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. *Human molecular genetics* 25, R106-R114.

Wijayarathne, A.L., and McDonnell, D.P. (2001). The human estrogen receptor-alpha is a ubiquitinated protein whose stability is affected differentially by agonists, antagonists, and selective estrogen receptor modulators. *The Journal of biological chemistry* 276, 35684-35692.

Wiley, C.D., Liu, S., Limbad, C., Zawadzka, A.M., Beck, J., Demaria, M., Artwood, R., Alimirah, F., Lopez-Dominguez, J.A., Kuehnemann, C., *et al.* (2019). SILAC Analysis Reveals Increased Secretion of Hemostasis-Related Factors by Senescent Cells. *Cell reports* 28, 3329-3337 e3325.

Williams, C., Edvardsson, K., Lewandowski, S.A., Strom, A., and Gustafsson, J.A. (2008). A genome-wide study of the repressive effects of estrogen receptor beta on estrogen receptor alpha signaling in breast cancer cells. *Oncogene* 27, 1019-1032.

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* 4, 1310.

Wu, L., Wang, Y., Liu, Y., Yu, S., Xie, H., Shi, X., Qin, S., Ma, F., Tan, T.Z., Thiery, J.P., *et al.* (2014). A central role for TRPS1 in the control of cell cycle and cancer development. *Oncotarget* 5, 7677-7690.

Wu, S.Y., Thomas, M.C., Hou, S.Y., Likhite, V., and Chiang, C.M. (1999). Isolation of mouse TFIID and functional characterization of TBP and TFIID in mediating estrogen receptor and chromatin transcription. *The Journal of biological chemistry* 274, 23480-23490.

Wu, Y., Peng, Y., Wu, M., Zhang, W., Zhang, M., Xie, R., Zhang, P., Bai, Y., Zhao, J., Li, A., *et al.* (2016). Oncogene FOXK1 enhances invasion of colorectal carcinoma by inducing epithelial-mesenchymal transition. *Oncotarget* 7, 51150-51162.

Xu, G., Chhangawala, S., Cocco, E., Razavi, P., Cai, Y., Otto, J.E., Ferrando, L., Selenica, P., Ladewig, E., Chan, C., *et al.* (2020). ARID1A determines luminal identity and therapeutic response in estrogen-receptor-positive breast cancer. *Nature genetics* 52, 198-207.

Xue, J.H., Zheng, M., Xu, X.W., Wu, S.S., Chen, Z., and Chen, F. (2011). Involvement of REST corepressor 3 in prognosis of human hepatitis B. *Acta Pharmacol Sin* 32, 1019-1024.

Yang, L., Hamilton, S.R., Sood, A., Kuwai, T., Ellis, L., Sanguino, A., Lopez-Berestein, G., and Boyd, D.D. (2008). The previously undescribed ZKSCAN3 (ZNF306) is a novel "driver" of colorectal cancer progression. *Cancer research* 68, 4321-4330.

Yates, L.R., Knappskog, S., Wedge, D., Farmery, J.H.R., Gonzalez, S., Martincorena, I., Alexandrov, L.B., Van Loo, P., Haugland, H.K., Lilleng, P.K., *et al.* (2017). Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer cell* 32, 169-184 e167.

Ye, B., Yang, G., Li, Y., Zhang, C., Wang, Q., and Yu, G. (2020). ZNF143 in Chromatin Looping and Gene Regulation. *Front Genet* 11, 338.

Yin, J.W., and Wang, G. (2014). The Mediator complex: a master coordinator of transcription and cell lineage development. *Development* 141, 977-987.

Yoder, B.J., Wilkinson, E.J., and Massoll, N.A. (2007). Molecular and morphologic distinctions between infiltrating ductal and lobular carcinoma of the breast. *Breast J* 13, 172-179.

Yu, C., and Huang, L. (2018). Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Analytical chemistry* 90, 144-165.

Yu, Q., Paulo, J.A., Naverrete-Perea, J., McAlister, G.C., Canterbury, J.D., Bailey, D.J., Robitaille, A.M., Huguet, R., Zabrouskov, V., Gygi, S.P., *et al.* (2020). Benchmarking the Orbitrap Tribrid Eclipse for Next Generation Multiplexed Proteomics. *Analytical chemistry* 92, 6478-6485.

Yu, S., Kim, T., Yoo, K.H., and Kang, K. (2017). The T47D cell line is an ideal experimental model to elucidate the progesterone-specific effects of a luminal A subtype of breast cancer. *Biochemical and biophysical research communications* 486, 752-758.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes & development* 25, 2227-2241.

Zhang, B., Chambers, K.J., Faller, D.V., and Wang, S. (2007). Reprogramming of the SWI/SNF complex for co-activation or co-repression in prohibitin-mediated estrogen receptor regulation. *Oncogene* 26, 7153-7157.

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., *et al.* (2014). Proteogenomic characterization of human colon and rectal cancer. *Nature* 513, 382-387.

Zhang, L., Cui, J., Leonard, M., Nephew, K., Li, Y., and Zhang, X. (2013). Silencing MED1 sensitizes breast cancer cells to pure anti-estrogen fulvestrant in vitro and in vivo. *PloS one* 8, e70641.

Zhang, X., Krutchinsky, A., Fukuda, A., Chen, W., Yamamura, S., Chait, B.T., and Roeder, R.G. (2005). MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. *Molecular cell* 19, 89-100.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome biology* 9, R137.

Zheng, J., Chen, X., Yang, Y., Tan, C.S.H., and Tian, R. (2020). Mass Spectrometry-Based Protein Complex Profiling in Time and Space. *Analytical chemistry*.

Zheng, W., Long, J., Gao, Y.T., Li, C., Zheng, Y., Xiang, Y.B., Wen, W., Levy, S., Deming, S.L., Haines, J.L., *et al.* (2009). Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nature genetics* 41, 324-328.

Zhou, L., Ng, H.K., Drautz-Moses, D.I., Schuster, S.C., Beck, S., Kim, C., Chambers, J.C., and Loh, M. (2019). Systematic evaluation of library preparation methods and sequencing platforms for high-throughput whole genome bisulfite sequencing. *Scientific reports* 9, 10383.

Zhu, W., Smith, J.W., and Huang, C.M. (2010). Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* 2010, 840518.

Zwart, W., Theodorou, V., Kok, M., Canisius, S., Linn, S., and Carroll, J.S. (2011). Oestrogen receptor-co-factor-chromatin specificity in the transcriptional regulation of breast cancer. *The EMBO journal* 30, 4764-4776.

Papers published during PhD

1. **Papachristou, E.K.**, Kishore, K., Holding, A.N., Harvey, K., Roumeliotis, T.I., Chilamakuri, C.S.R., Omarjee, S., Chia, K.M., Swarbrick, A., Lim, E., et al. (2018). A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. *Nature communications* 9, 2311.
2. Glont, S.E., **Papachristou, E.K.**, Sawle, A., Holmes, K.A., Carroll, J.S., and Siersbaek, R. (2019b). Identification of ChIP-seq and RIME grade antibodies for Estrogen Receptor alpha. *PloS one* 14, e0215340.

Data availability



1. For the publication *Papachristou et al.*, the RNA-seq data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE104872. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD007968.
2. For the publication *Glont et al.*, the ChIP-seq and RIME data are available from GEO (Accession number GSE128208) or PRIDE (Accession number PXD012930).

ARTICLE

DOI: 10.1038/s41467-018-04619-5

OPEN

A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes

Evangelia K. Papachristou¹ , Kamal Kishore¹, Andrew N. Holding¹, Kate Harvey², Theodoros I. Roumeliotis³, Chandra Sekhar Reddy Chilamakuri¹, Soleilmane Omarjee¹, Kee Ming Chia², Alex Swarbrick^{2,4}, Elgene Lim^{2,4}, Florian Markowetz¹ , Matthew Eldridge¹, Rasmus Siersbaek¹, Clive S. D'Santos¹ & Jason S. Carroll¹

Understanding the dynamics of endogenous protein-protein interactions in complex networks is pivotal in deciphering disease mechanisms. To enable the in-depth analysis of protein interactions in chromatin-associated protein complexes, we have previously developed a method termed RIME (Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins). Here, we present a quantitative multiplexed method (qPLEX-RIME), which integrates RIME with isobaric labelling and tribrid mass spectrometry for the study of protein interactome dynamics in a quantitative fashion with increased sensitivity. Using the qPLEX-RIME method, we delineate the temporal changes of the Estrogen Receptor alpha (ER α) interactome in breast cancer cells treated with 4-hydroxytamoxifen. Furthermore, we identify endogenous ER α -associated proteins in human Patient-Derived Xenograft tumours and in primary human breast cancer clinical tissue. Our results demonstrate that the combination of RIME with isobaric labelling offers a powerful tool for the in-depth and quantitative characterisation of protein interactome dynamics, which is applicable to clinical samples.

¹Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge CB2 0RE, UK. ²Garvan Institute of Medical Research, Darlinghurst, Sydney NSW 2010, Australia. ³Wellcome Trust Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK. ⁴St Vincent's Clinical School, UNSW, Sydney NSW 2052, Australia. These authors contributed equally: Kamal Kishore, Andrew N. Holding. Correspondence and requests for materials should be addressed to R.S. (email: Rasmus.Siersbaek@cruk.cam.ac.uk) or to C.S.D'S. (email: Clive.D'Santos@cruk.cam.ac.uk) or to J.S.C. (email: Jason.Carroll@cruk.cam.ac.uk)

Deciphering the role and the organisation of dynamically regulated protein networks is critical for the accurate molecular characterisation of biological systems¹. Over the last decade, the advancements made in mass spectrometry-based proteomics have enabled the rapid analysis of complex protein samples obtained from co-immunoprecipitation assays, providing a powerful tool for the study of protein interactions and protein complexes². In this regard, the first systematic efforts to generate human protein interactome maps using yeast two-hybrid^{3–5} have been recently complemented by studies utilising large-scale Affinity Purification followed by Mass Spectrometry analysis (AP-MS)^{6,7}. Additionally, the integration of AP-MS with quantitative approaches has enabled the study of stoichiometric changes in protein complexes⁸. More recently, the use of chemical crosslinking combined with mass spectrometry has provided information about endogenous protein assemblies in a proteome-wide scale⁹.

Gene regulation relies on the coordinated action of transcription factors and co-regulator complexes that control transcriptional activation at promoters or enhancers. To gain insight into the complex interactions between such regulators, the combination of Chromatin Immunoprecipitation (ChIP) with mass spectrometry has been used to study the composition of chromatin-associated complexes^{10–12}. In line with this strategy we have previously developed RIME (Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins)¹³, a method which has several advantages for the analysis of protein interactomes¹⁴. RIME provides a sensitive and rapid approach for the identification of protein complexes from low amounts of starting material and importantly involves purification of endogenous protein, rather than the use of exogenous tagged approaches.

In the present study, we have established a modified RIME assay to monitor the dynamics of chromatin-associated complexes using a quantitative multiplexed workflow (quantitative Multiplexed Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins or qPLEX-RIME). Specifically, we combine RIME with isobaric labelling using Tandem Mass Tags (TMT-10plex)^{15,16}, peptide fractionation and MultiNotch MS3 analysis¹⁷. This combination allows the simultaneous analysis of multiple conditions and biological replicates with high sensitivity in a single experiment. Additionally, we have developed a data analysis workflow termed quantitative Multiplexed analyzer (qPLEXanalyzer) that permits statistical analysis of the quantitative interactome data and the identification of differential interactions.

As a proof-of-concept, we apply the qPLEX-RIME method to discover the temporal changes of Estrogen Receptor alpha (ERα) interactors in breast cancer cells treated with 4-hydroxytamoxifen (OHT) and to identify the ERα interactome in human patient-derived xenograft (PDX) tumours and in human breast cancer tissues. Our data demonstrate that the qPLEX-RIME method combines multiplexity with quantitative accuracy and increased sensitivity, to enable the in-depth characterisation of dynamic changes in chromatin-associated protein complexes in vitro and in vivo.

Results

The qPLEX-RIME workflow. The qPLEX-RIME approach combines the RIME method^{13,14} with multiplex TMT chemical isobaric labelling^{15,16} to study the dynamics of chromatin-associated protein complexes. The workflow starts with a two-step fixation procedure using disuccinimidyl glutarate (DSG) and formaldehyde (FA) that has been previously applied in combination with ChIP assays to capture transient interactions more efficiently^{12,18}. A specific antibody against the target protein is

used for immunoprecipitation, followed by proteolysis, TMT-10plex peptide labelling and fractionation. The main steps of the qPLEX-RIME method are shown in Fig. 1. The main utility of the qPLEX-RIME method is the quantification of changes in the composition of protein complexes in response to cell perturbation and/or in variable genomic backgrounds (e.g. different cell lines or mutated conditions) using multiple biological replicates in a single experiment. Also proteins that are significantly and specifically associated with the bait protein can be discovered in the same analysis using appropriate negative controls, such as IgG pull-downs. For the downstream data analysis, we have developed a comprehensive bioinformatics workflow (qPLEXanalyzer) that includes data processing, visualisation, normalisation and differential statistics. In addition to the qPLEXanalyzer R package, the complete qPLEX-RIME and full proteome data sets of this work are included in the qPLEXdata R package. Both packages can be found at (<https://doi.org/10.5281/zenodo.1237825>) and a detailed description of the pipeline and the applications is provided in Supplementary Notes 1 and 2.

Characterisation of the ERα interactome in MCF7 cells. We first applied qPLEX-RIME to assess whether we could successfully identify the ERα interactome in asynchronous MCF7 breast cancer cells. To this end, we performed ERα qPLEX-RIME pull-downs in five independent biological replicates. An equal number of matched IgG control samples were prepared. In this experiment we used single crosslinking with FA, to permit a comparison with previously published approaches¹³. In addition to the qPLEX-RIME, we included a standard non-quantitative ERα RIME experiment with matched IgG controls (Supplementary Data 1).

The qPLEX-RIME raw data processing quantified 2955 proteins across the multiplexed set of all positive and negative samples at peptide false discovery rate (FDR) <1% (Supplementary Data 2). To test the efficiency of the method in capturing and quantifying previously described ERα-associated proteins, we compiled a list of known ERα interactors from BioGRID¹⁹ and STRING²⁰ resources. For BioGRID, we used only a subset of 386 proteins identified by high-throughput assays that are similar to the approach used here and for STRING we used only experimental associations (383 proteins, score > 200). Noteworthy, only 37 proteins were common between the two reference subsets. The qPLEX-RIME method identified 295 (76%) and 171 (45%) of the known ERα-associated proteins from BioGRID and STRING, respectively, of which 225 (58%) and 154 (40%) showed positive enrichment at adj. *p*-value < 0.1 (Limma moderated *t*-test) (Fig. 2a). Specifically, we found known co-regulators (e.g. EP300, NCOA3, CBP, NR1P1, TRIM24, GREB1, RARα, NCOR2 and HDACs^{13,21–25}), ERα-associated pioneer factors (e.g. FOXA1²⁶ and AP-2γ²⁷), and putative pioneer factors (e.g. GATA-3²⁸) with significant enrichment in the ERα samples (Fig. 2b).

