

INSTITUTE FOR WOMENS HEALTH

Department of Women's Cancer

Cancer Proteomics Group



**UCL**

# **Functional Characterisation of Copine III (CPNE3) in ERBB2 Overexpressing Breast Cancer**

**Tendayi Samuriwo**

Department of Women's Cancer

University College London

PhD in Cancer Proteomics and Mass Spectrometry

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

from University College London

## Declaration

I, Tendayi Samuriwo confirm that the work presented in this doctoral thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis and all work done in collaboration has been declared in the Declaration and specified in the text.

It is not substantially the same as any that I have submitted for any qualification at University College London or any other University or similar institution except as declared in the Declaration and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any qualifications at University College London or any other University or similar institution except as declared in the Declaration and specified in the text.

It does not exceed the prescribed word limit of 80,000 words exclusive of tables, references and appendices.

### List of Collaborations

Real time cell adhesion assays for the electrochemical impedance xCELLigence RTCA DP cell adhesion experiments were performed at the ImpedanCELL facility at Laboratoire BioTICLA, Centre François Baclesse.

Signature:

A solid black rectangular box used to redact the signature of the author.

Tendayi Samuriwo

September 2022

## Abstract

**BACKGROUND:** ERBB2/HER2 is amplified in a significant proportion of human breast cancers, where it is correlated to poor prognosis for the patient and therapeutic resistance. However, little is known about the underlying mechanisms and their effect on tumour progression. In the ongoing attempt to elucidate the downstream signalling mechanisms of ERBB2 overexpressing breast cancer, a poorly characterised calcium-binding protein called Copine III (CPNE3) has been identified as potentially associated with ERBB2-dependent transformation.

**METHODS:** Two mass spectrometry-based proteomic methods, tandem mass tagging (TMT) LC-MS/MS and label-free LC-MS/IMS/MS were implemented to discover candidate biomarkers regulated by siRNA-mediated gene knockdown of CPNE3 in ERBB2 overexpressing HMLECs. A combination of real-time cell adhesion assays, mass spectrometry based proteomic workflows, statistical analysis, biological network construction, causal and functional enrichment analysis revealed the functional role of these CPNE3 regulated candidate biomarkers. Novel clinical biomarkers of HER2-positive status were confirmed by mapping candidate biomarkers to data from two breast cancer patient cohorts using Pearson correlation and potential signalling mechanisms were identified by evaluating phosphopeptide enrichment.

**RESULTS:** The downregulation of adhesion related proteins ITGA6 and ITGB4 was shown to correlate with the overexpression of CPNE3 in ERBB2 over-expressing HMLECs. However, the knockdown of both CPNE3 and ERBB2 did not reverse the expression pattern. Cell adhesion assays demonstrated that ERBB2 over-expressing HMLECs adhere to an adherent surface more readily than parental HMLECs and mass spectrometry-based proteomic profiling, PCA and k-means clustering revealed a link between CPNE3 and components of ribonucleoprotein complexes that form in the early stages of cell spreading. In addition, CPNE3 expression regulates several proteins such as CANX, CS, HIST1H4A and PYGM that are downstream effectors or targets of OXPHOS.

**CONCLUSION:** CPNE3 is proposed as a marker for adaptive mechanosensing related metabolic reprogramming. Moreover, suggesting a role for CPNE3 in glucose homeostasis of breast cancer during malignant transformation.

This work is dedicated to my most loving mother, Margaret Ntombizodwa Samuriwo who passed on  
Monday 19<sup>th</sup> July 2021, at the age of 75 and now rests with Our Lord.



## Acknowledgements

Primarily, I would like to thank my initial primary supervisor, the late Dr John Timms for giving me the opportunity to carry out the work presented in this thesis. It would not have been possible in its entirety without his guidance and fellowship during the difficult times. Above all, I would like to thank him for his support and patience during the early days of my project up to the period of his battle with mesothelioma. It was a difficult time and I could not have achieved the eventual completion of this work without his assistance and fellowship.

I would like to thank Prof Alexey Zaikin and Dr Harry Whitwell for their invaluable guidance and support during all stages of my research. Alexey as my initial secondary supervisor and eventual primary supervisor following the passing of John and to Harry for his fellowship during the initial stages of my research and eventual assumption of the role of secondary supervisor for my project. Furthermore, a debt of gratitude to Alexey for providing an opportunity to discuss machine learning and systems biology during my research and to Harry for sharing his immense knowledge of Proteomics and providing access to alternative Mass Spectrometry facilities following the passing of John. I would also like to thank Dr Anna Kleyman and Richard Gunu for their occasional assistance around the laboratory and my colleagues in the Cancer Proteomics Group, Anna Kazarian, Emily Davies Vittersø, Harvey Johnston, Joy Cuenco, Nuno Rocha Nene, Oleg Blyuss and Stella Irungu, for their fellowship and wonderful memories. I am truly grateful for the invaluable support received from the Institute for Women's Health and Department for Women's Cancer, Prof Anne Lanceley, Dr Sioban Sen Gupta, Ms Marcia Jacks and Dr Angela Poulter truly kept my project on course within their respective roles. Especially during the most difficult periods following the loss of John and my mother.

Finally, I would like to thank my parents, Margaret and Emmanuel for the support and understanding throughout my PhD studies. It's a pity we won't celebrate the end of this journey with you dear mother but suffice it to say that my academic journey and life are a testament to your motherly fortitude.

## Table of Contents

<b>Title page</b> .....	1
<b>Signed declaration</b> .....	2
<b>Abstract</b> .....	3
<b>Acknowledgements</b> .....	5
<b>Table of contents</b> .....	6
<b>List of figures</b> .....	10
<b>List of tables</b> .....	13
<b>Abbreviations</b> .....	17
<b>1 Introduction</b> .....	25
1.1 Breast cancer epidemiology .....	25
1.1.1 Statistics .....	25
1.1.2 Risk factors .....	25
1.1.3 Malignancy and breast neoplasms .....	26
1.2 The ErbB receptor family .....	28
1.2.1 Receptor tyrosine kinases .....	28
1.2.2 Origin, structure and ligands .....	29
1.2.3 Signal transduction .....	31
1.2.4 Signal attenuation .....	33
1.2.5 ErbB expression in the mammary gland .....	34
1.3 ERBB2 over-expressing cell systems .....	35
1.4 Passive in vitro cell adhesion assays .....	36
1.4.1 End point cell adhesion assays .....	38
1.4.2 Real time cell adhesion assays .....	39
1.5 Proteomics .....	40
1.5.1 Proteomic workflows .....	40
1.5.2 High-performance/pressure liquid chromatography (HPLC) .....	41
1.5.3 Mass spectrometry .....	41
1.5.3.1 Soft ionization of analytes .....	42

1.5.3.2	Linear trap quadripole mass analyzers (LTQ).....	43
1.5.3.3	Orbitrap mass analyser.....	44
1.5.3.4	Ion mobility spectrometry (IMS) .....	44
1.5.3.5	Time of flight mass spectrometry (TOF MS) .....	45
1.5.4	Quantitative proteomics .....	45
1.6	Background research .....	48
1.7	Candidates .....	53
1.7.1	ARHGDIB .....	53
1.7.2	CACYBP.....	55
1.7.3	CANX .....	56
1.7.4	CPNE3.....	56
1.7.5	CS .....	58
1.7.6	EIF5A .....	58
1.7.7	HIST1H4A .....	59
1.7.8	ITGA6.....	60
1.7.9	ITGB4.....	61
1.7.10	KPNA2.....	62
1.7.11	PYGM.....	62
1.7.12	SSRP1.....	63
1.8	Impact statement.....	64
<b>2</b>	<b>Materials and methods .....</b>	<b>65</b>
2.1	Cell culture .....	65
2.2	Small interfering RNA (siRNA) reverse transfection .....	65
2.3	Protein extraction and sample preparation .....	66
2.4	Western blotting .....	66
2.5	Adhesion assays .....	67
2.6	De-adhesion assays.....	67
2.7	Real-time cell adhesion or spreading assays.....	68
2.8	TMT tagging and HPLC fractionation.....	68

2.9	Mass spectrometry (LC-MS/MS) and protein identification .....	69
2.10	Label-Free sample preparation .....	71
2.11	Mass spectrometry (LC-MS/IMS/MS) and protein identification .....	72
2.12	Dimensionality reduction by PCA and k-means clustering.....	73
2.13	Clinical data.....	75
2.14	Creating partial compendia of clinical data .....	75
2.15	Statistical analysis and network construction .....	76
2.16	Functional enrichment analysis.....	76
2.17	Causal analysis .....	76
2.18	Protein interaction analysis .....	77
2.19	Bioinformatics analysis of phosphoproteomic data .....	77
2.20	Logistic regression analysis of selected candidate biomarkers .....	78
<b>3</b>	<b>Evaluating the effect of gene expression on HMLEC phenotype.....</b>	<b>79</b>
3.1	Chapter Introduction.....	79
3.2	The effect of ERBB2 expression on HMLEC cell phenotype .....	79
3.2.1	siRNA-dependent knockdown of candidate proteins.....	79
3.2.2	Assessing the effect of ERBB2 overexpression using end point assays .....	80
3.3	The effect of CPNE3 expression on HMLEC cell phenotype .....	83
3.3.1	siRNA-dependent knockdown of CPNE3 expression.....	83
3.3.2	Assessing the effect of CPNE3 and ERBB2 expression using a real time cell adhesion and spreading assay .....	84
3.4	Mass Spectrometry based proteomic profiling .....	86
3.4.1	TMT-LC-MS/MS based protein quantification .....	86
3.4.2	Label-Free protein quantification.....	92
3.5	Cluster-based method for the evaluation of mass spectrometry data.....	96
3.5.1	Principal component analysis of mass spectrometry data.....	96
3.5.2	K-Means clustering of mass spectrometry data .....	99
3.6	Chapter Conclusion.....	105
<b>4</b>	<b>Clinical validation of CPNE3-related gene expression.....</b>	<b>108</b>

4.1	Chapter Introduction.....	108
4.2	Discovery of HER2 candidate biomarkers in invasive ductal carcinoma (IDC) patients.....	108
4.3	Independent verification of candidate biomarkers.....	122
4.4	ErbB2 specific phosphosignalling in breast cancer patients .....	139
4.5	Developing predictive models for HER2-positive breast cancer .....	143
4.5.1	Model 1: A predictive model of HER2 positive status.....	143
4.5.2	Model 2: A predictive model of HER2-positive and HER2 enriched status.....	148
4.6	Chapter Conclusion.....	154
<b>5</b>	<b>Discussion</b> .....	155
5.1	Chapter Introduction.....	155
5.2	Validating the role of CPNE3 in cell adhesion.....	157
5.3	Characterisation of CPNE3 mediated cell signalling.....	160
5.4	Combining multiple biomarkers in logistic regression models.....	164
5.5	Future prospects .....	165
5.6	Conclusions.....	166
<b>6</b>	<b>References</b> .....	168
<b>7</b>	<b>Appendix</b> .....	195
7.1	Mass spectrometry data tables .....	195
7.2	Supplementary data tables and code files .....	207

## List of Figures

Figure 1.1.1: Female breast morphology and breast malignancies .....	26
Figure 1.2.1: Phosphorylation of receptor tyrosine kinase (RTK) .....	28
Figure 1.2.2: Structural representation of a typical ERBB receptor. ....	31
Figure 1.2.3: Overview of ERBB signalling .....	33
Figure 1.4: Stages of passive <i>in vitro</i> cell adhesion.....	37
Figure 1.5.1: Mass spectrometry principle .....	42
Figure 1.5.2: LTQ Orbitrap Mass Spectrometer .....	43
Figure 1.5.3: Schematic Depictions of SRM, DDA and DIA Analyses.....	47
Figure 1.7.1: Biological features of biomarker candidates for ERBB2 overexpressing breast cancer .....	54
Figure 1.7.5: Expression of CPNE3 correlates with ERBB2 and ERBB3 status in a panel of HMLECs and breast tumour lines.....	57
Figure 2.11.1: Schematic representation of the mass spectrometer processes for various acquisition modes available on Synapt G2-Si mass spectrometer .....	73
Figure 3.2.1: Western blot confirmation of target expression of candidates in non-transfected cells (NTC), non-targeting siControl transfected cells and siRNA-targeted cells.....	80
Figure 3.2.2: De-adhesion assay .....	81
Figure 3.2.3: Adhesion assay .....	82
Figure 3.3.1: Western blot analysis of the expression of ITGA6 and ITGB4 in response to CPNE3 knockdown in C3.6 cells .....	83

Figure 3.3.2: RTCA assay analysis of real time adhesion and spreading of a layer of cells on an adherent surface .....	84
Figure 3.3.3: Duration of cell spreading period for HMLECs. ....	85
Figure 3.5.1: Explained variance of mass spectrometry data.....	96
Figure 3.5.2: PCA of TMT mass spectrometry data .....	97
Figure 3.5.3: PCA of Label Free mass spectrometry data.....	98
Figure 3.5.4: Optimal number of clusters as determined by elbow method.....	100
Figure 3.5.5: Visualisation of CPNE3 containing clusters.....	101
Figure 4.2.1: Pair plot of ERBB2 vs CPNE3 expression for ERPR, HER2 and TN for 40 female breast cancer patients.....	109
Figure 4.2.2: Heatmap of Pearson correlations for 4 protein candidates in the discovery cohort.....	111
Figure 4.2.3: Heatmap of subtype Pearson correlations for 36 proteins of interest in the discovery cohort.....	114
Figure 4.2.4: Heatmap of TNM staging Pearson correlations for 36 proteins of interest in the discovery cohort.....	118
Figure 4.2.5: Cytoscape network analysis of proteins associated with HER2 positive status in the discovery cohort.....	120
Figure 4.3.1: Pair plot of ERBB2 vs CPNE3 expression for Basal-like, HER2 positive, Luminal A and B subtypes of 75 female breast cancer patients.....	122
Figure 4.3.2: Heatmap of Pearson correlations for 4 protein candidates in the validation cohort.....	125
Figure 4.3.3: Heatmap of subtype Pearson correlations for 45 proteins of interest in the validation cohort.....	126

Figure 4.3.4: Heatmap of TNM staging Pearson correlations for 45 proteins of interest in the validation cohort.....	129
Figure 4.3.5: Cytoscape network analysis of proteins associated with HER2 positive status in the validation cohort.....	130
Figure 4.3.6: Cytoscape network analysis of proteins associated with HER2 positive Stage IIA in the validation cohort .....	133
Figure 4.3.7: Cytoscape network analysis of proteins associated with HER2 positive Stage IIB/IIIA in the validation cohort. ....	136
Figure 4.5.1: Pair plot of candidate biomarker combinations correlated with HER2 positive status in the discovery cohort (n=40).....	145
Figure 4.5.2: Pair plot of candidate biomarker combinations correlated with HER2 positive status in the validation cohort (n=75).....	146
Figure 4.5.3: Receiver operator curve analysis of multivariate logistic regression model 1 with individual or combined proteins as predictor variables.....	147
Figure 4.5.4: Pair plot of candidate biomarkers correlated with both HER2 positive and HER2-enriched status in the discovery cohort (n=40) .....	149
Figure 4.5.5: Pair plot of candidate biomarker combinations correlated with both HER2 positive and HER2-enriched status in the validation cohort (n=75). ....	150
Figure 4.5.6: Receiver operator curve analysis of multivariate logistic regression model 2 with individual or combined proteins as predictor variables.....	151



## List of Tables

Table 1.6.1: Table of selected gene product changes of interest in AGR2, CPNE3, ERBB2 and STARD10 siRNA knockdowns versus control siRNA from SILAC LC-MS/MS profiling of SKBR3 cells.....	50
Table 1.6.2: Table of differentially regulated proteins (>2 fold) in ERBB2-overexpressing human mammary luminal epithelial (HMLEC) cells (C3.6) versus parental HMLEC cells (HB4a) determined by SILAC LC-MS/MS profiling .....	51
Table 2.4: The table contains a list of antibodies used for western blotting and their corresponding dilutions.....	67
Table 3.4.2: Candidates displaying similar patterns of expression in response to CPNE3 knockdown in C3.6 and SKBR3 cells .....	88
Table 3.4.3: Top 10 enriched categories based on an analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown using the Panther Database .....	89
Table 3.4.4: Top 10 enriched categories based on an analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown using the Reactome Database .....	90
Table 3.4.5: Proteins with a causal relevance to breast cancer or the cell adhesion phenotype based on a Qiagen Ingenuity Pathway Analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown.....	91
Table 3.4.7: Top 10 enriched categories based on an analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown using the Panther Database .....	93
Table 3.4.8: Top 10 enriched categories based on an analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown using the Reactome Database .....	94
Table 3.4.9: Proteins with a causal relevance to breast cancer or the cell adhesion phenotype based on a Qiagen Ingenuity Pathway Analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown.....	95
Table 3.5.2: Top 10 enriched categories based on an analysis of proteins with a strong positive correlation to CPNE3 using K-Means clustering on TMT data and the Panther database .....	102

Table 3.5.3: Top 10 enriched categories based on an analysis of proteins with a strong positive correlation to CPNE3 using K-Means clustering on TMT data and the Reactome database .....	103
Table 3.5.5: Top 10 enriched categories based on an analysis of proteins with a strong positive correlation to CPNE3 using K-Means clustering on Label-Free data and the Panther Database .....	104
Table 3.5.6: Top 10 enriched categories based on an analysis of proteins with a strong positive correlation to CPNE3 using K-Means clustering on Label-Free data and the Reactome database .....	105
Table 4.2.1: Characteristics of female breast cancer patients in the discovery and validation cohorts .....	110
Table 4.2.2: The table lists the total number of candidate biomarkers from the discovery cohort .....	119
Table 4.2.3: The table lists the top 10 enriched KEGG Pathways based on an analysis of cluster 1 of the network of proteins associated with HER2 positive status in the discovery cohort .....	121
Table 4.2.4: The table lists the top 10 enriched Reactome Pathways based on an analysis of cluster 1 of the network of proteins associated with HER2 positive status in the discovery cohort .....	121
Table 4.3.1: The table lists the top 10 enriched KEGG Pathways based on an analysis of cluster 1 of the network of proteins associated with HER2 positive status in the validation cohort .....	131
Table 4.3.2: The table lists the top 10 enriched Reactome Pathways based on an analysis of cluster 1 of the network of proteins associated with HER2 positive status in the validation cohort .....	131
Table 4.3.3: The table lists the top 10 enriched KEGG Pathways based on an analysis of cluster 1 of the network of proteins associated with HER2 positive Stage IIA in the validation cohort .....	134

Table 4.3.4: The table lists the top 10 enriched Reactome Pathways based on an analysis of cluster 1 of the network of proteins associated with HER2 positive Stage IIA in the validation cohort.....	134
Table 4.3.5: The table lists the top 10 enriched KEGG Pathways based on an analysis of cluster 1 of the network of proteins associated with HER2 positive Stage IIB/IIIA in the validation cohort.....	137
Table 4.3.6: The table lists the top 10 enriched Reactome Pathways based on an analysis of cluster 1 of the network of proteins associated with HER2 positive Stage IIB/IIIA in the validation cohort.....	137
Table 4.3.7: The table lists the STRING interactions with a confidence score cut-off $\geq 0.90$ (high confidence) for significantly up/down-regulated proteins.....	138
Table 4.4.1: ERBB2-related phosphorylation in the validation cohort. 1pST, 2pST and 3pST refer to singly, doubly and triply phosphorylated peptide ratios for the indicated sites.....	140
Table 4.5.1: The table lists candidate biomarkers correlated with HER2 positive/HER2 enriched status .....	148
Table 4.5.2: Summary of logistic regression values for candidate biomarkers predicting HER2 status.....	153
Appendix Table 3.4.1: 52 proteins were identified as up/down regulated in response to CPNE3 knockdown.....	196
Appendix Table 3.4.6: 40 proteins identified as up/down regulated in response to CPNE3 knockdown.....	199
Appendix Table 3.5.1: 52 proteins identified by PCA and K-means clustering of TMT LC-MS/MS data. ....	201
Appendix Table 3.5.4: 44 proteins identified by PCA and K-means clustering of Label Free LC-MS/MS data.. ....	204
Appendix 1.1..... (Passive_in_vitro_cell_adhesion_assays.xlsx)	
Appendix 1.2..... (TMT_CPNE3_knockdown_protein_groups.xlsx)	

Appendix 1.3.....	(Label_free_CPNE3_knockdown_protein_groups.xlsx)
Appendix 1.4.....	(Enriched_pathways_TMT_CPNE3_knockdown.xlsx)
Appendix 1.5.....	(Enriched_pathways_label_free_CPNE3_knockdown.xlsx)
Appendix 1.6.....	(PCA_k-means_clustering.xlsx)
Appendix 2.1.....	(Partial_compendia_clinical_data.xlsx)
Appendix 2.2.....	(Statistical_analysis_and_network_construction.xlsx)
Appendix 2.3.....	(Enriched_pathways_clinical_networks.xlsx)
Appendix 2.4.....	(Protein_interaction_analysis.xlsx)
Appendix 2.5.....	(Phosphoproteome_enrichment_analysis.xlsx)
Appendix 3.1.....	(Model1_discovery_set_logistic_regression.ipynb)
Appendix 3.2.....	(Model1_validation_set_logistic_regression.ipynb)
Appendix 3.3.....	(Model2_discovery_set_logistic_regression.ipynb)
Appendix 3.4.....	(Model2_validation_set_logistic_regression.ipynb)

## Abbreviations

### Chemicals and reagents

ACN	Acetonitrile
BCA	Bicinchoninic acid
DMSO	Dimethyl sulfoxide
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
FA	Formic acid
FCS	Foetal calf serum
H <sub>2</sub> O	Water
HCL	Hydrogen Chloride
HEPES	N-[2-hydroxyethyl] piperazine-N'[2-ethanesulphonic acid]
MeOH	Methanol
MTT	3-(4, 5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NaCl	Sodium chloride
NP40	Nonidet P40
PBS	Phosphate buffered saline
PVDF	Polyvinylidene difluoride
SDS	Sodium dodecyl sulphate
TBS-T	Tris buffered saline with Tween 20
TEAB	Tetraethylammonium bromide
TFA	Trifluoroacetic acid
Tris	Tris-(hydroxymethyl) aminomethane

## General

AUC	Area under curve
2D-DIGE	2D difference gel electrophoresis
Ca <sup>2+</sup>	Calcium
CCS	Collision cross section
cDNA	Complementary DNA
CI	Cell index
CID	Collision-induced dissociation
CPTAC	Clinical proteomic tumour analysis consortium
DC	Direct current
DCIS	Ductal carcinoma in situ
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-mesenchymal transition
ER	Oestrogen receptor
ESI	Electrospray ionisation
FDR	False discovery rate
FFT	Fast Fourier-transform algorithm
FNA	Fine needle aspiration
HCD	Higher energy C-trap dissociation
HDDDA	High-definition data dependent acquisition
HDMS	High-definition mass spectrometry

HDMS <sup>E</sup>	MALDI Imaging High Definition MSE
HILIC	Hydrophilic interaction chromatography
HMLEC	Human mammary luminal epithelial cells
HPLC	High-performance/pressure liquid chromatography
ID	Intrinsic dimension
IDC	Invasive ductal carcinoma
IMS	Ion mobility spectrometry
IPA	Ingenuity pathway analysis
iTRAQ	Isobaric tags for relative and absolute quantification
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LCIS	Lobular carcinoma in situ
LOOCV	Leave-one-out-cross-validation
LTQ	Linear trap quadrupole
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionisation
MF	Microfilament
MMTV-LTR	Mouse mammary tumour virus long terminal repeat
MRM	Multiple reaction monitoring
MS	Mass spectrometry/Full ion scan
MS <sup>E</sup>	High and low collision energy mass spectrometry
MS/MS	Tandem mass spectrometry/Fragmentation ion scan
NPC	Nuclear pore complex
NTC	Non transfected cells

ORA	Over-representation analysis
OXPHOS	Oxidative phosphorylation
PAGE	Polyacrylamide gel electrophoresis
PCA	Principal component analysis
PR	Progesterone receptor
PTB	Phosphotyrosine-binding domain
PTM	Post translational modification
qRT-PCR	Quantitative real-time polymerase chain reaction
RF	Radiofrequency
RMSE	Root mean squared error
ROC	Receiver operating characteristic
RPLC	Reversed-phase liquid chromatography
RT-PCR	Reverse transcription polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIC	Spreading initiation centre
siCtrl	Non-targeting scrambled control siRNA
SILAC	Stable isotope labelling by amino acids in cell culture
siRNA	Small interfering RNA
SRM	Selected reaction monitoring
SSE	Sum of squared error
SV40	Simian virus 40
TCA	Tricarboxylic acid
TCGA	The cancer genome atlas



TME	Tumour microenvironment
TMT	Tandem mass tags
TNM	Tissue, node, metastasis
TNR	True negative rate
TOF MS	Time of flight mass spectrometry
TPR	True positive rate
TN	Triple negative
WEBGESTALT	WEB-based Gene SeT AnaLysis Toolkit

## Genes, proteins, enzymes, inhibitors and growth factors

AEBSF	4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride
AGR2	Anterior gradient protein 2 homolog
Akt/PKB	Protein kinase B
AR	Androgen receptor
ARHGDIB	Rho GDP-dissociation inhibitor 2
BRCA1	Breast cancer susceptibility genes 1
BSA	Bovine serum albumin
BTC	Betacellulin
CacyBP	Calcyclin-binding protein
CANX	Calnexin
CDK1	Cyclin-dependent kinase 1
CK2	Casein kinase 2
CPNE3	Copine III
CS	Citrate synthase, mitochondrial
CTNNB1	Catenin beta-1
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
eIF5A	Eukaryotic translation initiation factor 5A-1
EPR	Epiregulin
ER	Oestrogen receptor
ErbB2/Her2	Receptor tyrosine-protein kinase erbB-2
ErbB3	Receptor tyrosine-protein kinase erbB-3

ErbB4	Receptor tyrosine-protein kinase erbB-4
ERK	Extracellular signal-Regulated Kinase
FAK	Focal adhesion kinase
Grb2	Growth factor receptor bound protein-2
HB-EGF	Heparin-binding EGF
HIST1H4A	Histone H4
HRG	Heregulin $\beta$ 1
HRP	Horseradish peroxidase
IFN	Interferon
ITGA6	Integrin alpha 6
ITGB4	Integrin beta 4
KPNA2	Importin subunit alpha-1
LET-23	Lethal complementation group 23
Lin-3	Abnormal cell lineage 3
MAPK	Mitogen activated protein kinases
NDRG1	Protein NDRG1
NRG	Neuregulin
NRG1	Neuregulin 1
NRG2	Neuregulin 2
NRG3	Neuregulin 3
NRG4	Neuregulin 4
PAK2	p21-activated kinase 2
PI3K	Phosphatidylinositol 3 kinase

PR	Progesterone receptor
PTB	Phosphotyrosine-binding domain
PYGM	Glycogen phosphorylase
RAGE	Receptor of advanced glycation end products
RAS	Ras GTPase
RTK	Receptor tyrosine kinase
SH2	Src homology-2 domain
Sos	Son of sevenless
SSRP1	Structure specific recognition protein
STARD10	PCTP-like protein
P53	Cellular tumour antigen p53

# Chapter 1

## 1. Introduction

### 1.1 Breast cancer epidemiology

#### 1.1.1 Statistics

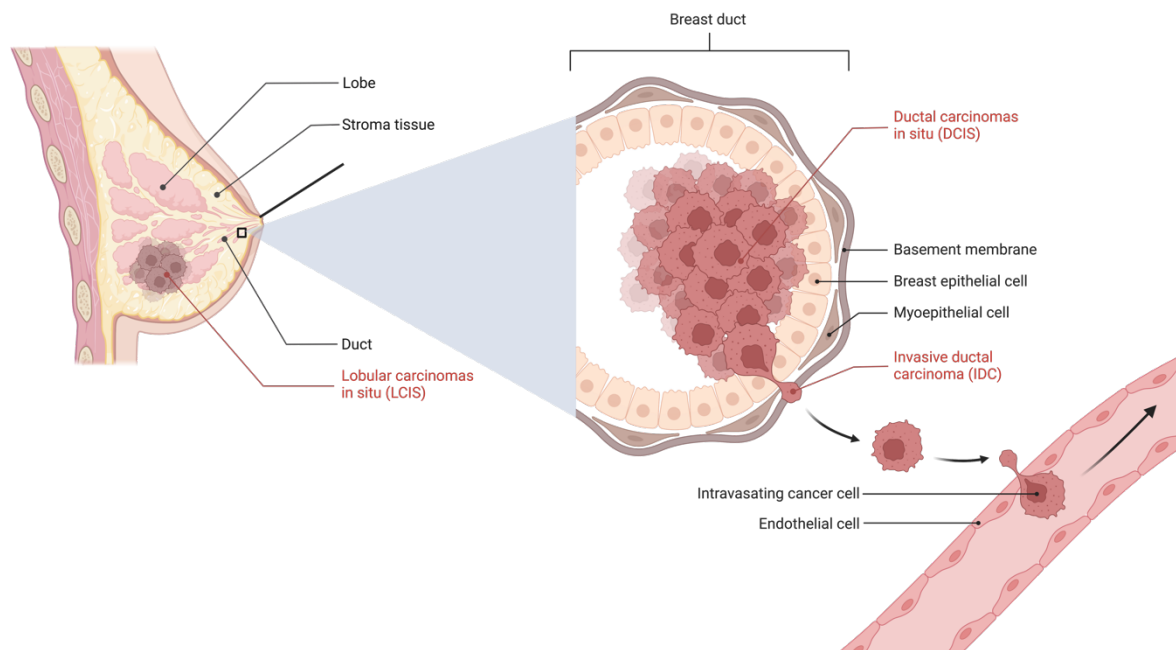
In women, breast cancer has a one in eight lifetime risk of incidence (CRUK 2006) and is the most prevalent malignancy in the United Kingdom (CRUK 2007). Despite an escalating rate of incidence, five-year survival rates in Britain are as high as 80%, attributed largely to the screening programme initiated in 1988 and dramatic advances in therapeutics (CRUK 2006).

#### 1.1.2 Risk factors

The risk of breast cancer incidence is related to multiple factors. The sex of the individual confers the highest risk of developing breast cancer since 1 in 8 women and 1 in 870 men will develop breast cancer in their lifetime. Age contributes greatly to the risk of breast cancer incidence although it is the most frequently diagnosed malignancy in women under the age of thirty-five (CRUK 2007). Early onset of the menstrual cycle, late age at first pregnancy, nulliparity and late age at menopause have all been reported to contribute to an elevated risk (Kelsey, Gammon et al. 1993). Furthermore, the use of exogenous hormones such as oral contraceptives and hormone replacement therapy may also affect tumourigenicity (Chen, 2008). The lifestyle of the individual contributes significantly to the risk of breast cancer development. Physical activity may play a protective role in premenopausal and overweight postmenopausal women (Singletary & Gapstur, 2001). However, obesity displays the opposite relationship to breast cancer incidence in both groups (Carmichael & Bates, 2004). High alcohol intake significantly increases the risk of malignancy (Teng et al., 2008). Familial mutations account for 5-10% of all cancers; the BRCA1 tumour suppressor is frequently mutated in hereditary breast cancer and confers a lifetime risk of 60-80% (Burnett & Kennedy, 1954). Nevertheless, an increase in the effectiveness of breast cancer screening methods such as mammography, which remains the main screening tool for breast cancer, has seen a reduction in breast cancer mortality

rates by about 20 to 35% in women aged 50 to 79 years and slightly less in women aged 40 to 49 years at 14 years of follow-up (Elmore et al., 2005).

### 1.1.3 Malignancy and breast neoplasms



**Figure 1.1.1** Female breast morphology and breast malignancies. Breast malignancies can be categorised based on mammary duct or lobular acini origin and whether the cancer cells have left the primary site. Invasive ductal carcinoma (IDC) acquires the ability to undergo basal extrusion via the basement membrane and intravasates into the surrounding blood capillaries gaining the ability to metastasise from the primary site to other body tissues. The figure was created using BioRender ([www.biorender.com](http://www.biorender.com)).

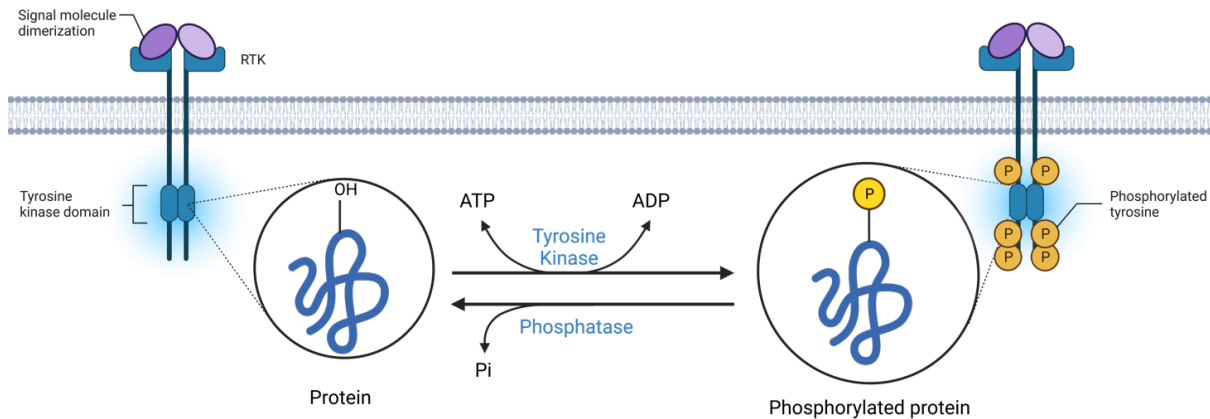
Malignancy is the condition where normal human cells undergo molecular changes that lead to aberrant cell division to form a mass of tissue or neoplasm that has the ability to spread to other sites in the body. Breast cancer is a malignant condition of breast cells that may invade adjacent tissues or propagate to other organs of the body. It includes carcinomas that originate from epithelial cells and sarcomas that arise from fat/muscle tissues or blood vessels. A carcinoma is a malignancy that develops from epithelial cells, particularly those found in a tissue that lines the inner and outer parts of the body. The female breast is a common source of carcinomas and consists of lobules which produce milk, ducts which deliver milk to the nipple, and stroma tissues. Ductal and lobular

carcinomas in situ (DCIS and LCIS respectively) are non-invasive breast malignancies wherein tumour cells are confined to the mammary ducts or lobular acini (Figure 1.1.1). Invasive carcinomas are classically adenocarcinomas, which arise from mammary glandular epithelial cells and are typically of ductal (85%) as opposed to lobular (10%) origin. Direct invasion to the skin and nipple is common whilst distal site metastases via the lymph nodes and blood circulation frequently establish secondary tumours in the lungs, bones, liver, adrenal glands and brain.