ERα was one of the most significantly enriched proteins identified with 19 unique peptides (Fig. 2c), which is consistent with previously published ERα RIME experiments¹⁴. A comparison between the non-quantitative RIME and the qPLEX-RIME data showed that 302 of the 323 (93%) proteins identified as ERα-specific in the non-quantitative ERα-RIME pull-down analysis were also identified by qPLEX-RIME with significant enrichment over the IgG controls (mean fold-change of 2.5). Notably, the application of qPLEX-RIME achieved overall better peptide coverage for the overlapping ERα-associated proteins compared to the non-quantitative RIME method (Fig. 2d). Additionally, qPLEX-RIME identified 124 more known BioGRID and STRING interactors compared to the non-quantitative RIME

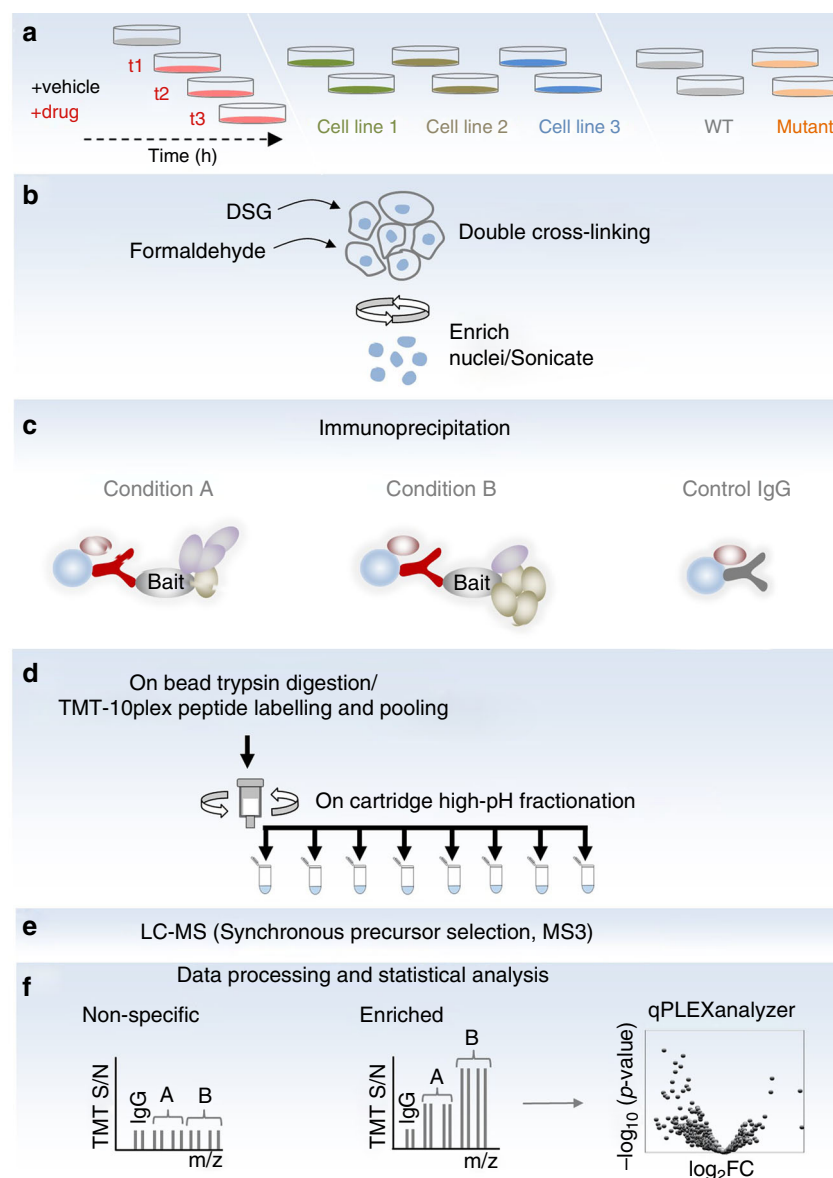


Fig. 1 The qPLEX-RIME workflow. Proteins in treated and un-treated cell cultures at different time points or in variable genomic backgrounds (e.g. different cell lines or mutated conditions) are double-crosslinked and cell nuclei are isolated and sonicated (**a**, **b**). Target protein complexes are immunoprecipitated and subjected to on-bead trypsin digestion (**c**, **d**). The generated peptides are labelled using different TMT reagents and pooled in a single mixture, which is fractionated using Reversed-Phase cartridges (**d**). Peptide fractions are analysed with the MultiNotch MS3 method (**e**) followed by data processing and statistical analysis using novel analytic suite (qPLEXanalyzer) (**f**)

analysis (175 proteins > 2-fold and adj. p -value < 0.01 (Limma moderated t -test) in qPLEX-RIME versus 51 proteins in non-quantitative RIME). Importantly, using the qPLEX-RIME we identified a number of novel ER α -associated candidate proteins. We validated the interactions of CBX3 (HP1 γ), NIPBL and FOXK1 with ER α , using Proximity Ligation Assay (PLA)²⁹ (Supplementary Fig. 1a). A GFP negative control was used to monitor for non-specific interactions (Supplementary Fig. 1b). Treatment of the MCF7 cells with the Selective ER α Degrader (SERD) Fulvestrant³⁰ (Supplementary Fig. 2) disrupted the above interactions demonstrating the specificity of the PLA assay and validating the interactors discovered by qPLEX-RIME (Supplementary Fig. 1a and c).

To test whether our quantitative pipeline can be widely used to study interactors of different bait proteins, we performed qPLEX-RIME experiments on three additional factors following the same experimental design as above. For these and all subsequent

experiments described in this study, we adopted the double crosslinking approach as a comparison between single and double crosslinking for ER α qPLEX-RIME data showed that the latter increases the pull-down efficiency of known and previously validated ER α interactors, including FOXA1, NR2F2 and NCOA2 (Supplementary Fig. 3a and Supplementary Data 3).

Firstly, the qPLEX-RIME method was applied to explore the interactome of CBP (CREB-binding protein) and NCOA3 (SRC-3); two well-characterised co-activators of nuclear receptors³¹. We identified 1437 and 1135 proteins for CBP and NCOA3, respectively, in the two multiplexed sets of bait and IgG pull-downs at peptide FDR < 1% (Supplementary Data 4 and 5). Both bait proteins were highly enriched in the target pull-downs compared to the IgG controls (CBP: $\log_2\text{fold-change}$ = 3.2 and NCOA3: $\log_2\text{fold-change}$ = 3.39) with a high number of unique peptides (44 unique peptides for CBP and 36 unique peptides for NCOA3) (Fig. 3a, b). Known interactors of CBP and NCOA3

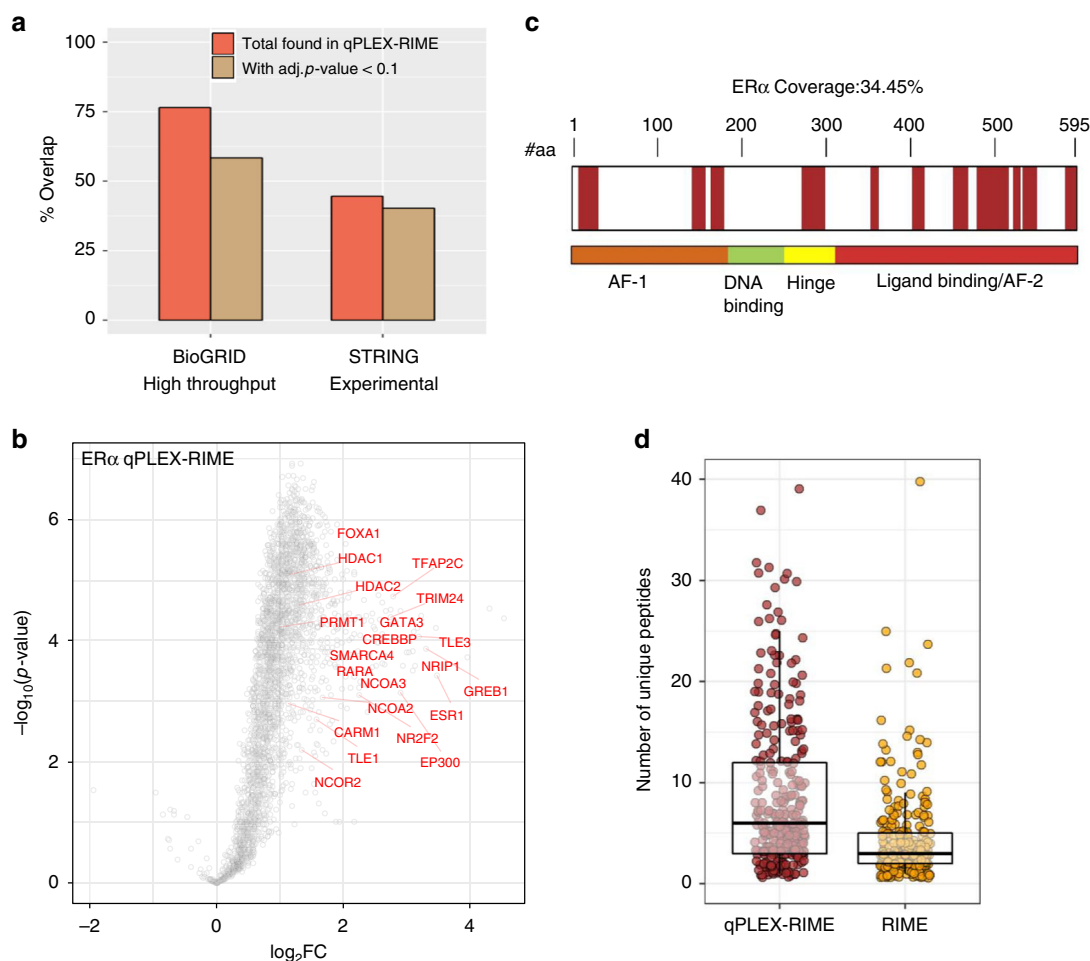


Fig. 2 Application of qPLEX-RIME for the identification of ERα-specific interactors in breast cancer cell lines. **a** Bar plots illustrating the overlap of the qPLEX-RIME data with known ERα-associated proteins from BioGRID and STRING databases. **b** Volcano plot summarising the quantitative results of the ERα qPLEX-RIME. ERα and several of its known significantly enriched interactors are labelled. **c** Sequence coverage of the ERα protein in the qPLEX-RIME analysis. **d** Boxplots of the number of unique peptides for the overlapping ERα-associated proteins between the non-quantitative RIME and the qPLEX-RIME method. Centre line shows the median, bounds of box correspond to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than $1.5 \times IQR$ (inter-quartile range)

were identified including EP300, p160 co-activators, arginine methyltransferases and the ERα complex³¹ (Fig. 3a, b). We also identified several members of the SWI/SNF chromatin remodelling complex, such as SMARCA4 (BRG1), SMARCE1 (BAF57), SMARCB1 (BAF47) and SMARCC2 (BAF170). Additionally, in the CBP qPLEX-RIME experiment we captured the association of CBP with the transcription factor JunB³², as well as with subunits of the mediator complex³³, which are known to associate with enhancer regions as well. Interestingly, in addition to other co-activators, we also found a strong enrichment of co-repressors such NCORs and HDACs in both data sets. This suggests that both co-activators and co-repressors are part of the same complex, which is consistent with previous findings demonstrating extensive co-localisation of co-repressors and co-activators by ChIP-seq³⁴.

Secondly, we studied the interactome of phospho-RNA polymerase II (POLR2A) using an antibody that recognises the phosphorylated serine-5, which serves as a platform for assembly of factors that regulate transcription initiation, elongation, termination and mRNA processing³⁵. We identified 1442 proteins across all multiplexed samples (Supplementary Data 6) and the bait protein was one of the top enriched proteins ($\log_2\text{fold-change} = 4.2$), identified with 96 unique peptides

(Fig. 3c). A list of known polymerase II-associated factors were also observed, such as subunits of the SWI/SNF complex, proteins of the mediator complex, initiation and elongation factors^{36,37} that are highlighted in Fig. 3c. A comparison of the interactomes of the four bait proteins (ERα, CBP, NCOA3 and POLR2A) showed significant numbers of uniquely identified interactors as well as partial overlap (Supplementary Fig. 3b). To examine whether the overlapping proteins are more likely due to the common underlying biology of the four baits rather than technical bias, we made a venn diagram using a random selection of proteins identified in the four qPLEX-RIME experiments, without considering enrichment relative to the IgG pull-downs (Supplementary Fig. 3c). This analysis showed a smaller number of proteins in the intersection of the four bait proteins indicating small contribution of technical factors to the observed overlap.

Taken together, our data demonstrate a gain in sensitivity using the qPLEX-RIME method that can lead to the identification of interacting proteins with statistical robustness and can be widely used for the characterisation of different interactomes.

Study of ERα complex dynamics upon OHT treatment. To investigate the dynamics of the ERα complex assembly upon

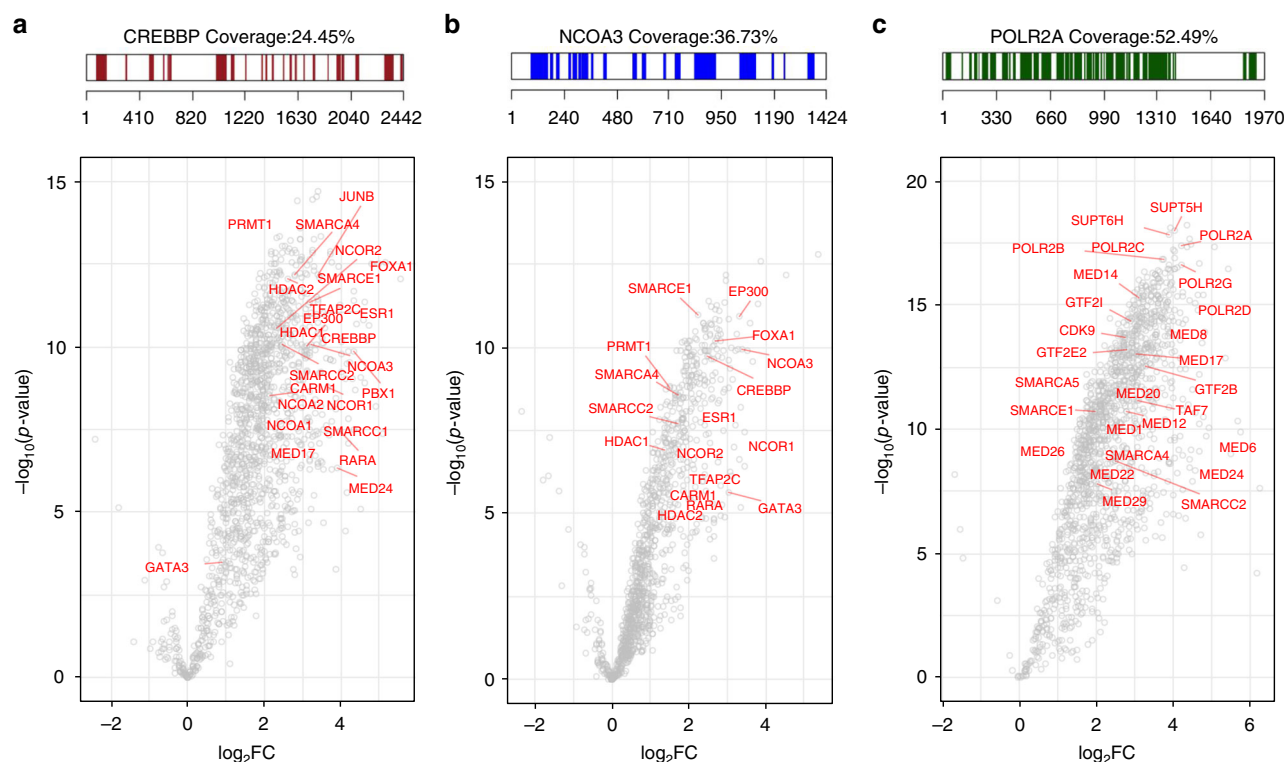


Fig. 3 Application of qPLEX-RIME in CREBBP, NCOA3 and POLR2A. **a** Peptide sequence coverage of the CREBBP (CBP) protein in the qPLEX-RIME analysis (top panel). The volcano plot summarises the quantitative results of the CBP qPLEX-RIME and several of its known interactors are highlighted in red font (bottom panel). **b** Peptide sequence coverage of the NCOA3 protein in the qPLEX-RIME analysis (top panel). The volcano plot summarises the quantitative results of the NCOA3 qPLEX-RIME and several of its known interactors are highlighted in red font (bottom panel). **c** Peptide sequence coverage of the phospho-polymerase II (POLR2A) protein in the qPLEX-RIME analysis (top panel). The volcano plot summarises the quantitative results of the POLR2A qPLEX-RIME and several of its known interactors are highlighted in red font (bottom panel)

treatment with the Selective Estrogen Receptor Modulator (SERM) 4-hydroxytamoxifen (OHT), we performed three qPLEX-RIME experiments (3×10 plex) using independently prepared biological replicates. MCF7 cells were crosslinked after treatment with 100 nM OHT for 2 h, 6 h and 24 h or after 24 h of vehicle (ethanol) treatment. Two biological replicates of each condition were included in each experiment, resulting in a total of six replicates per time point. Additionally, MCF7 cells were treated with OHT or ethanol and crosslinked after 24 h treatment in each experiment to be used for control IgG pull-downs, to enable discrimination of non-specific binding.

To confirm that the drug treatment was successful, we performed RNA-seq analysis of six biological replicates using matched OHT treated samples. The mRNA data revealed transcriptional repression of a number of known ERα target genes at 6 h and 24 h, confirming the response to the drug treatment. Specifically, at 24 h treatment the expression of *PGR*, *PDZK1*, *TFF1*, *AREG*, *PKIB*, *SIAH2*, *MYB*, *HEY2*, *FOS*, *GREB1* and *TFF3*^{38–43} was significantly inhibited compared to the vehicle treatment (\log_2 Fold-Change < -0.5 , adj. p -value < 0.05 , Limma moderated t -test) (Supplementary Fig. 4a and Supplementary Data 7).

MultiNotch MS3 analysis of the qPLEX-RIME samples quantified 1105 proteins (FDR $< 1\%$) across all three replicate experiments. Of these, 412 proteins were significantly enriched in ERα pull-downs compared to IgG samples (\log_2 Fold-Change > 1 , adj. p -value < 0.01 , Limma moderated t -test) (Supplementary Data 8). Total ERα levels changed upon OHT treatment (Supplementary Fig. 4b), indicating that altered levels of antigen may influence the amount of purified proteins. Our data showed that this resulted in a significant dependency of the quantified

proteins on the amount of ERα pulled down (Supplementary Fig. 5a). To correct for this effect, we applied a linear regression approach^{15,44} using the ERα profile as the independent variable and the profile of any other protein as the dependent variable. The advantage of this approach is that proteins with strong dependency on the target protein are subjected to significant correction, whereas proteins with small dependency on the target protein are only slightly corrected. Two such examples, of known ERα interactors before and after correction are shown in Supplementary Fig. 5b. Finally, using the quantification values corrected for the abundance of ERα, we found 249 specific proteins with altered profile in the interactome in at least one time point ($|\log_2$ Fold-Change| > 0.5 , adj. p -value < 0.05 , Limma moderated t -test) allowing for a comprehensive mapping of the dynamic organisation of the ERα complex in response to OHT treatment.

Dissociation and recruitment of co-factors upon OHT treatment. We next interrogated the significant changes observed in the ERα interactome at each time point during OHT-mediated growth inhibition. After 2 h treatment with OHT, a significant loss of 12 proteins was observed including known ERα co-activators, such as NCOA3 (AIB1/SRC-3) and CREBBP (CBP) (Fig. 4a). These proteins have been associated with histone acetylation and activation of gene transcription^{23,45} and their loss in the ERα interactome upon OHT treatment is consistent with previous studies showing that OHT binding blocks access of co-activators⁴⁶. We also observed a significant loss of the interaction between ERα and NRIP1 (RIP140) protein. NRIP1 can act as a corepressor or as a coactivator⁴⁷ with previous evidence

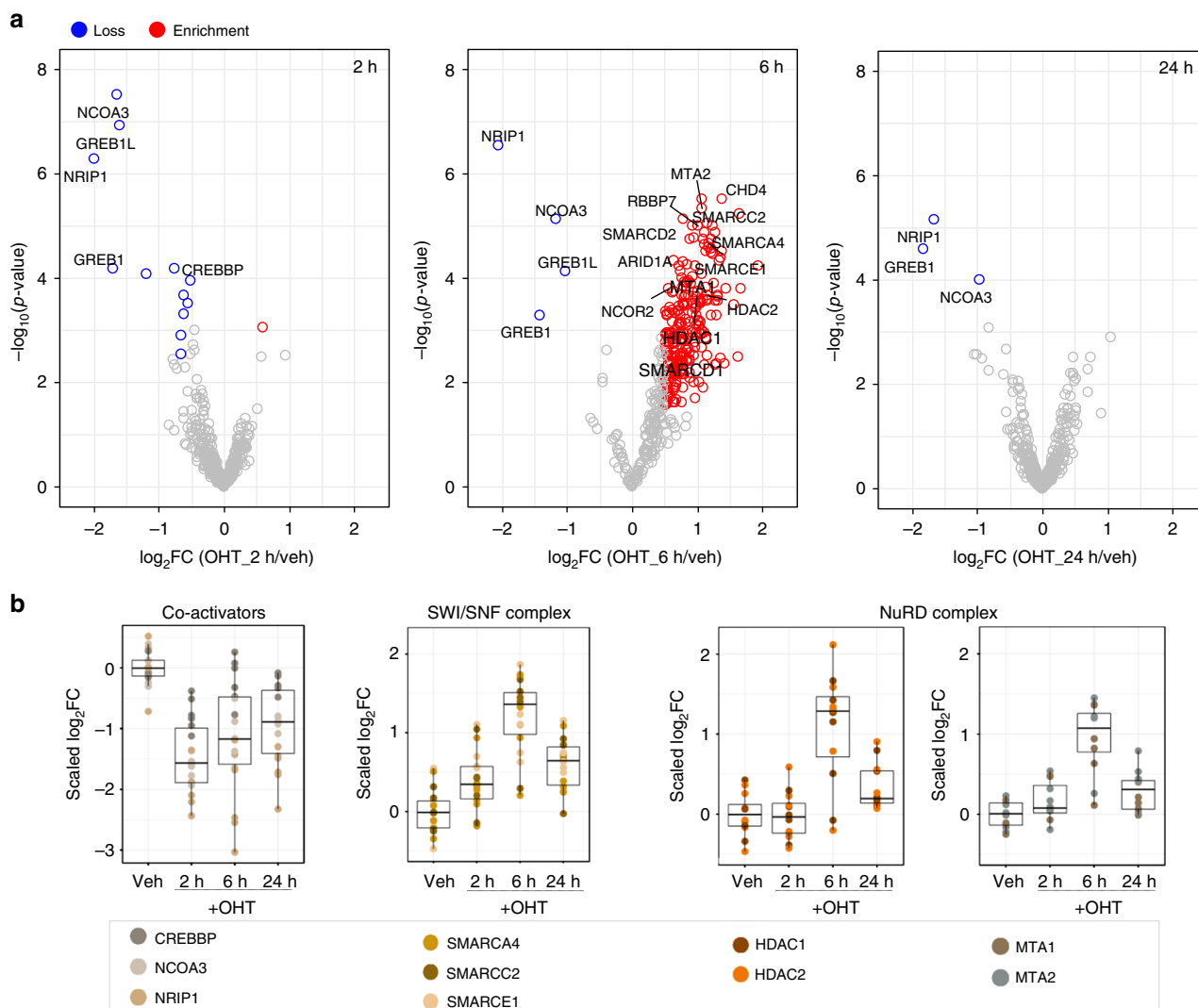


Fig. 4 Temporal profiling of the ER α interactome following treatment of MCF7 cells with OHT. **a** Volcano plots highlighting enriched or lost proteins in the ER α interactome upon OHT treatment for 2 h, 6 h and 24 h. **b** Boxplots illustrating the loss of ER α co-activator proteins CREBBP, NCOA3 and NRIP1 at 2 h and the enrichment of several subunits of the SWI/SNF and NuRD complexes at 6 h (left to right). Quantitative values are normalised so that the median of the vehicle treated samples is zero (centered around the median of vehicle). Centre line shows the median, bounds of box correspond to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than $1.5 \times \text{IQR}$ (inter-quartile range)

suggesting that NRIP1 is required for ER α -complex formation and ER α -mediated gene expression⁴⁷. The quantification profile of these key ER α co-activator proteins across all biological replicates is shown in Fig. 4b. Furthermore, OHT treatment resulted in the loss of GREB1 and its paralog gene product GREB1L. Loss of GREB1 upon OHT treatment has been previously described¹³, but here we report the loss of both proteins simultaneously.

After 6 h treatment with OHT, 237 specific interactors showed significant enrichment ($\log_2\text{Fold-Change} > 0.5$, adj. $p\text{-value} < 0.05$, Limma moderated $t\text{-test}$) compared to the vehicle treatment, whilst NCOA3, NRIP1, GREB1 and GREB1L remained at decreased levels in the interactome ($\log_2\text{Fold-Change} < -0.5$, adj. $p\text{-value} < 0.05$, Limma moderated $t\text{-test}$) (Fig. 4a). Notably, there was an enrichment in the recruitment of several components of the NuRD (Nucleosome Remodelling and Deacetylase) complex, e.g. HDAC1/2^{24,48} and the signature components MTA1/2^{24,48}, as well as an enrichment of the co-repressor NCOR2 (SMRT)^{25,49} (Fig. 4b). Consistently, NURD complex and NCOR2 has been previously shown by ChIP to be

recruited to promoter regions of ER α target genes following OHT treatment^{23,24}. Additionally, we found enriched subunits of the ATP-dependent chromatin remodelling complex SWI/SNF, which is known to regulate both gene activation and gene repression^{50,51}. Detected components included SMARCC2 (BAF170), SMARCE1 (BAF57) and SMARCA4 (BRG1) (Fig. 4b). SMARCA4 protein, which was previously shown to be required for repression of ER-mediated transcription⁵⁰, was one of the top enriched SWI/SNF proteins. The loss of NCOA3 and CBP at 2 h and the enrichment of SMARCC2 (BAF170) and HDAC1 at 6 h was validated with PLA assays (Supplementary Fig. 6a and b).

At 24 h we observed an almost complete restoration of the ER α complex, with the exception of the NCOA3, NRIP1 and GREB1 proteins, which were still decreased ($\log_2\text{Fold-Change} < -0.5$, adj. $p\text{-value} < 0.05$, Limma moderated $t\text{-test}$) (Fig. 4a). Taken together, our results indicate that the inhibitory effect of OHT peaks at 6 h, where ATP-dependent remodelling and corepressor complexes may coordinate to create a transcriptionally inactive chromatin environment.

Identification of net changes in the ER α complex. Our data suggest that treatment of MCF7 cells with OHT triggers significant changes in the composition of ER α interactome. To assess whether the changes identified by the qPLEX-RIME analysis are specific changes in interactions or result from changes in total protein levels, we performed timecourse whole proteome quantification in matched samples under the same conditions (vehicle, 2 h, 6 h and 24 h, four biological replicates each) (Supplementary Data 9). We confirmed the OHT up-regulation of ER α protein levels (\log_2 Fold-Change: 2 h 0.21, 6 h 0.5 and 24 h 1), which was not due to an increase in gene transcription. This is consistent with previous reports demonstrating increased ER α stability in the presence of OHT⁵². A comparison between the qPLEX-RIME results and the total proteome data confirmed that the changes detected in the ER α complex upon OHT treatment represent changes in protein recruitment as the respective total protein and mRNA levels remained unchanged (Fig. 5a). GREB1 was the only ER α interactor with decreased mRNA and total protein levels at 24 h treatment. This is consistent with GREB1 being an ER α target gene^{13,43} and explains the decreased association between ER α and GREB1 at this late time point.

Downstream k-means clustering of the most variable proteins (adj. p -value < 0.05, Limma moderated t -test) across the three time points in the total proteome, identified clusters of up- and downregulated proteins (Fig. 5b). Gene Set Enrichment Analysis of the clusters, performed in Perseus software⁵³, displayed an overrepresentation of genes related to estrogen response and tamoxifen resistance (Fig. 5c). Our findings also revealed the downregulation of proteins involved in cell cycle⁵⁴ (Supplementary Fig. 7a), in line with the antiproliferative effects of OHT²⁴. Overall, significant changes in gene expression were observed already at 6 h coinciding with pronounced changes in the ER α interactome. As expected, the most significant changes in the total proteome were observed at the later time point (24 h). These results confirm that shuffling of ER α -associated proteins is not typically due to global changes in protein levels. The low mRNA-to-protein correlation at 2 h and 6 h and the respective strong correlation at 24 h are shown in Supplementary Fig. 7b and c. We conclude that our qPLEX-RIME data in combination with the total proteome measurements delineate both the local molecular events in the ER interactome and the associated downstream global effects.

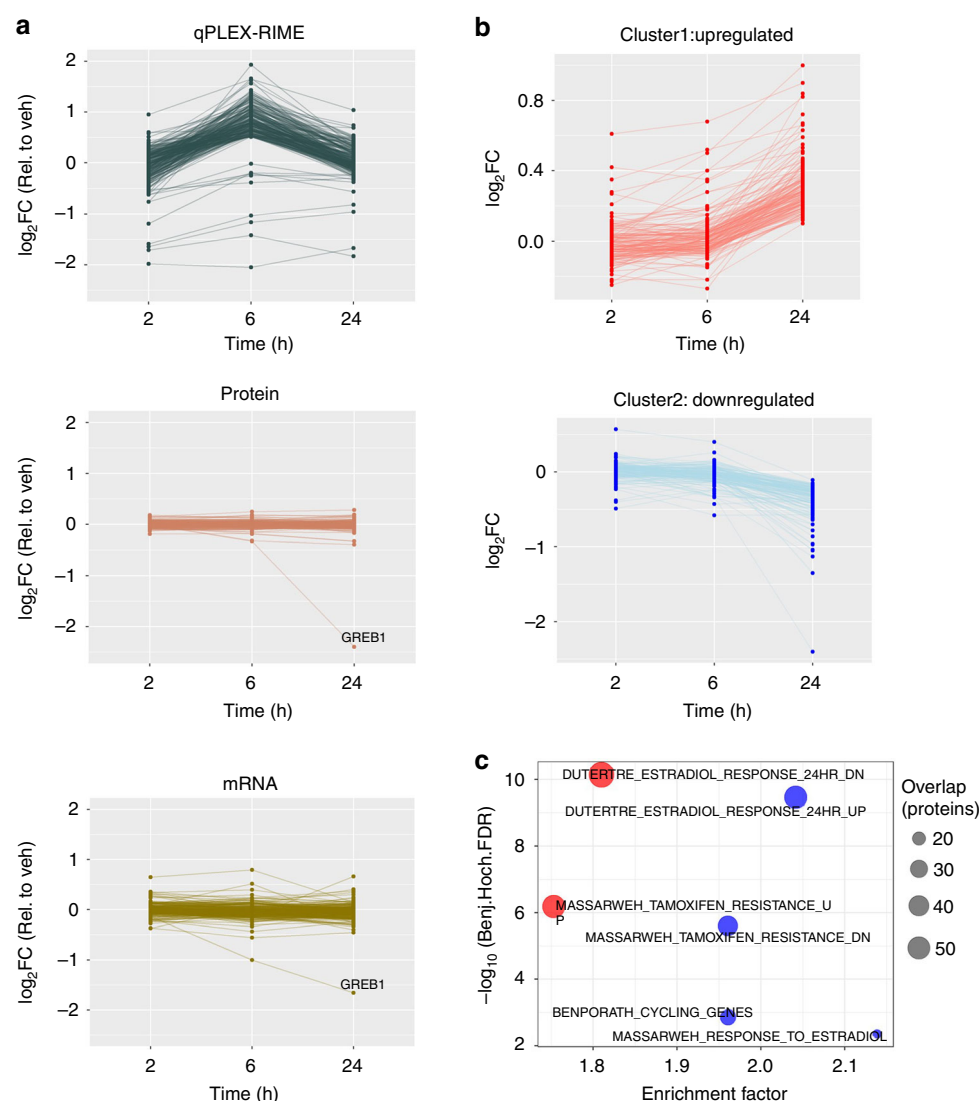


Fig. 5 Comparison of qPLEX-RIME data with total proteome analysis and RNA-seq data. **a** Line plots of the significantly enriched or lost proteins in the qPLEX-RIME data (top panel), their respective profiles in the total proteome analysis (middle panel) and in the RNA-seq analysis (bottom panel). **b** Line plots representing two k-means clusters of down- and upregulated proteins identified in total proteome analysis (top 10% most variable proteins with at least one significant change, adj. p -value < 0.05). **c** Gene set enrichment analysis for the down- and upregulated protein clusters