Oestrogen receptor (ER), progesterone receptor (PR) and epidermal growth factor receptor 2 (ERBB2/HER2) are frequently expressed in breast cancer and are effective both as prognostic indicators and for directing treatment. Breast cancers termed triple negative (TN) do not exhibit protein expression of any of these receptors (Anders & Carey, 2008). The receptors may also confer common expression patterns, for instance a general inverse correlation is observed between ERBB2 expression and the ER with or without the PR (Konecny et al., 2003). About 20%-30% of breast cancers overexpress ERBB2, which induces aberrant cell growth and proliferation. This type of cancer is correlated with a poor long-term prognosis and has a high rate of metastasis to other sites (Kallioniemi et al., 1991). More importantly, the ligand-unbound ERBB2 and kinase-deficient ERBB3 represent the most mitogenic signalling pair among ERBB family receptors, making ERBB2-related breast cancer signalling a key area of research, both for elucidation of the mechanisms of signalling and the identity of therapeutic targets.

## 1.2 The ERBB receptor family

### 1.2.1 Receptor tyrosine kinases



**Figure 1.2.1** Phosphorylation of receptor tyrosine kinase (RTK). Regulation of receptor tyrosine kinase (RTK) protein phosphorylation by protein kinase activities catalyse the addition of phosphate to the tyrosine residues on proteins and phosphatase activities remove phosphate from the proteins. The figure was created using BioRender ([www.biorender.com](http://www.biorender.com)).

Phosphorylation of proteins is a widespread chemical modification and has been known to exist since the nineteenth century (Venerando et al., 2017). The molecule responsible for the transfer of a phosphate group (ATP) remained elusive until 1954 (Burnett & Kennedy, 1954). The first mechanism was described as an ATP-mediated transfer of a  $^{32}\text{P}$  labelled phosphate group to serine residues on the protein casein (Burnett & Kennedy, 1954). Krebs and Fischer later studied the activity of a metabolic enzyme called glycogen phosphorylase and determined that its activity is mediated by an ATP- catalysed reversible phosphorylation (KREBS & FISCHER, 1956). Protein phosphorylation was later shown to decrease enzyme activity of glycogen synthase in rabbit muscle tissue (FRIEDMAN & LARNER, 1963). Protein kinases and phosphatases that mediate ATP-catalysed reversible protein phosphorylation are now recognized to regulate a wide range of cellular mechanisms (Figure 1.2.1) (Krebs, 1983). Approximately 1.7% of genes in the human genome code for protein kinases, which are a family of 518 proteins comprised of 478 members unique to eukaryotes (Roskoski, 2004). This eukaryotic protein kinase superfamily consists of 90 protein tyrosine kinases (of which 58 are



receptors and 32 are non-receptors) and 43 tyrosine kinase-like kinases (Manning et al., 2002; Robinson et al., 2000). Despite a phosphoserine: phosphothreonine: phosphotyrosine ratio of 3000: 300: 1 existing in many cell types, tyrosine phosphorylation is an essential mediator of cellular signal transduction (Roskoski, 2004).

Cell to cell signalling may occur via the secretion and reception of biomolecules which elicit diverse intracellular and extracellular responses. Various extracellular signals are incapable of passively entering the cytoplasm due to the impermeable nature of the plasma membrane. Receptor tyrosine kinases play a pivotal role in interpreting and transducing the chemical message from the external environment across the plasma membrane to initiate intracellular signalling. Common intracellular signalling events triggered by receptor tyrosine kinases are associated with cellular development, differentiation, proliferation, motility and survival. Effective signalling is controlled by reversible phosphorylation which is tightly regulated by the opposing activities of protein kinases and phosphatases. The human genome encodes 107 protein tyrosine phosphatases of which 38 are strictly tyrosine specific and are of receptor or non-receptor subtypes (Alonso et al., 2004). Tyrosine phosphatases often antagonise receptor tyrosine kinase signalling either via the dephosphorylation of receptors or target substrates. Protein tyrosine phosphatases may also agonise signalling through the inactivation of inhibitory signalling components (Ostman et al., 2006).

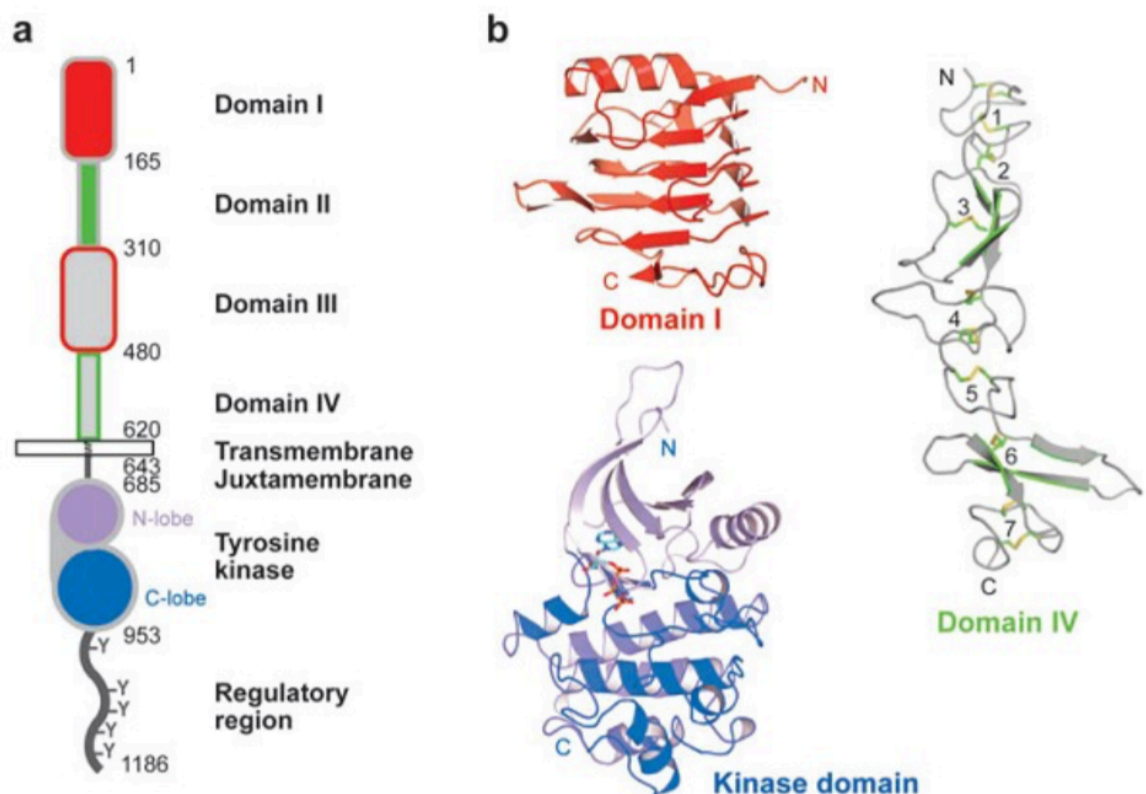
### 1.2.2 Origin, structure and ligands

The family of ERBB receptors is a group of type I receptor tyrosine kinases (RTKs) with structural homology. The family includes, epidermal growth factor receptor (EGFR), ERBB2 (HER2), ERBB3 (HER3) and ERBB4 (HER4) receptors. EGFR was one of the first RTKs described as a ligand-regulated protein kinase. It largely serves as the pioneering model for the structure of ERBB family RTKs. Consequently, all ERBB receptors consist of extracellular domains, a transmembrane domain, a juxtamembrane region, an intracellular tyrosine kinase domain and a regulatory region comprised of a non-catalytic carboxyl-terminal tail with numerous tyrosine phosphorylation sites (Figure 1.2.2) (Bazley & Gullick, 2005; Ferguson, 2008).

The ERBB receptor was originally observed in *Caenorhabditis elegans* as let-23 (lethal complementation group 23)(Aroian et al., 1990) with a sole ligand lin-3 (abnormal cell lineage 3)(Hill & Sternberg, 1992). The ERBB family demonstrates evolutionary conservation and the vertebrate forms are homologous variants of the ERBB in *C. elegans*. However, the vertebrate ERBB family demonstrates more complexity and diversity. The ancestral vertebrate ERBB family is thought to have arisen from a gene duplication event, which initially produced a set of precursors to the present day ERBB family of receptors. The ERBB family ligands are also thought to have undergone a similar diversification to yield the current EGFR and ERBB3/4 ligands (Stein & Staros, 2000, 2006). Moreover, the ERBB receptors have varying kinase activity with EGFR and ERBB4 demonstrating potent kinase activity and intact ligand binding. The catalytic region of the ERBB3 kinase domain has a point mutation at a key residue which results in a lack of kinase activity (Guy et al., 1994). In contrast, ERBB2 has potent kinase activity but lacks ligand binding capability. Therefore, ERBB2 heterodimerises with the other receptors which in turn activates signalling; the ERBB2/ERBB3 dimerisation pair being the most potent signalling activator.

Signalling activation in the ERBB receptor family is stimulated by the binding of small protein growth factor ligands. All ERBB ligands have a consensus EGF-like core domain, comprised of sixty amino acids with six cysteine residues forming intramolecular disulphide bonds to sufficiently confer binding specificity and basic activation (R. C. Harris et al., 2003; J. T. Jones et al., 1999).

Accordingly, they are categorized as per their binding specificity and ERBB receptor affinity. The ligands are divided into three groups, an epidermal growth factor (EGF) associated group which binds specifically to EGFR, a group of neuregulins (NRGs) that undergo alternative splicing to produce multiple isoforms that interact with ERBB3 and ERBB4 (Warren & Landgraf, 2006) and a group that demonstrates dual binding specificity to EGFR or ERBB4 and includes growth factors such as heparin-binding EGF-like growth factor (HB-EGF), epiregulin (EPR) and  $\beta$ -cellulin (BTC). ERBB2 is an orphan receptor with no known ligands (Klapper et al., 1999). The EGFR receptor has an affinity for EGF, amphiregulin, BTC, HB-EGF, and transforming growth factor  $\alpha$  (TGF- $\alpha$ ). The ERBB3 receptor has an affinity for neuregulin-1 and the ERBB4 receptor binds to neuregulins, EPR and HB-EGF (Yarden & Slivkowski, 2001).



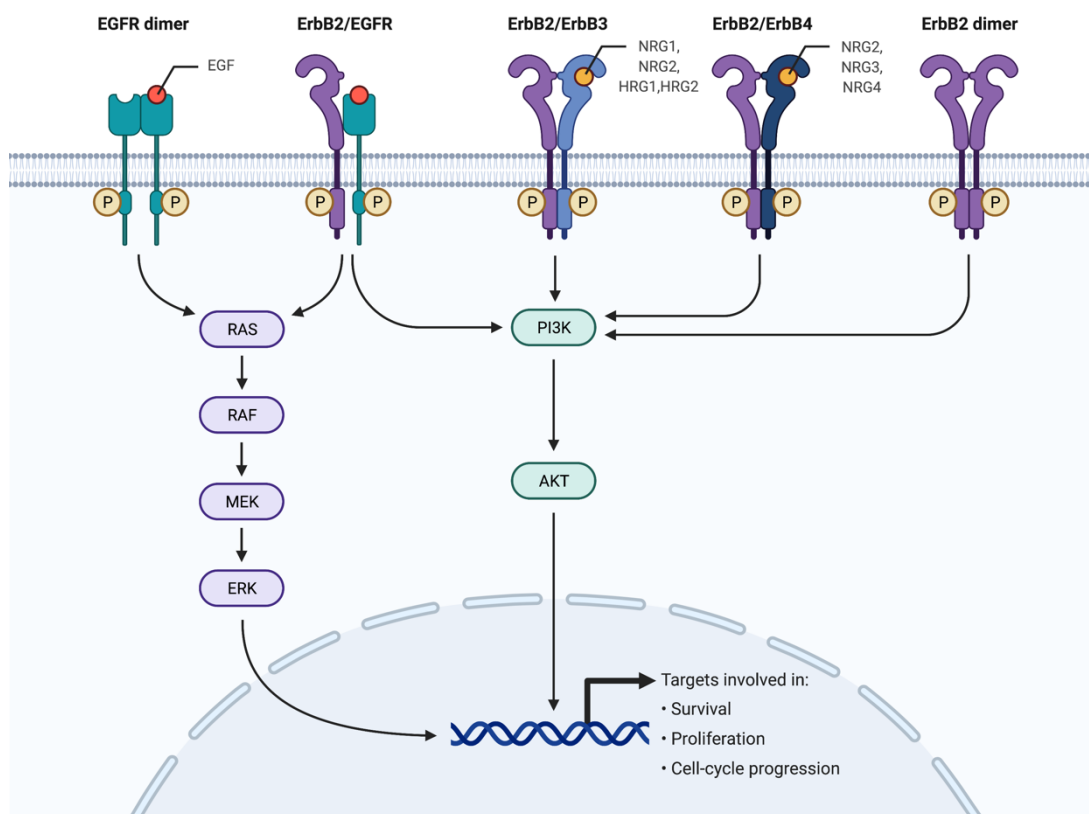
**Figure 1.2.2** Structural representation of a typical ERBB receptor. A. The EGFR receptor consists of four extracellular domains I-IV (L1, CR1, L2, CR2), transmembrane span, juxtamembrane region, kinase domain, and a C-terminal tail. B. Ribbon diagrams representing the secondary and tertiary structures of EGFR domains. Domain I and III adopt a  $\beta$ -helix fold. Domain II and IV adopt extended structures comprising a series of disulfide-bonded modules. The inactive kinase is shown with the ATP analogue (AMP-PNP) in stick representation (Ferguson, 2008).

### 1.2.3 Signal transduction

The ERBB receptors naturally exist as monomeric units spanning the plasma membrane and have dormant cytoplasmic kinase domains capable of intrinsic auto/cross-phosphorylation of tyrosine residues. The receptors hetero/homodimerise in response to ligand binding which is widely recognized as the principle mechanism of activating the auto/cross phosphorylation of tyrosine residues in the carboxyl-terminal tails (Hazan et al., 1990; Margolis et al., 1989). The EGFR, ERBB3 and ERBB4 receptors have intramolecular interactions between domains II and IV (CR1 and CR2) which maintain the dimerisation loop in an inactively tethered receptor conformation (Bouyain et al., 2005; Cho & Leahy, 2002; Ferguson et al., 2003). Ligand binding between extracellular domains I and III (L1 and L2) induces conformational changes and disrupts the dimerisation loop. This leads to an

untethered conformation and induces an interaction with another proximal ligand bound ERBB receptor (Garrett et al., 2002; Ogiso et al., 2002). The structure of ERBB2 closely resembles that of an untethered receptor and as such is capable of hetero/homodimerising in a ligand independent manner (Cho et al., 2003; Garrett et al., 2003). ERBB2 is however homodimerization deficient and prefers to heterodimerise with the other ERBB receptors. Thus, it interacts with its binding partners to promote signalling (Graus-Porta et al., 1997; Tzahar et al., 1996).

When a ligand binds to an ERBB receptor, the conformational change leads to kinase activation and phosphorylation of the receptor and downstream adaptor proteins which leads to recruitment and activation of signalling molecules which enhance signal transduction cascades (Figure 1.2.3). This can occur through the interaction of a phosphotyrosine docking site located on the carboxyl terminal tail of the receptor with SH2 domain and phosphotyrosine-binding domain (PTB) containing proteins (Schlessinger, 1994). All combinations of active receptor pairs can activate the GRB2-SOS-Ras pathway. However, the PI3K-Akt signalling pathway is mainly coupled to the EGFR homodimer and ERBB3 in its hetero-complex and ERBB4 in both its dimer complexes (Gschwind et al., 2004; Rubin & Yarden, 2001). Other major pathways involved in intracellular signalling are the stress activated protein kinase cascade, PLC $\gamma$ -PKC pathway and cell migration related to Vav-Rac-JNK which are EGFR activated (Rubin & Yarden, 2001). The EGFR homodimer binds and activates the widest range of downstream signalling proteins. In contrast the ERBB3 receptor is kinase-deficient with low complexity, but a highly potent interaction partner for adaptors. Intracellular signalling is transduced to the nucleus and activates transcription factors which regulate gene transcription (Schaeffer et al., 1998). Through this and other mechanisms, the receptors feed into pathways regulating cellular metabolism and energy control, biosynthesis, the cytoskeleton, adhesion and molecular transport (Figure 1.2.3).



**Figure 1.2.3** Overview of ERBB signalling. The extracellular ligand binding domains of the various receptors are shown to bind to either epidermal growth factor (EGF), neuregulin 1 (NRG1), neuregulin 2 (NRG2), neuregulin 3 (NRG3), neuregulin 4 (NRG4), heregulin 1 (HRG 1) or heregulin 2 (HRG 2). The figure was created using BioRender ([www.biorender.com](http://www.biorender.com)).

#### 1.2.4 Signal attenuation

Mammalian cells have developed complex regulatory mechanisms to turn off intracellular signalling, which prevents deregulated signalling and aberrant cell proliferation. This negative regulation process is essentially achieved in three ways: dephosphorylation of tyrosine residues, internalisation and degradation of activated receptors, and modulation of receptor kinase activity. Phosphatases, such as density-enhanced phosphatase-1 (DEP-1) and protein tyrosine phosphatase 1B (PTP1B), can dephosphorylate the phosphorylated tyrosine residues at receptor tail region, such that adaptor protein recruitment to the cell surface receptors is diminished (Tarcic et al., 2009; Yip et al., 2010).

Downregulation of signalling by receptor internalization occurs in two steps; receptor degradation and recycling. EGFR naturally resides in caveolae on the cell membrane, whereupon it exits caveolae following ligand binding and enters clathrin-coated vesicles. Clathrin-coated vesicles have been

implicated in several distinct intracellular transport steps. The clathrin-mediated endocytic pathway plays an important role in the selective uptake of proteins at the plasma membrane of eukaryotic cells. Hence, the EGFR undergoes endocytosis through endosomes. The endosome differentiates between unliganded EGFR, which is sent back to cell surface, whereas a ligand occupied receptor is tagged by the E3 ubiquitin ligase Cbl and directed to degradation (Levkowitz et al., 1998).

Modulation of receptor kinase activity can also be used to switch off oncogenic signalling. A good example is the binding of Hsp90 to ERBB2, which blocks phosphorylation of ERBB2 at residue Tyr877 within the receptor kinase domain (W. Xu et al., 2007). When Hsp90 binds to ERBB2, it stabilizes the receptor, and this may explain the correlation between poor prognosis and elevated ERBB2 and Hsp90 levels in breast and other cancers. Accordingly, ERBB2 overexpressing cancers are highly sensitive to Hsp90 inhibition (Pick et al., 2007).

#### 1.2.5 ERBB expression in the mammary gland

ERBB expression has been investigated extensively in knockout mice which have made an invaluable contribution towards functional assignment. Studies conducted with knockout mice have shown that the ERBB receptors are expressed in the mammary gland prior to puberty but are not phosphorylated. EGFR and ERBB2 phosphorylation are initiated during ductal morphogenesis at puberty, whilst ERBB3 and ERBB4 receptors are not (Sebastian et al., 1998). Similarly, the EGFR and ERBB2 receptors are preferentially expressed in mouse mammary ducts and alveoli, whilst the expression of ERBB3 and ERBB4 receptors is mostly alveolar. Early evidence suggested that all ERBB receptors are expressed and phosphorylated during murine pregnancy and lactation stages (Schroeder & Lee, 1998). These observations were substantiated by rat studies that confirmed the functional presence of EGFR and ERBB2 during the entire process of mammary development, whilst the ERBB3 and ERBB4 receptors were limited to pregnancy and lactation (Darcy et al., 1999, 2000). These studies point toward an essential role for EGFR during ductal morphogenesis (Andrechek et al., 2005; Jackson et al., 2004; Sebastian et al., 1998; Wiesen et al., 1999; Xie et al., 1997), and place ERBB2, ERBB3 and ERBB4 as essential receptors in lobuloalveolar formation and lactation (F. E. Jones et al., 1999; F. E. Jones & Stern, 1999; Qu et al., 2006; Tidcombe et al., 2003).

While EGFR and ERBB2 have been detected throughout all stages of mammary development (Darcy et al., 1999, 2000) and the presence of EGFR and ERBB2 in these tissues correlates with an early observation that their overexpression is associated with various types of invasive cancers in humans, unlike ERBB3 and ERBB4 (Verbeek et al., 1997), there still remain unanswered questions regarding the mechanisms that drive tumorigenesis and the role played by ERBB family members. Breast cancer development is known to involve several signalling pathways which are associated with patterning and morphogenesis during mammary gland development (S. Y. Lin et al., 2000). Both EGFR and ERBB2 have been shown to play a role in activating several well-known intracellular signalling pathways such as the PI3K-Akt and GRB2-SOS-Ras signalling pathways (Figure 1.2.3) (Gschwind et al., 2004; Rubin & Yarden, 2001). However, the downstream signalling mechanisms and their interaction partners remain largely unknown. Furthermore, there remains a need to elucidate the functional role of potential downstream signalling partners in bringing about neoplastic phenotypes in mammary tissue cells.

### 1.3 ERBB2 Overexpressing Cell Systems

The role of gene expression in disease progression can be studied in relevant cell models that reflect similar molecular characteristics to those of the normal cell type from which a malignancy originates. Cancer cell lines are generally regarded as a good model of primary cells as they retain the molecular features of cancer cells *in vivo* (Lacroix & Leclercq, 2004), are generally more robust and easier to work with than primary cells but are susceptible to genotypic and phenotypic transformation. Manipulation of normal cells may enable the characterisation of the underlying molecular alterations responsible for the development of carcinomas. Most breast cancers originate from ductal luminal epithelial cells. HB4a is an established cell line derived from normal human mammary luminal epithelial cells immortalised with a temperature-sensitive (TS) mutant of the SV40 large T antigen. The HB4a cells exhibit a non-transformed phenotype, and retain luminal epithelial cell markers (Stamps et al., 1994). However, analysis of mRNA profiles and transcriptome found that the miRNA profile of HB4a is quite different from that of normal breast tissue. Nevertheless, HB4a still carries a similar mRNA expression profile to that of normal breast tissue and is an excellent molecular model

for proteomic studies (Git et al., 2008). Human mammary luminal epithelial clones expressing different levels of ERBB2 were developed by transfecting HB4a cells with c-ERBB2 cDNA under the control of the MMTV-LTR (R. A. Harris et al., 1999). Selected clones (C3.6 and C5.2) assumed a hyperproliferative state and exhibited anchorage-independent proliferation and are thus relevant models for the study of ERBB2-specific transformation. In contrast, ERBB2 overexpressing cell lines such as BT474 and SkBr3, derived from primary tumours are also useful *in vitro* models for the study of cellular transformation in cancer research (Lasfargues et al., 1978; Trempe, 1976).

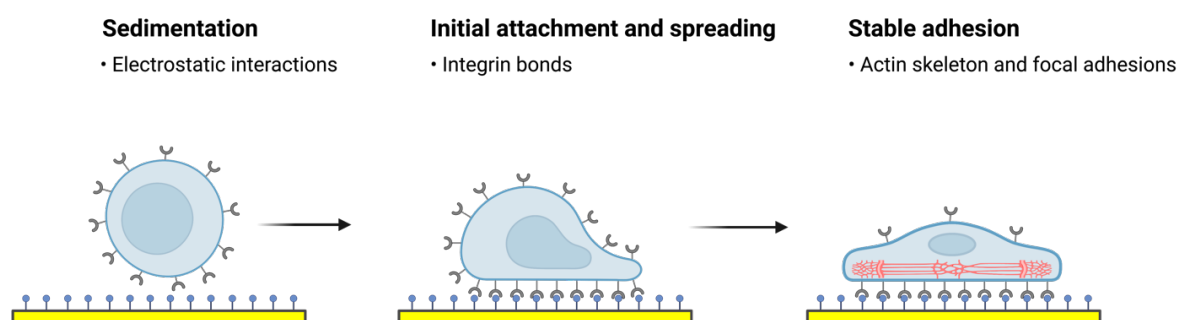
Cancer cell line models, although useful for understanding basic cellular mechanisms, have intrinsic and unsurmountable limitations. These limitations include the fact that they are cultured in 2D flat dishes, growing in cell culture media, and lacking matching tumour microenvironment (TME) components. In contrast, the TME is a complex network of cells, extracellular matrix, and signalling molecules that surround and support the growth of tumour cells. It is a dynamic and constantly evolving environment that plays a crucial role in tumour development and progression. The TME includes a variety of cell types, such as cancer-associated fibroblasts, immune cells, endothelial cells, and pericytes, as well as cytokines, chemokines, growth factors, and extracellular matrix proteins. These components interact with each other to regulate tumour cell behaviour, such as proliferation, migration, invasion, and resistance to therapy. The TME can also create a barrier for immune cells to enter and attack the tumour, leading to immune evasion and immune suppression (Anderson & Simon, 2020; Baghban et al., 2020). Overall, understanding the TME is critical for developing effective cancer therapies that can target the tumour and its surrounding microenvironment.

#### 1.4 Passive *in Vitro* Cell Adhesion Assays

Anchorage-independent proliferation is widely considered a hallmark of oncogenic transformation and has been shown to be a powerful predictor of tumourigenic and metastatic potential of non-malignant cells (Freedman, 1974). Studies involving several cell types have demonstrated that anchorage-independent proliferation is initiated by multiple genetic changes that inactivate the Rb and p53 tumour suppressor pathways while activating the Ras signalling pathways (Hahn & Weinberg, 2002). During anchorage-independent proliferation, cell-matrix and cell-cell adhesions enable cells to stick to



an extracellular matrix (ECM) characterised by passive mechanical aspects such as its bulk, local stiffness, viscoelasticity, ligand density and topography (Arnold et al., 2004), or to adjacent cells in a passive manner (Baum & Georgiou, 2011; Leckband & de Rooij, 2014; Yap et al., 2018). Since HB4a-derived ERBB2 overexpressing cell line models such as C3.6 and C5.2 are known to assume anchorage-independent growth (Lasfargues et al., 1978; Trempe, 1976), the effect of protein expression on cell adhesion can be assessed using passive *in vitro* cell adhesion assays. Passive *in vitro* cell adhesion is the cell adhesion process that occurs in a static medium culture, e.g., culture flasks, petri dishes. Static *in vitro* cell adhesion is characterised by three stages: sedimentation, attachment, flattening and spreading of the cell body on its substrate, and the organization of the actin skeleton with the formation of focal adhesion between the cell and its substrate (LeBaron & Athanasiou, 2000). The attachment of the cell body to its substrate involves specific integrin-mediated adhesion and starts with the binding of a single receptor-ligand pair (S. Hong et al., 2006), followed by subsequent receptor-ligand bonds that increase the total adhesion strength (Taubenberger et al., 2007). Cells continue to flatten, spread and increase contact area on the substrate (W. Huang et al., 2003). Finally, the cells spread beyond the projected area of the un spread cell due to the reorganisation of the actin skeleton and the formation of focal adhesions between the cell and its substrate (S. Hong et al., 2006) (Figure 1.4).



**Figure 2.4** Stages of passive *in vitro* cell adhesion. The three-stage process involves the sedimentation of cells, initial attachment and spread mediated by integrin bonds and the formation of stable adhesion through actin skeleton reorganisation and focal adhesions. The figure was created using BioRender ([www.biorender.com](http://www.biorender.com)).

### 1.4.1 End Point Cell Adhesion Assays

The MTT assay is a cell-based endpoint assay widely used to assess cell viability or cell adhesion by measuring the proportion of cells that attach to a plate after a given period. The MTT assay relies on the presence of functional mitochondria as the cellular reduction of tetrazolium salts to insoluble purple formazan crystals by the action of mitochondrial reductase (Mosmann, 1983a). The MTT reaction mixture is added towards the end of the incubation for a given assay and the plate is incubated for an additional time period. The formazan crystals are then dissolved in an acid/alcohol solution and the plate is read on a multi-well scanning spectrophotometer. Percentage cell adhesion is determined by comparing the optical density values of the treated cells with the optical density values of the control cells and the results are presented as a percentage of adherent cells (Miki et al., 1993). The MTT assay has several advantages that are favourable for cell adhesion assays, such as ease of implementation, rapidity of performance and reproducibility of the results (Colangelo et al., 1992; Iselt et al., 1989; Jiao et al., 1992; Pieters et al., 1988). However, one of the limitations of the MTT assay lies in the underlying assumption that mitochondrial activity is proportional to the formation of formazan in the mitochondria.

Hitherto, the location of formazan formation and its intracellular transportation has remained controversial (Ghasemi et al., 2021). Several biomolecules such as ascorbic acid, cysteine, dihydrolipoic acid, glutathione, glutathione S-transferase, and tocopherols have been found to reduce MTT, suggesting that the formazan produced during the MTT assay represents more than mitochondrial activity (Stockert et al., 2018). Moreover, the limitations of the MTT assay have been revealed in several studies (Ghazali et al., 2020; Kok et al., 2019; Y. Liu et al., 1997; Mor-Yossef Moldovan et al., 2020; Mosmann, 1983b; Stockert et al., 2018; Surin et al., 2017) that highlight various confounding factors to be considered when designing, performing, analysing, and interpreting the assay results (G. Diaz et al., n.d.; Gokduman & Gok, 2020; Y. Liu et al., 1997; Małaczewska et al., 2021; Nikoloff et al., 2021; Patra & Gupta, 2020; Stockert et al., 2012). Ultimately, the MTT assay is an endpoint assay that provides a “snapshot” of the adhesion process at the time the experiment is stopped and does not capture the real-time cell dynamics of passive *in vitro* cell adhesion as comprehensively as real time cell adhesion assays such as the xCELLigence RTCA system (S. Hong

et al., 2006; Khalili & Ahmad, 2015; Martinez-Serra et al., 2014; Stefanowicz-Hajduk & Ochocka, 2020).

#### 1.4.2 Real Time Cell Adhesion Assays

Unlike endpoint approaches, real-time assay systems allow for the tracking of cell dynamics throughout all the stages of passive *in vitro* cell adhesion. This is particularly effective for assessing characteristics which may otherwise be missed using endpoint-based methods; such as the flattening and spreading of the cell body, periodic changes in the actin skeleton and formation of focal adhesions between the cell and its substrate. The xCELLigence system is a real-time assay system that integrates microelectronics and cell biology and is suitable for uninterrupted monitoring of biological processes of living cells. It uses specially designed microtitre plates containing interdigitated gold microelectrodes to non-invasively monitor the viability of cultured cells. The electrodes measure the electrical impedance of the cell population in each well and it provides quantitative real-time information about the status of the cells (Roshan Moniri et al., 2015).

Microelectrode biosensor technology is a favourable method for assessment of passive *in vitro* cell adhesion because it reveals the temporal profile of the response in real-time, which enables both acute responses and longer-term responses to be profiled within the same assay (Kho et al., 2015). The monitoring of changes in the adhesion and morphology of target cells using xCELLigence microelectrode technology has been shown to be a convenient way to continuously determine cell number and cell adhesion of cells undergoing apoptosis due to NK cell-mediated cytotoxic effects (Glamann & Hansen, 2006). Moreover, the xCELLigence RTCA system has been used to analyse cell adhesion in MDA-MB-231 (triple-negative human breast adenocarcinoma) cells and HEK293 (human embryonic kidney) cells to demonstrate that the silencing of SHANK proteins increased integrin activation and promoted cell adhesion (Lilja et al., 2017). Therefore, continuous monitoring of passive *in vitro* cell adhesion by the xCELLigence system is a useful real-time cell adhesion method to distinguish between the sedimentation, initial attachment and spreading, and stable adhesion stages of passive *in vitro* cell adhesion (S. Hong et al., 2006; Khalili & Ahmad, 2015; Martinez-Serra et al., 2014; Stefanowicz-Hajduk & Ochocka, 2020).

## 1.5 Proteomics

Proteomics is the large-scale study of proteins, under a specific set of experimental conditions and often employs a method called mass spectrometry (MS; see section 1.5.3). Different techniques can be applied to identify a catalogue of the proteins present in a complex mixture (Washburn et al., 2001), characterise post-translational modifications (PTMs) (Mann & Jensen, 2003), or determine the relative abundance of the proteins (Griffin et al., 2003; Gygi & Aebersold, 2000; McIlwain et al., 2012). Protein samples can be analysed using two different MS strategies; “top-down” and “bottom-up” approaches (VerBerkmoes et al., 2002). Top down proteomics analyses intact proteins in a mixture and uses their accurate masses and fragment ions to yield a protein identity (Reid & McLuckey, 2002). The bottom-up approach involves a protein digestion step to yield peptides which are then fragmented and detected via tandem mass spectrometry (MS/MS) and their identities determined from sequence level data (Hunt et al., 1986). The identified peptide sequences are then assigned to the proteins they originate from, through a process called protein inference (Nesvizhskii & Aebersold, 2005).

### 1.5.1 Proteomic workflows

Proteomic workflows are often designed to aid the detection of low abundance protein species to maximize protein identification and coverage. This is achieved by reducing the sample complexity prior to mass analysis, which involves fractionating the sample either at the protein or peptide level. Gel-based protein fractionation techniques are widely used in proteomics and include sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) or 2D-difference gel electrophoresis (2D DIGE). Alternatively, OFFGEL electrophoresis may be used to fractionate proteins according to their isoelectric point where the separated proteins are recovered in the liquid phase. Following fractionation, proteins are enzymatically digested into peptides which makes them more amenable to analysis by mass spectrometry. The enzyme trypsin is often used, which cleaves after the commonly occurring amino acids lysine and arginine. Peptides can be further separated using ion-exchange chromatography, hydrophobicity (normal phase, HILIC) or ion exchange chromatography and/or resolved using HPLC (high-performance/pressure liquid chromatography).

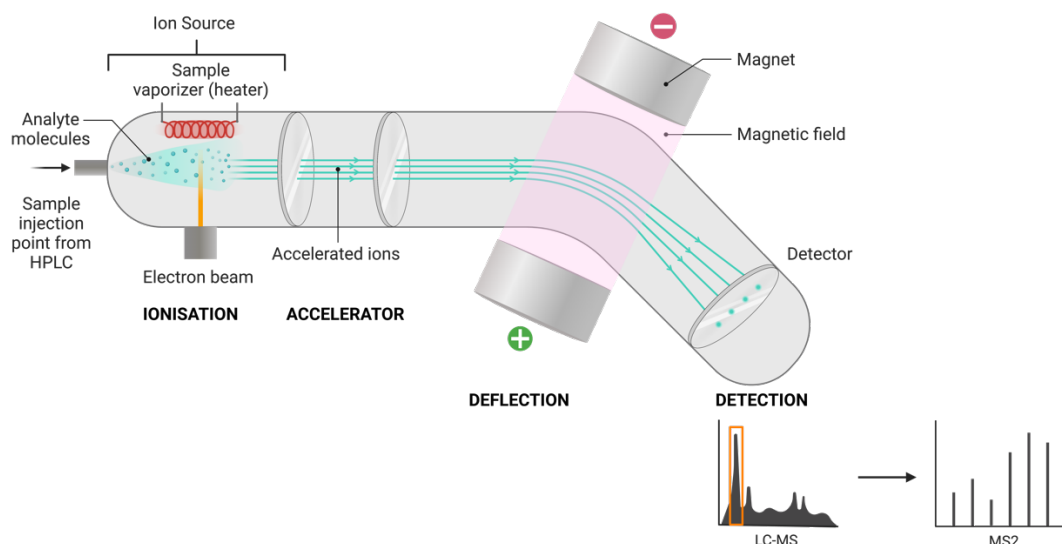
### 1.5.2 High-performance/pressure liquid chromatography (HPLC)

HPLC is an optimised analytical form of column chromatography. The technique is used to resolve molecular species based on their interactions with a column-bound stationary phase. Instead of allowing the solvent to drop through the column under gravity, it is forced through the column under high pressure. The passage of analyte molecules through the column is retarded by specific chemical or physical interactions with an immobilised chromatographic support called the stationary phase. Bound peptides are sequentially eluted using a graduated mobile phase. The specific point at which a peptide elutes from the column is known as the retention time and is a unique property of the analyte molecule. RPLC (reversed-phase liquid chromatography) separates peptides according to their hydrophobicity. The stationary phase typically consists of C18 alkyl chains. HPLC can be directly interfaced to an MS for direct analysis as peptides are sequentially eluted in order of their hydrophobicity using a linear gradient of organic solvent, typically with flow rates of 200-500 nL/min (nano-flow) or the use of capillary flow. HILIC (hydrophilic interaction chromatography) separates proteins according to their hydrophilicity and is a popular orthogonal technique to RPLC. Peptides are bound to a hydrophilic stationary phase and are eluted using a linear gradient of increasing polarity. HPLC is an invaluable fractionation tool due to its high resolution, reproducibility and compatibility with mass spectrometry (MS). In proteomic analyses, RPLC or HILIC can be coupled directly to a mass spectrometer in a technique known as liquid chromatography tandem mass spectrometry (LC-MS/MS).

### 1.5.3 Mass spectrometry

Mass spectrometry (MS) has become a fundamental technique in protein analysis. MS is the scientific field focusing on the separation and analysis of molecules by accurate molecular mass measurement. In principle, MS offers a comprehensive technique for the identification of molecules in a mixture by measuring the mass-to-charge ratio ( $m/z$ ) of gas-phase ions. Mass spectrometers are comprised of an ion source that converts analyte molecules into gas-phase ions, a mass analyser comprised of an

accelerator and an electromagnetic deflector that separates ionised analytes based on their  $m/z$  ratio, and a detector that records the number of ions at each  $m/z$  value (Figure 1.5.1).



**Figure 1.5.1** Mass spectrometry principle. This schematic provides a general overview of sample injection from HPLC, sample ionisation and acceleration, deflection and detection of sample gas phase ions. The figure was created using BioRender ([www.biorender.com](http://www.biorender.com)).

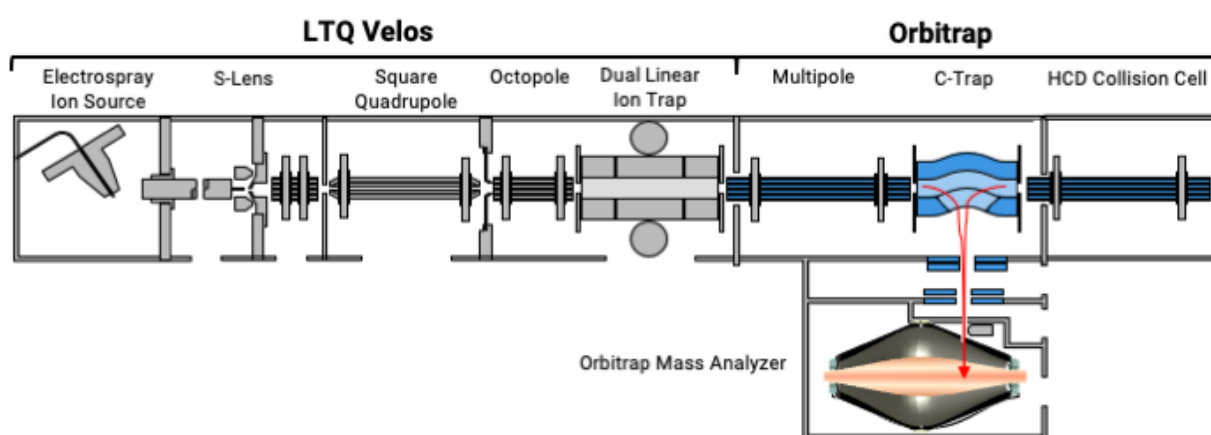
### 1.5.3.1 Soft ionization of analytes

Two methods of ionising analytes collectively termed as soft ionisation, revolutionised protein analysis using MS. Electrospray ionisation (ESI) is a soft ionisation technique used to convert analyte molecules into gas-phase ions (Fenn et al., 1989). The analyte is ionised by the source which is held at a positive or negative potential depending on the polarity of interest. Peptides are typically analysed in the positive ion mode and are sprayed into the source in acidified organic solvent to assist ionisation. As peptides are sprayed (usually from an LC column), a Taylor cone of highly charged droplets is formed (Taylor, 1964). Solvent evaporates from the droplets and the ions are transferred to the gas phase. Heat and gas are applied to assist droplet desolvation which is believed to occur via the combined effects of the charged residue and/or ion evaporation models. In the charged residue model, solvent evaporation initiates a sequence of Rayleigh instabilities (coulomb fissions) which ultimately produce individual gas phase ions (Dole et al., 1968). In the ion

evaporation model, desolvation decreases the droplet size until the surface field strength is sufficient to allow ion desorption (Iribarne, 1976).

ESI is the ionisation method of choice for LC-MS/MS analyses due to its compatibility with RPLC and ability to produce multiply charged ions which extend the mass range of the analyser (Karas & Hillenkamp, 1988).

#### 1.5.3.2 Linear trap quadrupole (LTQ) mass spectrometry



**Figure 1.5.2** LTQ Orbitrap Mass Spectrometer. The figure shows the structure of an LTQ Orbitrap mass spectrometer composed of an ion source where the ions are produced and transferred into the linear ion trap or via the C-trap into the Orbitrap mass analyzer. This figure is reproduced with permission of Thermo Fisher Scientific ([www.thermofisher.com](http://www.thermofisher.com))

The linear quadrupole ion trap (LTQ) is a low-resolution mass analyser which features fast scan times and high sensitivity. The LTQ is based on trapping ions using radiofrequency (RF) electric fields to form a "trap". Typically, ions in the LTQ are focused to the ion trap by a series of multipoles (quadrupoles). Multipoles consist of four parallel metal rods to which a radiofrequency (RF) voltage is applied. The same RF voltage is placed on opposing poles so that one set is positive whilst the other is negative. Ions are sequentially attracted/repelled by the poles and focus into a concise beam which moves forwards towards the trap. Radial trapping is achieved essentially as previously described, whilst ions are trapped in the axial direction by distinct direct current (DC) voltages. Changing the RF and direct current (DC) voltages facilitates the manipulation of the ions. These

manipulations include isolation of an ion with a specific mass to charge ratio, activation by putting kinetic energy into the ions with specific  $m/z$  to cause fragmentation by collision with a gas (helium is usually in the trap) (Figure 1.5.2). The helium serves as an inert dampening/buffer gas whose particles frequently collide with the ions and control their motion in a net loss of kinetic energy. The mass analysis is done in the analyser by selectively ejecting ions out of the trap into a pair of electron multiplier in the detector. The particles are transferred to a multiplier and create an electron cascade as they move towards the anode which results in a measurable current that is proportional to the ion intensity. This generates a mass spectrum which is the signal intensity of the ions at each value on the  $m/z$  scale.

#### 1.5.3.3 Orbitrap mass analyzers

The orbitrap is an ion trap mass analyser that consists of a coaxial inner spindle-like electrode that traps ions in an orbital motion and an outer barrel-like electrode, which generates an electrostatic field. Ions are trapped within the electrostatic field and subsequently orbit around the central electrode and oscillate in the axial direction. The  $m/z$  of the ions determines how frequently they oscillate along the axial trajectory. Oscillating ions induce an alternating current that is detected by the outer electrodes. This time-domain signal is converted to a  $m/z$  spectrum using the fast Fourier-transform (FFT) algorithm. The oscillation frequency can be read with high precision and thus the  $m/z$  measurement is extremely accurate. The orbitrap provides high resolution and mass accuracy. The LTQ Orbitrap XL (Thermo Fisher Scientific Ltd.) and other hybrid mass spectrometers, feature both the high resolution and mass accuracy of the orbitrap, combined with the speed and sensitivity of the LTQ and are increasingly popular instruments for high-throughput proteomic analysis (Figure 1.5.2).

#### 1.5.3.4 Ion mobility spectrometry (IMS)

Ion mobility spectrometry (IMS) is a bioanalytical technique often coupled with MS and HPLC to detect ionised molecules in the gas phase based on their mobility in a carrier buffer gas (Kabir & Donald, 2017). The coupling of ion mobility and mass spectrometry (IM-MS) is more favourable than



MS-only in lipidomics and metabolomics due to increased peak capacity and separation power of isomers and isobars. Furthermore, IM-MS has superior quality spectra acquisition and a higher confidence level of identification due to reduced chemical noise and inclusion of the collision cross section (CCS) as a quasi-orthogonal property respectively (Baker et al., 2014; Hines & Xu, 2019; Paglia et al., 2015, 2021). CCS values are considered important distinguishing characteristics of ions in the gas phase and can be determined empirically or derived computationally with a known 3D structure of a molecule. The structural identification confidence and accuracy of glycans has drastically been improved by adding CCS values of glycans and their fragments to databases.

#### 1.5.3.5 Time of flight mass spectrometry (TOF MS)

Time of flight mass spectrometry (TOF MS) is a mass spectrometry technique that determines the mass over charge ratio of detected ions by measuring their time of flight in an electric field of known strength (Wolff & Stephens, 1953). The TOF mass analyser can be coupled with an ion mobility (IMS) section and a soft ionisation source such as ESI, in a method known as electrospray ionisation ion mobility time-of-flight mass spectrometry (ESI-IMS-TOF MS) (Chakraborty et al., 2009). The method is developed based on a simple and flexible modification to a commercial TOF MS instrument with off-the-shelf components, thus enabling the gas-phase ion-molecule reaction of the ESI-IMS to be effectively coupled to the high-performance TOF MS without sacrificing any of the high-performance attributes of the original instrumentation.

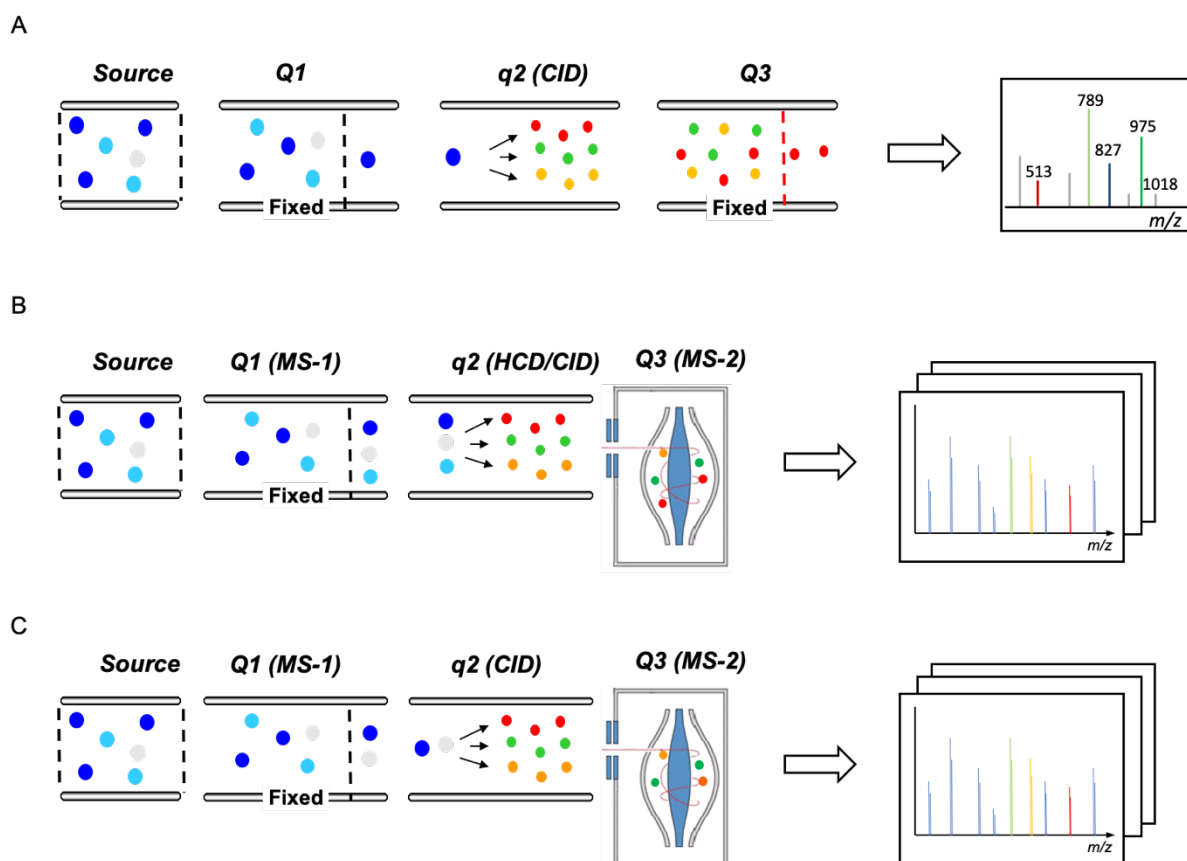
#### 1.5.4 Quantitative proteomics

Quantitative proteomics refers to the determination of the relative abundance of proteins in different cell. Protein quantification is typically achieved by incorporating mass spectrometry with stable isotope labelling by amino acids in cell culture (SILAC) (Rigbolt & Blagoev, 2010) or by chemical modification of the peptides using isobaric tags (TMT/ITRAQ) (Thompson et al., 2003; Unwin, 2010) or by label free methods that rely on measurement of chromatographic peak integrals, ion intensities or spectral counting (Levin & Bahn, 2010).

Isobaric tags for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT) are chemical labelling technologies that utilize isobaric tagging to enable multiple samples to be analysed in a single experiment (Sinclair & Timms, 2011). The isobaric tagging typically targets free amines at either the protein or peptide level. The isobaric tags all have an identical precursor mass addition but yield different reporter ions when fragmented during MS/MS analysis. The different reporter group intensities in the low mass region of the MS/MS spectrum are then used to derive the relative peptide and thus protein abundance ratios between the different experimental conditions (Unwin, 2010). Utilising isobaric tagging to analyse multiple samples in this manner reduces the number of missing peptide quantification values and is known as multiplexing. In addition, a deeper proteome coverage is achieved for multiple samples in a reasonable amount of experimental time. Therefore, isobaric tags such as TMT tags are highly attractive for analysis of proteoform studies that require robust quantification of single peptides across experimental conditions. Nevertheless, isobaric tagging has a high cost of reagent and several studies have successfully endeavoured to mitigate the cost by reducing the amount of TMT reagent used in each multiplexed experiment to address a variety of biological questions (Zecha et al., 2019).

Several label-free quantitative proteomic methods are available and can be applied to achieve an absolute or relative quantification. In this strategy, the proteins or peptides are not labelled, and the method relies on analysing the samples sequentially and discretely. Thus, the MS mode of operation defines the analysis and distinguishes the label-free methods. Obtaining a label-free relative quantification of proteins, involves either a spectral counting or peptide signal intensity method to estimate the relative peptide and thus protein abundance ratios in different samples. A label-free absolute quantification of proteins can be achieved using SRM (selected reaction monitoring) or MRM (multiple reaction monitoring) technologies. In a typical SRM/MRM experiment the mass spectrometer is set to exclusively monitor a specific set of transitions (precursor and product ions) from peptides unique to a protein of interest. These are then normalized against spiked heavy isomer peptide standards of the selected transitions to enable absolute peptide and thus protein quantification (Levin & Bahn, 2010). SRM-based methods have several limitations involving sample complexity and the number of transitions required to increase sensitivity and quantification capacity (Calvo et al., 2011).

In addition, low resolution analyzers (both Q1 and Q2) give rise to greater co-isolation of interferences along with the precursor ion (Figure 1.5.3) (Gallien et al., 2011). Unlike SRM, data-independent acquisition (DIA) (Venable et al., 2004) uses sequential window acquisition to fragment and quantify all precursor and product ions within a sample (Bruderer et al., 2015; Egertson et al., 2013; Geiger et al., 2010; Gillet et al., 2012) and offers high reproducibility and quantitation (Figure 1.5.3) (Selevsek et al., 2015).



**Figure 1.5.3** Schematic Depictions of SRM, DDA and DIA Analyses. A) In SRM, a single target precursor peptide ion is isolated in the Q1 quadrupole and then fragmented in q2 by collision-induced dissociation (CID). Specific, pre-selected product ions (typically three) are then detected and analysed in Q3. B) In DDA, the top #N precursors are isolated in Q1 based on mass-to-charge ( $m/z$ ) values and then fragmented together in q2 by CID or Higher-energy C-trap dissociation (HCD). The highly complex mixture of all product ions is then analysed in Q3 by an Orbitrap HR/AM or TOF mass analyzer. The analysis is repeated until the full  $m/z$  range has been covered in a stepwise fashion. C) In DIA, all precursors falling within a defined  $m/z$  window are isolated in Q1 and then fragmented together in q2 by CID or HCD. The highly complex mixture of all product ions is then analysed in Q3 by an Orbitrap HR/AM or TOF mass analyzer. The analysis is repeated until the full  $m/z$  range has been covered in a stepwise fashion (Hu et al., 2016; Shi et al., 2016; J. Li et al., 2021).

Advances in mass spectrometry have resulted in the development of a high-definition mass spectrometry (HDMS) technology with DIA and IMS function. HDMS improves peptide identification and protein sequence coverage in complex biological samples. Furthermore, an MS<sup>E</sup> implementation allows the collision energy to be alternated between low energy and high energy ramp in order to produce precursor- and product-ion spectra, respectively. The addition of IMS prior to peptide fragmentation aligns the precursor and its fragment ions in mobility drift time as well as chromatographic retention time to provide accurate coordinates for assigning precursor/product ion relationships. When ion mobility separation is used in conjunction with the MS<sup>E</sup> the workflow is referred to as HDMS<sup>E</sup> (H. Wang & Hanash, 2015). Due to analysis of all product ions in Q3, HDMS<sup>E</sup> is prone to detector saturation which may affect the dynamic range of protein quantitation. Several studies have demonstrated that this can be overcome using the “top3” method of protein quantitation to sum the three most intense peptides for each protein (Shliaha et al., 2013).

## 1.6 Background research

Amplification of the tyrosine kinase receptor ERBB2/HER2 in breast cancer is correlated with disease progression, metastasis and poor therapeutic response (Ross & Fletcher, 1999; Slamon et al., 1987). The mechanisms of downstream ERBB2 signalling and their effects on tumour progression remain ambiguous (Yarden & Sliwkowski, 2001) and thus the elucidation of pathways involved in ERBB2-dependent breast cancer is essential to understanding the biology of breast cancer and potentially finding biomarkers. Protein and gene expression profiling of cell models and breast tissues in the Timms laboratory previously identified gene products potentially implicated in ERBB2-dependent transformation (Durán et al., 2008; Gharbi et al., 2002; Timms et al., 2002; White et al., 2004; Worthington et al., 2010, 2017). Candidate proteins were selected from the profiling studies with respect to their relationship with ERBB2 expression and/or ERBB growth factor-dependent modulation. Interest was placed on candidates that are poorly characterised functionally or where a role in breast cancer is ambiguous. Candidates were validated and functionally characterised by analysing the effects of their siRNA-mediated knockdown on invasion, proliferation, adhesion and anchorage-independent growth in ERBB2 over-expressing breast cancer cell lines. The effects on global protein expression of knocking down the candidates (CPNE3, AGR2, STARD10 and ERBB2

itself) was also assessed using a quantitative stable isotope labelling of amino acids in cell culture (SILAC) combined with high resolution LC-MS/MS. Bioinformatics analysis of data using gene ontology, pathway and interaction enrichment was used to define the possible functional consequences of the siRNA-mediated knockdown, to identify novel targets and to link changes with cellular phenotype (**Table 1.6.1 and Table 1.6.2**). Finally, novel and specific ERBB2 signalling targets and putative sites of phosphorylation were determined using a combination of SILAC labelling, phosphopeptide enrichment and LC-MS/MS (Worthington et al., 2017). This research has revealed hitherto unknown components of the ERBB signalling network and suggests novel roles in cancer development for several poorly characterised gene products.

**Table 1.6.1** Table of selected gene product changes of interest in AGR2, CPNE3, ERBB2 and STARD10 siRNA knockdowns versus control siRNA from SILAC LC-MS/MS profiling of SKBR3 cells. Gene symbol, knockdown group and ratios of expression change for each knockdown versus non-targeting siRNA control are presented. The table below shows results of a phenotype analysis using cell-based assays where the numbers represent % versus the siRNA control. For example, siAGR2 transfected cells displayed 33% proliferation compared to siControl cells. The list represents potential candidates for validation and functional characterisation (Worthington, 2012).

Gene Name	Knockdown Group	Ratio siAGR2/ Ctrl	Ratio siCPNE3/ Ctrl	Ratio siERBB2/ Ctrl	Ratio siSTARD10/ Ctrl
AKR1C1	N/S	4.17		3.00	
AKR1C3	N/S		3.20	6.15	5.94
AP3D1	AGR2	0.65			
AIFM2	ERBB2			8.40	
CAD	STARD10	1.50		1.54	1.56
CAMK1G	STARD10				2.48
CAPS	STARD10				0.37
CAPS	ERBB2	0.67		0.65	
CAT	ERBB2	1.50		1.82	
ALCAM	ERBB2			1.69	
CHMP4A	ERBB2			0.34	
CHMP4B	CPNE3		1.54		
CKB	ERBB2	0.51		0.40	
CDK1	ERBB2			0.54	
DIEXF	AGR2	0.63			
EPS8L1	STARD10				1.59
EPPK1	CPNE3	0.64	0.62		
FER1L4	STARD10				5.94
FER1L6	CPNE3		15.15		
GRB7	ERBB2			1.53	
KPNA2	CPNE3		0.67		
ITGAV	CPNE3		5.44		
KRT10	AGR2	25.32			
KPRP	AGR2	0.04			
LAMC1	ERBB2			37.03	
LANCL1	STARD10	0.60		0.66	0.55
LRRTM3	ERBB2			0.00	
LIMA1	ERBB2			1.86	1.62
IQSEC3	AGR2	5.61			
LCN2	ERBB2			3.27	
PPL	ERBB2			1.52	
PBLD	ERBB2			4.05	
PI4KA	STARD10	1.57			1.64
PITPNB	STARD10				1.56
PLEC	ERBB2			1.82	
PCNA	ERBB2			0.61	
FAM49A	AGR2	2.35			
NDRG1	STARD10				0.49
RANGAP1	CPNE3		0.67		
RSU1	STARD10				1.54
RAB9A	ERBB2	1.63		1.93	
ARHGDIB	N/S	2.32	1.62	2.01	2.00
SMARCC1	ERBB2			0.64	
SMARCA5	ERBB2			0.63	
STX5	AGR2	0.65			
STX8	STARD10				1.67
TRIM16	STARD10		1.70		2.22

Phenotype analysis		% vs siCtrl			
		siAGR2	siCPNE3	siERBB2	siSTARD10
Adhesion		105	87	82	105
Anchorage-independent growth		68	97	45	118
Invasion		51	139	37	178
Proliferation		33	148	26	56

**Table 1.6.2** Table of differentially regulated proteins (>2 fold) in ERBB2-overexpressing human mammary luminal epithelial (HMLEC) cells (C3.6) versus parental HMLEC cells (HB4a) determined by SILAC LC-MS/MS profiling (Worthington et al., 2017).

Protein Name	Gene Name	Ratio (C3.6/HB4a)	Significance (*B-H FDR)	Function
Plastin-2	LCP1	29.09	*	cytoskeletal actin-binding, immune response
Transgelin	TAGLN	8.12	*	cell contraction
Single-stranded DNA-binding protein mitochondrial	SSBP1	7.08	*	DNA replication
Keratin type II cytoskeletal 5	KRT5	6.56	*	cytoskeletal intermediate filament protein
Interleukin-1 receptor antagonist protein	IL1RN	6.32	*	immune response, interleukin-1 receptor antagonist activity
3-hydroxyisobutyryl-CoA hydrolase mitochondrial	HIBCH	6.16	*	amino acid catabolic process
Aldehyde dehydrogenase family 1 member A3	ALDH1A3	6.09	*	retinal metabolic process
Interleukin-18	IL18	5.76	*	immune response, positive regulator of cytokine production
Keratin type II cytoskeletal 4	KRT4	4.97	*	cytoskeletal intermediate filament protein
Alpha-2-macroglobulin-like protein 1	A2M11	4.60	*	endopeptidase inhibitor activity
Zyxin	ZYX	4.34	*	cellular adhesion, signal transduction
Pirin	PIR	4.07	*	transcriptional regulation
Copine-3	CPNE3	3.95	*	vesicle-mediated transport
Phosphopantothienoylcysteine decarboxylase	PPCDC	3.90	*	glycolysis
Keratin type I cytoskeletal 13	KRT13	3.82	*	cytoskeletal intermediate filament protein
Acetyl-CoA acetyltransferase mitochondrial	ACAT1	3.76	*	amino acid catabolic process, ketone body and lipid metabolic process
PDZ and LIM domain protein 5	PDLM5	3.75	*	dendritic spine morphogenesis
LIM domain only protein 7	LMO7	3.74	*	unknown, ubiquitin ligase activity
Histone H1.4	HIST1H1E	3.54	*	nucleosome assembly
Histone H1.2	HIST1H1C	3.54	*	nucleosome assembly
Carboxymethylglutaminase homolog	CMBL	3.51	*	drug metabolism
Annexin A2	ANXA2	3.44	*	positive regulation of vesicle fusion
Calcineurin-like phosphoesterase domain-containing protein 1	CPPED1	3.39	*	unknown, hydrolase activity
Glutathione S-transferase kappa 1	GSTK1	3.34	*	unknown, glutathione transferase activity
Uncharacterized protein C12orf69	C12orf69	3.31	*	unknown, possible membrane protein
Vacuolar protein sorting-associated protein 13C	VPS13C	3.30	*	unknown, possibly protein transport
Beta-lactamase-like protein 2	LACTB2	3.25	*	unknown, hydrolase activity
Radixin	RDX	3.23	*	unknown, possible cytoskeletal actin-membrane linker
creatine kinase U-type mitochondrial	CKMT1B	3.20	*	creatine metabolic process
Sorbitol dehydrogenase	SORD	3.20	*	fructose biosynthetic process
Coatomer subunit gamma-2	COPG2	3.18	*	vesicle-mediated transport
Serpin B5	SERPINB5	2.99	*	tumour suppressor activity
Myosin phosphatase Rho-interacting protein	MPRIP	2.93	*	cytoskeletal actin-regulation
Proteasome subunit beta type-5	PSMB5	2.74	*	proteasomal degradation, antigen processing, cell cycle regulation, DNA damage response, apoptosis, immune response
cDNA FLJ56073 highly similar to Lipin-1	LPIN1	2.71	*	fatty acid metabolic process
Haloacid dehalogenase-like hydrolase domain-containing protein 2	HDHD2	2.67	*	unknown, hydrolase activity
Electron transfer flavoprotein subunit alpha mitochondrial	ETFA	2.65	*	respiratory electron transport chain
Tripartite motif-containing protein 29	TRIM29	2.61	*	transcriptional regulation
E3 ubiquitin-protein ligase NEDD4	NEDD4	2.58	*	signal transduction, proteasomal degradation
Vacuolar protein sorting-associated protein 4B	VPS4B	2.54	*	vesicle-mediated transport
60S ribosomal protein L14	RPL14	2.52	*	protein synthesis (ribosome constituent)
Transportin-3	TNPO3	2.48	*	protein transport
Isoform M2 of pyruvate kinase isozymes M1/M2	PKM2	2.47	*	cell death, transcriptional regulation, glycolysis
14-3-3 protein zeta/delta	YWHAZ	2.45	*	signal transduction, anti-apoptosis
Electron transfer flavoprotein subunit beta	ETFB	2.44	*	respiratory electron transport chain
Annexin A4	ANXA4	2.36	*	signal transduction, anti-apoptosis
Tropomyosin alpha-1 chain	TPM1	2.36	*	cytoskeletal structural protein, muscle contraction
Programmed cell death protein 4	PDCD4	2.36	*	apoptosis, inhibition of translation, negative regulation of cell cycle, tumour suppressor activity
Peptidyl-prolyl cis-trans isomerase B	PIIB	2.34	*	protein folding
Asparaginyl-tRNA synthetase cytoplasmic	NARS	2.32	*	unknown, tRNA ligase activity
Aldose reductase	AKR1B1	2.32	*	carbohydrate metabolic process, response to stress, carbonyl reduction
Filamin-C	FLNC	2.31	*	cell junction assembly
Platelet-activating factor acetylhydrolase IB subunit beta	PAFAH1B2	2.30	*	unknown, 1-alkyl-2-acetyl-glycerophosphocholine esterase activity
Tumor protein D52	TPD52	2.28	*	secretion
Protein AHNK2	AHNK2	2.27	*	unknown
Ras-related protein Rab-2A	RAB2A	2.27	*	vesicle-mediated transport
Niban-like protein 1	FAM129B	2.26	*	unknown, may function in invasion/anti-apoptosis
Oligoribonuclease mitochondrial	REXO2	2.25	*	nucleotide metabolic process
Galectin-3	LGALS3	2.24	*	immune response
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex mitochondrial	DLAT	2.23	*	glycolysis
Quinone oxidoreductase	CRYZ	2.23	*	xenobiotic catabolic process
Proteasome subunit beta type-6	PSMB6	2.18	*	proteasomal degradation, antigen processing/presentation, cell cycle regulation, DNA damage response, apoptosis, immune response
Deoxyuridine 5-triphosphate nucleotidohydrolase mitochondrial	DUT	2.17	*	DNA replication
Acyl-CoA synthetase family member 2 mitochondrial	ACSF2	2.17	*	fatty acid metabolic process
Plastin-3	PLS3	2.14	*	cytoskeletal actin-bundling
Vesicle-fusing ATPase	NSF	2.13	*	vesicle-mediated transport
Similar to Zinc finger CCH domain-containing protein 15	ZC3H15	2.11	*	unknown, possibly anti-degradation
Syntenin-1	SDCBP	2.08	*	cytoskeletal actin-organisation, axon guidance
Protein NDRG1	NDRG1	2.07	*	tumour suppressor activity, vesicle-mediated transport, cell cycle regulation, apoptosis, growth/differentiation
Ankyrin	RAI14	2.04	*	unknown, potential cytoskeletal protein
Ubiquitin-conjugating enzyme E2 H	UBE2H	2.03	*	proteasomal degradation
Plastin-1	PLS1	2.00	*	cytoskeletal actin-bundling
Histone H2A type 2-B	HIST2H2AB	1.99	*	nucleosome assembly