Application of qPLEX-RIME in clinical tumour material. To test whether qPLEX-RIME can be used to capture chromatin-associated protein–protein interactions in cancer specimens, we conducted an ER α qPLEX-RIME experiment using three independent ER positive human PDX tumours (HCI-003, HCI-005, HCI-006) that have been previously described⁵⁵. Cryosections (30 μ m) of each tumour were double-crosslinked and each tumour was split into ER α and matched IgG pull-downs (Fig. 6a). The MultiNotch MS3 analysis identified 2319 proteins (FDR < 1%) across all multiplexed samples with highly reproducible profiles (Supplementary Fig. 8a and Supplementary Data 10). This analysis successfully recovered and quantified ER α

(log₂Fold-Change = 1.72, adj. *p*-value = 0.026, Limma moderated *t*-test, unique peptides = 3) using an unbiased mass spectrometry approach in tissue. In addition, many validated and known ER α interactors were discovered from the qPLEX-RIME conducted in PDX material, including CBP^{23,56}, NCOA2⁵⁶, HDAC1²⁴, GREB1¹³, SMARCE1 (BAF57)⁵¹, SMARCA4 (BRG1)⁴⁵ and NCOA5 (CIA)⁵⁷ (Supplementary Fig. 8b). Sequence analysis of the qPLEX-RIME data showed that 60% of the significant interactors were identified with at least one unique human peptide (i.e. a peptide that does not align to the mouse proteome), indicating that the proteins identified above were primarily from the human cancer cells. Consistently, the tumour samples showed high

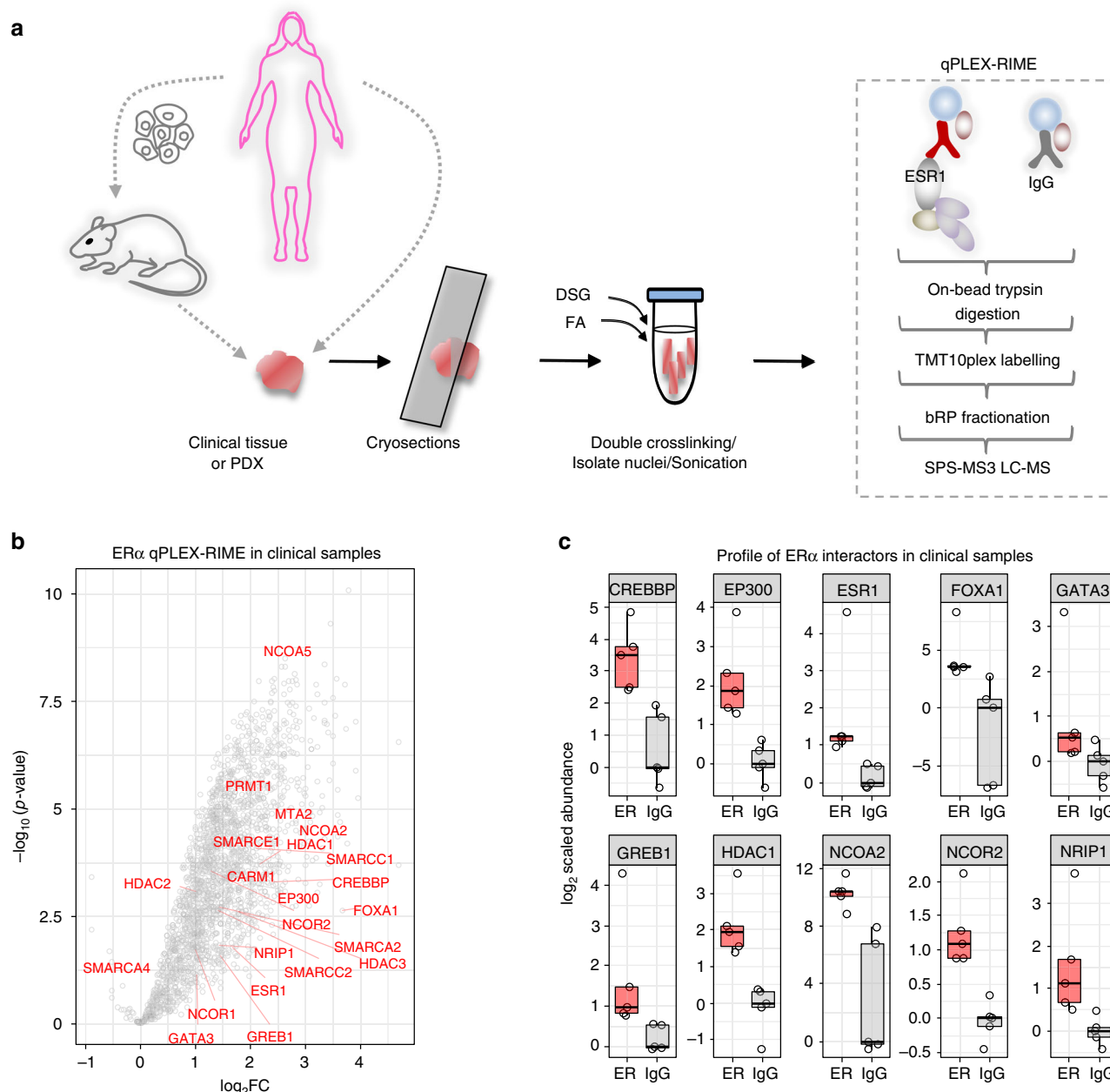


Fig. 6 Characterisation of ER α interactors from in vivo samples. **a** Schematic representation of the ER α qPLEX-RIME workflow in human xenograft tissues or in human breast cancer tumours. **b** Volcano plot summarising the quantitative results of the ER α interactome in human breast cancer tumours. Several well-known ER α interactors are labelled. **c** Boxplots illustrating the enrichment of selected known ER α interactors in the ER α samples compared to IgG controls in human breast cancer tissues. The log₂ values are normalised so that the median of IgGs is zero. Centre line shows the median, bounds of box correspond to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than 1.5 \times IQR (inter-quartile range)

cellularity and positive staining for human ER α exclusively in the cancer cells and not in the stroma (Supplementary Fig. 8c).

Prompted by the successful application of qPLEX-RIME in PDX tumours we sought to test the sensitivity of our method in human cancer clinical tissues, collected from surgery. To this end, we performed an ER α qPLEX-RIME experiment in five independent human breast cancer tumours (ER α positive, PR positive, Her2 negative and Grade2/Grade3). Approximately 60 sections (30 μ m) were obtained per sample, which were double-crosslinked and split for ER α and matched IgG pull-downs (Fig. 6a). The analysis successfully recovered ER α with excellent coverage (17 unique peptides), as well as 2191 proteins (FDR < 1%) that were quantified in all samples combined (Supplementary Data 11). These included well-described ER α interactors such as FOXA1, GATA3, GREB1, EP300, CBP, HDACs, NCORs and NCOA2 and subunits of the SWI/SNF complex (Fig. 6b). The enrichment of several ER α interactors in the bait samples compared to IgG control samples is illustrated in Fig. 6c.

Our data highlights the method's sensitivity and ability to identify endogenous protein networks from heterogeneous human tumour samples. Importantly, we report the identification of interactors from human tumour tissue material in an unbiased manner.

Discussion

Here, we describe qPLEX-RIME, a proteomic method which enables comprehensive mapping of endogenous protein interactomes with high sensitivity and statistical robustness. The qPLEX-RIME approach integrates the well-established RIME immunoprecipitation method with advanced high-resolution quantitative multiplexed mass spectrometry analysis. The method can be utilised to discriminate enriched bona fide binding partners from contaminant proteins and to delineate the dynamics of chromatin-associated protein complexes with in-depth protein detection, reproducible quantification and increased sample throughput. The filtering criteria for the prioritisation of the best candidates depend on the type of experiment and the biological question. For bait proteins where very little is known about their interactome, we recommend the use of more stringent specificity criteria in terms of enrichment fold-change, *p*-value and number of unique peptides in combination with additional filtering based on functional annotations. When the focus is on the dynamic changes of interactomes, the prioritisation of the candidates mostly relies on their robust quantitative profiling across different conditions.

The multiplexed analysis of our pipeline eliminates the need to compare multiple data obtained by individual LC-MS runs, thereby increasing the quantification coverage in very low abundant protein interactors that are stochastically captured between independent replicate runs⁵⁸. The ability to combine the labelled peptides derived from multiple samples increased the sensitivity of the method and enabled the characterisation of the ER α interactome in clinical tumours. Whilst interactors have previously been detected from clinical material, this required targeted mass spectrometry-based approaches and has not been done in an unbiased manner before¹³. Additionally, the use of isobaric labelling resolves the difficulties encountered with cell lines that are not compatible with stable-isotope labelled culture media and provides a means for quantitative analysis for clinical samples that are not amenable to in vivo isotopic labelling techniques. Importantly, our isobaric-labelling data demonstrated high reproducibility with previously published SILAC data¹³ (Supplementary Fig. 7d), confirming the accurate quantification obtained by the MultiNotch MS3 level mass spectrometry analysis.

Here, we focused on ER α , the major driving transcription factor in luminal breast cancer⁵⁶, which can be targeted by tamoxifen, a drug used for the treatment of ER⁺ breast cancer⁵⁹. Although many ER α interactors involved in ER-mediated gene expression have been discovered^{23,45} our knowledge about their relevance at the tissue level and the impact of tamoxifen on their global association with ER α remains limited. The quantitative data obtained by the qPLEX-RIME experiments has provided us with a list of ER α -associated proteins with significant enrichment over the IgG samples. These include transient, indirect or weak interactions, as it is known that ER α associates with a number of different co-activators rapidly in a cyclic fashion^{23,45}. As such the final readout of the crosslinking-based qPLEX-RIME method represents the sum of these interactions. Among these, we validated the interactions between ER α and three proteins; namely CBX3, NIPBL and FOXK1. CBX3 protein is a member of the HP1 protein family, a group of proteins that have been implicated in gene regulation, DNA replication and nuclear architecture⁶⁰, whereas NIPBL is a core subunit of the highly conserved protein complex cohesin that has an important role in chromatin structure, gene expression, and DNA repair⁶¹. The transcription factor FOXK1 belongs to the forkhead family and has an important role in tumorigenesis^{62,63}. These findings demonstrate that the gain in sensitivity obtained by qPLEX-RIME can reveal novel ER α interactors. Collectively, we identified a compendium of 253 proteins with consistent presence in all MCF7 data sets (Fig. 7). Importantly, our data show that the vast majority of these ER α -associated proteins (83%) can now be studied either in PDX or in human clinical tissues validating the relevance of these factors in vivo. Additionally, our qPLEX-RIME data on three additional factors, the CREBBP, NCOA3 and the phosphorylated form of POLR2A, highlight the wide applicability of our pipeline.

The application of qPLEX-RIME targeting ER α at multiple time points after OHT treatment, revealed a dynamic change in ER α co-regulators following drug treatment recapitulating and expanding the existing knowledge of OHT mechanism. After 2 h OHT treatment, we observed a loss of important transcriptional co-activators, such as NCOA3 and CBP, whereas at 6 h we observed enrichment on the recruitment of two well-conserved chromatin remodelling complexes, namely the NuRD and the SWI/SNF complex. This coincided with the enrichment of the basal corepressor NCOR2, which assists in the recruitment of HDAC proteins⁶⁴. At the latest time point of 24 h, we observed a restoration of the ER α complex, which may be linked to the half-life of OHT. The exceptions were NRIP1, GREB1 and NCOA3. Interestingly, NCOA3 is amplified in breast cancer²² and its expression levels have been associated with the effectiveness of tamoxifen treatment⁶⁵. Further, ChIP-seq analysis has revealed that a number of binding sites of NCOA3 are associated with genes with a predictive value for breast cancer patient outcome⁵⁶, supporting an important role of this co-regulator in tamoxifen response.

Our timecourse data indicate a switch between activation and repression of transcription in response to OHT treatment. This transition engages a two-step process with the immediate loss of co-activators, followed by the recruitment of co-repressors and ATP-chromatin remodelling complexes that may act cooperatively or in a sequential manner to accomplish transcriptional repression. The integration of qPLEX-RIME data with global protein and mRNA analysis provides a comprehensive view of the activity of a transcription-associated complex over time. A proposed model of OHT mechanism is depicted in Supplementary Fig. 9.

The qPLEX-RIME method can be used to monitor any dynamic changes of interest and importantly can be applied to clinical samples to study tumour evolution, treatment response or

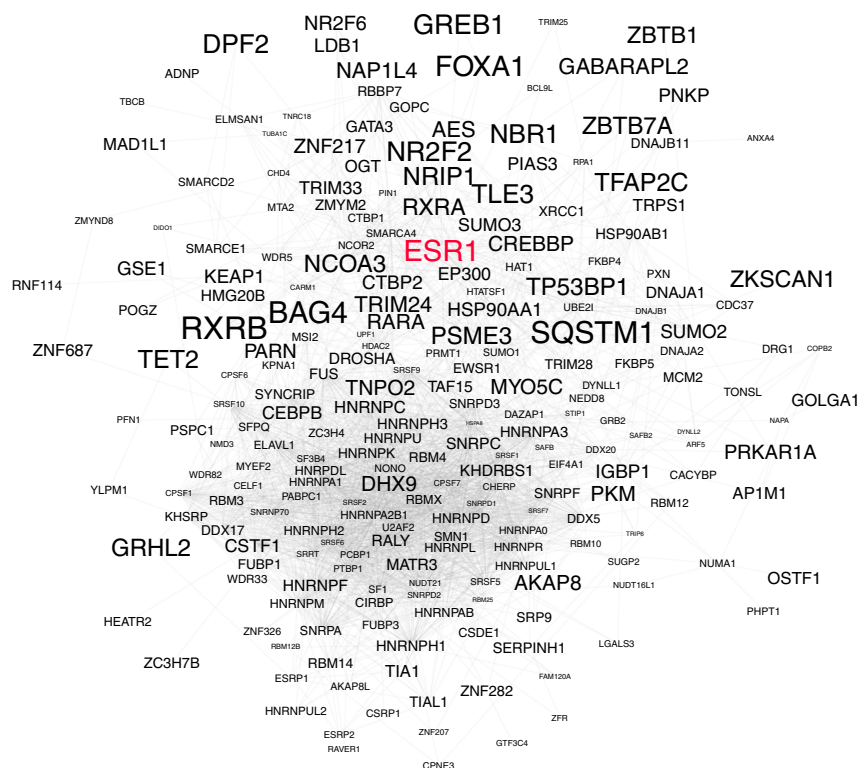


Fig. 7 Most frequently enriched ER α interactors. STRING network of 253 ER α interactors identified consistently across all the qPLEX-RIME analyses performed in MCF7 cells. The font size increases proportionally to the average fold-change enrichment of these proteins across all the ER α samples compare to IgG controls

numerous other biological and clinical questions. It provides a robust tool for the quantitative analysis of complexes that can be applied to generate comprehensive endogenous protein–protein interaction maps.

Methods

Cell lines and cell treatments. ER α -expressing MCF7 breast cancer cells were cultured in Dulbecco's Modified Eagle Medium DMEM (Gibco, Thermo Scientific, Leicestershire, UK, ref. 41965-239). Media was supplemented with 10% foetal bovine serum (FBS), 50 U/ml penicillin, 50 μ g/ml streptomycin and 2 mM L-glutamine. MCF7 cells were obtained from ATCC and they were tested for mycoplasma contamination. Also, the MCF7 cells were genotyped by short-tandem repeat genetic profiling using the PowerPlex_16HS_Cell Line panel and analysed using Applied Biosystems Gene Mapper ID v3.2.1 software by the external provider Genetica DNA Laboratories (LabCorp Specialty Testing Group). For the cell treatments, 4-Hydroxytamoxifen (Sigma-Aldrich, #HG278) or Fulvestrant (Selleckchem, #S1191) were used at final concentration 100 nM.

Whole cell lysate preparation and western blot analysis. Cell pellets were reconstituted in 100 μ l RIPA buffer (Thermo Scientific, #89901) that was supplemented with protease inhibitors (Roche). 25 μ g protein from each sample was loaded on the gel (Invitrogen 4–12%) and the Precision Plus, ProteinTM dual colour Standards Protein molecular weight marker (Bio-Rad, #161-0974) was used for the determination of protein sizes. The proteins were transferred onto a nitrocellulose membrane using the iBlot[®] 2 Dry Blotting System (Invitrogen) followed by one hour blocking using Odyssey[®] Blocking Buffer (Li-Cor, 927-40000). The membrane was immunoblotted with ER α antibody (Novocastra #6045332, 1:100) and beta-actin (Cell signalling #4970, 1:1000). Detection of the ER was achieved using the IRDye[®] 800 CW Goat anti-Mouse (926-32210, Li-Cor Biosciences) diluted to 1:5000, while the loading control was detected using the IRDye 680LT Goat anti-Rabbit (926-68071, Li-Cor Biosciences) diluted to 1:15000. All antibodies were diluted in Odyssey Buffer contained 0.1% Tween. Supplementary Fig. 10 shows the uncropped scan of the blot.

RNA-seq analysis. Cells were washed twice with cold Phosphate buffered saline (PBS) and harvested using 350 μ l of lysis buffer (RLT). Total RNA was extracted using the RNeasy[®] kit (Qiagen, #74106) according to the manufacturer's instructions. The extracted RNA was quantified using a NanoDrop[®] ND-1000

Spectrophotometer (Thermo Scientific). For the library preparation, the Illumina TruSeq Stranded mRNA Library Prep Kit High Throughput was used according to the manufacturer's instructions and two lanes of 50 bp single-end reads were run on HiSeq 4000. Reads were aligned to the human genome version GRCh37.75 using TopHat v2.1.0⁶⁶. Read counts were obtained using feature Counts function in Subread v1.5.2⁶⁷ and read counts were normalised and tested for differential gene expression using the DESeq2 workflow⁶⁸. Multiple testing correction was applied using the Benjamini–Hochberg method.

RIME analysis. MCF7 cells (2×10^6) were grown in complete media. The media was replaced with PBS containing 1% FA (Thermo #28908) and crosslinked for 10 min. For the double crosslinking cells were incubated in PBS containing 2 mM DSG (disuccinimidyl glutarate- Santa Cruz Biotechnology, #sc-285455A) for 20 min followed by incubation in 1% FA for 10 min. Crosslinking was quenched by adding glycine to a final concentration of 0.1 M. For the performance of RIME experiments, 50 μ l of Dynabeads[®] Protein A (Invitrogen) and 5 μ g of specific antibody were used for each sample. The antibodies used were: Rabbit polyclonal ER α antibody (Santa Cruz, sc-543), rabbit polyclonal SRC3 antibody (Bethyl laboratories, A300-347A), rabbit polyclonal CBP antibody (Diagenode, C15410224), rabbit polyclonal RNA polymerase II (phospho S5) antibody (Abcam, ab5131) and rabbit IgG antibody (Santa Cruz, sc-2027 or Abcam, ab171870). For nuclear extraction the cell pellet was resuspended in LB1 buffer (50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 and 0.25% Triton X-100) followed by rotation mixing for 10 min at 4 °C. Then, nuclei were pelleted and resuspended in LB2 buffer (10 mM Tris-HCL (pH 8.0), 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA) and rotated at 4 °C for 5 min. The samples were resuspended in LB3 buffer (10 mM Tris-HCL (pH 8), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate and 0.5% N-lauroylsarcosine). Chromatin was sheared by sonication (Diagenode) to produce DNA fragments of 100–1,000 bp. The bead-bound antibody and chromatin were incubated overnight at 4 °C. The next day the beads were washed 10 times with 1 ml ice-cold RIPA buffer and twice with 500 μ l 100 mM AMBIC (ammonium bicarbonate).

Proximity ligation assay. Cells were fixed and permeabilised by the addition of ice-cold methanol (–20 °C) for 3 min followed by three washing steps with cold PBS. PLA was carried out according to manufacturer's instructions (Sigma Aldrich, #DUO92007). The following primary antibodies were used for the PLA assay: ER α (Santa Cruz, sc-543 or sc-8002, 1:250) HP1 γ (Santa Cruz, sc-365085, 1:400), NIPBL (Santa Cruz, sc-374625, 1:200), FOXK1 (Santa Cruz, sc-373810, 1:200), GFP

(Abcam, ab1218, 1:200), NCOA3 (Bethyl Laboratories, A300-347A, 1:200), CBP (Bethyl Laboratories, A300-363A, 1:200), BAF170 (Santa Cruz, sc-17838, 1:200), HDAC1 (Santa Cruz, sc-81598, 1:200) and the incubation on the coverslips was performed for 1 h at 37 °C. For the single PLA recognition experiment two ERα antibodies (Santa Cruz, sc-543, 1:800 and Invitrogen, MA5-13191, 1:1200) were used in combination. The secondary proximity probes (Sigma Aldrich, Rabbit-PLUS, #DUO92002 and Mouse-MINUS, #DUO92004) were incubated for 1 h at 37 °C. The Leica DFC340FX microscope was used and images were captured with Leica Imaging software. DAPI and PLA fluorescence were captured at high resolution for a total of 8 separate observation fields. Cell numeration and PLA labelling were carried out using Image J software. Cells and red PLA dots were counted using the 'Analyze Particles' function. For each condition at least 200 cells were imaged and analysed. Then, the average value of number of spots per nucleus was calculated. All statistical analyses were carried out by performing Student's *t*-test.

Immunofluorescence. Cells were fixed and permeabilised with ice-cold (−20 °C) methanol for 3 min and after fixation cells were blocked in PBS-5 % (w/v) Bovine Serum Albumin (BSA) (Sigma-Aldrich) for 30 min at room temperature. The primary ERα antibody (Santa Cruz, sc-543, 1:250) was diluted in blocking solution (PBS-5 % (w/v) BSA) and incubated on coverslips for 1 h at 37 °C. Afterwards, the coverslips were washed four times in washing buffer (PBS-0.5% Tween). Secondary antibody conjugated to Alexa Fluor 488 (Invitrogen, #A-21206, 1:500) was diluted in blocking solution and incubated on coverslips for 1 h at 37 °C in the dark. Coverslips were then washed again three times in washing buffer and once in PBS.

PDX propagation and tissue collection. Viable frozen PDX tumour tissue was propagated in immune-compromised mice. Briefly, 1 mm³ tumour pieces were implanted into the 4th mammary pad of NSG mice. All mice were supplemented with estrogen, using silastic E2 pellets (made in-house) inserted into the dorsal scruff. Twice weekly standard monitoring and tumour measurement was conducted. Once tumours reached appropriate size, ~1000 mm³, mice were sacrificed by cervical dislocation under deep, isoflurane-induced anaesthesia. The tumours were resected, diced and processed by either snap freezing in liquid nitrogen, fixing in 10% neutral buffered formalin solution for subsequent paraffin embedding, embedding in OCT, or viably freezing in FCS supplemented with 5% DMSO.

Sample preparation of clinical tumour material. Clinical samples were cryo-sectioned in 30 μm slices using the Leica CM 3050 S cryostat. Tissue sections were fixed in a two-step procedure by adding 2 mM DSG for 25 min. In the same suspension of tissue sections, 1% FA was added for another 20 min without removal of the DSG. Crosslinking was quenched by the addition of glycine to a final concentration of 0.25 M. Samples were centrifuged for 3 min at 2500 g and the supernatant was discarded. Tissue pellets were washed twice with cold PBS and resuspended in 6 ml LB3 buffer (10 mM Tris-HCl (pH 8), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, and 0.5% N-lauroylsarcosine), followed by tip sonication for 12–20 cycles (30 s on, 1 min off) depending on the tumour size. The downstream processing was performed as described above (see RIME method section) and the tissue samples were separated in two parts for the performance of ERα and IgG RIME pull-down assays. Patient and patient-derived tissues used in this work were collected under protocol X13-0133, HREC/13/RPAH/187. HREC approval was obtained through the SLHD (Sydney Local Health District) Ethics Committee (Royal Prince Alfred Hospital) zone), and site-specific approvals were obtained for all additional sites. Written consent was obtained from all patients prior to collection of tissue and clinical data stored in a de-identified manner, following pre-approved protocols. All animal procedures were carried out in accordance to relevant national and international guidelines and animal protocols approved by the Garvan/St Vincent's Animal Ethics Committee (Animal ethics number 15/10).

Immunohistochemistry. FFPE blocks from PDX tumours were sectioned at 4 μm onto Superfrost Plus slides. Immunohistochemistry was carried out using the Leica Bond Autostainer. Sections underwent dewaxing, heat induced antigen retrieval (Leica reagent ER2, 30 mins), and primary and secondary antibody incubations, using ERα antibody (ab108398, Abcam, 1:500) and the EnVision + Rabbit secondary system, respectively. Sections were counterstained with haematoxylin.

Trypsin digestion and TMT labelling. A volume of 10 μL trypsin solution (15 ng/μL) (Pierce) in 100 mM AMBIC was added to the beads followed by overnight incubation at 37 °C. A second digestion step was performed the next day for 4 h. After proteolysis the tubes were placed on a magnet and the supernatant solution was collected after acidification by the addition of 2 μL 5% formic acid. The resultant peptides were cleaned with the Ultra-Micro C18 Spin Columns (Harvard Apparatus) according to manufacturer's instructions. The peptide samples were dried with speedvac, reconstituted in 100 μL 0.1 M TEAB (triethylammonium bicarbonate) and labelled using the TMT-10plex reagents (Thermo Fisher) with a randomised design. The peptide mixture was fractionated with Reversed-Phase cartridges at high pH (Pierce #84868). Nine fractions were collected using different elution solutions in the range of 5–50% ACN.

For the total proteome analysis 200 μL of 0.1 M TEAB, 0.1% SDS buffer was added to each cell pellet followed by probe sonication and boiling at 95 °C. Protein concentration was estimated with Bradford assay (BIO-RAD-Quick start) according to manufacturer's instructions. For each sample, 90 μg of total protein were reduced for 1 h at 60 °C by the addition of 2 μL 50 mM tris-2-carboxyethyl phosphine (TCEP, Sigma). Cysteines were blocked for 10 min on the bench with the addition of 1 μL 200 mM methyl methanethiosulfonate (MMTS, Sigma). For peptide generation, trypsin (Pierce #90058) solution was added at ratio protein/trypsin ~30:1 for overnight digestion at 37 °C. The next day peptides were allowed to react with the TMT-10plex reagents (Thermo Scientific) for one hour. The reaction was quenched with 8 μL of 5% hydroxylamine (Thermo Scientific) and the labelled samples were mixed and dried with speedvac concentrator. The TMT mix was reconstituted and fractionated on a Dionex Ultimate 3000 system at high pH using the X-Bridge C18 column (3.5 μm 2.1 × 150 mm, Waters) with 1% gradient. UV signal was recorded at 280 and 215 nm and fractions were collected in a peak dependent manner.

LC-MS analysis. Peptide fractions were analysed on a Dionex Ultimate 3000 UHPLC system coupled with the nano-ESI Fusion Lumos (Thermo Scientific). Samples were loaded on the Acclaim PepMap 100, 100 μm × 2 cm C18, 5 μm, 100 Å trapping column with the ulPickUp injection method using the loading pump at 5 μL/min flow rate for 10 min. For the peptide separation the EASY-Spray analytical column 75 μm × 25 cm, C18, 2 μm, 100 Å column was used for multi-step gradient elution. Mobile phase (A) was composed of 2% acetonitrile, 0.1% formic acid and mobile phase (B) was composed of 80% acetonitrile, 0.1% formic acid. The elution method at flow rate 300 nL/min included the following: for 95 min gradient up to 45% (B), for 5 min gradient up to 95% (B), for 8 min isocratic 95% (B), for 2 min down to 5% (B), for 10 min isocratic equilibration 5% (B) at 40 °C. For the clinical sample analysis, a longer gradient separation was used as follows: for 160 min gradient up to 40% (B), for 10 min gradient up to 95% (B), for 8 min isocratic 95% (B), for 2 min down to 5% (B), and for 10 min isocratic equilibration 5% (B). The Lumos was operated in a data-dependent mode for both MS2 and SPS-MS3 methods. The full scans were performed in the Orbitrap in the range of 380–1500 *m/z* at 120 K resolution. The MS2 scans were performed in the ion trap with collision energy 35%. Peptides were isolated in the quadrupole with isolation window 0.7 Th. The 10 most intense fragments were selected for Synchronous Precursor Selection (SPS) HCD-MS3 analysis with MS2 isolation window 2.0 Th. The HCD collision energy was set at 55% and the detection was performed with Orbitrap resolution 60k and in scan range 110–400 *m/z*.

Data processing and interpretation. The collected CID tandem mass spectra were processed with the SequestHT search engine on the Proteome Discoverer 2.1 software for peptide and protein identifications. The node for SequestHT included the following parameters: Precursor Mass Tolerance 20 ppm, Fragment Mass Tolerance 0.5 Da, Dynamic Modifications were Oxidation of M (+15.995 Da), Deamidation of N, Q (+0.984 Da) and Static Modifications were TMT6plex at any N-Terminus, K (+229.163 Da) for the quantitative data. Methylthio at C (+45.988) was included for the total proteome data. The Reporter Ion Quantifier node included a TMT 6plex (Thermo Scientific Instruments) Quantification Method, for MS3 scan events, HCD activation type, integration window tolerance 20 ppm and integration method Most Confident Centroid. The consensus workflow included S/N calculation for TMT intensities and the level of confidence for peptide identifications was estimated using the Percolator node with decoy database search. Strict FDR was set at *q*-value < 0.01.

Bioinformatics Analysis. We developed an R package (qPLEXanalyzer) to perform downstream data analysis. All analyses were performed using only unique peptides identified with high confidence (peptide FDR < 1%) across all experiments. Peptide-level signal-to-noise (S/N) TMT values were corrected for equal loading across samples using different normalisation approaches based upon the experiment type. For the regression-based correction, unique peptides were aggregated and proteins identified in all the experiments were kept for further analysis. The normalisation on the bait protein level was carried out at protein level using log₂ row-mean scaled values. To filter-out non-specific proteins, a limma-based differential analysis was performed comparing ER and IgG control samples. In the regression analysis, the ERα profile was used as the independent variable (*x*) and the profile of any other protein as the dependent variable (*y*) excluding the IgG controls. The residuals of the $y = ax + b$ linear model represent the protein quantification profiles that are not driven by ERα amount in the pull-down. More details on the normalisation methods used can be found in Supplementary Note 2. The identification of differentially bound proteins was carried out using the limma-based analysis. A multiple testing correction was applied on *p*-value using the Benjamini-Hochberg method to control the FDR.

Code availability. The qPLEXanalyzer and qPLEXdata R packages are available at (<https://doi.org/10.5281/zenodo.1237825>). Both pipelines are described in detail in Supplementary Notes 1 and 2.