Cullin-4B	CUL4B	1.98	DNA damage response, cell cycle progression, proteasomal degradation
Adenylyl cyclase-associated protein 1	CAP1	1.98	axon guidance, regulation of cell polarity, signal transduction
Uveal autoantigen with coiled-coil domains and ankyrin repeats	UACA	1.97	apoptosis
Actin cytoplasmic 2	ACTG1	1.95	cytoskeletal structural protein
Inorganic pyrophosphatase	PPA1	1.95	diphosphate metabolic process, tRNA aminoacylation
Glucosamine 6-phosphate N-acetyltransferase	GNPNAT1	1.93	N-linked glycosylation
COP9 signalosome complex subunit 7a	COPS7A	1.93	cullin deneddylation
Sorting nexin-1	SNX1	1.91	vesicle-mediated transport
Adenylyl cyclase-associated protein 1	CAP1	1.91	axon guidance, regulation of cell polarity, signal transduction
Dihydropyrimidinase-related protein 2	DPYSL2	1.88	signal transduction, regulation of cell polarity, axon guidance
Quinone oxidoreductase PIG3	TP53I3	1.88	apoptosis by oxidative stress
Tripartite motif-containing protein 16	TRIM16	1.87	differentiation, transcriptional regulation
Xaa-Pro dipeptidase	PEPD	1.86	proteolysis
Interferon-inducible double stranded RNA-dependent protein kinase activator A	PRKRA	1.86	immune response, apoptosis, inhibition of translation
Puromycin-sensitive aminopeptidase	NPEPPS	1.86	proteolysis, antigen processing/presentation
Programmed cell death protein 6	PDCD6	1.85	apoptosis
Eukaryotic translation initiation factor 3 subunit J	EIF3J	1.85	protein synthesis
RNA polymerase II-associated protein 1	RPAP1	1.84	transcriptional regulation
3-ketoacyl-CoA thiolase mitochondrial	ACAA2	1.83	cholesterol biosynthetic process, anti-apoptosis
Sorcin	SRI	1.82	muscle contraction, signal transduction
LDLR chaperone MESD	MESDC2	1.81	signal transduction, protein folding
Keratin type II cytoskeletal 6A	KRT6A	1.80	cytoskeletal intermediate filament protein
Nucleoside diphosphate kinase A	NME1	1.79	nucleoside triphosphate synthesis, proliferation/differentiation, signal transduction, anti-apoptosis
Activated RNA polymerase II transcriptional coactivator p15	SUB1	1.78	transcriptional regulation
Calpain-2 catalytic subunit	CAPN2	1.76	protease involved in cytoskeletal remodelling and signal transduction
Ribosomal L1 domain-containing protein 1	RSL1D1	1.73	protein synthesis (ribosome constituent)
Signal transducer and activator of transcription 1-alpha	STAT1	0.51	cytokine signal transduction, transcriptional regulation
Very long-chain specific acyl-CoA dehydrogenase mitochondrial	ACADVL	0.49	fatty acid beta-oxidation
Peptidyl-prolyl cis-trans isomerase FKBP4	FKBP4	0.49	protein folding, regulation of microtubule dynamics
cDNA FLJ54365 highly similar to DNA replication licensing factor MCM4	MCM4	0.48	DNA replication, cell cycle regulation
Proteasome activator complex subunit 1	PSME1	0.46	proteasomal degradation, antigen processing/presentation, cell cycle regulation, DNA damage response, apoptosis, immune response (IFN $\gamma$ -induced)
Bifunctional purine biosynthesis protein PURH	ATIC	0.46	purine base metabolic process
DNA replication licensing factor MCM6	MCM6	0.46	DNA replication, cell cycle regulation
DNA replication licensing factor MCM2	MCM2	0.45	DNA replication, cell cycle regulation
RuvB-like 2	RUVBL2	0.45	DNA damage repair, transcriptional regulation, protein folding
60 kDa heat shock protein mitochondrial	HSPD1	0.44	protein chaperone/re-folding, stress response, immune response
DNA replication licensing factor MCM7	MCM7	0.44	DNA replication, cell cycle regulation
Protein TFG	TFG	0.42	unknown, possibly positive regulation NFkB cascade
Alkaline phosphatase placental type precursor	ALPP	0.42	unknown, alkaline phosphatase activity
Ras-related protein Rab-6A	RAB6A	0.41	vesicle-mediated transport
Cellular tumor antigen p53	TP53	0.41	DNA damage response, tumour suppressor, apoptosis, negative regulation of cell cycle, transcriptional regulation
Transferrin receptor protein 1	TFRC	0.40	iron-homeostasis
Abhydrolase domain-containing protein 10 mitochondrial	ABHD10	0.40	unknown, peptidase activity
TUBB6 protein	TUBB6	0.40	cytoskeletal microtubule protein, protein folding
D-3-phosphoglycerate dehydrogenase	PHGDH	0.40	amino acid metabolic process
Leukocyte elastase inhibitor	SERPINB1	0.40	endopeptidase inhibitor activity
Reticulocalbin-1	RCN1	0.40	unknown, may regulate Ca $^{2+}$ binding activities in ER
Dipeptidyl peptidase 1	CTSC	0.40	proteolysis
Ornithine aminotransferase mitochondrial	OAT	0.39	amino acid metabolic process
Glycine dehydrogenase [decarboxylating] mitochondrial	GLDC	0.38	amino acid metabolic process
Coiled-coil-helix-coiled-coil-helix domain-containing protein 6	CHCHD6	0.38	mitochondrial maintenance, DNA damage response
Nuclear autoantigenic sperm protein	NASP	0.38	DNA replication, cell cycle progression, proliferation
Proteasome subunit beta type-8	PSMB8	0.37	proteasomal degradation, antigen processing/presentation, cell cycle regulation, DNA damage response, apoptosis, immune response (IFN $\gamma$ -induced)
Proteasome activator complex subunit 2	PSME2	0.37	proteasomal degradation, antigen processing, cell cycle regulation, DNA damage response, apoptosis, immune response (IFN $\gamma$ -induced)
Tryptophanyl-tRNA synthetase cytoplasmic	WARS	0.36	tRNA aminoacylation, immune response (IFN $\gamma$ -induced), angiogenesis, negative regulation of proliferation
cDNA FLJ52712 highly similar to Tubulin beta-6 chain	TUBB6	0.36	cytoskeletal microtubule protein, protein folding
Secernin-1	SCRN1	0.34	exocytosis
Uncharacterized protein C11orf73	C11orf73	0.34	unknown, possibly secretion
Structural maintenance of chromosomes flexible hinge domain-containing protein 1	SMCHD1	0.33	unknown, possibly chromatin organisation
Interferon-induced double-stranded RNA-activated protein kinase	EIF2AK2	0.33	immune response (IFN $\alpha$ -induced), inhibition of translation
Stromal cell-derived factor 2	SDF2	0.32	glycosylation
Histone H1x	H1FX	0.32	nucleosome assembly
Cytosol aminopeptidase	LAP3	0.30	proteolysis
Phosphoserine aminotransferase	PSAT1	0.30	amino acid metabolic process
Hydroxymethylglutaryl-CoA synthase cytoplasmic	HMGCS1	0.29	cholesterol biosynthetic process
Argininosuccinate synthase	ASS1	0.29	urea cycle
Fascin	FSCN1	0.29	cytoskeletal actin-regulation, proliferation, migration
Integrin beta-4	ITGB4	0.28	epithelial cell laminin receptor, hemidesmosome assembly, migration
SAM domain and HD domain-containing protein 1	SAMHD1	0.28	immune response (IFN $\gamma$ -induced)
Signal transducer and activator of transcription 2	STAT2	0.28	cytokine signal transduction, transcriptional regulation
Serpin H1	SERPINH1	0.23	protein chaperone/re-folding
Endoplasmic reticulum aminopeptidase 1	ERAP1	0.23	antigen processing/presentation, immune response (IFN $\gamma$ -induced), angiogenesis, differentiation
Serine protease HTRA1	HTRA1	0.21	proteolysis
Integrin alpha-6	ITGA6	0.21	epithelial cell laminin receptor, hemidesmosome assembly, migration
Complement C3 fragment	C3	0.18	inflammatory response, complement cascade, signal transduction
Proteasome subunit beta type-9	PSMB9	0.15	proteasomal degradation, antigen processing/presentation, cell cycle regulation, DNA damage response, apoptosis, immune response (IFN $\gamma$ -induced)



Basal cell adhesion molecule	BCAM	0.13	*	laminin $\alpha$ -5 receptor
Interferon-induced GTP-binding protein Mx1	MX1	0.13	*	immune response (IFN-induced), apoptosis
Superoxide dismutase [Mn] mitochondrial	SOD2	0.13	*	redox homeostasis
Retinal dehydrogenase 1	ALDH1A1	0.12	*	retinal metabolic process, xenobiotic metabolic process, aldehyde metabolic process
Intercellular adhesion molecule 1	ICAM1	0.11		cellular adhesion, heterophilic cell-cell adhesion, integrin $\alpha$ -L/ $\beta$ -2 ligand, immune response
Interferon-induced 35 kDa protein	IFI35	0.10		immune response (IFN $\gamma$ -induced)
N-myc-interactor	NMI	0.09	*	immune response (IFN $\gamma$ -induced), transcriptional regulation
Fatty acid-binding protein epidermal	FABP5	0.08	*	fatty acid transport, differentiation

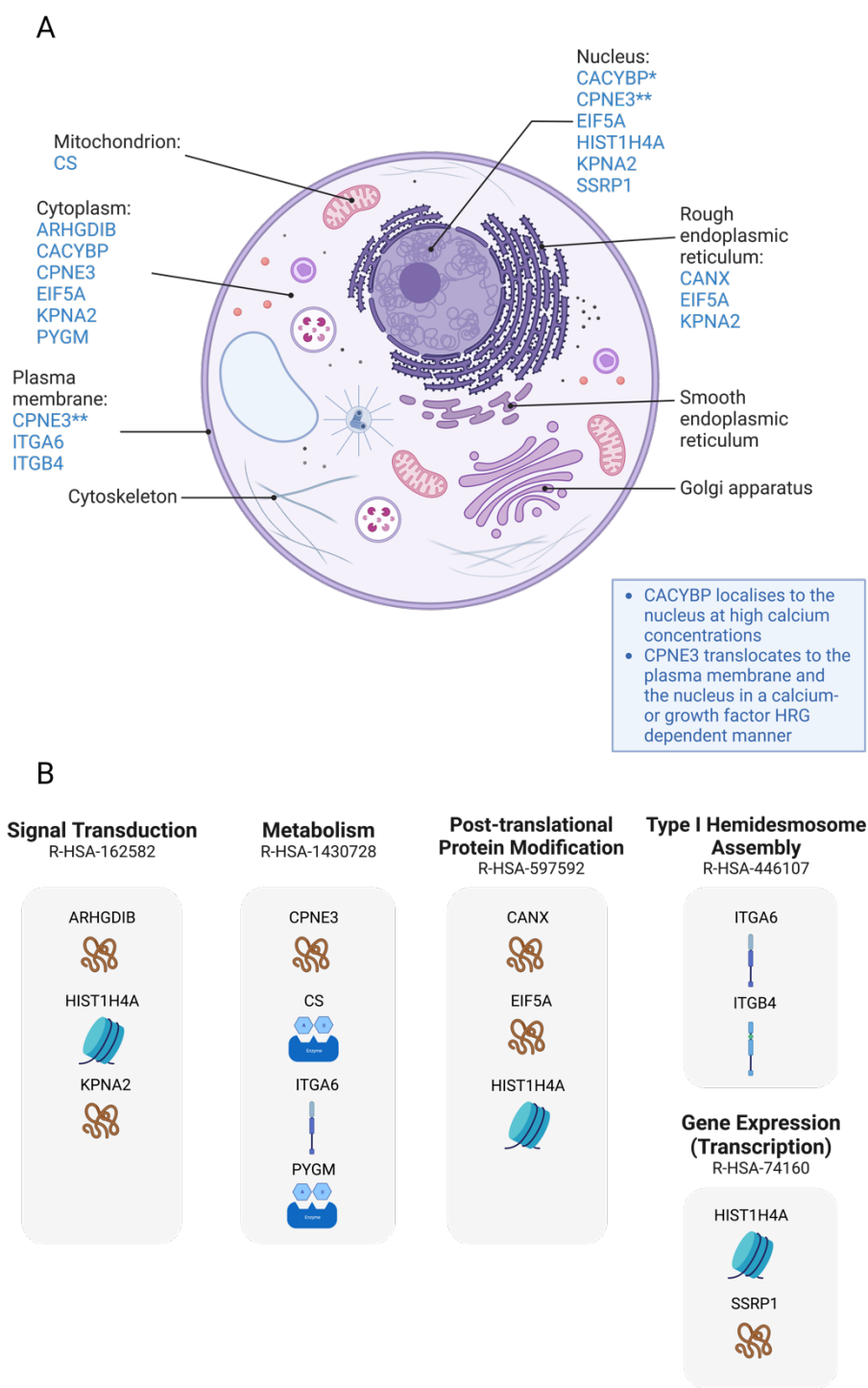
## 1.7 Candidates

### 1.7.1 Rho GDP-dissociation inhibitor 2 (ARHGDIB)

ARHGDIB is a member of the Rho (or ARH) protein family and other Ras-related small GTP binding proteins. The GTP-binding proteins are active only in the GTP-bound state and ARHGDIB is a GDP-dissociation inhibitor (GDI), one of at least 3 classes of proteins that tightly regulate cycling between the GTP-bound and GDP-bound states (Hoffman et al., 2000). ARHGDIB regulates the GDP/GTP exchange reaction of the Rho proteins by inhibiting the dissociation of GDP from Ras-like GTPases, and the subsequent binding of ATP to them (Adra et al., 1993; Scherle et al., 1993). It also plays a key role in regulating the reorganisation of the actin cytoskeleton mediated by Rho family members (Leffers et al., 1993). A recent study found that the depletion of ARHGDIB decreased the migration of breast cancer cells and linked ARHGDIB to epithelial-mesenchymal transition (X. Wang et al., 2020). This makes ARHGDIB a possible drug target for a therapeutic mechanism of action that targets the reorganisation of the actin cytoskeleton during migration of breast cancer cells.

ARGHDIB is located in the cytoplasm and has a functional role in signal transduction (Figure 1.7.1). This role in signal transduction is thought to regulate cytoskeletal organisation through phosphorylation of ARHGDIB on the tyrosine 24 (Y24) residue by Src kinase (Y. Wu et al., 2009). The role of Src as an upstream regulator of ARHGDIB is plausible because Src has been shown to be a non-receptor protein tyrosine kinase with a key role in regulating cell-to-matrix adhesion, migration, and junctional stability (Frame, 2004). Furthermore, activation of the ERBB2 receptor and subsequent downstream signalling have been described as leading to Src up-regulation and activation to mediate breast cancer invasion and metastasis (Tan et al., 2005). Interestingly, SILAC LC-MS/MS profiling of ERBB2 overexpressing SKBR3 cells found that ARHGDIB was upregulated >1.5 fold following the

siRNA-mediated knockdown of AGR2, CPNE3, ERBB2 and STARD10 versus control siRNA (Table 1.6.1) (Worthington, 2012).



**Figure 1.7.1** Biological features of biomarker candidates for ERBB2 overexpressing breast cancer. A) The figure shows the subcellular location of proteins presented as candidate biomarkers in this study retrieved from UniProtKB. B) This schematic provides an overview of the functional intersection of candidate biomarkers and groups them under common Reactome pathways. The figure was created using BioRender ([www.biorender.com](http://www.biorender.com)).

### 1.7.2 Calcyclin-binding protein (CacyBP)

CacyBP is primarily a cytoplasmic protein at low calcium concentrations. However, it has been shown to localize to both the nucleus and cytoplasm after a retinoic acid (RA) induction and calcium increase in neuroblastoma cells. Moreover, the nuclear fraction may be phosphorylated (Figure 1.7.1) (J. Wu et al., 2003). CacyBP is known to interact with members of the S100 family, tubulin, and Siah-1 and Skp-1 (Schneider et al., 2007). S100 is a family of low molecular weight proteins that contains more than 20 family members and comprises the largest subfamily of EF-hand calcium - binding proteins (J. Wu et al., 2003). The interaction of CacyBP and S100 proteins is particularly interesting because members of the S100 family have been shown to regulate  $\text{Ca}^{2+}$  homeostasis, protein phosphorylation, and degradation by interacting with several other targets, such as RAGE, p53, Jab-1, and matrix metalloproteinases, affecting cell proliferation, metastasis and many other biological events related to cancer progression (Santamaria-Kisiel et al., 2006). CacyBP may play a role in cytoskeletal reorganisation by interacting with S100 proteins that regulate all three major constituents of cytoplasmic cytoskeleton, i.e. MTs, IFs and microfilaments (MFs), and tropomyosin and myosin (Donato, 2001).

CacyBP/SIP expression has also been linked with the clinical progression of breast cancer, where it is evident in more advanced clinical progression of breast cancer (N. Wang et al., 2010). However, in a separate study that investigated the relationship between CacyBP and COX-2 in breast cancer, the knockdown of CacyBP gene using siRNA enhanced the proliferation and invasion ability of breast cancer cells, which was dependent on COX-2 expression suggesting that it may play a role as a tumour suppressor (Nie et al., 2010). More recent studies have shown that the overexpression of CacyBP and not the knockdown, enhance the proliferation and invasion ability of certain cancers. For instance, the overexpression of CacyBP in NSCLC cell lines results in differential expression of epithelial-mesenchymal transition (EMT) markers including E-cadherin, N-cadherin, Snail1, Vimentin and it may promote the proliferation and invasion of NSCLC cells by regulating the Akt signalling pathway (Y. J. Xu et al., 2021). ERBB2 and COX-2 expression is known to be upregulated in NSCLC lung cancer and ERBB2 has been shown to elevate COX-2 expression through the activation of MEK/ERK pathway, which subsequently induced cell proliferation and invasion via AKT pathway (Chi

et al., 2016). This is particularly interesting because ERBB2 amplification in breast cancer is known to activate the PI3K/Akt pathway independent of ERBB3. The resulting hyperactivation of this PI3K/Akt signalling cascade is associated with resistance to ERBB2 targeting therapies (Carmona et al., 2016; Ruiz-Saenz et al., 2018).

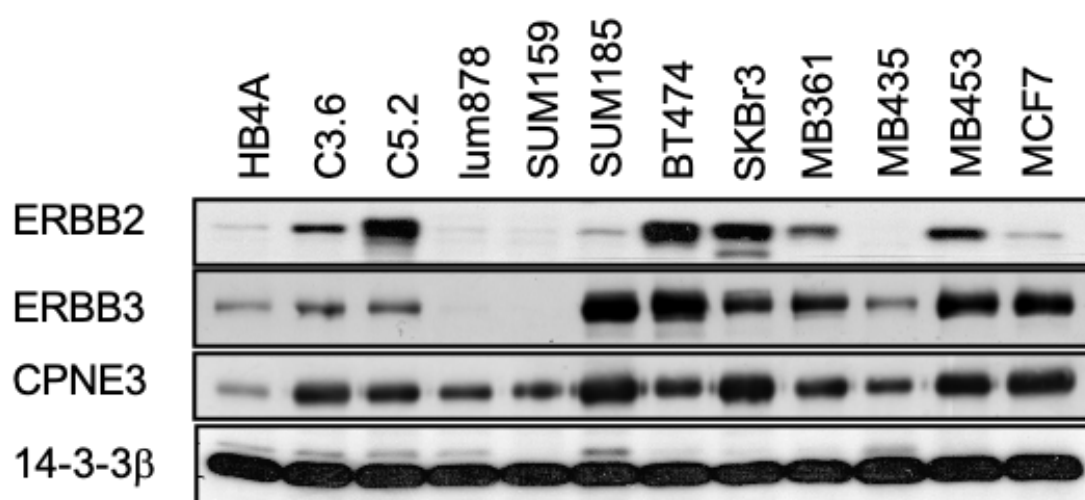
### 1.7.3 Calnexin (CANX)

CANX is a member of the calreticulin family and the gene encodes a member of the calnexin family of molecular chaperones. The calcium-binding protein interacts transiently with newly synthesized N-linked glycoproteins in the endoplasmic reticulum, facilitating post-translational protein modifications such as protein folding and assembly (Figure 1.7.1). It may play a central role in the quality control of protein folding by retaining incorrectly folded protein subunits within the endoplasmic reticulum for degradation (Kleizen & Braakman, 2004; Olsen et al., 2013). CANX is also associated with partial T-cell antigen receptor complexes that escape the endoplasmic reticulum of immature thymocytes and it may function as a signalling complex regulating thymocyte maturation (Okazaki et al., 2000; Wiest et al., 1995). Low expression or defective CANX may lead to a higher risk of developing brain metastasis due to the defects in T cell-based immunosurveillance in primary breast cancer patients (Y. Liu et al., 2012). Furthermore, upon viral infection, the levels of SMAR1 have been shown to undergo a significant increase that results in a reduced expression of CANX and an increase in MHC I antigen presentation (Alam et al., 2019). Stimulation of HER2-positive breast cancers by heregulin-beta1 (HRG) has been shown to induce a rapid redistribution of CANX from vesicle-like structures in the cell cytoplasm to the perinuclear area and to the cell membrane. The co-localisation and physical interaction of CANX with the HER2 growth factor receptor and an increase in CANX protein levels linked to the progressive stages of human breast cancer have also been shown (F. Li et al., 2001).

### 1.7.4 Copine III (CPNE3)

CPNE3 (Copine III) is a predicted calcium-dependant, phospholipid binding protein first isolated from *Paramecium tetraurelia*. CPNE3 binds to calcium, phospholipids, and inositol polyphosphates via two

C2 domains (Creutz et al., 1998). CPNE3 was later found to share sequence homology to kinases but lacks conserved residues required for kinase activity (Caudell et al., 2000). CPNE3 is located in the cytoplasm and translocates to the plasma membrane or the nucleus in a calcium or growth factor HRG dependent manner (Figure 1.7.1). It was shown to interact with ERBB2 and promote cell migration (Heinrich et al., 2010). Furthermore, CPNE3 expression has been shown to correlate with ERBB2 and ERBB3 overexpression in a panel of HMLECs and breast tumour cells (Figure 1.7.2). Additionally, the expression of CPNE3 was found to be higher in the ERBB2 overexpressing C3.6 clone compared to the parental HB4a cell line. Treatment with growth factors induced a moderate increase in CPNE3 expression in the HB4a cells (Bertani, 2005; Gharbi et al., 2002; White et al., 2004). A study found JAB1 as a direct interactor of CPNE3 and that binding of CPNE3 to ERBB2 correlates with Jab1 overexpression in SKBR3 cells. JAB1 expression also led to the activation of protein kinase B (AKT) and phosphatidylinositol 3 (PI3) kinase which are key mediators of downstream ERBB2 signalling. This suggests a possible role for the CPNE3 and Jab1 interaction in regulating downstream ERBB2 signalling pathways (Choi et al., 2016). In our laboratory, knockdown of CPNE3 in SKBR3 cells increased cellular invasiveness and proliferation suggesting a tumour suppressor function for CPNE3 (Worthington, 2012).



**Figure 1.7.5** Expression of CPNE3 correlates with ERBB2 and ERBB3 status in a panel of HMLECs and breast tumour lines. The candidates are shown alongside 14-3-3β as a loading control (Worthington, 2012).

### 1.7.5 Citrate synthase, mitochondrial (CS)

CS is a member of the citrate synthase family located in the mitochondria (Figure 1.7.1) and the enzyme catalyses the synthesis of citrate from oxaloacetate and acetyl coenzyme A during the Krebs tricarboxylic acid cycle (Usher et al., 1994). CS is found in nearly all cells capable of oxidative metabolism and unlikely to be a plausible drug target for cancer. To date, little is known about the role of CS in ERBB2 overexpressing breast cancers. Nevertheless, the downregulation of CS expression in Embryonic kidney 293-T (HEK293T) cells has been linked to low levels of ATP production, excessive superoxide formation and cell apoptosis (Cai et al., 2017). Furthermore, RNAi-mediated CS knockdown in human cervical carcinoma cells has been shown to induce severe defects in respiratory activity and marked decreases in ATP production, but great increases in glycolytic metabolism. This change in cell metabolism accelerated cancer cell metastasis and proliferation for *in vitro* assays and *in vivo* tumour xenograft models.

Citrate, a substance related to *de novo* fatty acid synthesis and tricarboxylic acid (TCA) cycle and synthesised by CS, has a pivotal role in cell survival. The potential for CS to greatly increase glycolytic metabolism while reducing the citrate contribution to the TCA cycle links CS expression to the Warburg effect, a phenomenon where cancer cells shift energy production from the TCA cycle and OXPHOS in the mitochondria to a less efficient process in aerobic glycolysis (Alfarouk, 2016; Peng et al., 2019). The aerobic glycolysis consists of a high level of glucose uptake, glycolysis and lactic acid fermentation that takes place in the cytosol, not the mitochondria, even in the presence of abundant oxygen (Alfarouk, 2016). The Warburg effect is also known to be directly linked to tumour malignancy via induction of CS expression dependant morphological changes characteristic of the epithelial-mesenchymal transition (EMT) in human cervical carcinoma cells (C.-C. Lin et al., 2012). CS's role in metabolic reprogramming and morphological changes during EMT of cervical cancer cells make it an interesting potential biomarker for ErbB2 overexpressing breast cancer, particularly where the metabolic reprogramming is facilitated by downstream signalling of ErbB2.

### 1.7.6 Eukaryotic translation initiation factor 5A-1 (eIF5A)

EIF5A is a member of the eIF-5A family and it has an important function as an mRNA binding protein involved in mRNA turnover, and is located in the cytoplasm, nucleus and rough endoplasmic reticulum (Figure 1.7.1). Its subcellular location is linked to a suggested role in translation of a specific subset of mRNAs, particularly, those involved in the cell cycle progression (G1/S transition) (H. A. Kang & Hershey, 1994; Park et al., 1993, 1997). EIF5A is linked to the Post Translational Protein modification Reactome Pathway (R-HSA-597592) along with two other candidates CANX and HIST1H4 (Figure 1.7.1). This role in post translational modification of proteins contributes to actin dynamics (Hofmann et al., 2001), mRNA decay and maintenance of cell integrity during stress response (Galvão et al., 2013; Hoque et al., 2017; Smeltzer et al., 2021). However, its role in malignancy has not been well established. Recently, EIF5A2 has been shown to play an important role in doxorubicin chemoresistance (Y. Liu et al., 2015), to be an effective drug target for reducing cell viability by inhibiting c-Myc expression *in vitro* (Shah et al., 2016) and identified as a candidate target gene of miR-375 which may be a potential key regulator of EIF5A2 in human breast cancer cells (J. Liu et al., 2019). Nonetheless, the expression of eIF5A in breast cancer tissues and its association with the clinicopathology of patients with breast cancer have not been reported (Ning et al., 2020).

#### 1.7.7 Histone H4 (HIST1H4A)

HIST1H4A is a member of the histone H4 family located in the nucleus (Figure 1.7.1) and is responsible for nucleosome structure of the chromosomal fiber in eukaryotes. Two molecules of HIST1H4A form an octamer with two molecules of the other core histones (H2A, H2B and H3), around which approximately 146 bp of DNA is wrapped in repeating units, called nucleosomes (Bellard et al., 1976). Nucleosomes wrap and compact DNA into chromatin, limiting DNA accessibility to the cellular machineries which require DNA as a template (Aragay et al., 1988; Moore et al., 1997; Simpson, 1976). HIST1H4A thereby plays a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability, and is linked to the Post Translational Protein modification Reactome Pathway (R-HSA-597592) along with CANX and EIF5A (Figure 1.7.1). ERBB2/HER2-overexpressing breast cancer cells have been found to contain significantly higher levels of acetylated

and phosphorylated histone H3, and acetylated histone H4 associated with the HER2 promoter (Mishra et al., 2001). Histone H4 is associated with arginine-rich fractions (F2a1) and undergoes more active histone acetylation than in the lysine rich fractions (F1, F2b) (JOHNS & BUTLER, 1962; Marzluff & McCarty, 1970; Wilhelm & McCarty, 1970). This increase in levels of acetylated histone H4 in association with the HER2 promoter, is likely to contribute to post translational modifications that regulate nucleic acids and mRNA turnover in association with EIF5A (Figure 1.7.1).

### 1.7.8 Integrin alpha 6 (ITGA6)

The integrins are a family of integral membrane proteins by which cells attach to extracellular matrix (ECM) proteins and in some cases, mediate cell-cell adhesion by forming interactions with the cell cytoskeleton (Figure 1.7.1) (Hynes, 1987, 1992; Pinkstaff et al., 1999). ITGA6 is a receptor for glycoproteins of the extracellular matrix of epithelial cells known as laminin and a major component of highly specialised integrin-mediated epithelial attachment structures called hemidesmosomes (Borradori & Sonnenberg, 1999; Ieguchi et al., 2010). ITGA6 is linked to the Type I Hemidesmosome Assembly Reactome Pathway (R-HSA-446107) along with ITGB4 and Metabolism Pathway Reactome Pathway (R-HSA-1430728) along with other candidates CPNE3, CS and PYGM (Figure 1.7.1). Studies involving mouse models have revealed the role of ITGA6 in many developmental processes such as cortical and retinal lamination, apical ectodermal ridge formation and organogenesis (De Arcangelis et al., 1999; Georges-Labouesse et al., 1998). Of greater interest is the role played by ITGA6 in altering survival signalling processes mediated by neuregulin with a switch from dependence on the PI3K pathway to the mitogen-activated protein kinase (MAPK) pathways (Colognato et al., 2002). The integrin heterodimer ITGA6/B4 ( $\alpha 6\beta 4$ ) has been found to play a specific role in breast cancer progression and metastasis (Diaz et al., 2005; Friedrichs et al., 1995; J. L. Jones et al., 1997; Mukhopadhyay et al., 1999) and there is growing evidence that cross-talk between HER2 and  $\alpha 6\beta 4$  promotes tumour aggressiveness. This evidence points towards a cross talk between receptor tyrosine kinases and the Src/FAK complex activated by TGF- $\beta$ .

The activation of Src-FAK has been shown to integrate ErbB receptor and integrin signalling to induce cell migration and survival during breast cancer progression (Carraway & Sweeney, 2006; Fan et al.,



2011; H. Wang et al., 2010; S. E. Wang et al., 2009). The knockdown of ITGA6 expression was found to impair FAK-mediated PI3K signalling to Akt by abolishing the formation of a functional complex comprised of receptor tyrosine kinase ERBB2, FAK and integrin (S. E. Wang et al., 2009). The cross talk between HER2 and  $\alpha 6 \beta 4$  has been shown by the phosphorylation of Src-FAK and integrin-FAK signalling has been linked to resistance to either trastuzumab or lapatinib (X. H. Yang et al., 2010). Integrin  $\alpha 6 \beta 4$  is known to amplify downstream pathways such as PI3K, AKT, MAPK, and the Rho family small GTPases by cooperating with growth factor receptors including EGFR, ERBB-2, and c-Met. Moreover, it upregulates and activates key tumour-promoting transcription factors such as the NFATs and NF- $\kappa$ B (Stewart & O'Connor, 2015). Inhibition of HER2/integrin signalling has been shown to suppress *in vitro* and *in vivo* breast tumour growth (P. Gupta & Srivastava, 2014). SILAC-based LC-MS/MS profiling of HMLECs found that ITGA6 was downregulated in response to ERBB2 overexpression (Worthington et al., 2017).

#### 1.7.9 Integrin beta 4 (ITGB4)

ITGB4 is a component of the alpha-6/beta-4 integrin and is primarily expressed in epithelial cells, where it plays a role in adhesion and is recognized as a receptor for most of the known laminins (Mercurio, 1995). ITGB4 is a structurally distinct integrin subunit with an intracellular domain of approximately 1000 amino acids located at the plasma membrane (Figure 1.7.1) (Hemler et al., 1989). The intracellular domain of the ITGB4 links the beta 4 subunits to intermediate filaments in hemidesmosomes (Wilhelmsen et al., 2006). ITGB4 is predominantly associated with hemidesmosome organization and function and plays a pivotal role in modulating tumour progression (Bachelder et al., 1999; Chao et al., 1996; Jauliac et al., 2002; Lipscomb et al., 2005; O'Connor et al., 1998; Rabinovitz & Mercurio, 1997; Santoro et al., 2003; Shaw et al., 1997; Wilhelmsen et al., 2006). ITGB4 is linked to the Type I Hemidesmosome Assembly Reactome Pathway (R-HSA-446107) along with ITGA6 (Figure 1.7.1). SILAC LC-MS/MS profiling of HMLECs showed that like its binding partner ITGA6, ITGB4 was downregulated in response to ERBB2 overexpression in HMLECs (Worthington et al., 2017).

#### 1.7.10 Importin subunit alpha-1 (KPNA2)

KPNA2 is a member of the importin alpha family located in the cytoplasm, rough endoplasmic reticulum or nucleus, and functions in nuclear protein import as an adapter protein for nuclear receptor importin beta (KPNB1) (Figure 1.7.1). It binds specifically and directly to substrates containing either a simple or bipartite NLS motif. The docking of the importin substrate complex to the nuclear pore complex (NPC) is mediated by KPNB1 through binding to nucleoporin FxFG repeats and the complex is subsequently translocated through the pore by an energy requiring, Ran-dependent mechanism (Görlich, 1998; Lange et al., 2007). The three components of the translocated complex are separated by binding of Ran to KPNB1 at the nucleoplasmic side of the NPC. KPNA2 and KPNB1 are re-exported from the nucleus to the cytoplasm where GTP hydrolysis releases Ran from importin. The directionality of nuclear import is thought to be conferred by an asymmetric distribution of the GTP- and GDP-bound forms of Ran between the cytoplasm and nucleus (Chook & Blobel, 2001; Goldfarb et al., 2004). KPNA2 is linked to the Signal Transduction Reactome Pathway (R-HSA-162582) along with ARHGDIB and HIST1H4A (Figure 1.7.1) which alludes to a potential mechanism involving GDP dissociation and nuclear import. Increased expression of KPNA2 has been shown to predict unfavourable prognosis in ovarian cancer patients and to promote colorectal cancer development by activating the PI3K/AKT pathway (Cui et al., 2021; B. Huang et al., n.d.). A separate study concluded that the silencing of KPNA2 expression inhibits the proliferation, migration and invasion of breast cancer cells by blocking NF- $\kappa$ B signalling and c-Myc nuclear translocation *in vitro* (Duan et al., 2020). The downregulation of KPNA2 was previously observed following specific knock-down of target candidate CPNE3 in SKBR3 breast cancer cells (Worthington, 2012) (Table 1.6.1) and KPNA2 may function as a downstream signalling candidate in ERBB2 overexpressing breast cancer.

#### 1.7.11 Glycogen phosphorylase (PYGM)

PYGM is an allosteric enzyme located in the cytoplasm that catalyses the rate-limiting step in glycogen catabolism, the phosphorolytic cleavage of glycogen to produce glucose-1-phosphate and plays a central role in maintaining cellular and organismal glucose homeostasis. PYGM is linked to the Metabolism Reactome Pathway (R-HSA-1430728) along with CPNE3, CS and ITGA6 (Figure

1.7.1). The glycogen phosphorylase isoforms, PYGB and PYGM, have been shown to be expressed at very low levels compared to PYGL in EMT-derived breast cancer cells. The study demonstrated that EMT induces multiple metabolic changes such as enhanced glycolysis in HER2-positive, epithelial BT- 474 and ER $\alpha$ -positive, epithelial MCF-7 breast cancer cells, suggesting that breakdown of glycogen is promoted during EMT in order to provide carbons for glycolysis (Kondaveeti et al., 2015). While cells get less energy from each molecule of glucose during aerobic glycolysis, the process is faster and has a higher rate of energy turnover than would be obtained per time unit through the TCA cycle as long as the extracellular glucose levels are maintained at a high enough concentration (Cox E & Bonner J T, 2001). Low expression of PYGM could be serve as a marker for EMT and upregulation of aerobic glycolysis in breast cancer cells (Kondaveeti et al., 2015).

#### 1.7.12 Structure specific recognition protein (SSRP1)

SSRP1 is a component or subunit of the chromatin transcriptional elongation factor FACT and forms the heterodimer complex along with SUPT16H located in the nucleus (Figure 1.7.1). SSRP1 is typically associated with the cell differentiation stage and elevated SSRP1 has been found in many metastasised tumours making SSRP1 a potential prognostic marker (Garcia et al., 2013; Gurova et al., 2013). SSRP1 is linked to the Gene Expression (Transcription) Reactome Pathway (R-HSA-74160) along with HIST1H4A (Figure 1.7.1). High SSRP1 is known to correlate with markers of poor prognosis in breast cancer such as negative hormone receptor status and presence of HER2 (Attwood et al., 2017). Nevertheless, SSRP1 is not an established marker or drug target for ERBB2 overexpressing breast cancer.

## 1.8 Impact statement

The work presented in this thesis addresses the issue of therapeutic resistance in the ERBB2/HER2 overexpressing subset of breast cancer patients. In particular, the elucidation of downstream signalling mechanisms responsible for the malignancy of HER2 positive breast cancers by functionally characterising gene products associated with ERBB2-dependent breast cancer and identifying potential biomarkers linked to therapeutic resistance mechanisms. The work presented herein focused on CPNE3, a poorly characterised gene product which the Timms lab have shown to be linked to the overexpression of ERBB2. The change in CPNE3 gene expression along with adhesion related proteins was previously observed in HB4a/C3.6 HMLECs and SKBR3 breast cancer cells following specific knock-down of ERBB2. It can be hypothesized that the altered gene products are somehow involved in generating the observed altered cellular phenotypes and play a role in ERBB2-mediated signalling.

This study has successfully unravelled an underlying mechanism driving ERBB2-mediated therapeutic resistance, by functionally characterising CPNE3's role in cellular adhesion. Moreover, the study introduces a novel method or workflow to discover and validate CPNE3 linked gene expression changes by combining siRNA mediated knockdown of CPNE3 in an ERBB2 overexpressing cell line model, mass spectrometry based global protein expression profiling and a biostatistical method to create functional networks using clinical data from patient cohorts of all breast cancer subtypes. The evaluation of clinical data provided a real-world model that will better inform on the relevance of changes observed in the cell line profiling experiment, overcoming the commonly cited drawbacks regarding the limited clinical translation of discoveries made using cell line models.

Identifying biomarkers that predict changes in cell adhesion is an important first step in preventing treatment failure for breast cancer. This thesis presents the discovery of such biomarkers related to changes in cell adhesion or metabolic reprogramming and differentially regulated by CPNE3 expression in cell line models and HER2 positive breast cancer patients. Moreover, suggesting a role in glucose homeostasis for the hitherto poorly characterised downstream signalling candidate CPNE3.