Data availability. RNA-seq data have been deposited in NCBI's Gene Expression Omnibus⁶⁹ and are accessible through GEO Series accession number [GSE104872](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104872). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁷⁰ partner repository with the data set identifier [PXD007968](https://www.ebi.ac.uk/pride/archive/projects/PXD007968). All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

Received: 2 November 2017 Accepted: 3 May 2018

Published online: 13 June 2018

References

- Vidal, M., Cusick, M. E. & Barabasi, A. L. Interactome networks and human disease. *Cell* **144**, 986–998 (2011).
- Ewing, R. M. et al. Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3**, 89 (2007).
- Stelzl, U. et al. A human protein–protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
- Rolland, T. et al. A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).
- Rual, J. F. et al. Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173–1178 (2005).
- Hein, M. Y. et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
- Huttlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).
- Collins, B. C. et al. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat. Methods* **10**, 1246–1253 (2013).
- Liu, F., Rijkers, D. T., Post, H. & Heck, A. J. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12**, 1179–1184 (2015).
- Rafiee, M. R., Girardot, C., Sigismondo, G. & Krijgsvelde, J. Expanding the circuitry of pluripotency by selective isolation of chromatin-associated proteins. *Mol. Cell* **64**, 624–635 (2016).
- Wierer, M. & Mann, M. Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. *Hum. Mol. Genet.* **25**, R106–R114 (2016).
- Engelen, E. et al. Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat. Commun.* **6**, 7155 (2015).
- Mohammed, H. et al. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep.* **3**, 342–349 (2013).
- Mohammed, H. et al. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat. Protoc.* **11**, 316–326 (2016).
- Roumeliotis, T. I. et al. Genomic determinants of protein abundance variation in colorectal cancer cells. *Cell Rep.* **20**, 2201–2214 (2017).
- McAlister, G. C. et al. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal. Chem.* **84**, 7469–7478 (2012).
- McAlister, G. C. et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
- Nowak, D. E., Tian, B. & Brasier, A. R. Two-step cross-linking method for identification of NF- κ B gene network by chromatin immunoprecipitation. *Biotechniques* **39**, 715–725 (2005).
- Chatr-Aryamontri, A. et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **45**, D369–D379 (2017).
- Szklarczyk, D. et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).
- Tsai, W. W. et al. TRIM24 links a non-canonical histone signature to breast cancer. *Nature* **468**, 927–932 (2010).
- Anzick, S. L. et al. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science* **277**, 965–968 (1997).
- Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843–852 (2000).
- Liu, X. F. & Bagchi, M. K. Recruitment of distinct chromatin-modifying complexes by tamoxifen-complexed estrogen receptor at natural target gene promoters in vivo. *J. Biol. Chem.* **279**, 15050–15058 (2004).
- Jepsen, K. et al. Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell* **102**, 753–763 (2000).
- 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).
- Tan, S. K. et al. AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. *EMBO J.* **30**, 2569–2581 (2011).
- Theodorou, V., Stark, R., Menon, S. & Carroll, J. S. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res.* **23**, 12–22 (2013).
- Soderberg, O. et al. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995–1000 (2006).
- Wardell, S. E., Marks, J. R. & McDonnell, D. P. The turnover of estrogen receptor alpha by the selective estrogen receptor degrader (SERD) fulvestrant is a saturable process that is not required for antagonist efficacy. *Biochem. Pharmacol.* **82**, 122–130 (2011).
- Lonard, D. M. & O'Malley, B. W. Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation. *Mol. Cell* **27**, 691–700 (2007).
- Bannister, A. J., Oehler, T., Wilhelm, D., Angel, P. & Kouzarides, T. Stimulation of c-Jun activity by CBP: c-Jun residues Ser63/73 are required for CBP induced stimulation in vivo and CBP binding in vitro. *Oncogene* **11**, 2509–2514 (1995).
- Acevedo, M. L. & Kraus, W. L. Mediator and p300/CBP-steroid receptor coactivator complexes have distinct roles, but function synergistically, during estrogen receptor alpha-dependent transcription with chromatin templates. *Mol. Cell. Biol.* **23**, 335–348 (2003).
- Siersbaek, R. et al. Dynamic rewiring of promoter-anchored chromatin loops during adipocyte differentiation. *Mol. Cell* **66**, 420–435 e425 (2017).
- Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes & Dev.* **20**, 2922–2936 (2006).
- Orphanides, G. & Reinberg, D. RNA polymerase II elongation through chromatin. *Nature* **407**, 471–475 (2000).
- Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11**, 394–403 (2004).
- Frasor, J. et al. Selective estrogen receptor modulators: discrimination of agonistic versus antagonistic activities by gene expression profiling in breast cancer cells. *Cancer Res.* **64**, 1522–1533 (2004).
- Frasor, J. et al. Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. *Endocrinology* **144**, 4562–4574 (2003).
- Frasor, J., Danes, J. M., Funk, C. C. & Katzenellenbogen, B. S. Estrogen down-regulation of the corepressor N-CoR: mechanism and implications for estrogen derepression of N-CoR-regulated genes. *Proc Natl Acad. Sci. USA* **102**, 13153–13157 (2005).
- Caliceti, C. et al. 17beta-estradiol enhances signalling mediated by VEGF-A-delta-like ligand 4-notch1 axis in human endothelial cells. *PLoS ONE* **8**, e71440 (2013).
- May, F. E. & Westley, B. R. Expression of human intestinal trefoil factor in malignant cells and its regulation by oestrogen in breast cancer cells. *J. Pathol.* **182**, 404–413 (1997).
- Ghosh, M. G., Thompson, D. A. & Weigel, R. J. PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. *Cancer Res.* **60**, 6367–6375 (2000).
- McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).
- Metivier, R. et al. Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763 (2003).
- Shiau, A. K. et al. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927–937 (1998).
- Rosell, M. et al. Complex formation and function of estrogen receptor alpha in transcription requires RIP140. *Cancer Res.* **74**, 5469–5479 (2014).
- Lai, A. Y. & Wade, P. A. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat. Rev. Cancer* **11**, 588–596 (2011).
- Chen, J. D. & Evans, R. M. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* **377**, 454 (1995).
- Zhang, B., Chambers, K. J., Faller, D. V. & Wang, S. Reprogramming of the SWI/SNF complex for co-activation or co-repression in prohibitin-mediated estrogen receptor regulation. *Oncogene* **26**, 7153–7157 (2007).
- Belandia, B., Orford, R. L., Hurst, H. C. & Parker, M. G. Targeting of SWI/SNF chromatin remodelling complexes to estrogen-responsive genes. *EMBO J.* **21**, 4094–4103 (2002).
- Wijayarathne, A. L. & McDonnell, D. P. The human estrogen receptor-alpha is a ubiquitinated protein whose stability is affected differentially by agonists,

- antagonists, and selective estrogen receptor modulators. *J. Biol. Chem.* **276**, 35684–35692 (2001).
53. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
 54. Whitfield, M. L. et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell.* **13**, 1977–2000 (2002).
 55. DeRose, Y. S. et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat. Med.* **17**, 1514–1520 (2011).
 56. Zwart, W. et al. Oestrogen receptor-co-factor-chromatin specificity in the transcriptional regulation of breast cancer. *EMBO J.* **30**, 4764–4776 (2011).
 57. Sauve, F. et al. CIA, a novel estrogen receptor coactivator with a bifunctional nuclear receptor interacting determinant. *Mol. Cell. Biol.* **21**, 343–353 (2001).
 58. Rauniyar, N. & Yates, J. R. 3rd Isobaric labeling-based relative quantification in shotgun proteomics. *J. Proteome Res.* **13**, 5293–5309 (2014).
 59. Shang, Y. & Brown, M. Molecular determinants for the tissue specificity of SERMs. *Science* **295**, 2465–2468 (2002).
 60. Eissenberg, J. C. & Elgin, S. C. The HP1 protein family: getting a grip on chromatin. *Curr. Opin. Genet. Dev.* **10**, 204–210 (2000).
 61. Hill, V. K., Kim, J. S. & Waldman, T. Cohesin mutations in human cancer. *Biochim. Biophys. Acta* **1866**, 1–11 (2016).
 62. Katoh, M. & Katoh, M. Identification and characterization of human FOXK1 gene in silico. *Int. J. Mol. Med.* **14**, 127–132 (2004).
 63. Wu, Y. et al. Oncogene FOXK1 enhances invasion of colorectal carcinoma by inducing epithelial-mesenchymal transition. *Oncotarget* **7**, 51150–51162 (2016).
 64. Underhill, C., Qutob, M. S., Yee, S. P. & Torchia, J. A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. *J. Biol. Chem.* **275**, 40463–40470 (2000).
 65. Osborne, C. K. et al. Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *J. Natl. Cancer Inst.* **95**, 353–361 (2003).
 66. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
 67. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
 68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
 69. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
 70. Vizcaino, J. A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, 11033 (2016).

Acknowledgements

The authors would like to thank the Genomics core, the staff in the Proteomics Core especially Valar Nila Roamio Franklin and Carmen Gonzalez Tejedo, the Bioinformatics Core especially Dominique Laurent Couturier and Rory Stark, the Histopathology Core especially Jo Arnold and Bev Wilson, the Biorepository Core and the Instrumentation Core facility at Cancer Research UK. We also would like to thank Igor Chernukhin for uploading the RNA-seq data to GEO. The authors would also like to thank Alana Welm for sharing the PDX models and Elena Provenzano for the evaluation of the IHC

staining. We acknowledge the support of the University of Cambridge and Cancer Research UK. We also would like to acknowledge the National Breast Cancer Foundation of Australia (Grant 14001) and the Personalised Medicine for Breast Cancer Biobank, funded by the Sydney Breast Cancer Foundation, National Health and Medical Research Council (NHMRC), TransBCR Centre for research excellence in translational breast cancer research and the McMurtrie family. The Fusion Lumos Orbitrap mass spectrometer was purchased with the support from a Wellcome Trust Multi-user Equipment Grant (Grant # 108467/Z/15/Z). Parts of this work were funded by CRUK core grant (grant numbers C14303/A17197, A19274) to F.M.; Breast Cancer Now Award (grant number 2012NovPR042) to F.M. R.S. is supported by a fellowship from the Novo Nordisk Foundation (NNF 14136). J.S.C. is supported by an ERC Consolidator award and CRUK funding.

Authors Contributions

E.K.P., R.S., C.S.D.S. and J.S.C., conceptualisation; E.K.P., A.N.H., R.S. and C.S.D.S., methodology; E.K.P., proteomic sample preparation and mass spectrometry; E.K.P., A.N.H. and R.S., RIME experiments; E.K.P., S.O. and R.S., validation assays; E.K.P. and R.S., in vivo experiments; E.K.P., K.K., A.N.H., T.I.R., M.E., R.S. and C.S.R.C., data analysis; K.K. and M.E., development of qPLEXanalyzer; E.K.P., K.K., A.N.H., F.M., M.E., R.S., C.S.D.S. and J.S.C., study design; K.H., K.M.C., A.S. and E.L., collection of human clinical samples and PDX models; R.S., C.S.D.S. and J.S.C., oversaw all experiments; E.K.P., R.S., C.S.D.S. and J.S.C., writing-original draft; All, writing-review, and editing.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-04619-5>.

Competing interests: The authors declare no competing interests.

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

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

RESEARCH ARTICLE

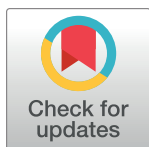
Identification of ChIP-seq and RIME grade antibodies for Estrogen Receptor alpha

Silvia-E. Glont[✉], Evangelia K. Papachristou[✉], Ashley Sawle[✉], Kelly A. Holmes, Jason S. Carroll[✉]*, Rasmus Siersbaek*

Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge, United Kingdom

✉ These authors contributed equally to this work.

* Rasmus.Siersbaek@cruk.cam.ac.uk (RS); Jason.Carroll@cruk.cam.ac.uk (JC)



OPEN ACCESS

Citation: Glont S-E, Papachristou EK, Sawle A, Holmes KA, Carroll JS, Siersbaek R (2019) Identification of ChIP-seq and RIME grade antibodies for Estrogen Receptor alpha. PLoS ONE 14(4): e0215340. <https://doi.org/10.1371/journal.pone.0215340>

Editor: Alessandro Weisz, Universita degli Studi di Salerno, ITALY

Received: February 13, 2019

Accepted: March 29, 2019

Published: April 10, 2019

Copyright: © 2019 Glont et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data is be available from GEO (Accession number GSE128208) or PRIDE (Accession number PXD012930).

Funding: All work was funded by Cancer Research UK. RS is supported by a fellowship from the Novo Nordisk Foundation (NNF15OC0014136). JSC is supported by an ERC Consolidator award. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Estrogen Receptor alpha (ER α) plays a major role in most breast cancers, and it is the target of endocrine therapies used in the clinic as standard of care for women with breast cancer expressing this receptor. The two methods ChIP-seq (chromatin immunoprecipitation coupled with deep sequencing) and RIME (Rapid Immunoprecipitation of Endogenous Proteins) have greatly improved our understanding of ER α function during breast cancer progression and in response to anti-estrogens. A critical component of both ChIP-seq and RIME protocols is the antibody that is used against the bait protein. To date, most of the ChIP-seq and RIME experiments for the study of ER α have been performed using the sc-543 antibody from Santa Cruz Biotechnology. However, this antibody has been discontinued, thereby severely impacting the study of ER α in normal physiology as well as diseases such as breast cancer and ovarian cancer. Here, we compare the sc-543 antibody with other commercially available antibodies, and we show that 06–935 (EMD Millipore) and ab3575 (Abcam) antibodies can successfully replace the sc-543 antibody for ChIP-seq and RIME experiments.

Introduction

In the last decades, there has been significant interest in studying Estrogen Receptor alpha (ER α) due to its causal role in more than three quarters of breast cancers[1]. Its key role in breast cancer progression makes ER α the major target for endocrine therapies, which have substantially improved patient survival. However, resistance to these therapies occurs in many patients[2], which leads to incurable metastatic disease. Therefore, it is important to understand the mechanisms underlying ER α action in cancer initiation as well as progression of the disease. In addition, ER α plays an important role in development[3] and other diseases such as ovarian cancer[4].

Our understanding of ER α -mediated gene transcription has evolved in recent years, due to delineation of ER α -chromatin binding mechanisms through ChIP-seq (chromatin immunoprecipitation followed by next generation sequencing) experiments[5–15]. It is now clear that

Competing interests: Jason Carroll is founder and CSO of Azeria Therapeutics. None of the work in this manuscript is related to the work in Azeria Therapeutics. The authors declare no additional competing interests and the declared funding does not alter our adherence to PLOS ONE policies on sharing data and materials.

differential binding of ER α to chromatin is associated with clinical outcome in primary ER α -positive breast tumours[5], suggesting that changes in ER α binding mediates the altered gene expression program that dictates endocrine responsiveness and clinical outcome. In addition to changes in binding to chromatin, ER α transcriptional activity can be modulated by its association with different co-regulators and other associated transcription factors. Our lab has previously developed a method termed RIME (Rapid Immunoprecipitation of Endogenous Proteins) for the study of protein complexes using mass spectrometry[16, 17]. A key component of ER α ChIP-seq and RIME assays is the antibody that specifically and with high sensitivity targets ER α . Most ChIP-seq and RIME experiments have been performed using the ER α antibody sc-543 from Santa Cruz Biotechnology[5, 9, 17–21]. This antibody has recently been discontinued, impacting the ability to study ER α function in breast cancer as well as in other diseases and physiological conditions. Here, we compare the sc-543 (Santa Cruz Biotechnology) with other commercially available antibodies using breast cancer cells as a model and demonstrate that 06–935 (EMD Millipore) and ab3575 (Abcam) antibodies can replace sc-543 in ChIP-seq and RIME assays.

Materials and methods

Cell culture

MCF7 cells were cultured in Dulbecco's Modified Eagle Medium DMEM (Gibco, Thermo Scientific) and MDA-MB-231 cells were grown in RPMI-1640 medium (Gibco, Thermo Scientific). Both media conditions were supplemented with 10% foetal bovine serum (FBS), 50 U/ml penicillin, 50 μ g/ml streptomycin and 2 mM L-glutamine. Cell lines were obtained from ATCC (Middlesex). For both ChIP-seq and RIME experiments, 2×10^6 cells were seeded in 15 cm² plates and collected at 80–90% confluency.