## Chapter 2

### 2. Materials and methods

#### 2.1 Cell culture

SKBR3 cells were cultured in 75cm<sup>2</sup> tissue culture flasks in DMEM/F-12 medium supplemented with 10% (v/v) foetal calf serum, 100 µg/ml streptomycin and 100 IU/ml penicillin (Gibco-Invitrogen Corp) in a humidified incubator at 37°C with 5% CO<sub>2</sub>. HB4a and C3.6 cells were cultured in 75cm<sup>2</sup> tissue culture flasks in 1640 RPMI medium supplemented with 10% (v/v) foetal calf serum, 2 mM L-glutamine, 100 µg/ml streptomycin, 100 IU/ml penicillin (Gibco-Invitrogen Corp), 5 µg/ml insulin and 5 µg/ml hydrocortisone (both Sigma) in a humidified incubator at 37°C with 5% CO<sub>2</sub>.

#### 2.2 Small interfering RNA (siRNA) reverse transfection

SKBR3 and C3.6 cells were incubated in an antibiotic-free medium overnight. The cells were subsequently reverse transfected with pools of siRNA targeting Copine III and ERBB2 or an ON-TARGET plus non-targeting scrambled control siRNA pool (Dharmacon RNA technologies). The reverse transfection was performed in 6-well plates as per the manufacturer's instructions using lipofectamine™ RNAi Max (Invitrogen) and diluting the siRNA with Opti-Mem® reduced serum medium (Invitrogen). A final concentration of 50 nM of siRNA was found to be effective and was typically used to transfect 2.5 x 10<sup>5</sup> cells. The cells were maintained in antibiotic-free medium, either 1640 RPMI supplemented with 10% (v/v) foetal calf serum, 2 mM L-glutamine, 5 µg/mL insulin and 5 µg/mL hydrocortisone (Sigma) for C3.6 cells or DMEM/F-12 medium supplemented with 10% (v/v) foetal calf serum for SKBr3 cells. Cells were incubated at 37°C and lysed with 200 µL/well of NP40 buffer after 96 hours. The knockdown of protein expression was confirmed by western blotting.

## 2.3 Protein extraction and sample preparation

Transfected cells were washed twice in ice-cold PBS prior to lysis in 200  $\mu$ L of NP40 lysis buffer (50 mM HEPES, 150 mM NaCl, 1% NP40 and 1 mM EDTA) supplemented with the protease and phosphatase inhibitors: AEBSF (100  $\mu$ g/mL), aprotinin (17  $\mu$ g/mL), leupeptin (4.8  $\mu$ g/mL), pepstatin (1  $\mu$ g/mL), okadaic acid (1  $\mu$ M), sodium orthovanadate (2 mM), BpVphen (5  $\mu$ M) and fenvalerate (5  $\mu$ M) or 8 M urea, 100 mM TEAB pH 8. Lysates were placed on ice for 20 minutes followed by centrifugation at 13,000 rpm and 4°C for 10 minutes to remove cell debris. A BCA assay was used to determine protein concentration using a microtitre plate and 2  $\mu$ L of cell lysate (in triplicate) plus 200  $\mu$ L of BCA reagent mixture (Thermo-Scientific). The microtiter plates were incubated for 30 minutes at 37°C per the manufacturer's instructions. Absorbance readings were taken at 562 nm on a microtitre plate spectrophotometer and protein quantification was determined against a standard curve of incremental dilutions of bovine serum albumin (BSA).

## 2.4 Western blotting

The samples were then diluted to equal protein concentrations followed by reduction and denaturation in sample buffer (50 mM Tris-HCL pH 6.8, 2% SDS (w/v), 6% (v/v) glycerol, 1%  $\beta$ -mercaptoethanol and 0.02% (w/v) bromophenol blue) at 96°C for 5 minutes. Proteins were resolved by SDS-PAGE on 4-12% NuPAGE® Novex Bis-Tris Gels with NuPAGE® MOPS SDS Running Buffer (Invitrogen) as per the manufacturer's instructions. The proteins were subsequently electroblotted onto polyvinylidene fluoride (PVDF) membranes (Immobilon P, Millipore) in a wet transfer tank containing a transfer buffer composed of 195 mM glycine, 25 mM Tris and 20% (v/v) methanol. The membranes were blocked overnight in a blocking solution of 5% BSA in TBS-T (50 mM Tris-HCL pH 8.0, 150 mM NaCl and 0.05% Tween-20) followed by a 1-hour incubation with primary antibody diluted in TBS-T. Membranes were washed for 3 x 15 minutes in TBS-T. An appropriate horseradish peroxidase (HRP)-coupled secondary antibody was diluted in TBS-T and applied to the membrane for 45 minutes. Membranes were washed in TBS-T as previously. The membrane-bound proteins were visualised on film using enhanced chemiluminescence reagent (ECL) (PerkinElmer Life Sciences) as per the manufacturer's instructions.

## 2.5 Adhesion assays

ERBB2 overexpressing HMLEC C3.6 and its parental cell line HB4a were subjected to adhesion assays on 96-well plates. Cells were trypsinised, washed twice in serum-free medium and seeded at

**Table 2.4.** The table contains a list of antibodies used for western blotting and their corresponding dilutions.

Antibody	Company/Source	Mono/Polyclonal	Dilution
CPNE3	Elizabeth Grimm lab	Rabbit pAb	1:1000
ERBB2	Santa Cruz	Rabbit pAb	1: 500
ITGA6	Anti-ITGA6 (ab105669 Abcam)	Rabbit mAb	1:100
ITGB4	Anti-ITGB4 Cell Signalling Technologies	Rabbit pAb	1:1000
NDRG1	Sigma-Prestige	Rabbit	1:500
Anti-Mouse Secondary	GE Healthcare	sheep anti mouse	1:5000
Anti-Rabbit Secondary	GE Healthcare	donkey anti-rabbit	1: 5000
Anti-Goat Secondary	Chemicon International	rabbit anti-goat	1: 15000

a cell density of  $2 \times 10^4$  per well into 96-well plates with complete medium. Cells were left to adhere for 10, 20, 30, 60 and 120 minutes at 37 °C before washing with PBS and adding a volume of 200  $\mu$ L of complete medium. The cells were seeded on either laminin-coated or non-coated tissue culture plates. The plates were incubated overnight at 37 °C. The medium was aspirated, and the cells were washed with PBS before adding a volume of 50  $\mu$ L of 1 mg/mL, 3-[4,5-dimethyl-thiazol-2-yl]-2,5-diphenyl-tetrazolium bromide (MTT) (Sigma) in PBS. The cells were incubated for 4 hours. DMSO (100  $\mu$ L) was added, the plates were shaken for 20 minutes at 22 °C, and the absorbance read at 540nm. Each time point was performed in quadruplicate per well and an average absorbance reading was calculated along with the standard error of the mean.

## 2.6 De-adhesion assays

ERBB2 overexpressing HMLEC C3.6 and its parental cell line HB4a were subjected to de-adhesion assays. Cells were trypsinised, washed twice in serum-free medium and seeded at a cell density of  $2 \times 10^4$  per well into 96-well plates with complete medium. Cells were left to adhere overnight at 37 °C

before washing with PBS and trypsinising at varying concentrations of trypsin. The cells were seeded on either laminin-coated or non-coated tissue culture plates. The concentrations of trypsin were 0.01, 0.02, 0.03, 0.04 and 0.06%. The plates were incubated for 10 minutes before the plates were knocked, detached cells aspirated, and the remaining cells washed gently with PBS. A volume of 50  $\mu$ L of 1 mg/mL, 3-[4,5-dimethyl-thiazol-2-yl]-2,5-diphenyl-tetrazolium bromide (MTT) (Sigma) in PBS was added to each well. The cells were incubated for 4 hours. DMSO (100  $\mu$ L) was added, the plates were shaken for 20 minutes at 22 °C, and the absorbance read at 540nm. Each time point was performed in quadruplicate per well and an average absorbance reading was calculated with standard error of the mean.

## 2.7 Real-time cell adhesion or spreading assays

Cells were plated in 16-wells E-Plate View (ACEA Biosciences/Agilent) and placed at a seeding density of 40000 cells per well onto an xCELLigence RTCA DP (ACEA Biosciences/Agilent) located inside a tissue culture incubator at 37°C and 5% CO<sub>2</sub>. The cells were allowed to obtain equilibrium according to the manufacturer's guidelines. The cell index (CI) values were measured automatically every 30 min over 12 h for adhesion. The cell adherence was determined by calculating the slope of the line between two given time points. Standard deviations of well replicates were analysed with the RTCA 2.1.0 Software. The rate of cell spread was determined by identifying the maxima or the largest CI where two given time points have a slope equal to zero or less than zero.

## 2.8 TMT tagging and HPLC fractionation.

Cell lysates (urea lysis buffer) from candidate siRNA-transfections were concentrated to approximately 50  $\mu$ L in a Viva Spin 5kDa column by centrifuging at 15,000xg (approximately 12.5 x1000rpm) at 4°C. A buffer exchange was conducted by adding an additional 450  $\mu$ L 0.1M TEAB. The final concentration of urea was < 1M. The vial was centrifuged to concentrate the solution to approximately 50  $\mu$ L.



Reduction of 100  $\mu\text{g}$  for each test sample protein was achieved by adding TCEP to a final concentration of 1mM. The samples were incubated at 56° for 45 minutes followed by centrifugation at 4°C for 1 minute to ensure all liquid was at the bottom of the tube and to chill the solution.

An alkylation step was performed by adding iodoacetamide at a final concentration of 7.5mM. The sample was left for 45 minutes in the dark. The protein was then digested by adding sequencing grade modified porcine trypsin (Promega) trypsin to the sample to give a 1:50 trypsin to protein ratio. The mixture was left overnight at 37°C before being dried in a vacuum centrifuge.

The digests 100  $\mu\text{g}$  protein were re-suspended in 50 nM TEAB (Thermo Scientific) by shaking in a Grant-bio shaker at 1130 rpm for 10 minutes. The TMT Label Reagents were equilibrated to room temperature immediately before use. 41  $\mu\text{L}$  of anhydrous acetonitrile was added to each 0.8mg vial. TMT reagents were left to dissolve for 5 minutes with occasional vortexing. The tube was centrifuged to gather the solution and were then added to the samples as follows: siCtrl-1: TMT126, siCPNE3-1: TMT127, siCtrl-2: TMT128, siCPNE3-2: TMT129, siCtrl-3: TMT130 and siCPNE3-3: TMT131.

The tags and samples were mixed and incubated at room temperature for 1 hour. A 50-fold dilution of 5% hydroxylamine was made up in 50 mM TEAB and 16  $\mu\text{L}$  added to the labelled vials. The individual TMT-labelled samples were pooled (total volume of 942  $\mu\text{L}$ ). The sample was desalted with Oasis HLB 1cc (30 mg) extraction cartridges as per the manufacturing protocol before high pH C-18 reverse phase HPLC. Fractionation was achieved using high pH, RP-LC. Samples were suspended in 20 mM ammonium formate at pH 8.4 and loaded onto a Poroshell 300 Extend-C18 column (2.1 x 75 mm, 5 mm-bead size, 300 Å-pore size, Agilent) using an Agilent 1100 HPLC system. 30 fractions were collected by eluting with a gradient of acetonitrile (3-45%) over 35 mins. Samples were dried to completion, re-solubilised in 200  $\mu\text{L}$  0.1% formic acid, dried down again and stored at -20°C to remove ammonium formate.

## 2.9 Mass Spectrometry (LC-MS/MS) and protein identification

The 30 fractions of the TMT 6plex-labelled sample were re-suspended to a concentration of ~0.2  $\mu\text{g}/\mu\text{L}$  in buffer A (0.1% formic acid) and sequentially analysed by nanoLC-MS/MS (LTQ Orbitrap XL) as described in Sinclair et al (Sinclair & Timms, 2011). In brief, peptides were separated by C18 RP-

LC on an Ultimate 3000 nano-liquid chromatography system (Dionex); 5  $\mu$ L samples were injected onto an Acclaim PepMap 100 C18 pre-column (5  $\mu$ m, 100  $\text{\AA}$ , 300  $\mu$ m i.d x 5 mm) (ThermoFisher Scientific) and washed for 3 min with 10% buffer B (ACN + 0.1% (v/v) FA) at a flow rate of 25  $\mu$ L/min and then peptides separated on an Acclaim PepMap 100 C18 Nano-LC column (3  $\mu$ m-bead size, 100  $\text{\AA}$ -pore size, 75  $\mu$ m i.d x 250 mm) (ThermoFisher Scientific) with a 90 min linear gradient of 10-50% buffer B at a flow rate of 300 nL/min. Samples were electrosprayed into an LTQ-Orbitrap XL instrument controlled using Xcaliber software (ThermoFisher Scientific). Precursor MS scans were acquired in the orbitrap using data-dependent acquisition (DDA) at a resolution of 60,000 at m/z 400, followed by top3 CID/HCD using a normalised collision energy of 40% for HCD and a resolution of 7,500 for detection of product ions in the orbitrap. Dynamic exclusion was enabled with a list size of 500, excluding for 15 seconds. Atmospheric polymethylcyclsiloxane was used as a lock mass (455.12003 m/z) for in-run calibration.

Raw data files were analysed using Proteome Discoverer V1.4 software (ThermoFisher Scientific) with database searching against the UniProtKB/SwissProt database (2014\_09; 546,439 sequence entries) using the Mascot search engine V2.4 (Matrix Science). For searching, taxonomy was human, enzyme was trypsin, MS tolerance was set to  $\pm$  10 ppm, MS/MS tolerance was set to 0.5 Da and one missed cleavage was allowed. TMT 6-plex modification of peptides and carbamidomethylation of cysteines were set as fixed modifications. Protein N-terminal acetylation, methionine oxidation and N/Q deamidation were set as variable modifications. Search result filters were as follows: only peptides with a score of  $>20$  and below the Mascot significance threshold filter of  $P < 0.05$  were included. Protein grouping was enabled such that when a set of peptides in one protein were equal to or completely contained within the set of peptides of another protein, the two proteins were put together into a protein group. The false discovery rate was calculated to be 2.3% based on searching a decoy database.

Reporter ion-based quantification was carried out using Proteome Discoverer Version 2.4 with the following ratios calculated:  $(127+129+131)/(126+128+130)$ , 126/128, 126/130, 127/126, 127/128, 127/129, 127/130, 127/131, 128/130, 129/126, 129/128, 129/130, 129/131, 131/126, 131/128 and 131/130. The reporter ion ratios were normalized on protein median with a minimum protein count of 20. A filtering system was applied to aid in defining a list of up and down-regulated proteins. Search

result filters were selected as follows; only peptides with a score >20 and below the Mascot significance threshold filter of  $p = 0.05$  were included and single peptide identifications required a score equal to or above the Mascot identity threshold (Sinclair & Timms, 2011).

While TMT labelling can offer greater accuracy and precision in quantitation, particularly for small quantitative differences, it requires more complex sample preparation and is more expensive than label free alternatives. Furthermore, several experiments have shown that it may suffer from variations in labelling efficiency (Hutchinson-Bunch et al., 2021; Zecha et al., 2019). On the other hand, label-free quantification is a low-cost alternative to labelled quantification for relative and absolute protein quantification in large-scale experiments. Label-free methods are generally well-suited for large-scale screening and biomarker discovery experiments and can detect large differences in protein expression. However, label-free methods may be less sensitive than labelled methods for detecting small quantitative differences and have a limited range of linear quantitative measurement. Additionally, label-free methods require careful experimental control to account for experimental variations and have been shown to suffer from missing values which can impact statistical analysis and downstream functional interpretation (Ball et al., 2023; Hamid et al., 2022; Rozanova et al., 2021; M. Wang et al., 2008).

## 2.10 Label-Free sample preparation

Reduction of 100  $\mu\text{g}$  for each test sample protein was achieved by adding TCEP to a final concentration of 10mM. The samples were incubated at 56° for 45 minutes followed by centrifugation at 4°C for 1 minute to ensure all liquid was at the bottom of the tube and to chill the solution. An alkylation step was performed by adding 2-chloroacetamide at a final concentration of 40mM. The sample was left for 45 minutes in the dark. The protein was then digested by adding sequencing grade modified porcine trypsin (Promega V5111) trypsin to the sample to give a 1:50 trypsin to protein ratio. The mixture was left overnight at 37°C before being dried in a vacuum centrifuge.

The digests of 100  $\mu\text{g}$  protein were resuspended in an aqueous solution free of any organic solvents and acidified with trifluoroacetic acid (TFA) to a concentration of 0.1% (v/v). Each sample was desalted with Oasis C18 spin tip desalting cartridge (Pierce 89873) and capacity filter (HLB 1cc (30

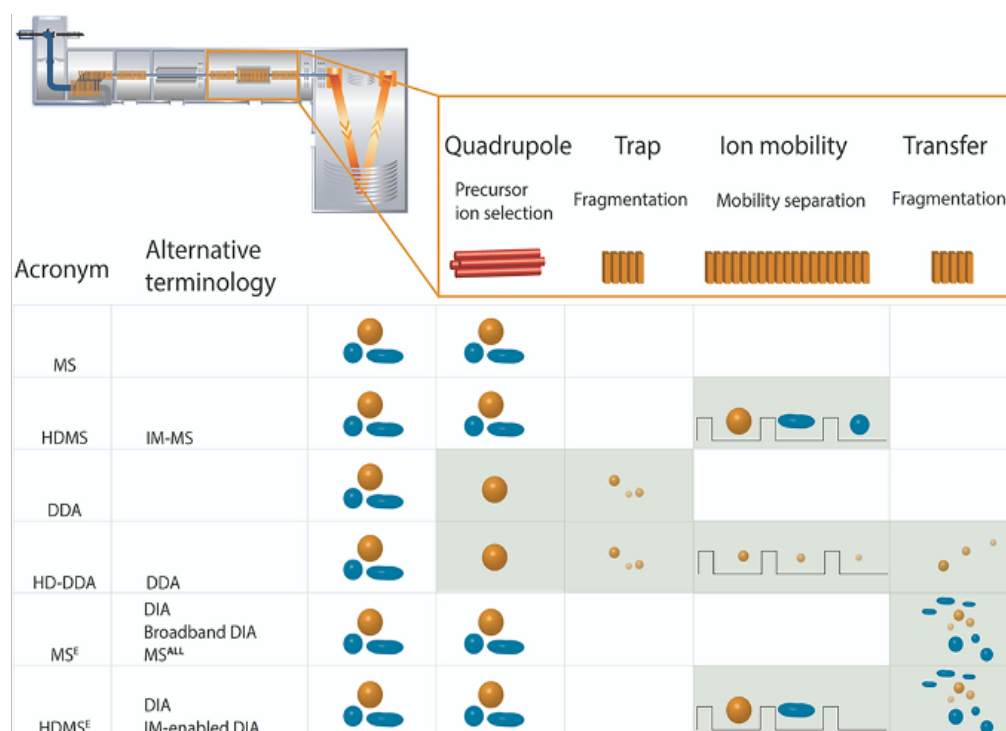
mg) extraction cartridges) as per the manufacturing protocol. Samples were dried to completion without pre-fractionation.

## 2.11 Mass Spectrometry (LC-MS/IMS/MS) and protein identification

The samples were re-suspended in a buffer comprised of solvent A1 (Water LC/MS Grade  $\geq 99.9\%$ ) and solvent B1 (Acetonitrile) at a ratio of 95% A1 and 5% B1 and sequentially analysed by LC-MS/IMS/MS (Waters Synapt G2 QTOF). In brief, peptides were separated by C18 RP-LC on an Acquity M+ (Waters Corporation); samples were injected at a flow rate of 0.5  $\mu\text{L}/\text{min}$  and washed for 5 min at a flow rate of 15  $\mu\text{L}/\text{min}$ . Samples were electrosprayed into an ion mobility separation (IMS) capable Waters Synapt G2 QTOF instrument controlled using MassLynx software (Waters Corporation). Analysis was performed on a Waters Synapt G2 QTOF operating in positive HDMS<sup>E</sup> mode (high and low fragmentation channel monitoring with IMS active) and TOF mass resolution in positive ions. In HDMS<sup>E</sup>, the mass spectrometer alternates between low-energy and high-energy scans to generate two separate data sets for each sample. The low-energy scan is used for intact precursor ion detection and quantification, while the high-energy scan is used for fragment ion detection and structural characterization (Claude et al., 2013; Britt et al., 2022). Furthermore, IMS in HDMS<sup>E</sup> provides an additional dimension of separation, improving the resolution of complex sample mixtures. By incorporating IMS and separating ions based on their CCS, HDMS<sup>E</sup> allows for more accurate identification and quantification of analytes in the presence of interferences or background noise (Figure 2.11.1) ( Gil-Solsona et al., 2021; Britt et al., 2022). Precursor MS scans were acquired between 50 to 2000 Da for 120 minutes at collision energies between 19v to 45v.

Raw data files were analysed using Progenesis Qi (Waters Corporation) for fragment ion searches and peak alignment across multiple alignments over the MZ range with 2 database searches against the default database for the Ion Accounting search algorithm (Waters Corporation) and the UniProtKB/SwissProt database (2014\_09; 546,439 sequence entries) using the Mascot search engine V2.4 (Matrix Science) respectively. Optional peak filtering was applied with FDR correction for experimental design. Output data from the Ion Accounting default database search was used for the

Mascot search as a data normalisation step to ensure that the SwissProt search was for human samples. The high false discovery rate for samples was adjusted using ANOVA FDR adjustment.



**Figure 2.11.1** Schematic representation of the mass spectrometer processes for various acquisition modes available on Synapt G2-Si mass spectrometer. The quadrupole can be used in three modes allowing all ions to pass through (MS, HDMS, MS<sup>E</sup>, HDMS<sup>E</sup>) or perform precursor ion selection based on intensity (DDA, HDDDA). CID fragmentation is applied in the trap for some acquisition modes (DDA, HDDDA). Mobility separation is performed next for IM-enabled modes either on precursor ions (HDMS, HDMS<sup>E</sup>) or CID fragments (HDDDA). In addition to the trap, CID fragmentation can be performed in the transfer for the high energy MS<sup>2</sup> experiment that makes up part of the MS<sup>E</sup> and HDMS<sup>E</sup> acquisition modes. The broadband DIA modes (MS<sup>E</sup> and HDMS<sup>E</sup>) and DDA provide both MS<sup>1</sup> and MS<sup>2</sup> data. The ion  $m/z$  is denoted by the colour of the precursor/product ion (orange/blue), and the shape of the precursor/product ion denotes the collision cross-section (CCS) Adapted with permission from the authors. (Britt et al., 2022).

## 2.12 Dimensionality reduction by PCA and k-means clustering

The comparison of proteomic profiles obtained from TMT LC-MS/MS and Label Free LC-MS/MS was achieved by reducing the dimensions of each dataset to an intrinsic dimension that describes protein expression changes with a high similarity to CPNE3 in Python 3.8. The TMT data was comprised of 9 sample combinations derived from all possible knockdown vs control combinations for CPNE3 knockdown sample tandem mass tags TMT<sup>6</sup>-127, TMT<sup>6</sup>-129 and TMT<sup>6</sup>-131 and control sample

tandem mass tags TMT<sup>6</sup>-126, TMT<sup>6</sup>-128 and TMT<sup>6</sup>-130. Similarly, the label free data was transformed to 9 sample combinations comprised of all possible knockdown vs control combinations for CPNE3 knockdown samples 1 to 3 and control samples 1 to 3. Both datasets were log<sub>2</sub>-transformed normalised to moderate the variance across the mean. The optimal number of components which capture the greatest amount of variance in the data was determined using StandardScaler() function and the principal component analysis (PCA) module in the Scikit-learn library. K-means clustering to identify the PCA component with proteins of high expression similarity to CPNE3 in all sample replicates was implemented using the k-means module in the Scikit-learn library. Principal components were first fit using the k-means algorithm and the optimal number of clusters determined for each data set.

The number of clusters to be used for the k-means clustering were determined by an iterative statistical technique for measuring the sum of the squared distances to the nearest cluster centre (sum of squared error) by running the k-means algorithm for a range of cluster values. The iterative statistical technique known as the elbow method, was used to determine the optimal number of clusters for k-means clustering. The elbow method involves the sum of squared error (SSE) values against the number of clusters and selecting the number of clusters at the point of the plot where the rate of SSE reduction begins to level off, forming an elbow-like shape. This point indicates that additional clusters do not significantly reduce SSE and can lead to overfitting, while using fewer clusters may result in underfitting. Therefore, the elbow method is useful in identifying the optimal number of clusters for k-means clustering that balance model complexity with the need for accurate clustering. However, it should be noted that the elbow method is not always definitive, and other factors such as domain knowledge and context should also be considered. Additionally, the elbow method may not be applicable in cases where missing values or noisy data are present.

To ensure that the entire range of features in our datasets was represented, a range of cluster values from n=1 to 15 were selected. A 2 component PCA visualisation of k-means clustered components was carried out and used to interpret the CPNE3 containing principal component for each dataset.

## 2.13 Clinical data

The sources and processing of clinical protein expression data used in the construction of biological networks reported here are detailed elsewhere. The data consists of protein expression profiles of 40 breast cancer patients derived through SILAC-based mass spectrometry (Tyanova et al., 2016) and 105 breast cancer patient profiles derived through iTRAQ-based mass spectrometry by the NCI clinical proteomic tumour analysis consortium (CPTAC) (Ellis et al., 2013; Mertins et al., 2016).

## 2.14 Creating partial compendia of clinical data

We call the total set of patient samples for which protein expression data is available the full compendium of protein expression data. However, for this analysis, we created partial compendia for each of the two data sets, a discovery cohort and a validation cohort (Table 4.2.1). Partial compendia were created to include only protein expression data for the 52 proteins identified as up/down-regulated in response to CPNE3 knockdown in HMLECs using TMT mass spectrometry (Appendix 2.1). 36 proteins were consistently quantified in all 40 samples of the discovery cohort. 16 proteins, AHSG, ALPP, HIST1H1A, HIST1H1C, HIST1H1D, HIST1H3A, MGST1, PNPLA7, PPP6R2, PYGM, RPL32, S100A10, SLC3A2, TIMELESS, UBC and ZFP28, were not quantified for all 40 patient samples. The 36 proteins were evaluated in the correlation analysis of the discovery cohort of 40 breast cancer patients. Because protein expression data and patient prognosis involve numerous uncertainties, the data set with 105 patients was partitioned on the basis of characterisation and quality control tests previously described (Mertins et al., 2016). 75 female patient samples were included in our validation cohort (Table 4.2.1). Genes with protein expression values for all 75 samples were included in the compendia. 45 proteins were consistently quantified in all samples. 7 proteins, ALPP, HIST1H3A, HMGA1, PNPLA7, PTGES, UBC and ZFP28, were omitted due to a high number of missing values for the samples or not detected at all during the quantification. The 45 proteins quantified for all 75 patients were used in the correlation analysis of the validation cohort.

## 2.15 Statistical analysis and network construction

The Pearson correlation coefficient  $r$  for each protein-protein association was computed using GraphPad Prism with statistical significance of the  $r$ -values determined using a two-tailed test with a 95% confidence interval. Heatmaps of the respective correlation matrix for each breast cancer subtype were visualised in GraphPad Prism software (version 9) (GraphPad Software, Inc) (Mitteer & Greer, 2022) and correlation networks constructed using Cytoscape software (version 3.9.1) (Shannon et al., 2003). Proteins included as nodes in the constructed networks were required to share at least one correlation with significance level  $p$ -value  $< 0.05$  for the visualised edges. Each network was used to identify the predominant functional cluster containing either a set of highly interconnected nodes or CPNE3 respectively and for functional enrichment analysis of proteins related to the cluster.

## 2.16 Functional enrichment analysis

Differentially expressed proteins (up/ down-regulated ( $>1.5$  fold)) were imported into WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) and mapped to the pathway functional database for either KEGG, Panther or Reactome databases using the Over-Representation Analysis (ORA) enrichment analysis method for Homo sapiens using the genome reference set (Liao et al., 2019). The significance of the association between the uploaded data set and the KEGG, Panther or Reactome pathway was determined for each gene set, by calculating an enrichment  $p$ -value using the Hypergeometric distribution or the Fisher exact test and corrected for multiple-hypothesis testing with a Benjamini-Hochberg FDR (Liao et al., 2019; J. Wang et al., 2013; Zhang et al., 2005).

## 2.17 Causal analysis

Differentially expressed proteins (up/ down-regulated ( $>1.5$  fold)) were imported into Ingenuity® pathway analysis (IPA) (QIAGEN) (Krämer et al., 2014) and mapped to the IPA knowledgebase to identify causal relevance to breast cancer, a role in cell adhesion or known biomarker application.



## 2.18 Protein interaction analysis

Differentially expressed proteins (up/ down-regulated ( $>1.5$  fold)) were uploaded to the STRING database (version 11.5) (Szklarczyk et al., 2021) and mapped to protein-protein interaction networks. Molecules were required to interact with a confidence score cut-off  $\geq 0.90$  (high confidence) for significantly up/down - regulated proteins discovered using the TMT-LC-MS/MS method and only interactions between proteins identified by the dataset were permitted.

## 2.19 Bioinformatics analysis of phosphoproteomic data

Bioinformatics analysis of phosphopeptides/proteins that were differentially regulated with ErbB2 over-expression (HER2 positive + HER2-enriched vs. HER2 negative + HER2-enriched) in the validation cohort was used to define the downstream phosphorylation networks potentially involved in oncogenic transformation. Differentially regulated phosphoproteins (containing at least one phosphosite with  $\geq 1.5$ -fold change) were identified using Clustal Omega (Sievers et al., 2011) sequence alignments of phosphopeptides extracted from the TCGA\_Breast\_BI\_Phosphoproteome.phosphopeptide.itraq file of the data used for the validation cohort. This represented a total of 45 proteins of which 11 were found to have phosphopeptide enrichment and assigned sites for modified residues on each peptide sequence. The phosphopeptide ratio was determined for HER2 positive + HER2-enriched vs. HER2 negative + HER2-enriched patients from the 75 patient validation cohort described previously for each of the peptide sequences from the 11 proteins found to contain a modified residue. Given the small number of 11 phosphoproteins differentially regulated for HER2 positive + HER2-enriched vs. HER2 negative + HER2-enriched patients there were a high number of significantly enriched ( $\geq 1.5$ -fold change) singly (1pST) or doubly phosphorylated (2pST) phosphopeptides (Table 4.4.1) (see Appendix 2.2.2 for more detailed results).

## 2.20 Logistic regression analysis of selected candidate biomarkers

Logistic regression analysis was implemented in Python 3.8 using candidate biomarkers identified by statistical analysis and network construction as the given set of input variables. A binary logistic regression model was implemented by creating an instance of the `LogisticRegression()` class in Scikit learn (Sklearn) and a 70:30 splitting ratio or train: test ratio, `x_train`, `x_test` for the independent features and `y_train`, `y_test` for the dependent variables using the sklearn library. The target variable or dependent variable was either HER2 positive (case) or HER2 negative (control). Model performance was evaluated using Leave-One-Out-Cross-Validation (LOOCV) and Youden's J statistic. LOOCV was implemented using sklearn's `model_selection` method and the Youden's J statistic was calculated by subtracting 1 from the sum of sensitivity ( $\text{True Positive} / (\text{True Positive} + \text{False Negative})$ ) and specificity ( $\text{True Negative} / (\text{True Negative} + \text{False Positive})$ ).

## Chapter 3

### 3. Evaluating the effect of gene expression on HMLEC phenotype

#### 3.1. Chapter Introduction

This chapter presents the findings of our work that investigates the functional role of CPNE3 in ERBB2 overexpressing breast cancer. Specifically, we sought to answer the following research questions: What is the effect of ERBB2 overexpression on cell adhesion? What is the role of CPNE3 expression on cell adhesion? To address these questions, we conducted experiments to assess the expression of ERBB2 and CPNE3 in HMLECs and their impact on cell adhesion. Additionally, we analysed global protein expression in ERBB2 overexpressing HMLECs using two mass spectrometry instrument platforms and quantification methods, the Orbitrap LTQ for TMT and the Waters Synapt QTOF for Label-Free LC-MS/MS, to identify proteins up/downregulated in response to CPNE3 knockdown. The mass spectrometry data from both instruments and methods were evaluated using a quantification method agnostic dimensionality reduction method, to identify proteins with a similar expression pattern to CPNE3 across all samples as an intrinsic dimension (id) of each data set. Our hypothesis posited that there is a positive correlation between ERBB2 overexpression and CPNE3 in ERBB2 overexpressing breast cancer, and CPNE3 is a potential downstream effector. Therefore, the method applied for dimensional reduction identifies a cluster of proteins with a strong positive correlation to CPNE3 following siRNA mediated knockdown. These proteins are the minimum number of parameters needed to generate a data description of the proteomic profiles generated by each method for comparison using functional annotation databases.

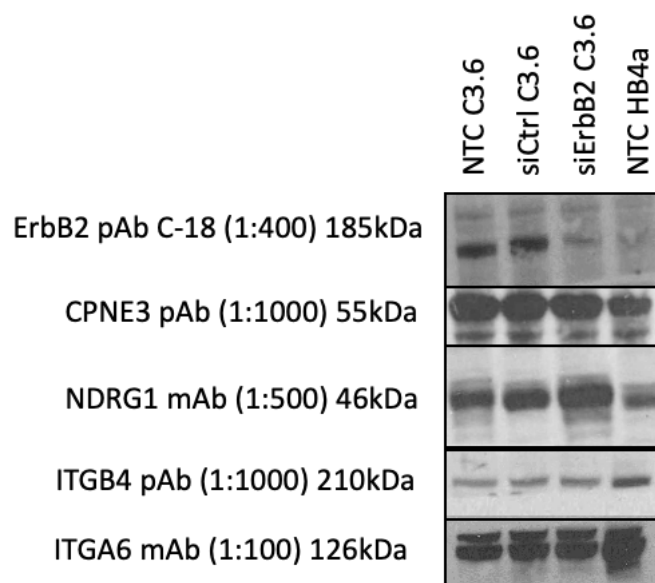
#### 3.2. The effect of ERBB2 expression on HMLEC cell phenotype

##### 3.2.1. siRNA-dependent knockdown of candidate proteins

To validate the role of ERBB2 in regulating the expression of downstream proteins in HMLECs, siRNA-mediated knockdown of ERBB2 was carried out in C3.6 cells and compared to the parental

cell line HB4a. The siRNA knockdown of target proteins was conducted by reverse transfection and a concentration of 50 nM of siRNA with lipofectamine produced an optimal knockdown of ERBB2 as determined by immunoblotting of lysates (Figure 3.2.1).

As shown in figure 3.2.1., the western blot confirms the differential expression of ERBB2, CPNE3, NDRG1, ITGB4 and ITGA6 in HB4a versus C3.6 cells as previously reported (Worthington et al., 2017). CPNE3, NDRG1, ITGA6 and ITGB4 were expected to exhibit an inverse change in expression in C3.6 cells following the knockdown of ERBB2. However, this was not the case and all candidates except for NDRG1, which was elevated in expression, remained unchanged by the siRNA knockdown of ERBB2 in C3.6 cells. This data suggests that ERBB2 either modulates the expression of these candidates through a long-term adaptive regulation of expression that cannot be reversed by acute knockdown of ERBB2 or it does not modulate them at all.

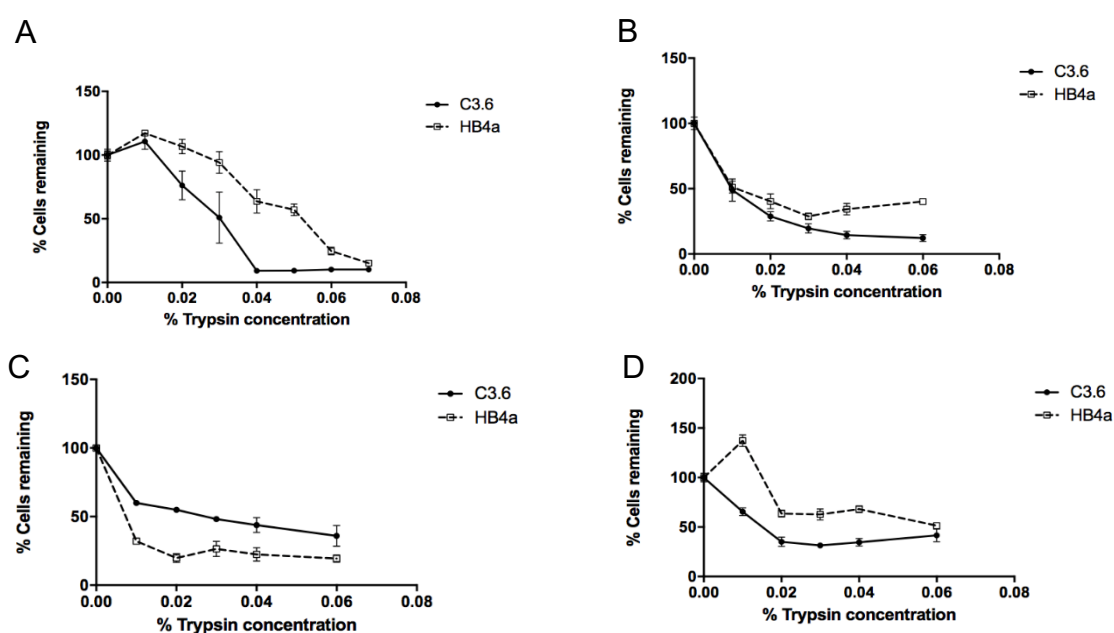


**Figure 3.2.1** Western blot confirmation of target expression of candidates in non-transfected cells (NTC), non-targeting siControl transfected cells and siRNA-targeted cells. No loading control was used and no probing for a housekeeping protein was done. Uniform signal intensity of proteins expected to share similar expression change demonstrates a consistent loading.

### 3.2.2. Assessing the effect of ERBB2 overexpression using end point assays

The differential expression of several integrin subunits and other cell adhesion molecules in the HMLEC system, suggest that ERBB2 expression may modulate adhesion by altering interactions with

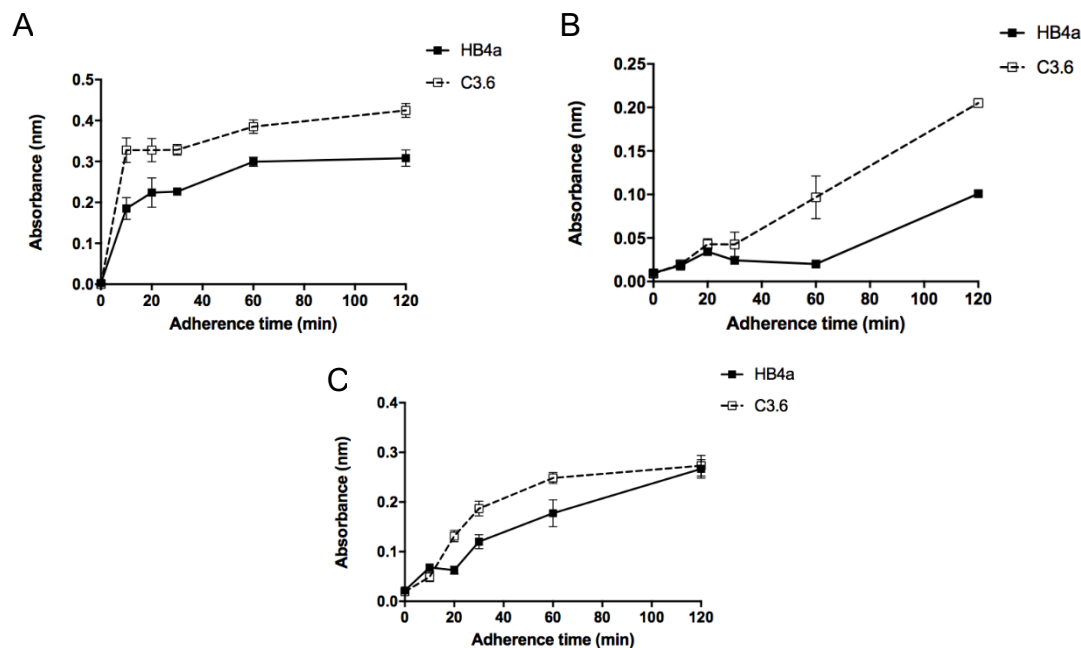
laminin in the basement membrane. Thus, an investigation of the cell adhesion of HB4a and its ERBB2 overexpressing derivative C3.6 were carried out on both plastic and laminin-coated tissue culture plates. Deadhesion assays were conducted to assess the detachment of cells under varying concentrations of trypsin. The results largely suggest that C3.6 cells detach more readily than the HB4a cells (Figure 3.2.2). The deadhesion assay was performed in triplicate (Figure 3.2.2 A-C) on tissue culture plastic, or in a single experiment on laminin-coated plates (Figure 3.2.2 D). However, the results are somewhat inconclusive as graph C displayed an inverse relationship, suggesting that HB4a cells detach more readily than C3.6 cells when seeded at 20,000 cells per well in a 96-well plate and incubated for 10 minutes with a range of trypsin concentrations. Furthermore, the difference in the rate of detachment between C3.6 and HB4a cells shown in graph B is not wide enough to accept the trend as significant. This might be due to subtle differences in deadhesion dynamics within each experiment and the limitations of end point assays. Conducting a series of additional assays using a real time analysis system was determined to be necessary.



**Figure 3.2.2** De-adhesion assay. C3.6 and HB4a cells were seeded at 20,000 cells per well in a 96-well plate and incubated for 10 minutes with a range of trypsin concentrations. A) no laminin B) no laminin C) no laminin D) laminin-coated plate. Errors bars represent standard error of the mean of technical replicates on each plate.

A time-dependent analysis of the attachment of cells was also conducted on both tissue culture plastic and laminin-coated surfaces. The results demonstrate that the C3.6 cells adhere at a greater

rate than HB4a cells (Figure 3.2.3). Nevertheless, the full dynamics of a time dependent cell adhesion process were not captured by these end point assays and a need to conduct further investigations using a real time analysis system was determined. Furthermore, the experimental procedure introduced inconsistencies involving knocking the plate to free loosely attached cells before washing the well. The differences in mechanical force applied to each well due to the direction of the force applied with each knock would introduce some technical differences. These are partially responsible for the fact that experiment A and B results in graphs wherein the absorbance of the two cell types doesn't resolve. However, it is much more likely due to the fact that C3.6 cells proliferate at a higher rate than HB4a which results in the C3.6 cells having a greater maximum absorbance at 120 minutes.

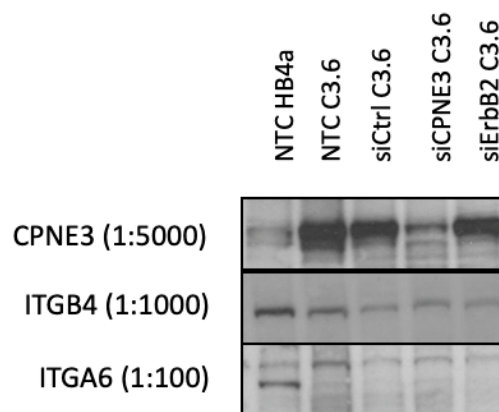


**Figure 3.2.3** Adhesion assay. C3.6 and HB4a cells were seeded at 20,000 cells per well and allowed to adhere for different times. A) no laminin B) no laminin C) laminin-coated plates. Errors bars represent standard error of the mean of technical replicates on each plate.

### 3.3. The effect of CPNE3 expression on HMLEC cell phenotype

#### 3.3.1. siRNA-dependent knockdown of CPNE3 expression

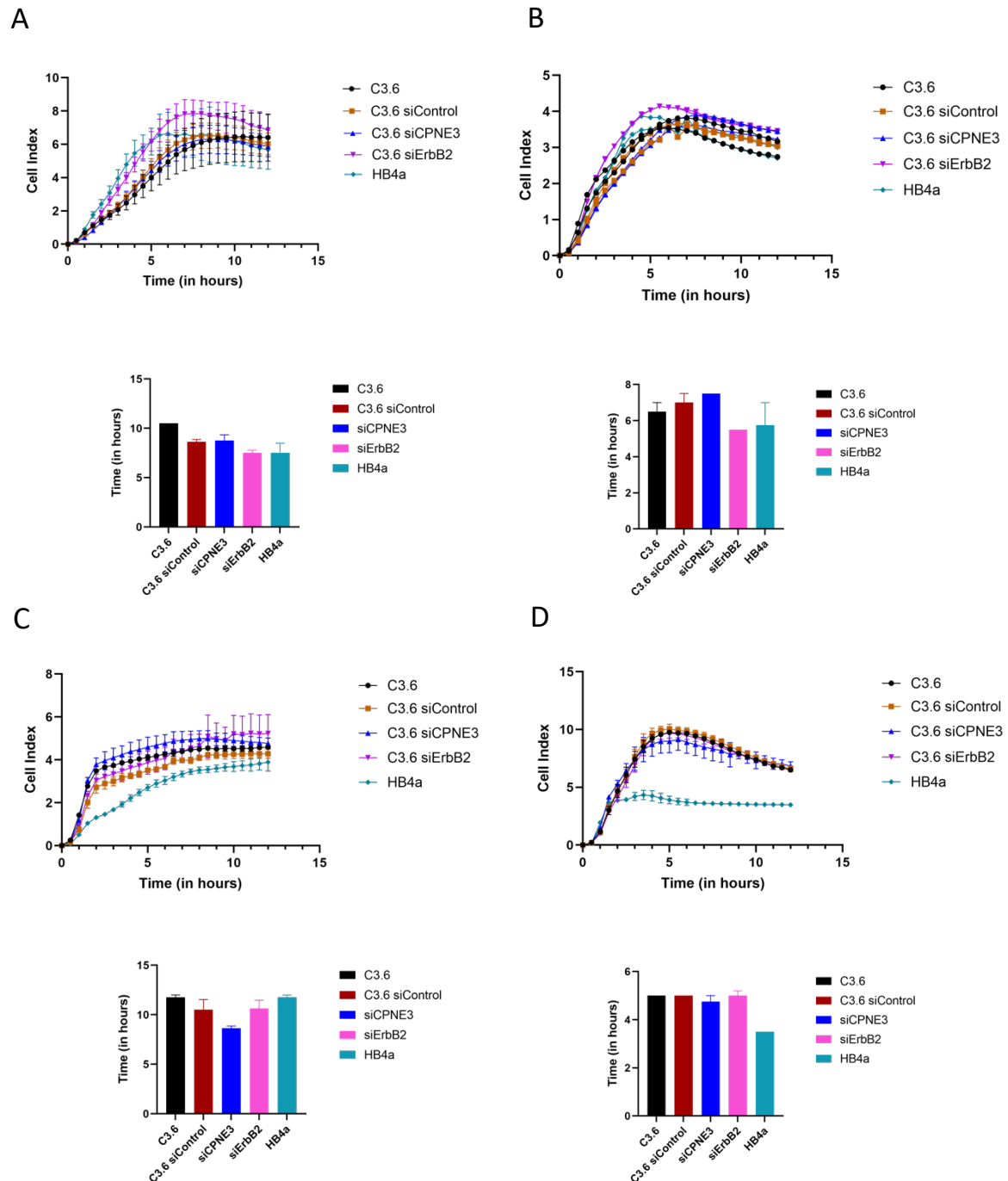
To evaluate the functional role of CPNE3 as a downstream signalling partner in ERBB2 mediated regulation of cell adhesion, C3.6 cells were reverse transfected with siRNA targeting the expression of CPNE3. A concentration of 50 nM of siRNA with lipofectamine incubated for 96 hours produced an optimal knockdown of ERBB2 and CPNE3 as determined by immunoblotting of lysates (Figure 3.2.1. & 3.3.1). The expression of ITGA6 and ITGB4 was then assessed under different conditions of both CPNE3 and ERBB2 expression in HMLEC cells using western blot analysis.



**Figure 3.3.1** Western blot analysis of the expression of ITGA6 and ITGB4 in response to CPNE3 knockdown in C3.6 cells. No loading control was used and no probing for a housekeeping protein was done. Uniform signal intensity of proteins expected to share similar expression change demonstrates a consistent loading.

As shown in Figure 3.3.1, the expression of ITGA6 and ITGB4 does not correlate to the expression of CPNE3 in C3.6 cells. Furthermore, all experimental conditions of transfected C3.6 cells demonstrated a uniform protein expression. Interestingly, the siRNA treated cells appear to share a uniform reduction in protein expression compared to the non-transfected C3.6 cells. As expected, the HB4a cells exhibit a greater expression of both ITGA6 and ITGB4 compared to the C3.6 cells, which is consistent with previously reported findings (Worthington et al., 2017).

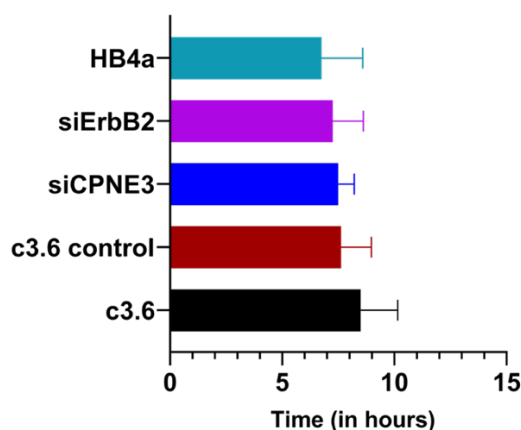
### 3.3.2. Assessing the effect of CPNE3 and ERBB2 expression using a real time cell adhesion and spreading assay



**Figure 3.3.2** RTCA assay analysis of real time adhesion and spreading of a layer of cells on an adherent surface. C3.6 and HB4a cells were seeded at 40,000 cells per well and assessed for 12 hours to assess duration of cell spreading. Histogram shows the maxima of each curve as the point which the cells stop spreading in units of time. Errors bars represent standard error of the mean of technical replicates on each well.



To further evaluate the potential role of ERBB2 overexpression in modulating cell adhesion, a real-time cell analysis (RTCA) of HB4a and C3.6 cells was carried out to evaluate the continuous change in cell proliferation or cell spreading over an assessment period of 12 hours. The cell dynamics of several gene expression conditions of C3.6 cells were assessed alongside HB4a using the ACEA xCELLingence RTCA DP system. The assay measured the impedance of electron flow as cells spread on an adherent surface with gold microelectrodes. The maxima of each graph (Figure 3.3.2), was found and is shown in histograms below each plot. The results were combined (Figure 3.3.3) and are inconclusive. While the knockdown of ERBB2 and CPNE3 initially appears to induce a higher rate of adhesion and spreading in C3.6 cells, which reach the maxima in a shorter duration. It is also clear that transfection with siControl RNA appears to induce a similar effect versus non transfected control.



**Figure 3.3.3** Duration of cell spreading period for HMLECs. Errors bars represent standard error of the mean of replicate experiments.

A separate method was applied to evaluate the cell adhesion and spreading dynamics illustrated by the cell impedance plots. In this case, the maxima of each plot were identified and taken to indicate the end of the cell spreading phase of the assessment. The average duration of cell spread was then determined from the maxima of all experiments as shown in Figure 3.3.3. This suggests that a knockdown of ERBB2 in c3.6 cells confers a slight shift in adhesion, which leads to a reduction in the duration of the cell spreading period towards one similar to HB4a. However, the standard error of the mean for all experimental replicates of the siERBB2 cells was not low enough to accept this observation as significant and according to a Welch's unpaired t-test had p-value > 0.05.

### 3.4. Mass Spectrometry based proteomic profiling







#### 3.4.1. TMT-LC-MS/MS based protein quantification

2D LC-MS/MS profiling with TMT-based quantification was undertaken to assess the effects of CPNE3 knockdown on global protein expression in C3.6 cells with the aim of further understanding the function of this poorly characterised gene product. Technical triplicates of CPNE3 knockdown and control cell lysates were compared in a 6-plex TMT experiment with 30 peptide fractions prepared by high pH C18 reversed-phase LC. A total of 827 filtered protein groups were identified with 1 or more peptides from 6,400 PSMs (peptide spectrum matches). Of these 641 protein groups were quantified from 3,883 PSMs. During TMT-based quantification experiments, some peptide ions were co-isolated with other ions, resulting in interference during quantification of the peptide of interest. Despite an acceptable TMT labelling efficiency (98.5%), the quantitative coverage was lower than expected and 1,758 PSMs (~70% of unused PSMs) were excluded from the quantification as the co-isolation interference was above the set threshold of 25%. The interference threshold of 25% was set as the limit above which PSMs were excluded from the quantification. Whilst the current method used sample pre-fractionation to reduce peptide co-elution and co-isolation, it appears to have been insufficient.

Along with confirmation that expression of CPNE3 was reduced, 51 other proteins were identified as up/down regulated ( $>1.5$ -fold,  $p$ -value $<0.05$ ) in response to CPNE3 knockdown (Appendix Table 3.4.1). Histones and other DNA binding proteins were prominent among the proteins identified with 11 out of 13 of these proteins being upregulated. Cadherin binding, molecular transport and calcium ion binding proteins were also prominent among the proteins identified by the profiling. Importin subunit alpha-1 or KPNA2 was found among the downregulated molecular transport proteins. Rho-related GTP-binding protein RhoC and Rho GDP-dissociation inhibitor 2 or ARHGDIB were identified as significantly upregulated.

The current study has successfully completed an initial assessment of the effect of knocking down CPNE3 on cell adhesion properties in C3.6 cells. Global protein expression was evaluated using TMT-LC-MS/MS based proteomic profiling (Appendix Table 3.4.1). The expression profiles were compared with a previously performed CPNE3 knockdown in SKBr3 cells (Worthington, 2012) and commonly changing proteins noted as targets for validation and functional characterisation. A total of 25 proteins were commonly differentially expressed in response to CPNE3 knockdown in both cell lines and are highlighted in yellow. However, only 3 of these, including CPNE3, had a fold-change greater than 1.5-fold in both C3.6 and SKBr3 cells. The proteins ARHGDIB, CPNE3 and KPNA2 showed a significant change in both C3.6 and SKBr3 cells (Table 3.4.2). ARHGDIB/Rho GDP-dissociation inhibitor 2 was upregulated in response to CPNE3 knockdown in both C3.6 and SKBr3 cells, whilst KPNA2/importin subunit alpha-1 was downregulated in both cell types.

**Table 3.4.2. Candidates displaying similar patterns of expression in response to CPNE3 knockdown in C3.6 and SKBR3 cells.** Average ratio for CPNE3kd vs Ctrl in C3.6 was compared to the SKBR3 ratio (Worthington, 2012) to identify commonly changing proteins.

Protein name	Gene name	Function	Average ratio CPNE3kd vs. Ctrl (C3.6)		Average ratio CPNE3kd vs. Ctrl (SKBr3)	
Rho GDP-dissociation inhibitor 2	ARHGDIB	GTPase activator activity, GTPase activity, Rac GTPase binding, Rho GDP-dissociation inhibitor activity		1.539		1.62
Copine-3	CPNE3	calcium-dependent phospholipid binding, calcium-dependent protein binding, protein serine/threonine kinase activity, receptor tyrosine kinase binding, RNA binding, transporter activity		0.588		0.27
Karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	KPNA2	histone deacetylase binding, nuclear localization sequence binding, protein transporter activity, RNA binding		0.550		0.67

Functional and pathway enrichment analysis of proteins upregulated or downregulated after CPNE3 knockdown was conducted using overrepresentation analysis (ORA) and the Panther Database. The search was limited to the Top 10 pathways with the highest level of significance. The search returned pathways based on 50 gene hits and 5 of these genes were associated with the top 10 panther pathways (Table 3.4.3). The angiogenesis pathway (P00005), was found to have the highest number of two genes linked to the gene set. The remainder of the top 10 pathways were each linked to one gene per pathway. Two of the five genes linked to the top 10 pathways, RHOC and CALML3, were associated with G-protein signalling. Rho related proteins are well-established regulators of the actin cytoskeleton and have been shown to play a role in modulation of gene expression and cellular proliferation (Ridley, 1995).

**Table 3.4.3** Top 10 enriched categories based on an ORA analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown using the Panther Database. P-value adjusted for multiple comparisons using the Benjamini-Hochberg FDR method.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio	FDR
P02723	Adenine and hypoxanthine salvage pathway	APRT	0.0158	62.944	0.987
P00051	TCA cycle	CS	0.026217	37.767	0.987
P02772	Pyruvate metabolism	CS	0.026217	37.767	0.987
P00008	Axon guidance mediated by Slit/Robo	RHOC	0.046775	20.981	1
P00005	Angiogenesis	CTNNB1, RHOC	0.060166	4.7806	1
P00028	Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction	CALML3	0.084348	11.444	1
P04398	p53 pathway feedback loops 2	CTNNB1	0.11591	8.2101	1
P04393	Ras Pathway	RHOC	0.17179	5.3952	1
P00016	Cytoskeletal regulation by Rho GTPase	RHOC	0.17405	5.3192	1
P00004	Alzheimer disease-presenilin pathway	CTNNB1	0.26249	3.372	1

A similar search was conducted using ORA and the Reactome Database and limited to the Top 10 Pathways with the highest level of significance (Table 3.4.4). The search returned pathways based on 50 gene hits and 18 of these genes were associated with the top 10 Reactome pathways. 9 out of the 10 pathways had a common link to 5 genes, HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D and HIST1H1E, and are related to apoptosis or senescence pathways. The exception was the infectious disease pathway (R-HSA-5663205), which was found to have the highest number of unique genes.

**Table 3.4.4** Top 10 enriched categories based on an ORA analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown using the Reactome Database. P-value adjusted for multiple comparisons using the Benjamini-Hochberg FDR method.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio	FDR
R-HSA-2559584	Formation of Senescence-Associated Heterochromatin Foci (SAHF)	<b>HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HMGA1</b>	2.86E-11	89.949	4.941E-08
R-HSA-140342	Apoptosis induced DNA fragmentation	<b>HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E</b>	1.25E-09	92.255	7.205E-07
R-HSA-211227	Activation of DNA fragmentation factor	<b>HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E</b>	1.25E-09	92.255	7.205E-07
R-HSA-2559586	DNA Damage/Telomere Stress Induced Senescence	<b>HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H4A, HMGA1</b>	3.37E-08	20.988	1.454E-05
R-HSA-75153	Apoptotic execution phase	<b>HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, CTNNB1</b>	6.50E-08	27.677	2.248E-05
R-HSA-2559583	Cellular Senescence	<b>HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H3A, HIST1H4A, HMGA1, UBC</b>	8.46E-08	11.071	2.437E-05
R-HSA-5663205	Infectious disease	CANX, CTNNB1, HMGA1, KPNA2, RANBP1, RPL19, RPL32, SLC25A6, SSRP1, UBC	2.83E-06	6.2792	6.985E-04
R-HSA-109581	Apoptosis	<b>HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, CTNNB1, UBC</b>	5.55E-06	9.9352	1.196E-03
R-HSA-5357801	Programmed Cell Death	<b>HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, CTNNB1, UBC</b>	6.23E-06	9.7619	1.196E-03
R-HSA-2262752	Cellular responses to stress	<b>HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H3A, HIST1H4A, HMGA1, PRDX3, UBC</b>	7.47E-06	5.6306	1.290E-03

The 52 proteins identified as up/down regulated in response to CPNE3 knockdown were further evaluated using Qiagen's Ingenuity Pathway Analysis to identify proteins that are causally relevant to breast cancer or the cell adhesion phenotype. 29 proteins were discovered to have a causal relevance to breast cancer (Table 3.4.5) and 5 proteins, ARHGDIB, CTNNB1, KPNA2, RAP1B and RHOC, found to have a functional link to cell adhesion. ARHGDIB, KPNA2 and RHOC have

previously been cited as potential biomarkers for a diagnostic application and CTNNB1 for a disease progression application.

**Table 3.4.5** Proteins with a causal relevance to breast cancer or the cell adhesion phenotype based on a Qiagen Ingenuity Pathway Analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown.

Gene Symbol	Molecule Type	Expr Fold Change	Disease or Function	Biomarker Application	Cell Adhesion Role
PTGES	enzyme	3.397	Breast cancer	none	none
CANX	other	3.129	Breast carcinoma	none	none
SLC25A6	transporter	2.995	Breast cancer	none	none
TPD52	other	2.083	Ductal breast carcinoma	none	none
SSBP1	other	1.992	Migration of breast cancer cell lines	none	none
VDAC2	ion channel	1.990	Breast cancer	none	none
RHOC	enzyme	1.973	Migration of breast cancer cell lines	diagnosis	Turnover of focal adhesions
RAP1B	enzyme	1.911	Breast cancer	none	Adhesion of cells
SMG1	kinase	1.691	Ductal breast carcinoma	none	none
RANBP1	other	1.677	Breast cancer	none	none
UBC	enzyme	1.650	Breast carcinoma	none	none
HMGA1	transcription regulator	1.638	Migration of breast cancer cell lines	none	none
ALPP	phosphatase	1.607	HER2-positive carcinoma of breast	none	none
CTNNB1	transcription regulator	1.586	Migration of breast cancer cell lines	disease progression	Quantity of focal adhesions
HIST1H1C	other	1.567	Cell proliferation of breast cancer cell lines	none	none
GSDMC	other	1.562	Breast carcinoma	none	none
ARHGDIB	enzyme	1.539	Migration of breast cancer cell lines	diagnosis	Adhesion of cells
CYCS	enzyme	1.492	Breast cancer	none	none
SERPINH1	other	0.677	Migration of breast cancer cell lines	none	none
EIF4G1	translation regulator	0.674	Ductal breast carcinoma	none	none
PYGM	enzyme	0.661	Ductal breast carcinoma	none	none
KARS1	enzyme	0.650	Breast cancer	none	none
PASK	kinase	0.649	Ductal breast carcinoma	none	none
RPL19	other	0.618	Breast cancer	none	none
RUNX1	transcription regulator	0.615	Breast cancer	none	none
CPNE3	kinase	0.588	Migration of breast cancer cell lines	none	none
KPNA2	other	0.550	Migration of breast cell lines	diagnosis	Adhesion of breast cell lines
SSRP1	transcription regulator	0.467	Cell death of breast cancer cell lines	none	none
EIF3F	translation regulator	0.451	Breast cancer	none	none

### 3.4.2. Label-Free protein quantification

A follow-up protein quantification method using label-free shotgun proteomics was used to further evaluate the effect of CPNE3 knockdown on global protein expression in C3.6 cells. Technical triplicates of CPNE3 knockdown and control cell lysates were compared in a label-free experiment and a total of 992 filtered protein groups were identified with 1 or more peptides. Of these 850 protein groups were quantified. In contrast to the TMT method, the quantification did not yield any PSMs since the spectra generated for DIA are of all the ions fragmented together, as opposed to individual precursor ions generated for DDA.

Along with confirmation that the expression of CPNE3 was reduced, 39 other proteins were identified as up/down-regulated ( $>1.5$ -fold,  $p$ -value  $< 0.05$ ) in response to CPNE3 knockdown (Appendix Table 3.4.6). Noteworthy was the presence of only one upregulated protein among the 39 identified proteins. The rest of the proteins were downregulated in response to CPNE3 knockdown. The current study has successfully completed a secondary assessment of the effect of knocking down CPNE3 on global protein expression in C3.6 cells. Global protein expression was evaluated using label-free proteomic profiling (Appendix Table 3.4.6). The expression profiles were compared with a previously performed CPNE3 knockdown in SKBr3 cells (Worthington, 2012) and commonly changing proteins noted as targets for validation and functional characterisation. None of the proteins identified using the label-free quantification method was found to be commonly expressed in response to CPNE3 knockdown in both cell lines.

Functional and pathway enrichment analysis of proteins upregulated or downregulated after CPNE3 knockdown was conducted using ORA and the Panther Database. The search was limited to the Top 10 pathways with the highest level of significance. The search returned pathways based on 39 gene hits and 8 of these genes were associated with the top 10 panther pathways (Table 3.4.7). The Huntington's disease pathway, P00029, was found to have the highest number of genes, consisting of two genes linked to the pathway gene set. The remainder of the top 10 pathways were each linked to



one gene per pathway. Four pathways, cell cycle, glycolysis, TGF-beta signalling and cadherin signalling are of particular interest to cancer. Actin-related protein 2 (ACTR2) was the most frequently occurring gene among the top 10 categories.

**Table 3.4.7** Top 10 enriched categories based on an ORA analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown using the Panther Database. P-value adjusted for multiple comparisons using the Benjamini-Hochberg FDR method.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio	FDR
P02771	Pyrimidine Metabolism	DPYSL2	0.039091	25.178	1
P04395	Vasopressin synthesis	PREP	0.042925	22.889	1
P00013	Cell cycle	RPA3	0.061889	15.736	1
P00024	Glycolysis	PFKM	0.065641	14.81	1
P00007	Axon guidance mediated by semaphorins	DPYSL2	0.073106	13.251	1
P00029	Huntington disease	ACTR2, DYNC1LI2	0.08903	3.9035	1
P00049	Parkinson disease	PSMA4	0.30321	2.829	1
P00052	TGF-beta signaling pathway	FKBP1A	0.30896	2.7668	1
P00004	Alzheimer disease-presenilin pathway	ACTR2	0.36684	2.248	1
P00012	Cadherin signaling pathway	ACTR2	0.46758	1.6456	1

A similar search was conducted using ORA and the Reactome Database and limited to the Top 10 Pathways with the highest level of significance (Table 3.4.8). The search returned pathways based on 39 gene hits and 9 of these genes were associated with the top 10 Reactome pathways. 9 out of the 10 pathways had a common link to 3 genes, MCM3, PSMA4 and PSMB6, and are related to cell cycle pathways. The exception was the metabolism of amino acids and derivatives pathway, R-HSA-71291, which was found to have the highest number of unique genes.

**Table 3.4.8** Top 10 enriched categories based on an ORA analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown using the Reactome Database. P-value adjusted for multiple comparisons using the Benjamini-Hochberg FDR method.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio	FDR
R-HSA-71291	Metabolism of amino acids and derivatives	CKMT1A, NAALAD2, OAT, PSMA4, PSMB6, RPS15, RPS21	9.69E-05	6.2397	0.1025
R-HSA-69002	DNA Replication Pre-Initiation	<b>MCM3, PSMA4, PSMB6, RPA3</b>	1.19E-04	15.521	0.1025
R-HSA-69239	Synthesis of DNA	<b>MCM3, PSMA4, PSMB6, RPA3</b>	4.33E-04	11.086	0.2152
R-HSA-69306	DNA Replication	<b>MCM3, PSMA4, PSMB6, RPA3</b>	5.54E-04	10.388	0.2152
R-HSA-69206	G1/S Transition	<b>MCM3, PSMA4, PSMB6, RPA3</b>	6.23E-04	10.071	0.2152
R-HSA-453279	Mitotic G1-G1/S phases	<b>MCM3, PSMA4, PSMB6, RPA3</b>	1.01E-03	8.854	0.2497
R-HSA-68867	Assembly of the pre-replicative complex	<b>MCM3, PSMA4, PSMB6</b>	1.11E-03	14.551	0.2497
R-HSA-68949	Orc1 removal from chromatin	<b>MCM3, PSMA4, PSMB6</b>	1.26E-03	13.936	0.2497
R-HSA-69242	S Phase	<b>MCM3, PSMA4, PSMB6, RPA3</b>	1.34E-03	8.1941	0.2497
R-HSA-69481	G2/M Checkpoints	<b>MCM3, PSMA4, PSMB6, RPA3</b>	1.57E-03	7.8527	0.2497

The 40 proteins identified as up/down-regulated in response to CPNE3 knockdown were further evaluated using Qiagen's Ingenuity Pathway Analysis to explore the proteins that are causally relevant to breast cancer or the cell adhesion phenotype. 18 proteins were discovered to have a causal relevance to breast cancer (Table 3.4.9) and 3 proteins, ACTA2, ACTN2 and PTBP1, were found to have a functional link to cell adhesion. None of the 18 proteins has previously been cited as potential biomarkers.

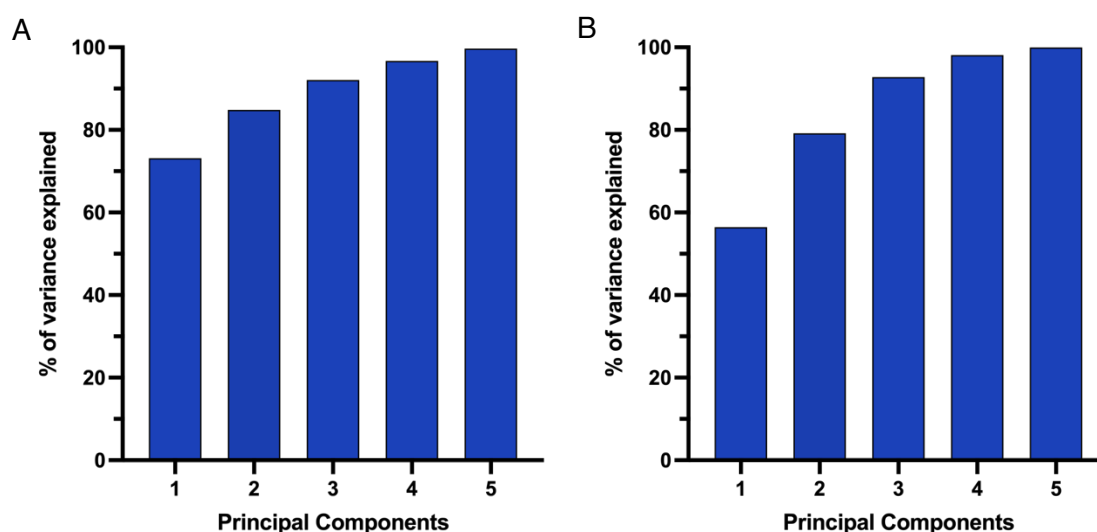
**Table 3.4.9** Proteins with a causal relevance to breast cancer or the cell adhesion phenotype based on a Qiagen Ingenuity Pathway Analysis of proteins significantly upregulated or downregulated after CPNE3 knockdown.