ChIP-Seq and RIME assays

The sc-543 (Santa Cruz), ab80922 (Abcam), ab3575 (Abcam), sc-514857 (C-3) (Santa Cruz Biotechnology), C15100066 (Diagenode) and 06–935 (EMD Millipore) antibodies were used for ChIP-qPCR. The sc-543, ab3575 and 06–935 antibodies were then used for ChIP-seq and RIME. For each ChIP, 10 μ g of each of the antibodies sc-543, 06–935 and ab3575 or the rabbit IgG ab37415 (Abcam) were used together with 100 μ l of Dynabeads Protein A (Invitrogen). The antibody and the beads were incubated overnight at 4°C with rotation. MCF7 cells were fixed for 10 minutes using 1% formaldehyde (Thermo, #28908) and quenched with 0.1M glycine. Cells were then washed and harvested in ice-cold PBS containing protease inhibitors (Roche). In order to enrich for the nuclear fraction, pellets were resuspended in Lysis Buffer 1 (50mM Hepes–KOH, pH 7.5, 140mM NaCl, 1mM EDTA, 10% Glycerol, 0.5% NP-40/Igepal CA-630, 0.25% Triton X-100) and rotated for 10 minutes, at 4°C. Cells were then pelleted, resuspended in Lysis buffer 2 (10mM Tris–HCl, pH8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA) and incubated for 5 minutes, at 4°C with rotation. For both ChIP-seq and RIME experiments, cells were pelleted, resuspended in 300 μ l Lysis buffer 3 (10mM Tris–HCl, pH 8, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na–Deoxycholate) and sonicated using the Bioruptor Pico sonicator (Diagenode, Liege, Belgium) for 10 cycles (30 seconds on, 30 seconds off). After sonication the samples were centrifuged at maximum speed for 10 minutes at 4°C and a small aliquot of supernatant was kept as input for ChIP-seq. The rest of the supernatant was added to the Protein A Dynabeads, which were incubated overnight with antibody. The next day, the beads for ChIP-seq were washed six times with RIPA buffer (150mM NaCl, 10mM Tris, pH 7.2, 0.1% SDS, 1% Triton X-100, 1% NaDeoxycholate), followed by one wash with TE (pH 7.4). Both ChIP samples and inputs were then de-crosslinked by adding 200 μ l

elution buffer (1% SDS, 0.1 M NaHCO₃) overnight at 65°C. After reverse crosslinking, DNA was purified using the phenol-chloroform-isoamyl DNA extraction method. ChIP-seq and the input libraries were prepared using the ThruPLEX Sample Prep Kit (Illumina). ERα ChIP-seq was performed in at least duplicates for each condition. For RIME, the antibody-bound beads incubated with the chromatin samples were washed 10 times with RIPA buffer and twice with 100mM AMBIC (ammonium bicarbonate) prior to mass spectrometry analysis.

Sample preparation, LC-MS/MS analysis and data processing

A 10μL trypsin solution (15ng/ul) (Pierce) prepared in 100mM AMBIC was added to the beads followed by overnight incubation at 37°C. The next day, trypsin solution was added for a second digestion step followed by incubation for 4h at 37°C. At the end of the second step digestion, the tubes were placed on a magnet and the supernatant solution was collected and acidified by the addition of 2μl 5% formic acid. The peptides were cleaned with the Ultra-Micro C18 Spin Columns (Harvard Apparatus) and were analysed in the Dionex Ultimate 3000 UHPLC system coupled with the Q-Exactive HF (Thermo Scientific) mass spectrometer. Samples were loaded on the Acclaim PepMap 100, 100μm × 2cm C18, 5μm, 100Å trapping column with the ulPickUp injection method at loading flow rate 5μL/min for 10 min. For the peptide separation the EASY-Spray analytical column 75μm × 25cm, C18, 2μm, 100 Å was used for multi-step gradient elution. Mobile phase (A) was composed of 2% acetonitrile, 0.1% formic acid, 5% dimethyl sulfoxide (DMSO) and mobile phase (B) was composed of 80% acetonitrile, 0.1% formic acid, 5% DMSO. The full scan was performed in the Orbitrap in the range of 400-1600m/z at 60K resolution. For MS2, the 10 most intense fragments were selected at resolution 30K. A 2.0Th isolation window was used and the HCD collision energy was set up at 28%. The HCD tandem mass spectra were processed with the SequestHT search engine on Proteome Discoverer 2.2 software. The node for SequestHT included the following parameters: Precursor Mass Tolerance 20ppm, Maximum Missed Cleavages sites 2, Fragment Mass Tolerance 0.02Da and Dynamic Modifications were Oxidation of M (+15.995Da) and Deamidation of N, Q (+0.984Da). The Minora Feature Detector node was used for label-free quantification and the consensus workflow included the Feature Mapper and the Precursor Ion Quantifier nodes using intensity for the precursor quantification. The protein intensities were normalized by the summed intensity separately for the IgG and ERα pull downs (within group normalization). The plots for ERα coverage were created using the qPLEXanalyzer tool[22]. Heatmaps and PCA plot were done with the Phantasus Web tool (<https://artyomovlab.wustl.edu/phantasus/>). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[23] partner repository with the dataset identifier PXD012930.

ChIP-seq data analysis

Reads were mapped to the GRCh38 genome using bwa version 0.7.12[24]. Prior to peak calling, reads were filtered according to four criteria: (1) only reads aligning to canonical chromosomes (1–22, X, Y, MT) were considered for further analysis; (2) read aligning in blacklisted regions were excluded[25]; (3) grey lists were generated using the R package GreyListChIP and reads aligned in these regions were excluded; (4) reads with a mapping quality of less than 15 were excluded. Peak calling was carried out on each ChIP sample with MACS2 version 2.1.1.20160309 using the relevant input sample[26]. Peaks with a q-value < 0.01 were accepted for further analysis. To create tag heatmaps, a consensus peak set was generated using the R package DiffBind[5, 16]. The consensus peak set was composed of any peak that was called in at least two samples. Motif analysis was carried out using AME[27] from the MEME suite

version 4.12.0[28] and the HOCOMOCO Human (v10) motif database[29]. Sequences for motif analysis for each sample were derived by selecting the top 1000 peaks by q-value from the MACS2 peak set and then extracting the genomic sequence 500 bases either side of the peak summits. A detailed description of the pipeline can be found in [S1 File](#). ChIP-seq data have been deposited in NCBI's Gene Expression Omnibus[30] and are accessible through GEO Series accession number GSE128208.

Results and discussion

ChIP-sequencing validates 06–935 and ab3575 as specific ER α antibodies

Given the discontinuation of anti-ER α antibody sc-543, we sought to validate alternatives for immunoprecipitation experiments. We first compared the established sc-543 (Santa Cruz Biotechnology) antibody with ab80922 (Abcam), ab3575 (Abcam), sc-514857 (C-3) (Santa Cruz Biotechnology), C15100066 (Diagenode) and 06–935 (Millipore). For this purpose, we used the ER α positive cell line MCF7 and performed ChIP-qPCR in biological duplicates ([S1 Fig](#)) to assess ER α binding at known target regions ([S1 Table](#)).

The ChIP-qPCR comparison suggested that 06–935 (Millipore) and ab3575 (Abcam) could successfully enrich ER α -bound chromatin at these selected loci and could therefore substitute for sc-543. We performed ChIP-seq to compare these three antibodies in MCF7 cells using IgG as a negative control. ER α ChIP-seq was performed in at least duplicates for each condition, using the same batch of chromatin, to ensure that antibodies could be directly compared. In addition, we included the ER α negative MDA-MB-231 cell line in order to assess non-specific binding by these antibodies. For MDA-MB-231, ChIP-seq was performed in biological triplicates.

We observed 6,031 ER α binding sites for sc-543 (Santa Cruz) antibody, 6,192 peaks for ab3575 (Abcam) and 6,552 for 06–935 (Millipore). Importantly, none of these binding sites were observed in the IgG negative control. The vast majority of sites identified in MCF7 cells by sc-543 overlapped with those detected by ab3575 and 06–935 ([Fig 1A](#)). Consistently, we found a strong correlation between the binding intensities for the three antibodies, which was similar to the correlation between replicates for the same antibody ([Fig 1B](#)). All three antibodies showed robust enrichment at binding sites compared to background and motif analysis identified the ER α response element (ERE) as highly significantly enriched at these sites ([Fig 1C](#)). Importantly, neither of the ab3575 and 06–935 antibodies showed any significant enrichment in the ER α negative cell line MDA-MB-231 ([Fig 1C](#)). In total, one peak was detected in ER-negative cells using ab3575, two peaks for 06–935 and 124 binding sites for sc-543, confirming the specificity of the antibodies. Examples of ER α binding to previously described ER α binding sites[16, 31] are illustrated in [Fig 1D](#). Taken together, this indicates that the ab3575 (Abcam) and 06–935 (Millipore) antibodies perform similarly to the sc-543 (Santa Cruz) antibody in ChIP-seq experiments, both in terms of sensitivity and specificity.

Validation of 06–935 and ab3575 antibodies using RIME

We next sought to evaluate the performance of ab3575 (Abcam) and 06–935 (Millipore) in RIME experiments to directly compare with the sc-543 (Santa Cruz) antibody, which has previously been successfully used in RIME experiments to explore the ER α interactome[9, 16, 22]. To this end, we tested the 06–935, ab3575 and sc-543 antibodies in two technical replicates each using MCF7 cells. IgG controls were also analysed to discriminate specific associations from non-specific interaction events.

To evaluate the pull-down efficiencies, we compared the sequence coverage of the bait protein obtained by the different antibodies. ER α was identified with a similar number of peptides ([Fig 2A](#)) across the three different pull-downs, confirming that all three antibodies achieve

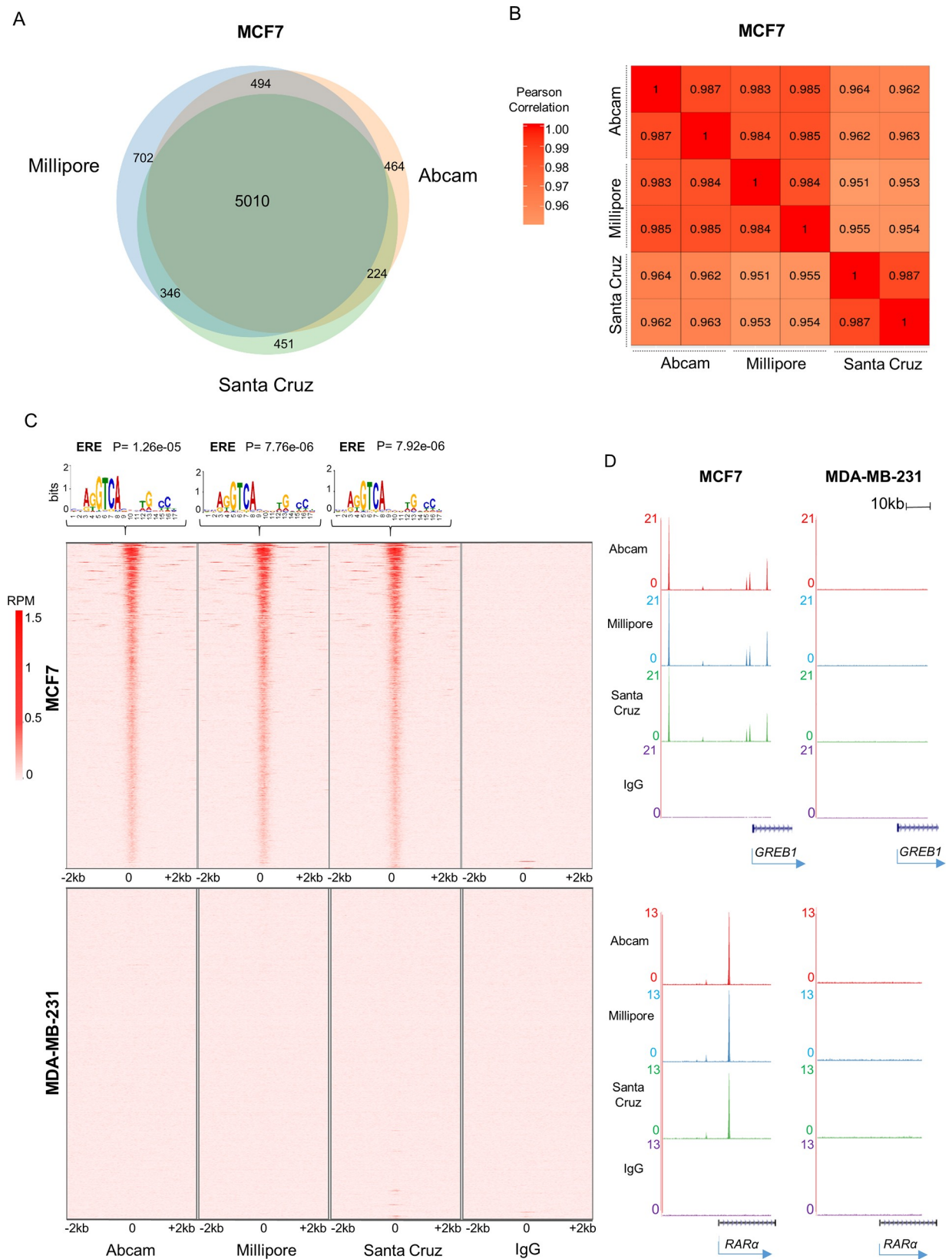


Fig 1. ChIP-seq comparison between Santa Cruz (sc-543), Millipore (06-935) and Abcam (ab3575) antibodies. A) Venn diagram showing the overlap between ER α binding sites for Santa Cruz (sc-543), Millipore (06-935) and Abcam (ab3575) antibodies in MCF7 cells. B) Pearson's correlation between each replicate of all three antibodies in MCF7 cells. C) Top: De novo motif analysis of ER α binding sites using MEME. Bottom: Heatmap of total number of ER α binding sites identified in both technical replicates of MCF7, and in all three biological replicates for MDA-MB-231, respectively. D) Examples of ER α - bound regions. Tag densities are shown as reads per million.