Gene Symbol	Molecule Type	Expr Fold Change	Disease or Function	Biomarker Application	Cell Adhesion Role
IGF2BP2	translation regulator	0.653	Breast carcinoma	none	none
FKBP1A	enzyme	0.650	HER2 negative breast cancer	none	none
MCM3	enzyme	0.634	Ductal breast carcinoma	none	none
PRDX5	enzyme	0.632	Breast cancer	none	none
CLNS1A	ion channel	0.611	Breast cancer	none	none
SND1	enzyme	0.611	Breast cancer	none	none
NONO	transcription regulator	0.607	Breast cancer	none	none
ARRDC3	other	0.603	Migration of breast cancer cell lines	none	none
ACTA2	other	0.561	Breast carcinoma	none	Formation of focal adhesions
NASP	other	0.555	Breast carcinoma	none	none
UQCRC2	enzyme	0.537	Ductal breast carcinoma	none	none
PTBP1	enzyme	0.535	Cell proliferation of breast cancer cell lines	none	Size of focal adhesions
PPP2R2B	phosphatase	0.529	Cell proliferation of breast cancer cell lines	none	none
DHX9	enzyme	0.438	Breast carcinoma	none	none
UCHL3	peptidase	0.437	Breast cancer	none	none
ACTN2	transcription regulator	0.421	Breast carcinoma	none	Adhesion of cells
CTNNA1	other	0.387	Ductal breast carcinoma	none	none
CPNE3	kinase	0.319	Migration of breast cancer cell lines	none	none

### 3.5. Cluster-based method for the evaluation of mass spectrometry data

#### 3.5.1. Principal component analysis of mass spectrometry data

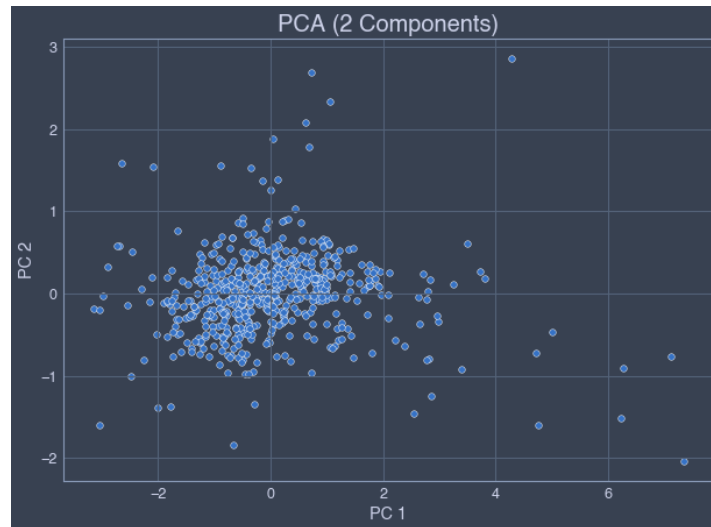
To identify potential downstream interaction partners and further characterise the functional role of CPNE3, principal component analysis (PCA) was used to reduce the collinearity of upregulated or downregulated proteins in response to CPNE3 knockdown. PCA was used to evaluate both the TMT and Label-Free mass spectrometry data and implemented using Python's Scikit-Learn library. The explained variance was used to determine the number of PCA components and to tell us how much information (variance) can be attributed to each of the selected principal components (Figure 3.5.1).



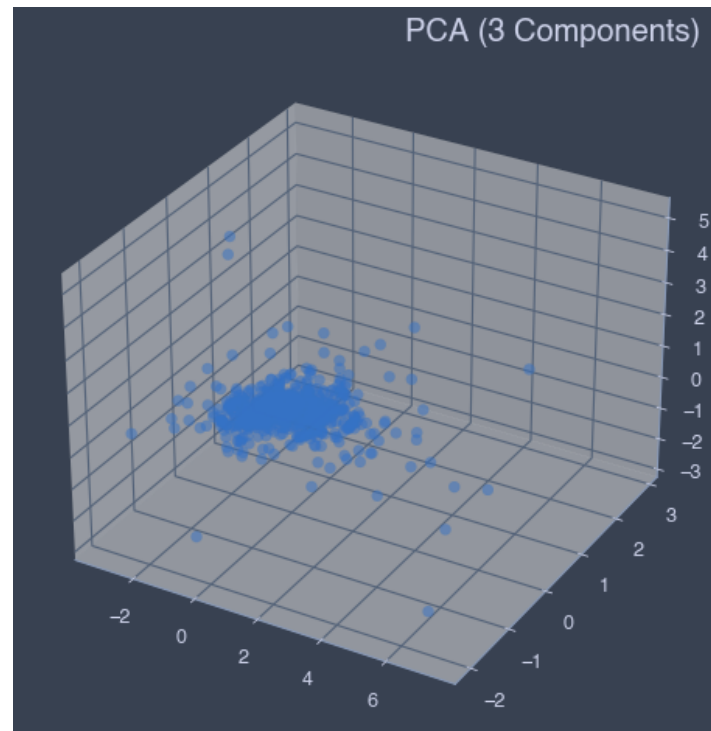
**Figure 3.5.1** Explained variance of mass spectrometry data. **A)** TMT **B)** Label Free

Given the results, it can be said that 84.85% of the variance of our TMT data is explained by the 2 Principal Components and 92.10% of the variance is explained by the 3 Principal Components (Figure 3.5.2). In addition, 79.20% of the variance of our Label Free Data is explained by the 2 Principal Components and 92.79% of the variance is explained by the 3 Principal Components (Figure 3.5.3).

A

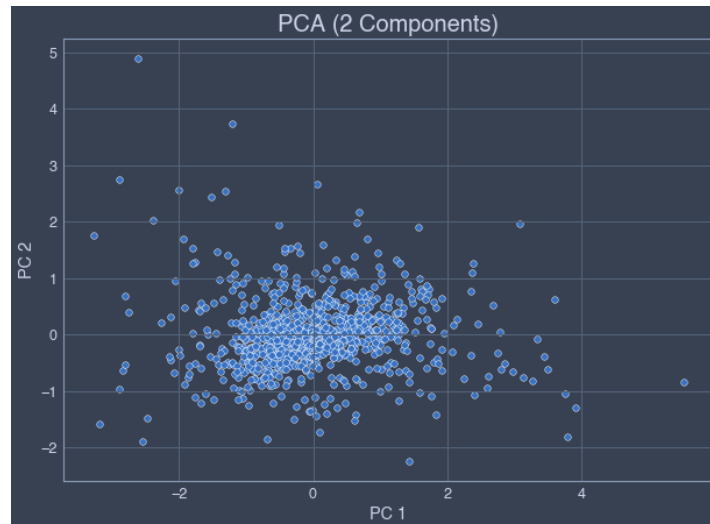


B

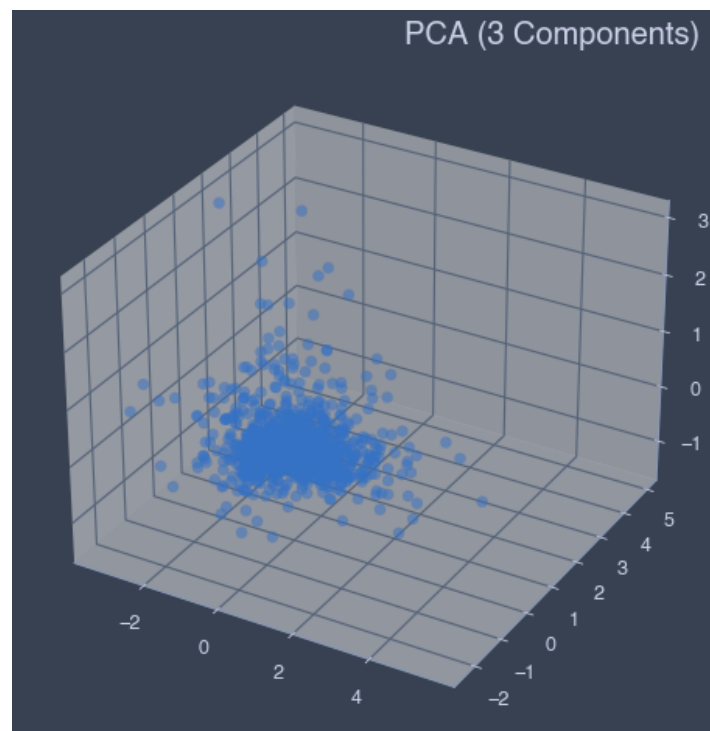


**Figure 3.5.2** Principal Component Analysis of log2 transformed TMT mass spectrometry data. Each data point represents an individual protein. **A)** A scatter plot of the first two principal components for the dataset. The first two principal components explain 84.85% of the total variation in the data. **B)** A scatter plot of the first three principal components for the dataset. The first three principal components explain 92.10% of the total variation in the data.

A



B



**Figure 3.5.3** Principal Component Analysis of log2 transformed Label Free mass spectrometry data. Each data point represents an individual protein. **A)** A scatter plot of the first two principal components for the dataset. The first two principal components explain 79.20% of the total variation in the data. **B)** A scatter plot of the first three principal components for the dataset. The first three principal components explain 92.79% of the total variation in the data.

In both cases, 3 Principal Component has higher explained variance than 2 Principal Component and the former presents clusters better than the latter. Both plots show lumped data points about the origin. Thus, it is more favourable to plot on a 3D plane. However, CPNE3 expression is

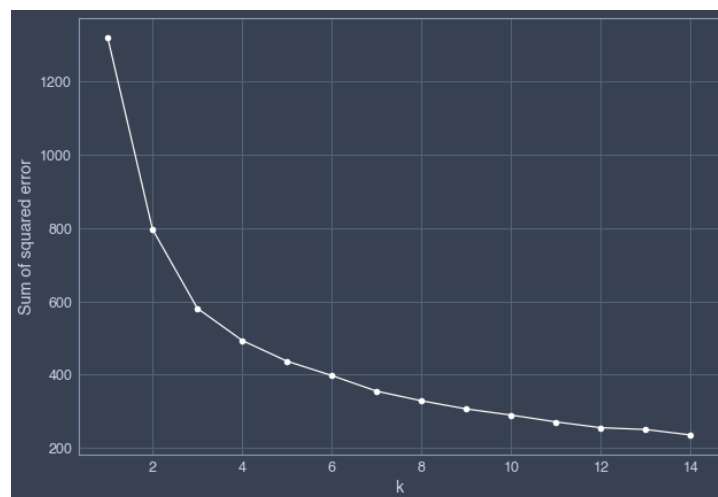
expected to fall within the lower quartile range of the data and its cluster can be visualised accurately on a 2 Principal Component plot or 2D plane.

### 3.5.2. K-Means clustering of mass spectrometry data

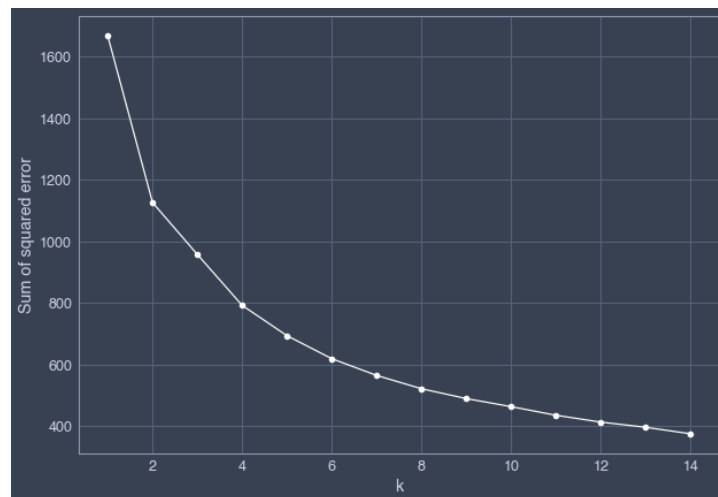
To uncover meaningful groups from our data and identify proteins with a strong positive correlation to CPNE3 expression, K-Means clustering was implemented using Python's Scikit-Learn library. The number of clusters to be used for the K-Means clustering were determined by an iterative statistical technique for determining the optimal number of clusters by running the K-Means algorithm for a range of cluster values. The range of cluster values was selected as n=15 to ensure that the full range of features in our datasets were represented by the technique. The optimal k was determined by elbow method and found to be k=6 for both datasets (Figure 3.5.4).

Cluster 4 was found to contain CPNE3 and 51 other proteins with a high expression similarity score for the TMT data (Appendix Table 3.5.1) and cluster 1 was found to contain CPNE3 and 41 other proteins with a high expression similarity score for the label-free data (Appendix Table 3.5.4)(Figure 3.5.5). Functional and pathway enrichment analysis of the proteins in cluster 4 was conducted using overrepresentation analysis (ORA) and the Panther Database. The search was limited to the Top 10 pathways with the highest level of significance. The search returned pathways based on 51 gene hits and 9 of these genes were associated with the top 10 panther pathways (Table 3.5.2). However, only three of the pathways had a p-value < 0.05, arginine biosynthesis, adenine and hypoxanthine salvage pathway and heme biosynthesis. The FDR values for the enrichment analysis was 1 for all pathways. Therefore, none of the enriched categories identified were significant enough to accept.

A



B



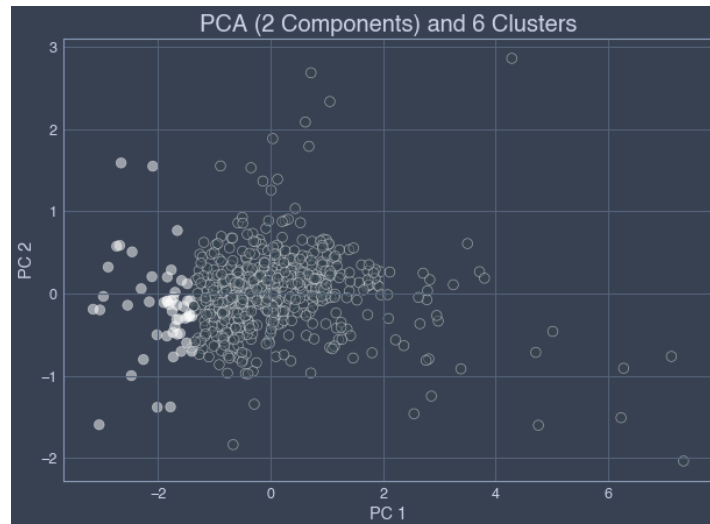
**Figure 3.5.4** Optimal number of clusters as determined by elbow method for TMT and Label Free mass spectrometry data. X-axis: number of clusters (k). Y-axis: sum of squared error.

**A)** Elbow graph shows 6 optimal clusters, determined by elbow method analysis on TMT data.

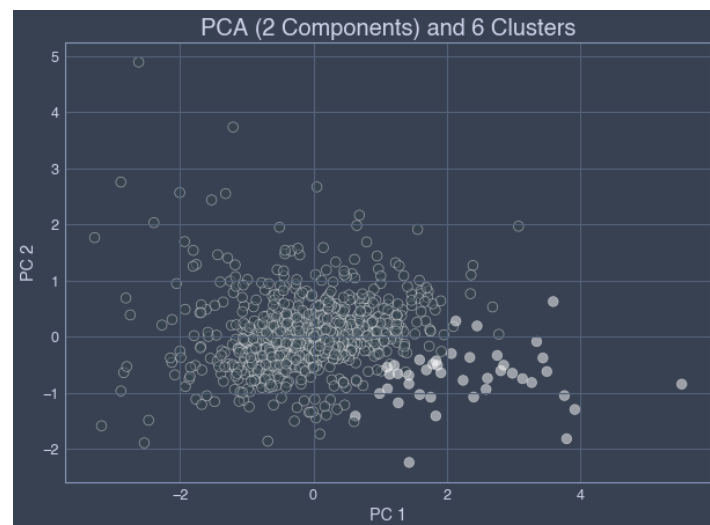
**B)** Elbow graph shows 6 optimal clusters, determined by elbow method analysis on Label Free data.



A



B



**Figure 3.5.5** Two-component PCA visualisations of TMT and Label Free data. Each point represents an individual protein and the component of interest is highlighted in white based on k-means clustering **A)** Two-component PCA visualisation of TMT data: The plot shows a two-dimensional representation of TMT-labelled proteomics data in the first two principal components. **B)** Two-component PCA visualisation of Label Free data: The plot shows a two-dimensional representation of Label Free proteomics data in the first two principal components.

**Table 3.5.2** Top 10 enriched categories based on an ORA analysis of proteins with a strong positive correlation to CPNE3 using K-Means clustering on TMT data and the Panther Database. P-value adjusted for multiple comparisons using the Benjamini-Hochberg FDR method.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio	FDR
P02728	Arginine biosynthesis	ASS1	0.02102	47.208	1
P02723	Adenine and hypoxanthine salvage pathway	APRT	0.02102	47.208	1
P02746	Heme biosynthesis	QARS	0.041652	23.604	1
P00017	DNA replication	XRN2	0.065241	14.908	1
P00026	Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	CLTA, PYGM	0.09476	3.7517	1
P00060	Ubiquitin proteasome pathway	UBA2	0.145393	6.4375	1
P00006	Apoptosis signaling pathway	MAP4K4	0.323819	2.6227	1
P00027	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	CLTA	0.348522	2.4004	1

A similar search was conducted to analyse proteins in cluster 4 using ORA and the Reactome Database and was limited to the Top 10 Pathways with the highest level of significance (Table 3.5.3). The search returned pathways based on 51 gene hits and 27 of these genes were associated with the top 10 Reactome pathways. All top 10 pathways had a p-value < 0.05 and FDR < 0.05, and 9 out of the 10 pathways were either related to mRNA processing or translation pathways. The exception was the influenza infection pathway (R-HSA-168254), which was found to have a link to KPNA2. The pathways with the highest enrichment ratio, L13a-mediated translational silencing of Ceruloplasmin expression (R-HSA-156827), GTP hydrolysis and joining of the 60S ribosomal subunit (R-HSA-72706), Eukaryotic Translation Initiation (R-HSA-72613) and Cap-dependent Translation Initiation (R-HSA-72737), are all related to translation and share the same genes RPL7, RPL6, RPL28, EIF4H, RPL27, EIF3F, RPL32 and EIF4G1.

**Table 3.5.3** Top 10 enriched categories based on an ORA analysis of proteins with a strong positive correlation to CPNE3 using K-Means clustering on TMT data and the Reactome database. P-value adjusted for multiple comparisons using the Benjamini-Hochberg FDR method.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio	FDR
R-HSA-8953854	Metabolism of RNA	LSM5, RPL7, RPL6, RPL28, SRSF3, RPL27, HNRNPD, HNRNPU, SNRPD3, SNRPB, RPL32, EIF4A3, DHX15, EIF4G1, NOP2, SF3B2, XRN2, POLR2B, GEMIN4	1.26E-11	6.4677	2.2E-08
R-HSA-72766	Translation	RPL7, RPL6, RPL28, EIF4H, RPL27, EIF3F, QARS, RPL32, TARS, EIF4G1, HARS, KARS	2.59E-09	9.4612	2.2E-06
R-HSA-72163	mRNA Splicing - Major Pathway	LSM5, SRSF3, HNRNPD, HNRNPU, SNRPD3, SNRPB, EIF4A3, DHX15, SF3B2, POLR2B	4.56E-09	12.537	2.6E-06
R-HSA-72172	mRNA Splicing	LSM5, SRSF3, HNRNPD, HNRNPU, SNRPD3, SNRPB, EIF4A3, DHX15, SF3B2, POLR2B	6.89E-09	12.012	3E-06
R-HSA-156827	L13a-mediated translational silencing of Ceruloplasmin expression	RPL7, RPL6, RPL28, EIF4H, RPL27, EIF3F, RPL32, EIF4G1	2.17E-08	16.536	6.7E-06
R-HSA-72706	GTP hydrolysis and joining of the 60S ribosomal subunit	RPL7, RPL6, RPL28, EIF4H, RPL27, EIF3F, RPL32, EIF4G1	2.33E-08	16.388	6.7E-06
R-HSA-72613	Eukaryotic Translation Initiation	RPL7, RPL6, RPL28, EIF4H, RPL27, EIF3F, RPL32, EIF4G1	3.76E-08	15.424	8.1E-06
R-HSA-72737	Cap-dependent Translation Initiation	RPL7, RPL6, RPL28, EIF4H, RPL27, EIF3F, RPL32, EIF4G1	3.76E-08	15.424	8.1E-06
R-HSA-72203	Processing of Capped Intron-Containing Pre-mRNA	LSM5, SRSF3, HNRNPD, HNRNPU, SNRPD3, SNRPB, EIF4A3, DHX15, SF3B2, POLR2B	6.85E-08	9.4418	1.3E-05
R-HSA-168254	Influenza Infection	RPL7, RPL6, RPL28, RPL27, CLTA, RPL32, KPNA2, POLR2B	2.79E-07	11.919	4.8E-05

Functional and pathway enrichment analysis of the proteins in cluster 1 was conducted using overrepresentation analysis (ORA) and the Panther Database. The search was limited to the Top 10 pathways with the highest level of significance. The search returned pathways based on 40 gene hits and 7 of these genes were associated with the top 10 panther pathways (Table 3.5.5). However, only one of the pathways had a p-value < 0.05.

**Table 3.5.5** Top 10 enriched categories based on an ORA analysis of proteins with a strong positive correlation to CPNE3 using K-Means clustering on Label-Free data and the Panther Database. P-value adjusted for multiple comparisons using the Benjamini-Hochberg FDR method.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio	FDR
P04396	Vitamin D metabolism and pathway	GC	0.030525	32.371	1
P00012	Cadherin signaling pathway	CTNNA1, CDH5	0.075995	4.2316	1
P02738	De novo purine biosynthesis	AK2	0.083461	11.561	1
P00057	Wnt signaling pathway	CTNNA1, CDH5	0.227315	2.2021	1
P00006	Apoptosis signaling pathway	EIF2S1	0.289868	2.9974	1
P00004	Alzheimer disease-presenilin pathway	CTNNA1	0.299044	2.8903	1
P00034	Integrin signalling pathway	ACTN2	0.413326	1.9501	1
P06959	CCKR signaling map	ACAT1	0.424976	1.8821	1

A similar search was conducted to analyse proteins in cluster 1 using ORA and the Reactome Database and was limited to the Top 10 Pathways with the highest level of significance (Table 3.5.6). The search returned pathways based on 40 gene hits and 17 of these genes were associated with the top 10 Reactome pathways. Three pathways had distinctly unique gene associations, branched-chain amino acid catabolism (R-HSA-70895), PTEN regulation (R-HSA-6807070) and neutrophil degranulation (R-HSA-6798695), which was found to have a link to CPNE3.

**Table 3.5.6** Top 10 enriched categories based on an ORA analysis of proteins with a strong positive correlation to CPNE3 using K-Means clustering on Label-Free data and the Reactome database. P-value adjusted for multiple comparisons using the Benjamini-Hochberg FDR method.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio	FDR
R-HSA-72163	mRNA Splicing - Major Pathway	DHX9, HNRNPU, HNRNPH1, HNRNPR	1.68E-03	7.6896	1
R-HSA-70895	Branched-chain amino acid catabolism	ACAT1, DLD	1.90E-03	30.591	1
R-HSA-72172	mRNA Splicing	DHX9, HNRNPU, HNRNPH1, HNRNPR	1.97E-03	7.3675	1
R-HSA-5218920	VEGFR2 mediated vascular permeability	CTNNA1, CDH5	3.02E-03	24.262	1
R-HSA-418990	Adherens junctions interactions	CTNNA1, CDH5	3.90E-03	21.321	1
R-HSA-72203	Processing of Capped Intron-Containing Pre-mRNA	DHX9, HNRNPU, HNRNPH1, HNRNPR	4.69E-03	5.7909	1
R-HSA-1500931	Cell-Cell communication	ACTN2, CTNNA1, CDH5	5.69E-03	8.1814	1
R-HSA-6807070	PTEN Regulation	PSMC6, REST, USP13	7.14E-03	7.5386	1
R-HSA-6798695	Neutrophil degranulation	CPNE3, EEF2, ARMC8, PAFAH1B2, ERP44	1.05E-02	3.6722	1
R-HSA-421270	Cell-cell junction organization	CTNNA1, CDH5	1.41E-02	10.994	1

### 3.6. Chapter Conclusion

In this chapter, we presented the results of our study that aimed to investigate the functional role of CPNE3 in ERBB2 overexpressing breast cancer. Our findings demonstrated that ERBB2 and CPNE3 knockdown had a subtle effect on cell adhesion in HMLECs, which has previously been demonstrated in SKBR3 cells (Worthington, 2012). While the effect of ERBB2 overexpression on global protein expression in ERBB2 overexpressing HMLECs has previously been shown (Worthington et al., 2017) and downregulation of cell adhesion proteins such as ITGA6 and ITGB4 in HMLECs that overexpress ERBB2 was identified, this study demonstrates that the downregulation of ITGA6 and ITGB4 is not reversed by siRNA mediated knockdown of ERBB2 or CPNE3 in HMLECs. Moreover, our analysis demonstrates the effect of siRNA mediated knockdown of CPNE3 on global protein expression in ERBB2 overexpressing HMLECs using two mass spectrometry methods to identify proteins that explain the functional role of CPNE3 in ERBB2 overexpressing breast cancer.

Our results are in agreement with previous studies by Worthington (2012) who demonstrated a partial reduction in cell adhesion for both ERBB2 and CPNE3 expression, and Worthington et al. (2017) who showed differential expression of cell adhesion-related proteins due to ERBB2 overexpression in HMLECs using SILAC MS/MS. The present study has evaluated global protein expression in response to CPNE3 knockdown in ERBB2 over expressing HMLECs and identified hitherto unknown biomarker candidates linked to CPNE3 expression. These candidates, ARHGDIB and KPNA2 are thought to be independent of disease progression as they are both evident following knockdown of CPNE3 in SKBR3 (Worthington, 2012), a metastatic breast cancer cell line. Thus, their link to CPNE3 expression is thought to be related to a functional role of CPNE3. Further evaluation of global protein expression using two separate methods and K-means clustering with PCA identified a cluster of proteins that showed a strong positive correlation with CPNE3 following siRNA mediated knockdown. These proteins were found to be the minimum number of parameters needed to generate a data description of the proteomic profiles generated by each method for comparison using functional annotation databases. A high expression similarity between CPNE3 and ribonucleoprotein and ribonucleoprotein associated proteins which are known to interact with actin during mRNA processing is evident in both datasets.

Moving forward, additional experiments could be conducted to further validate our findings with treatment of HMLEC cells with growth factors such as HRG and EGF and real time cell adhesion assays to assess cell adhesion and cell spreading on an adherent surface in response to CPNE3 knockdown in HMLECs. Phosphoproteomic profiling of HMLECs has previously identified a total of 50 phosphosites that were previously not known to be regulated by ErbB growth factor treatment and/or ErbB2 overexpression, with EGF stimulation being primarily enriched for chromosomal and chromatin-binding proteins and HRG stimulation being enriched for cytoskeletal proteins (Worthington et al., 2017). Investigating the role of CPNE3 expression under growth factor treatment using LC-MS/MS based phosphoproteomic profiling could provide important insights into the mechanisms driving the aggressive behaviour of this cancer subtype. Furthermore, exploring the signalling pathways that interact with CPNE3 in ERBB2 overexpressing cells could identify additional biomarkers or targets for therapeutic intervention.

In the next chapter, we further interpret and discuss the findings presented in this chapter in the context of their expression in tumours that develop under growth factor rich TME and explore the expression of proteins that correlate with CPNE3 in clinical patient samples. The TME consists of various immune cells, such as T cells, B cells, and macrophages, which have been shown to play a crucial role in modulating the tumour immune response. Moreover, growth factors, such as epidermal growth factor (EGF), fibroblast growth factor (FGF), and transforming growth factor-beta (TGF- $\beta$ ), are also present in the TME and can promote cancer cell survival, proliferation, angiogenesis, and invasion (Anderson & Simon, 2020; Baghban et al., 2020). The presence of these factors in the TME can result in the dysregulation of several cellular signalling pathways, including the ERBB2/HER2 signalling pathway, which has been implicated in the development and progression of several cancer types, including breast cancer (B. Zhou & Hung MC, 2003). This dysregulation of signalling pathways in response to the TME may influence the expression of the candidate proteins discovered and presented in this chapter following siRNA mediated knockdown of CPNE3 in HMLECs. Therefore, clinical validation of our findings using patient samples derived from a TME of the respective patients is essential to confirm the relevance and clinical significance of our results.

## Chapter 4

### 4. Clinical validation of CPNE3 related gene expression

#### 4.1. Chapter Introduction

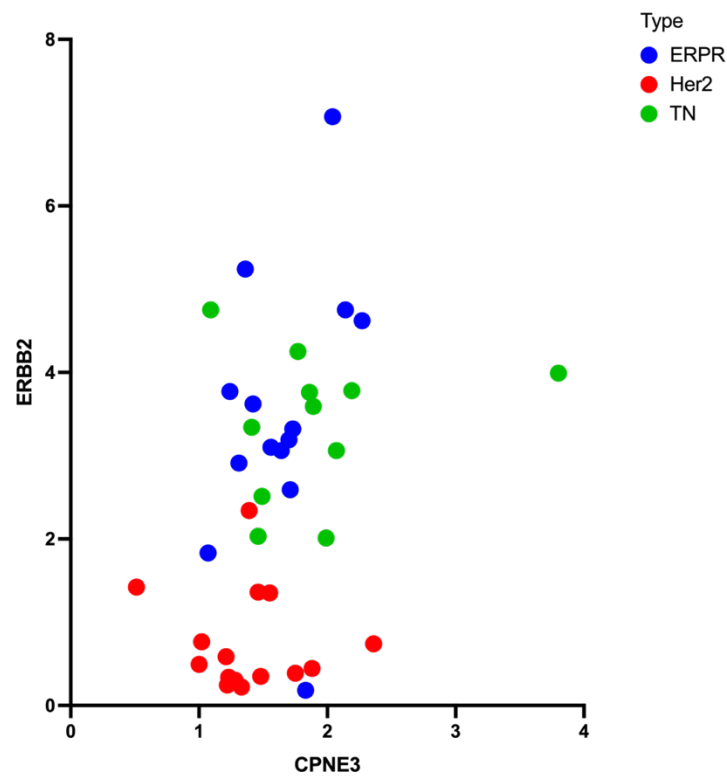
Although the role of ERBB2 and its overexpression in breast cancers is well documented (Slamon et al., 1987; Ross & Fletcher, 1999; Yu & Hung, 2000; Tan & Yu, 2007; Iqbal & Iqbal, 2014; Daemen & Manning, 2018; Marchiò et al., 2021), the underlying molecular mechanisms involved in disease development remain unclear. Mass spectrometry technology is a powerful tool for evaluating global protein expression changes and its usefulness in the study of breast cancer proteomics and identifying potential downstream candidates has been well established (see Chapter 1 and 3). However, few studies report the clinical validation of potential candidate proteins operating downstream, such as CPNE3 which is poorly characterised. This study addresses the functional characterisation and clinical validation of CPNE3 by performing a number of computational studies to evaluate the role of CPNE3 as a biomarker in HER2 positive breast cancer patients. These *in-silico* experiments were carried out using protein expression profiles of 40 breast cancer patients derived through SILAC based mass spectrometry (Tyanova et al., 2016) and 75 breast cancer patient protein expression profiles derived through iTRAQ based mass spectrometry by CPTAC (Ellis et al., 2013). This chapter describes the comparison of ERBB2 and CPNE3 expression patterns between HER2 or ERBB2 overexpressing patients and other breast cancer subtypes, and the subsequent harnessing of these expression patterns to develop a predictive model for HER2 diagnosis using mass spectrometry derived protein expression data. Such analysis may help to validate downstream signalling candidates in ERBB2 overexpressing breast cancer and their potential role as diagnostic biomarkers.

#### 4.2. Discovery of HER2 candidate biomarkers in invasive ductal carcinoma (IDC) patients

To demonstrate the validity of CPNE3 as a diagnostic or prognostic marker in HER2/ERBB2 overexpressing breast cancer and to discover potential clinical biomarkers from the previously



identified proteins (Appendix Table 3.4.1), a statistical evaluation of 40 breast invasive ductal carcinoma (IDC) patient samples comprised of HER2 positive (HER2), Estrogen Receptor and Progesterone Receptor positive (ERPR), and Triple Negative (TN) patients was carried out. The expression of HER2/ERBB2 and CPNE3 was evaluated in the samples and visualised with a pairwise plot (Figure 4.2.1).



**Figure 4.2.1** Pair plot of ERBB2 vs CPNE3 expression for ERPR, HER2 and TN for 40 female breast cancer patients.

Sparse clustering of the HER2 samples was observed in comparison to the ERPR and TN samples. However, the distribution of CPNE3 was narrower and non-specific compared to that of ERBB2. In addition, CPNE3 appears to make a minimal contribution to the overall clustering. This observation is interesting as treatment with growth factors has been shown to induce a moderate increase in CPNE3 expression in the HB4a cells, which do not overexpress ERBB2 (Bertani, 2005; Gharbi et al., 2002; White et al., 2004). Therefore, the distribution of CPNE3 in the 40 breast cancer patient samples suggests that a growth factor rich tumour microenvironment might affect the expected ratio of CPNE3 to ERBB2 expression. Furthermore, the patient samples are known to be largely distributed between clinical stages I, IIA and IIB (Table 4.2.1) and CPNE3 might undergo variable expression at different

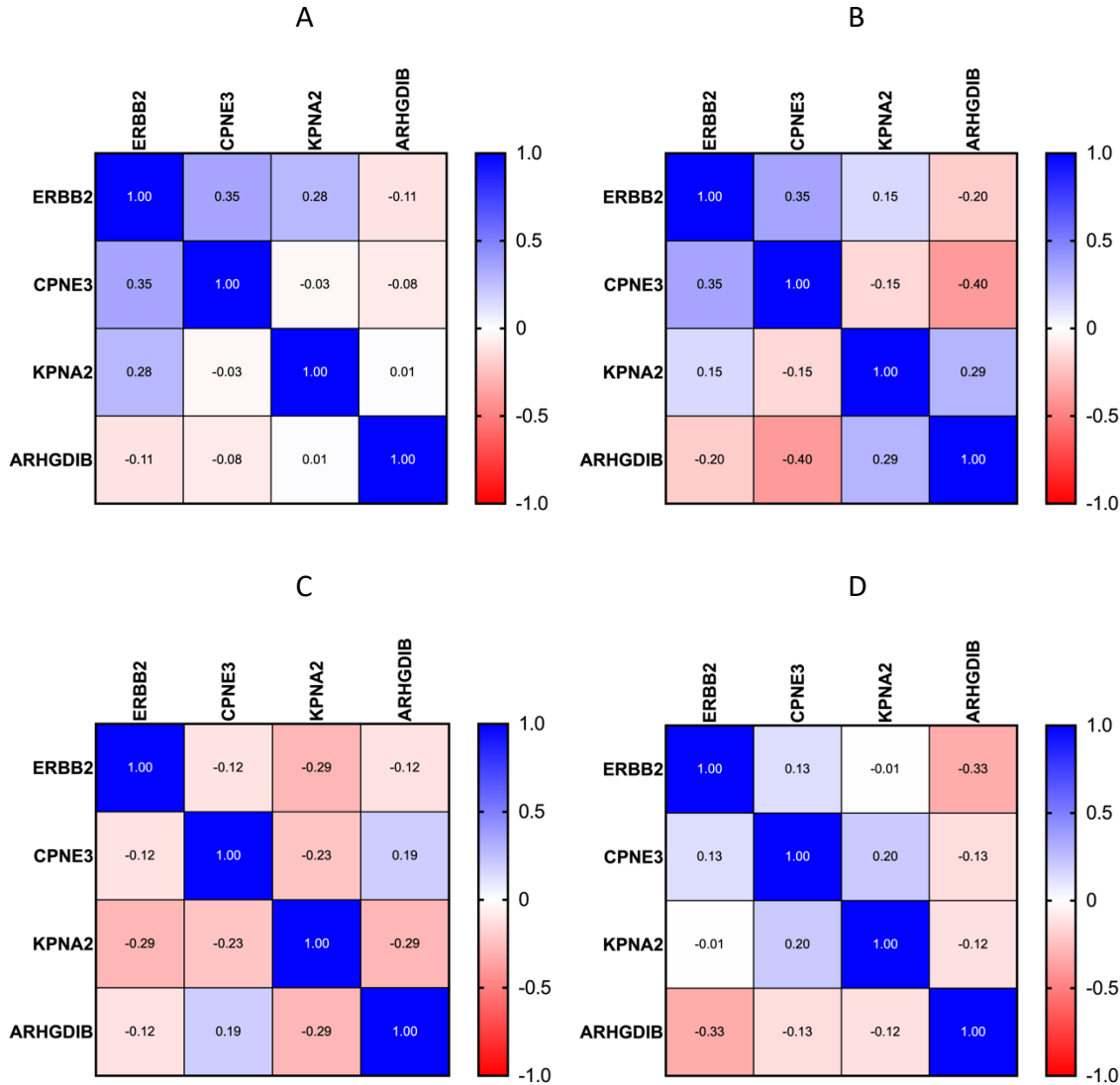
clinical stages. With this in mind, the data for the 40 breast cancer patient samples can be confidently used as an *in-silico* clinical discovery cohort for candidate proteins or biomarkers associated with CPNE3 expression and is herein referred to as the discovery cohort (Table 4.2.1).