<https://doi.org/10.1371/journal.pone.0215340.g001>

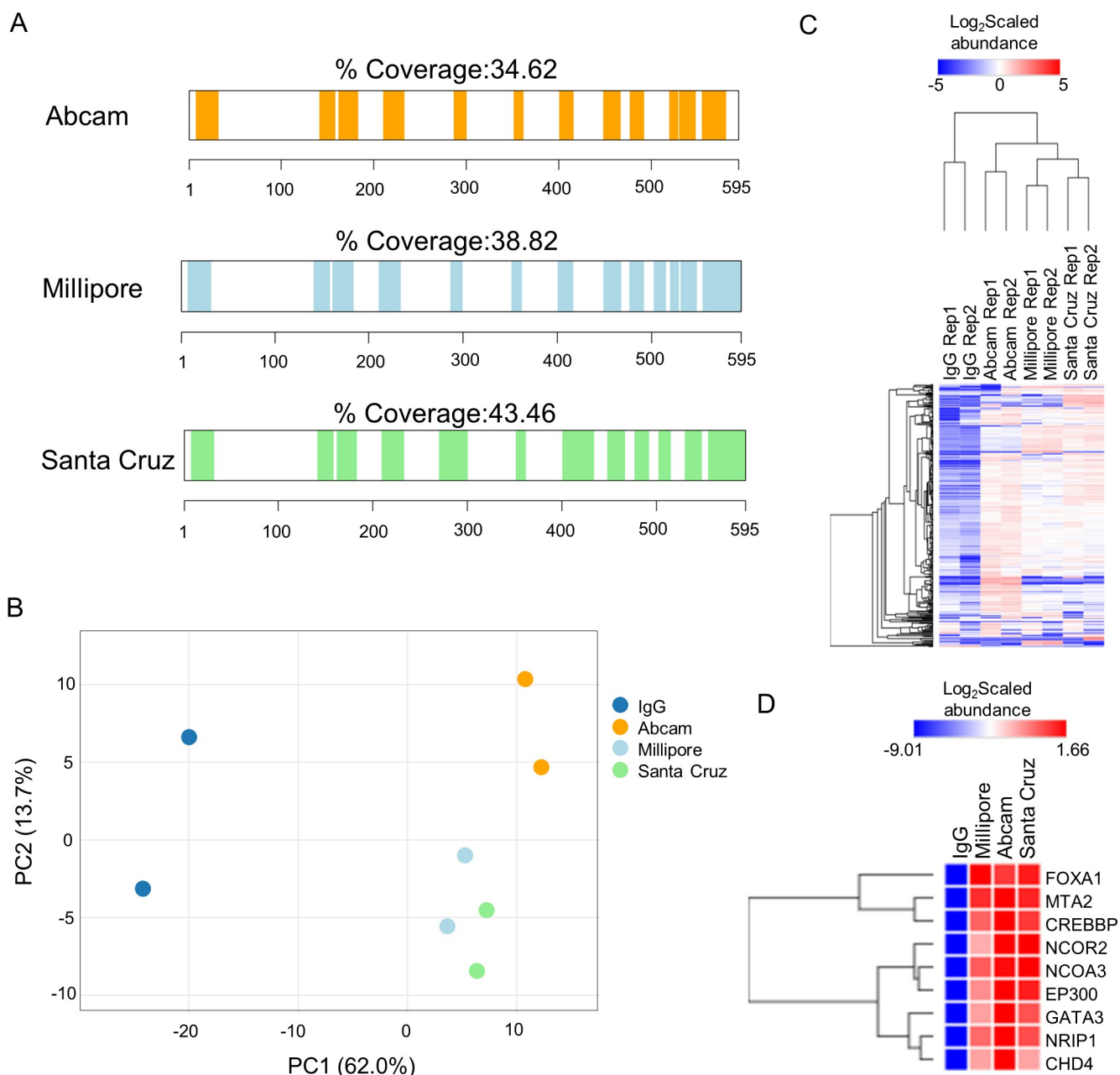


Fig 2. Comparison of RIME data between Santa Cruz (sc-543), Millipore (06-935) and Abcam (ab3575) antibodies. A) Protein sequence coverage of ER α achieved by the use of Abcam (ab3575), Millipore (06-935) and Santa Cruz (sc-543) antibodies in RIME. B) PCA plot of known ER α interactors (n = 319, BIOGRID and STRING databases) for the four different RIME pull-downs. C) Hierarchical clustering of the scaled intensities of known ER α interactors from BIOGRID and STRING databases (n = 319). D) Hierarchical clustering of well-characterized ER α interactors.

<https://doi.org/10.1371/journal.pone.0215340.g002>

efficient immunoprecipitation of the bait protein. Next, to compare the efficiency of the different antibodies to detect known ER α interactors, we used a label-free quantification method based on the Minora algorithm implemented in Proteome Discoverer 2.2 software (S2 File). The PCA plot using intensities of known ER α -associated proteins ($n = 319$, BIOGRID and STRING databases) across all four samples revealed a good separation between the ER α RIME samples and the IgG controls, indicative of high specificity of all antibodies (Fig 2B). Importantly, we identify only minor differences between the three antibodies, suggesting that they all efficiently pull down known ER α -associated proteins (Fig 2B and 2C). Specifically, amongst the known ER α interactors we identified FOXA1, GATA3 and members of the p160 family that were all highly enriched by all three antibodies (Fig 2D). Taken together, the three ER α antibodies perform similarly in RIME experiments, enriching for well-known key ER α interactors.

Conclusions

Genome-wide analyses of ER α -chromatin binding sites using ChIP-based methods have exponentially increased our knowledge of the role of ER α in breast cancer. Most of the published ChIP-seq and RIME studies for ER α have been performed using the sc-543 antibody from Santa Cruz Biotechnology [13, 16, 17, 19–21, 32] and the quality and specificity of sc-543 has made it the ‘golden standard’ for immunoprecipitation experiments. However, this antibody has recently been discontinued, which has significantly impacted our ability to study ER α biology. Here, we have assessed commercially available alternative antibodies. We demonstrate using ChIP-seq and RIME that the two antibodies 06–935 (Millipore) and ab3575 (Abcam) perform similarly to sc-543, in terms of sensitivity and specificity. We therefore propose that these antibodies can replace the sc-543 antibody for immunoprecipitation-based experiments such as ChIP-seq and RIME to explore ER α function.

Supporting information

S1 Fig. ER α antibody comparison by ChIP-qPCR. ChIP-qPCR analysis for ER α known binding sites was performed in MCF7 cells in biological duplicates. Results are shown as arbitrary units. Antibodies used: sc-543 (Santa Cruz Biotechnology), ab80922 (Abcam), ab3575 (Abcam), sc-514857 (C-3) (Santa Cruz Biotechnology), C15100066 (Diagenode) and 6–935 (EMD Millipore).

(TIF)

S1 File. Main steps of the ChIP-seq analysis. The file provides details for the main steps of the Bioinformatic analysis of the ChIP-seq data.

(PDF)

S2 File. Quantitative proteomics analysis results. The file contains the protein intensities across all the different RIME samples based on a label free quantification method using the Minora algorithm in Proteome Discoverer 2.2.

(XLSX)

S1 Table. ChIP-qPCR primers. Table listing the primers used for the ChIP-qPCR experiment.

(DOCX)

Acknowledgments

The authors would like to thank the Genomics core, the head of the Proteomics Core Clive D’Santos and the Bioinformatics Core especially Kamal Kishore.

Author Contributions

Conceptualization: Silvia-E. Glont, Evangelia K. Papachristou, Jason S. Carroll, Rasmus Siersbaek.

Data curation: Ashley Sawle.

Formal analysis: Silvia-E. Glont, Evangelia K. Papachristou, Ashley Sawle.

Funding acquisition: Jason S. Carroll.

Investigation: Silvia-E. Glont, Evangelia K. Papachristou, Kelly A. Holmes, Rasmus Siersbaek.

Project administration: Jason S. Carroll, Rasmus Siersbaek.

Supervision: Jason S. Carroll, Rasmus Siersbaek.

Visualization: Silvia-E. Glont, Evangelia K. Papachristou, Ashley Sawle, Rasmus Siersbaek.

Writing – original draft: Silvia-E. Glont, Evangelia K. Papachristou, Jason S. Carroll, Rasmus Siersbaek.

Writing – review & editing: Silvia-E. Glont, Evangelia K. Papachristou, Ashley Sawle, Kelly A. Holmes, Jason S. Carroll, Rasmus Siersbaek.

References

1. Ali S, Coombes RC. Estrogen receptor alpha in human breast cancer: occurrence and significance. *J Mammary Gland Biol Neoplasia*. 2000; 5(3):271–81. Epub 2004/02/20. PMID: [14973389](#).
2. Ali S, Coombes RC. Endocrine-responsive breast cancer and strategies for combating resistance. *Nat Rev Cancer*. 2002; 2(2):101–12. Epub 2003/03/15. <https://doi.org/10.1038/nrc721> PMID: [12635173](#).
3. Brisken C, O'Malley B. Hormone action in the mammary gland. *Cold Spring Harb Perspect Biol*. 2010; 2(12):a003178. Epub 2010/08/27. <https://doi.org/10.1101/cshperspect.a003178> PMID: [20739412](#); PubMed Central PMCID: PMCPMC2982168.
4. Voutsadakis IA. Hormone Receptors in Serous Ovarian Carcinoma: Prognosis, Pathogenesis, and Treatment Considerations. *Clin Med Insights Oncol*. 2016; 10:17–25. Epub 2016/04/08. <https://doi.org/10.4137/CMO.S32813> PMID: [27053923](#); PubMed Central PMCID: PMCPMC4814131.
5. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012; 481(7381):389–93. Epub 2012/01/06. <https://doi.org/10.1038/nature10730> PMID: [22217937](#); PubMed Central PMCID: PMCPMC3272464.
6. Harrod A, Fulton J, Nguyen VTM, Periyasamy M, Ramos-Garcia L, Lai CF, et al. Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer. *Oncogene*. 2017; 36(16):2286–96. Epub 2016/10/18. <https://doi.org/10.1038/ncr.2016.382> PMID: [27748765](#); PubMed Central PMCID: PMCPMC5245767.
7. Lupien M, Meyer CA, Bailey ST, Eeckhoute J, Cook J, Westerling T, et al. Growth factor stimulation induces a distinct ER(alpha) cistrome underlying breast cancer endocrine resistance. *Genes Dev*. 2010; 24(19):2219–27. Epub 2010/10/05. <https://doi.org/10.1101/gad.1944810> PMID: [20889718](#); PubMed Central PMCID: PMCPMC2947773.
8. Jeselsohn R, Cornwell M, Pun M, Buchwalter G, Nguyen M, Bango C, et al. Embryonic transcription factor SOX9 drives breast cancer endocrine resistance. *Proc Natl Acad Sci U S A*. 2017; 114(22):E4482–E91. Epub 2017/05/17. <https://doi.org/10.1073/pnas.1620993114> PMID: [28507152](#); PubMed Central PMCID: PMCPMC5465894.
9. Mohammed H, Russell IA, Stark R, Rueda OM, Hickey TE, Tarulli GA, et al. Progesterone receptor modulates ERalpha action in breast cancer. *Nature*. 2015; 523(7560):313–7. Epub 2015/07/15. <https://doi.org/10.1038/nature14583> PMID: [26153859](#); PubMed Central PMCID: PMCPMC4650274.
10. Kong SL, Li G, Loh SL, Sung WK, Liu ET. Cellular reprogramming by the conjoint action of ERalpha, FOXA1, and GATA3 to a ligand-inducible growth state. *Mol Syst Biol*. 2011; 7:526. Epub 2011/09/01. <https://doi.org/10.1038/msb.2011.59> PMID: [21878914](#); PubMed Central PMCID: PMCPMC3202798.
11. 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*. 2011; 43(1):27–33. Epub 2010/12/15. <https://doi.org/10.1038/ng.730> PMID: [21151129](#); PubMed Central PMCID: PMCPMC3024537.

12. Theodorou V, Stark R, Menon S, Carroll JS. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res.* 2013; 23(1):12–22. Epub 2012/11/23. <https://doi.org/10.1101/gr.139469.112> PMID: 23172872; PubMed Central PMCID: PMC3530671.
13. Stender JD, Kim K, Charn TH, Komm B, Chang KC, Kraus WL, et al. Genome-wide analysis of estrogen receptor alpha DNA binding and tethering mechanisms identifies Runx1 as a novel tethering factor in receptor-mediated transcriptional activation. *Mol Cell Biol.* 2010; 30(16):3943–55. Epub 2010/06/16. <https://doi.org/10.1128/MCB.00118-10> PMID: 20547749; PubMed Central PMCID: PMC2916448.
14. Tan SK, Lin ZH, Chang CW, Varang V, Chng KR, Pan YF, et al. AP-2gamma regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. *EMBO J.* 2011; 30(13):2569–81. Epub 2011/05/17. <https://doi.org/10.1038/emboj.2011.151> PMID: 21572391; PubMed Central PMCID: PMC3155293.
15. Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods.* 2009; 48(3):240–8. Epub 2009/03/12. <https://doi.org/10.1016/j.ymeth.2009.03.001> PMID: 19275939; PubMed Central PMCID: PMC24052679.
16. Mohammed H, D'Santos C, Serandour AA, Ali HR, Brown GD, Atkins A, et al. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep.* 2013; 3(2):342–9. Epub 2013/02/14. <https://doi.org/10.1016/j.celrep.2013.01.010> PMID: 23403292.
17. Mohammed H, Taylor C, Brown GD, Papachristou EK, Carroll JS, D'Santos CS. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat Protoc.* 2016; 11(2):316–26. Epub 2016/01/23. <https://doi.org/10.1038/nprot.2016.020> PMID: 26797456.
18. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell.* 2005; 122(1):33–43. Epub 2005/07/13. <https://doi.org/10.1016/j.cell.2005.05.008> PMID: 16009131.
19. Fournier M, Bourriquen G, Lamaze FC, Cote MC, Fournier E, Joly-Beauparlant C, et al. FOXA and master transcription factors recruit Mediator and Cohesin to the core transcriptional regulatory circuitry of cancer cells. *Sci Rep.* 2016; 6:34962. Epub 2016/10/16. <https://doi.org/10.1038/srep34962> PMID: 27739523; PubMed Central PMCID: PMC24052679.
20. Jeselsohn R, Bergholz JS, Pun M, Cornwell M, Liu W, Nardone A, et al. Allele-Specific Chromatin Recruitment and Therapeutic Vulnerabilities of ESR1 Activating Mutations. *Cancer Cell.* 2018; 33(2):173–86 e5. Epub 2018/02/14. <https://doi.org/10.1016/j.ccell.2018.01.004> PMID: 29438694; PubMed Central PMCID: PMC24052679.
21. Zwart W, Flach KD, Rudraraju B, Abdel-Fatah TM, Gojis O, Canisius S, et al. SRC3 Phosphorylation at Serine 543 Is a Positive Independent Prognostic Factor in ER-Positive Breast Cancer. *Clin Cancer Res.* 2016; 22(2):479–91. Epub 2015/09/16. <https://doi.org/10.1158/1078-0432.CCR-14-3277> PMID: 26369632.
22. Papachristou EK, Kishore K, Holding AN, Harvey K, Roumeliotis TI, Chilamakuri CSR, et al. A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. *Nat Commun.* 2018; 9(1):2311. Epub 2018/06/15. <https://doi.org/10.1038/s41467-018-04619-5> PMID: 29899353; PubMed Central PMCID: PMC24052679.
23. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 2019; 47(D1):D442–D50. Epub 2018/11/06. <https://doi.org/10.1093/nar/gky1106> PMID: 30395289; PubMed Central PMCID: PMC24052679.
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25(14):1754–60. Epub 2009/05/20. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168; PubMed Central PMCID: PMC24052679.
25. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489(7414):57–74. Epub 2012/09/08. <https://doi.org/10.1038/nature11247> PMID: 22955616; PubMed Central PMCID: PMC24052679.
26. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008; 9(9):R137. Epub 2008/09/19. <https://doi.org/10.1186/gb-2008-9-9-r137> PMID: 18798982; PubMed Central PMCID: PMC24052679.
27. McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics.* 2010; 11:165. Epub 2010/04/02. <https://doi.org/10.1186/1471-2105-11-165> PMID: 20356413; PubMed Central PMCID: PMC24052679.
28. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009; 37(Web Server issue):W202–8. Epub 2009/05/22. <https://doi.org/10.1093/nar/gkp335> PMID: 19458158; PubMed Central PMCID: PMC24052679.

29. Kulakovskiy IV, Vorontsov IE, Yevshin IS, Soboleva AV, Kasianov AS, Ashoor H, et al. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res.* 2016; 44(D1):D116–25. Epub 2015/11/21. <https://doi.org/10.1093/nar/gkv1249> PMID: 26586801; PubMed Central PMCID: PMC4702883.
30. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2013; 41(Database issue):D991–5. Epub 2012/11/30. <https://doi.org/10.1093/nar/gks1193> PMID: 23193258; PubMed Central PMCID: PMC3531084.
31. Ross-Innes CS, Stark R, Holmes KA, Schmidt D, Spyrou C, Russell R, et al. Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. *Genes Dev.* 2010; 24(2):171–82. Epub 2010/01/19. <https://doi.org/10.1101/gad.552910> PMID: 20080953; PubMed Central PMCID: PMC2807352.
32. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoutte J, et al. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet.* 2006; 38(11):1289–97. Epub 2006/10/03. <https://doi.org/10.1038/ng1901> PMID: 17013392.