**Table 4.2.1** Characteristics of female breast cancer patients in the discovery and validation cohorts.

Baseline variable	Discovery cohort		Validation cohort	
	Cases (n=15)	Controls (n=25)	Cases (n=6)	Controls (n=69)
<b>Sex (Female)</b>	15	25	6	69
<b>Mean age (Years)</b>			54.1 ± 10.7	58.9 ± 13.7
<b>Suspicious pathology</b>	0	0	0	1
<b>Clinical stage</b>				
0 Tis N0 M0	-	-	-	-
I T1 N0 M0	2	6	-	6
IIA T1 N1 M0, T2 N0 M0	3	12	3	24
IIB T2 N1 M0	8	6	2	20
IIIA T3 N1 M0	-	-	1	9
IIIB T4 N0-N3 M0	1	-	-	5
IIIC T0-T4 N3 M0	-	-	-	4
IV T0-T4 N0-N3 M1	-	-	-	1
n.d	1	1	-	-
<b>ER</b>				
Positive/Negative/NA	0/15/0	14/11/0	0/6/0	51/18/0
<b>PR</b>				
Positive/Negative/NA	0/15/0	14/11/0	0/6/0	41/28/0
<b>HER2/Neu</b>				
Positive/Negative/NA	15/0/0	2/23/0	6/0/0	11/67/1
<b>Histologic grading</b>				
G1/G2/G3/NA	0/3/12/0	2/6/17/0	0/0/0/6	0/0/0/69
<b>Molecular subtype</b>				
ERPR/HER2 positive/Triple negative/NA	0/15/0/0	14/0/11/0	0/6/0/0	51/0/18/0

For each patient group of the discovery cohort, the correlation of four protein candidates, ARHGDIB, CPNE3, ERBB2, and KPNA2 was evaluated using a Pearson correlation test. The P-values were determined using a two tailed test and a 95% confidence interval (Figure 4.2.2). The HER2 subtype was characterised by a negative correlation of ERBB2 to CPNE3 and the candidates displaying similar patterns of expression in response to CPNE3 knockdown in C3.6 and SKBR3 cells. In addition, only CPNE3 and KPNA2 were found to have a correlation pattern unique to the HER2

subtype of breast cancers. The Pearson  $r$  values for the HER2 subtype of breast cancers were  $r = -0.12$  and  $-0.29$  for CPNE3 and KPNA2 with respect to ERBB2 and were unlike those of the ERPR or TN patient groups, which either show positive correlations or no correlation.



**Figure 4.2.2** Heatmap of Pearson correlations for 4 protein candidates in the discovery cohort. A) All breast cancer patients B) ERPR patients C) HER2 patients D) TN patients

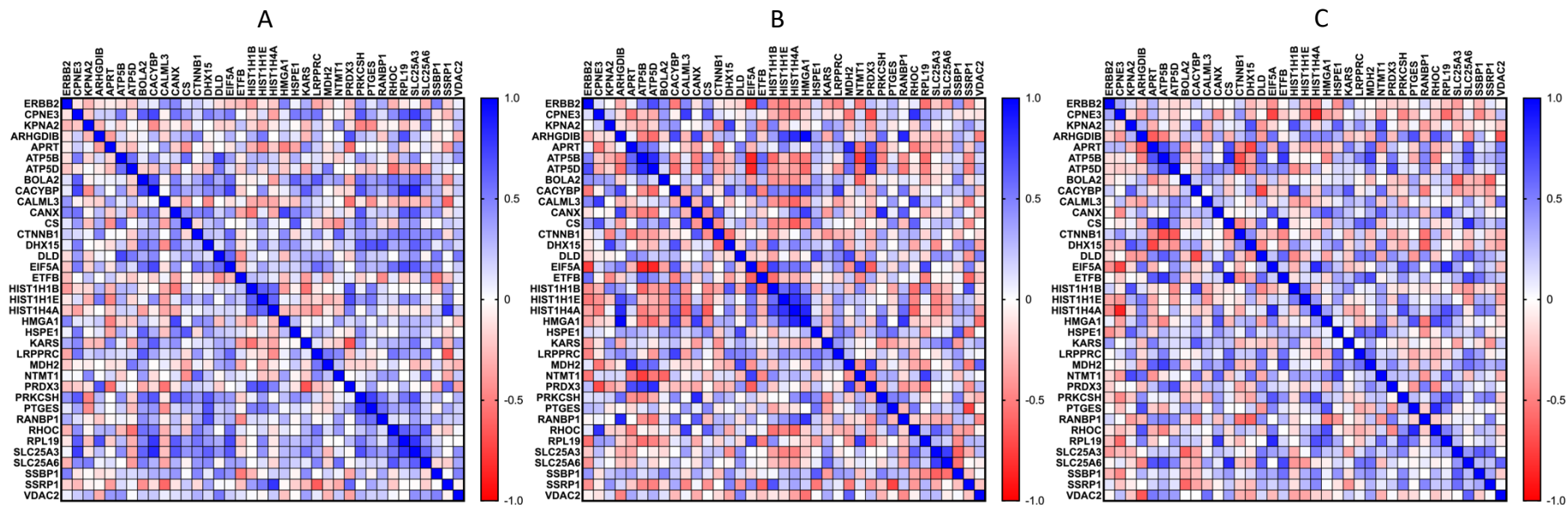
To confirm the expression of identified candidates and proteins of interest from the 52 proteins identified as up/down regulated in response to CPNE3 knockdown in HMLECs using TMT mass spectrometry, shown in Appendix Table 3.4.1, a Pearson correlation test was implemented on the discovery cohort comprised of 40 clinical breast cancer samples. P-values were determined using a two tailed test and a 95% confidence interval. Proteins with expression values in all 40 samples were compared across the three patient subtypes, ERPR, HER2 and TN (Figure 4.2.3).

Thirty-six proteins were consistently quantified in all samples. Sixteen proteins, AHSG, ALPP, HIST1H1A, HIST1H1C, HIST1H1D, HIST1H3A, MGST1, PNPLA7, PPP6R2, PYGM, RPL32, S100A10, SLC3A2, TIMELESS, UBC and ZFP28, were not quantified for all 40 patient samples. The thirty-six proteins were compared in the discovery cohort of 40 breast cancer patients and differences between the correlations of ERBB2 or CPNE3 and the other proteins were determined for ERPR, HER2 and TN patient subtypes.

The potential role of CPNE3 as a biomarker for HER2 positive breast cancer was evident in the correlation pattern demonstrated by the 36 proteins. The HER2 subtype was characterised by a high number of proteins that demonstrated a positive correlation with 26 proteins, APRT, ARHGDIB, ATP5B, BOLA2, CACYBP, CANX, CS, CTNNB1, DHX15, DLD, EIF5A, ETFB, HIST1H1E, HMGA1, HSPE1, KARS, LRPPRC, NTMT1, PRKCSH, PTGES, RANBP1, RHOC, RPL19, SLC25A3, SLC25A6 and VDAC2 (Figure 4.2.3A). However, only 10 proteins, ATP5B, BOLA2, CACYBP, CANX, EIF5A, HSPE1, LRPPRC, PRKCSH, RPL19 and SLC25A3, had statistically significant correlations ( $p$ -value  $< 0.05$ ). Each of these 10 proteins had a strong positive correlation with CPNE3 ( $r$ -value  $> 0.5$ ). The TN subtype was characterised by a median number of proteins that demonstrated a positive correlation, with 16 proteins, KPNA2, BOLA2, CALML3, CS, CTNNB1, DHX15, EIF5A, HSPE1, NTMT1, PRKCSH, PTGES, RANBP1, RHOC, RPL19, SLC25A3 and SLC25A6 (Figure 4.2.3B). Albeit, only 3 proteins, CALML3, CS and RHOC, had statistically significant correlations ( $p$ -value  $< 0.05$ ). Each of these 3 proteins had a strong positive correlation with CPNE3 ( $r$ -value  $> 0.65$ ). The ERPR subtype was characterised by a lower number of 8 proteins, APRT, CALML3, CTNNB1, KARS, LRPPRC, RANBP1, SLC25A6, and VDAC2, that demonstrated a positive correlation with CPNE3 (Figure 4.2.3C). None of these 8 proteins had statistically significant correlations ( $p$ -value  $< 0.05$ ) to CPNE3.

CPNE3 has previously been shown to have a positive correlation with ERBB2 overexpression in a panel of HMLECs and breast tumour cell lines (Worthington, 2012). It is also known to undergo a moderate increase in expression following treatment with growth factors in non ERBB2 overexpressing HB4a cells (Bertani, 2005; Gharbi et al., 2002; White et al., 2004). Thus, it's

expression may differ from the expected positive correlation in tissues exposed to varying concentrations of growth factors in each patient specific tumour microenvironment. To evaluate how this observation translates to a patient cohort, an inspection of the correlation between ERBB2 expression and the 36 proteins of interest, including CPNE3, was carried out. The HER2 subtype and the ERPR subtype were found to yield no proteins with a statistically significant positive correlation for ERBB2. The TN subtype was found to yield 1 protein, ATP5D with a statistically significant (p-value < 0.05) and a strong positive correlation (r-value > 0.65) to ERBB2. However, this observation is negligible due to the widely accepted profile of triple negative status which precludes the absence of progesterone receptor, estrogen receptor and HER2 receptor for all TN subtype breast cancers.



**Figure 4.2.3** Heatmap of subtype Pearson correlations for 36 proteins of interest in the discovery cohort. A) HER2 patients B) TN patients C) ERPR patients

Many of the HER2-positive patient tumours (> 50%) in the discovery cohort are known to be Stage IIB tumours (Table 4.2.1). This characteristic of our discovery cohort is particularly interesting since variable protein expression in malignant tumours is known to not only contribute to rapid growth and metastasis to various other tissues but can also lead to the development of drug resistance (Brabletz et al., 2001; G. P. Gupta & Massagué, 2006; Lee et al., 2017). The amplification of ERBB2/HER2 in breast cancer has long been known to correlate with disease progression, metastasis and poor therapeutic response (Ross & Fletcher, 1999; Slamon et al., 1987). Therefore, we evaluated the correlation of ERBB2 and CPNE3 to our proteins of interest for the various clinical stages of the discovery cohort (Figure 4.2.4).

It was not possible to determine the correlations for stages with less than 3 patient samples. The HER2 patient subtype was well represented in clinical stages IIA and IIB, whilst our control group comprised of ERPR and TN patient subtypes was well represented in clinical stages I, IIA and IIB (Table 4.2.1). The relevant data for each subtype is shown as a heatmap representation of the respective correlation matrix (Figure 4.2.4). HER2 Stage IIA was characterised by a strong negative correlation between ERBB2 and CPNE3 which is not consistent with our findings in the HMLECs or breast tumour cell lines. In addition, none of the 24 positive correlations with CPNE3 were statistically significant ( $p$ -value < 0.05) (Figure 4.2.4A). Unlike HER2 Stage IIA, the correlation between ERBB2 and CPNE3 was found to be positive for HER2 Stage IIB breast cancer patients (Figure 4.2.4B). HER2 Stage IIB was characterised by a high number of proteins that demonstrated a positive correlation with 28 proteins identified, APRT, ATP5B, ATP5D, BOLA2, CACYBP, CALML3, CANX, CS, CTNNB1, DHX15, DLD, EIF5A, ERBB2, HIST1H1E, HMGA1, HSPE1, KARS, LRPPRC, PRKCSH, PTGES, RANBP1, RHOC, RPL19, SLC25A3, SLC25A6, SSBP1, SSRP1 and VDAC2. However, only 7 proteins, ATP5B, CACYBP, DLD, KARS, PRKCSH, SLC25A3 and VDAC2, had statistically significant correlations ( $p$ -value < 0.05) (Figure 4.2.4B).

ERBB2 is widely known to be expressed in other breast cancer subtypes and its expression in relation to CPNE3 and disease progression of the ERPR subtype was assessed in clinical stages IIA and IIB. ERPR Stage IIA was characterised by a positive correlation between ERBB2 and CPNE3, and 12

proteins that demonstrated a positive correlation with CPNE3. However, none of the 12 positive correlations to CPNE3 were statistically significant ( $p\text{-value} < 0.05$ ) (Figure 4.2.4C). Similarly, ERPR Stage IIB was characterised by a positive correlation between ERBB2 and CPNE3, and 13 proteins, APRT, ATP5B, ATP5D, CALML3, DLD, ERBB2, KARS, MDH2, PRDX3, RANBP1, SLC25A3, SLC25A6 and VDAC2, that demonstrated a positive correlation with CPNE3. A noteworthy difference was the identification of 1 protein, PRDX3 with a statistically significant ( $p\text{-value} < 0.05$ ) and strong positive correlation ( $r\text{-value} > 0.65$ ) to CPNE3 (Figure 4.2.4D).

As expected, both stages of the TN subtype demonstrated a negative correlation between ERBB2 and CPNE3. Nevertheless, CPNE3's correlation to proteins of interest was still carried out for each of the relevant TN clinical stages. TN Stage I had 17 proteins that were found to have a positive correlation to CPNE3, but none were statistically significant correlations ( $p\text{-value} < 0.05$ ) (Figure 4.2.4F). TN Stage IIA had 20 proteins that demonstrated a positive correlation to CPNE3 and only 1 protein, HMGA1, with a statistically significant ( $p\text{-value} < 0.05$ ) and strong positive correlation ( $r\text{-value} > 0.65$ ) to CPNE3 (Figure 4.2.4E).

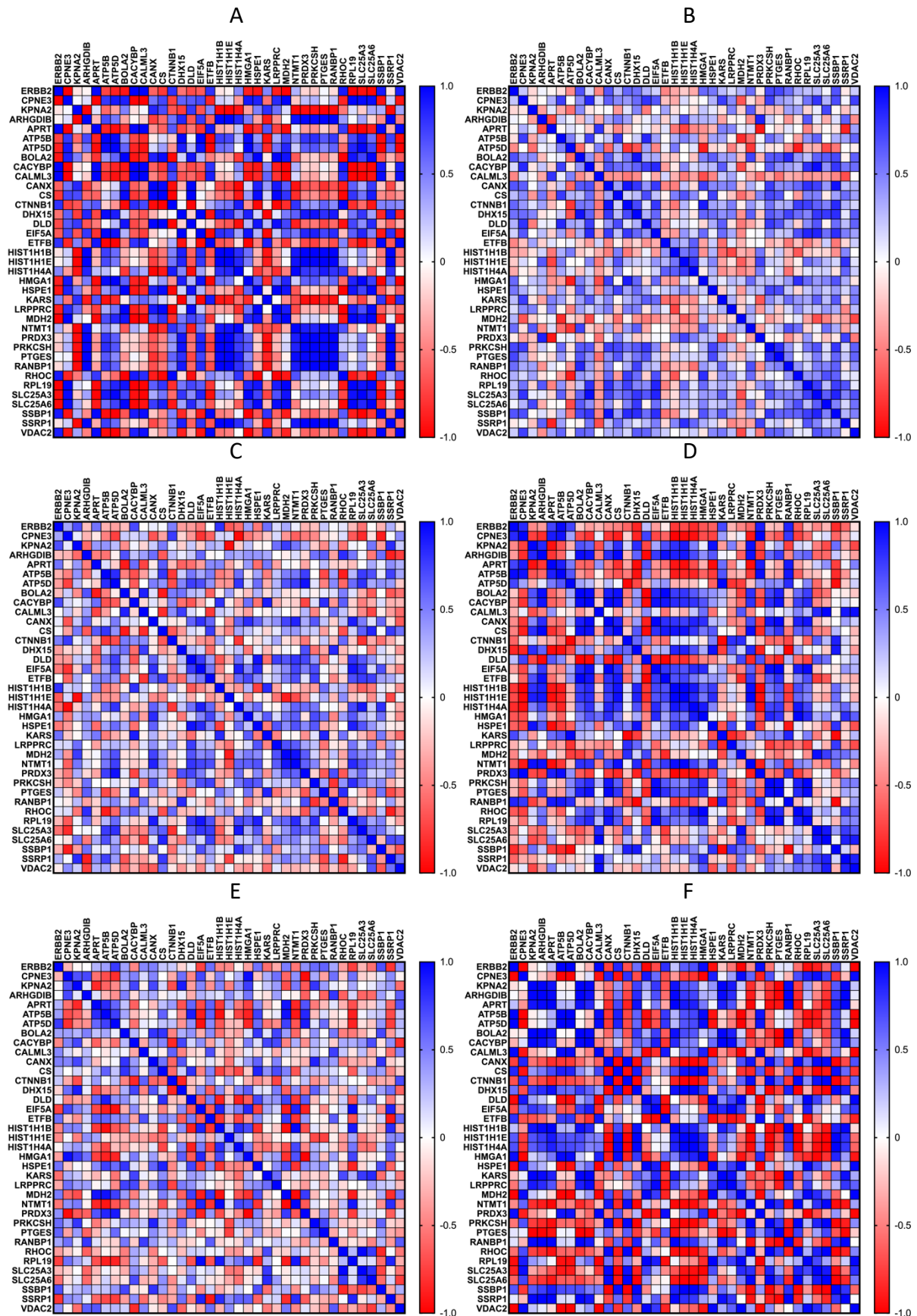
There were a total of 13 proteins, ATP5B, BOLA2, CACYBP, CANX, EIF5A, HSPE1, LRPPRC, PRKCSH, RPL19, SLC25A3, DLD, KARS and VDAC2, with a statistically significant ( $p\text{-value} < 0.05$ ) positive correlation to CPNE3 in HER2 patients vs control (ERPR and TN) and in Stage IIB of HER2 patients. The identified proteins are mostly comprised of enzymes, translational regulators and transport proteins and may be potential diagnostic or disease progression biomarkers in HER2 breast cancer (Table 4.2.2). Despite the inconsistent negative correlation between CPNE3 and ERBB2 in HER2 patients in the discovery cohort (Figure 4.2.2C and Figure 4.2.3A), the candidates are still of interest due to their statistically significant correlation ( $p\text{-value} < 0.05$ ) correlation to CPNE3 and their uniqueness to the HER2 patient subtype in the discovery cohort.

To identify potential interaction partners of CPNE3, a protein network was constructed from correlation data (Figure 4.2.5). The network was used to identify the predominant functional cluster containing either a set of highly interconnected nodes or CPNE3 respectively and to identify the



significant KEGG Pathways related to the cluster. The functional cluster (cluster 1) of the HER2 positive subtype in the discovery cohort (Figure 4.2.5), contained molecules with enzyme, kinase, translation regulator and transporter function. The enrichment of amino acid metabolism or glucose metabolism terms and pathways was observed. For example, citrate cycle (TCA cycle), glyoxylate and dicarboxylate metabolism, 2-Oxocarboxylic acid metabolism and pyruvate metabolism pathways were enriched KEGG terms related to glucose metabolism (Table 4.2.3).

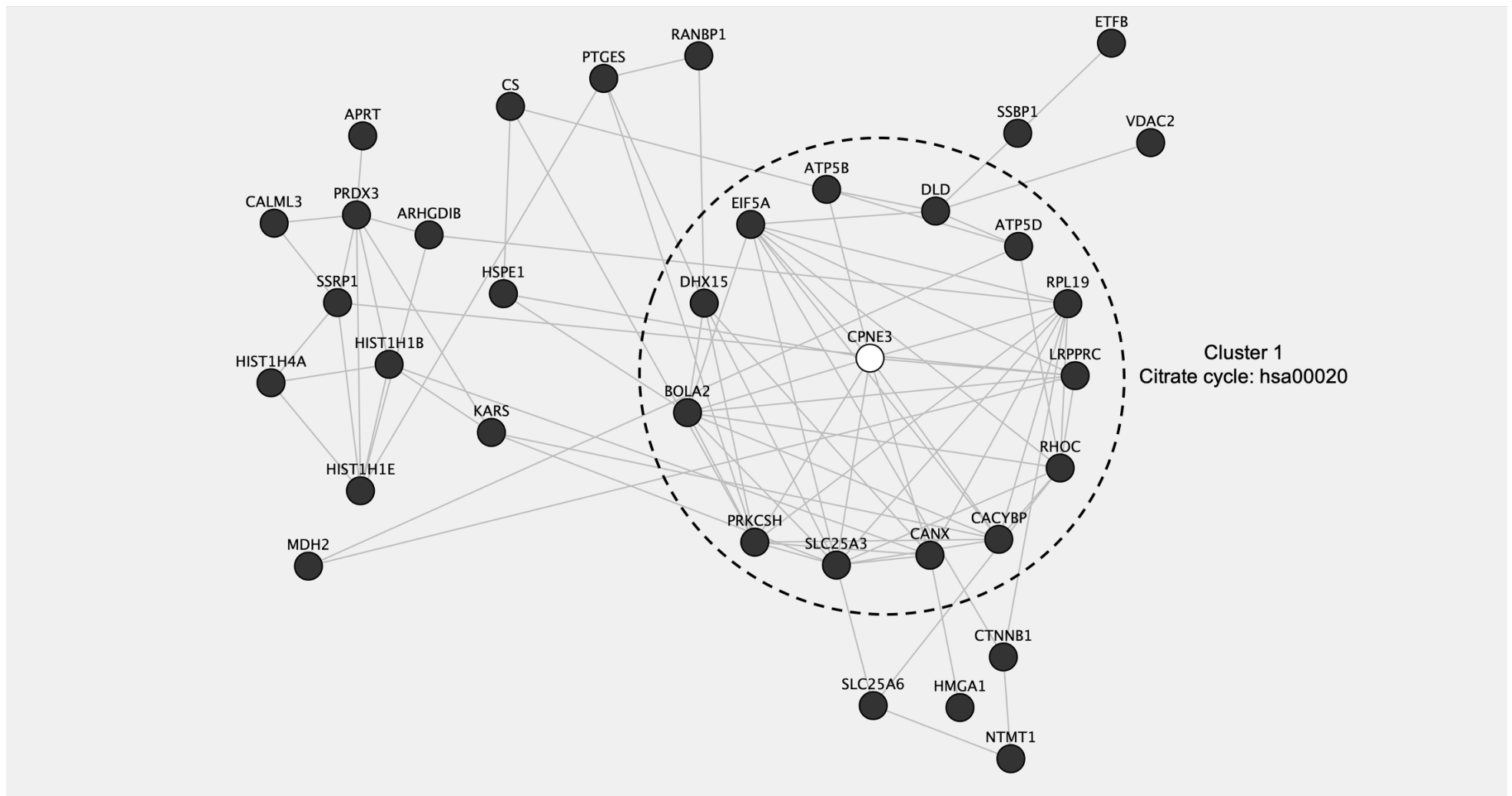
Unlike the KEGG pathways, the cluster was observed to be represented in glucose metabolism, interleukin signalling, viral infection, calreticulin cycle and RHO GTPase related Reactome terms, such as the citric acid cycle (TCA cycle) and RHO GTPases Activate Rhotekin and Rhophilins (Table 4.2.4). Six of the seven genes observed in the KEGG (Table 4.2.3) and Reactome (Table 4.2.4) pathway results, CANX, DLD, LRPPRC, PRKCSH, RHOC and RPL19 were found to have protein to protein interactions with a high confidence score ( $\geq 0.90$ ) when mapped to the STRING interactions database (Table 4.3.7). Moreover, the interactions shown in the constructed protein network from ATP5B to CS, CANX to PRKCSH, and ATP5B to ATP5D (Figure 4.2.5), were consistent with high confidence score ( $\geq 0.90$ ) STRING interactions (Table 4.3.7).



**Figure 4.2.4** Heatmap of TNM staging Pearson correlations for 36 proteins of interest in the discovery cohort. A) HER2 Stage IIA B) HER2 Stage IIB C) ERPR Stage IIA D) ERPR Stage IIB E) TN Stage IIA F) TN Stage I

**Table 4.2.2** The table lists the total number of candidate biomarkers from the discovery cohort. Highlighted proteins are not specific to HER2 positive patients.

Gene Symbol	Molecule Type	Potential Biomarker Application
ATP5B	Transporter	HER2 diagnosis/Disease progression (Stage IIB)
BOLA2	Enzyme	HER2 diagnosis
CACYBP	Translation regulator	HER2 diagnosis/Disease progression (Stage IIB)
CANX	Other	HER2 diagnosis
EIF5A	Translation regulator	HER2 diagnosis
HSPE1	Enzyme	HER2 diagnosis
LRPPRC	Other	HER2 diagnosis
PRKCSH	Enzyme	HER2 diagnosis/Disease progression (Stage IIB)
RPL19	Other	HER2 diagnosis
SLC25A3	Transporter	HER2 diagnosis/Disease progression (Stage IIB)
DLD	Enzyme	Disease progression (Stage IIB)
KARS	Enzyme	Disease progression (Stage IIB)
VDAC2	Ion channel	Disease progression (Stage IIB)



**Figure 4.2.5** Cytoscape network analysis of proteins associated with HER2 positive status in the discovery cohort. Cluster 1 represents the primary functional cluster containing CPNE3 or a set of highly interconnected nodes and KEGG pathway with the highest enrichment ratio. Edges represent strong correlations ( $r$ -value > 0.5).

**Table 4.2.3** The table lists the top 10 enriched KEGG Pathways based on an ORA analysis of cluster 1 of the network of proteins associated with HER2 positive status in the discovery cohort.

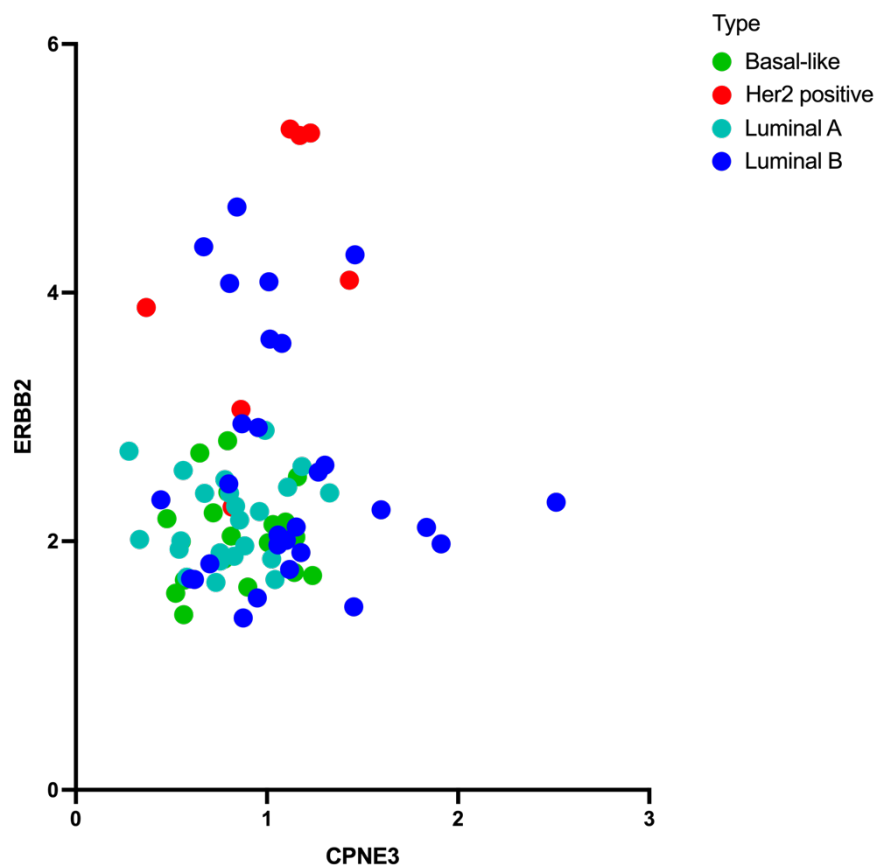
Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio
hsa04141	Protein processing in endoplasmic reticulum	CANX, PRKCSH	0.0069	15.09
hsa00020	Citrate cycle (TCA cycle)	DLD	0.0239	41.49
hsa00630	Glyoxylate and dicarboxylate metabolism	DLD	0.0239	41.49
hsa00640	Propanoate metabolism	DLD	0.0254	38.90
hsa00620	Pyruvate metabolism	DLD	0.0309	31.92
hsa00260	Glycine, serine and threonine metabolism	DLD	0.0317	31.12
hsa00280	Valine, leucine and isoleucine degradation	DLD	0.0380	25.93
hsa00010	Glycolysis / Gluconeogenesis	DLD	0.0534	18.31
hsa04918	Thyroid hormone synthesis	CANX	0.0580	16.82
hsa04612	Antigen processing and presentation	CANX	0.0603	16.17

**Table 4.2.4** The table lists the top 10 enriched Reactome Pathways based on an ORA analysis of cluster 1 of the network of proteins associated with HER2 positive status in the discovery cohort.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio
R-HSA-901042	Calnexin/calreticulin cycle	CANX, PRKCSH	2.59E-04	81.18
R-HSA-532668	N-glycan trimming in the ER and Calnexin/Calreticulin cycle	CANX, PRKCSH	4.73E-04	60.31
R-HSA-447115	Interleukin-12 family signaling	BOLA2, CANX	1.30E-03	36.39
R-HSA-168255	Influenza Life Cycle	CANX, RPL19	7.64E-03	14.76
R-HSA-5666185	RHO GTPases Activate Rhotekin and RhoGDI	RHOC	8.50E-03	117.27
R-HSA-168254	Influenza Infection	CANX, RPL19	8.82E-03	13.71
R-HSA-9020956	Interleukin-27 signaling	CANX	1.04E-02	95.95
R-HSA-1428517	The citric acid (TCA) cycle and respiratory electron transport	DLD, LRPPRC	1.11E-02	12.13
R-HSA-8984722	Interleukin-35 Signalling	CANX	1.13E-02	87.95
R-HSA-71064	Lysine catabolism	DLD	1.13E-02	87.95

### 4.3. Independent verification of candidate biomarkers

To validate candidate biomarkers associated with the clinical stages of HER2 breast cancer and confirm the clinical merit of proteins identified from the discovery cohort (Table 4.2.2), a statistical evaluation of 75 breast cancer patient samples comprised of Basal-like, HER2-enriched (HER2 positive), Luminal A and Luminal B patient subtypes was carried out. The expression of HER2/ERBB2 and CPNE3 was evaluated in the samples and visualised with a pairwise plot (Figure 4.3.1).



**Figure 4.3.1** Pair plot of ERBB2 vs CPNE3 expression for Basal-like, HER2 positive, Luminal A and B subtypes of 75 female breast cancer patients.

The majority of patients identified as HER2-enriched were found to cluster in the high ERBB2 expression ranges, as expected for HER2 positive breast cancers. However, only a fraction of the patient samples for the HER2-enriched subtype were HER2 positive and several appear to be clustered in the lower ERBB2 expression range among patients of the Basal-like, Luminal A and Luminal B subtypes. In addition, several Luminal B patients appear among HER-enriched patients in

the high ERBB2 expression range of the pairwise plot. Much like the plot of ERBB2 vs CPNE3 expression for the discovery cohort (Figure 4.2.1), the distribution of CPNE3 was narrower and non-specific compared to that of ERBB2. A noteworthy difference lies in the clinical stage distribution of the patient samples, as the 75 patient cohort has a wider distribution comprised of clinical stages I, IIA, IIB, IIIA, IIIB and IIIC (Table 4.2.1). CPNE3 expression level is known to be positively correlated with TNM clinical stage and CPNE3 has been identified as a novel metastasis-promoting gene in a quantitative proteomic analysis of NSCLC (H. Lin et al., 2013). Due to the wide clinical stage distribution of the 75-patient cohort, it is an ideal dataset to explore biomarkers correlated with TNM staging for breast cancer and to carry out independent verification of the candidate biomarkers of HER2 positive breast cancer identified using the discovery cohort (Table 4.2.2). It is herein referred to as the validation cohort (Table 4.2.1).

For each patient subtype within the validation cohort, the correlation of four protein candidates discovered using TMT mass spectrometry, ARHGDIB, CPNE3, ERBB2, and KPNA2 was evaluated using a Pearson correlation test. The p-values were determined using a two tailed test and a 95% confidence interval (Figure 4.3.2). The HER2 positive subtype of the HER2-enriched subtype was characterised by a positive correlation of ERBB2 to CPNE3 which was consistent with our hypothesis. Furthermore, KPNA2 was found to have a positive correlation, while ARHGDIB had a negative correlation with CPNE3 (Figure 4.3.2B). The expression of ARHGDIB and KPNA2 is consistent with the established expression of the candidate biomarkers in response to CPNE3 knockdown in C3.6 and SKBR3 cells (Table 3.4.2). The positive correlation between ERBB2 and CPNE3 was also observed for the Luminal A subtype. It is not uncommon for luminal subtypes to undergo subtype switching (Bastien et al., 2012; Daemen & Manning, 2018; Klebe et al., 2020) and a positive correlation between ERBB2 and CPNE3 could indicate an early shift towards HER2-enrichment. Previous studies have shown that treatment with growth factors can induce a moderate increase in CPNE3 expression in the HB4a cells, which are the parental human mammary luminal epithelial cell line for the C3.6 cells and do not overexpress ERBB2/HER2 (Bertani, 2005; Gharbi et al., 2002; White et al., 2004). Luminal subtype cancers may undergo differential expression of CPNE3 due to the presence of growth factors in the tumour microenvironment.

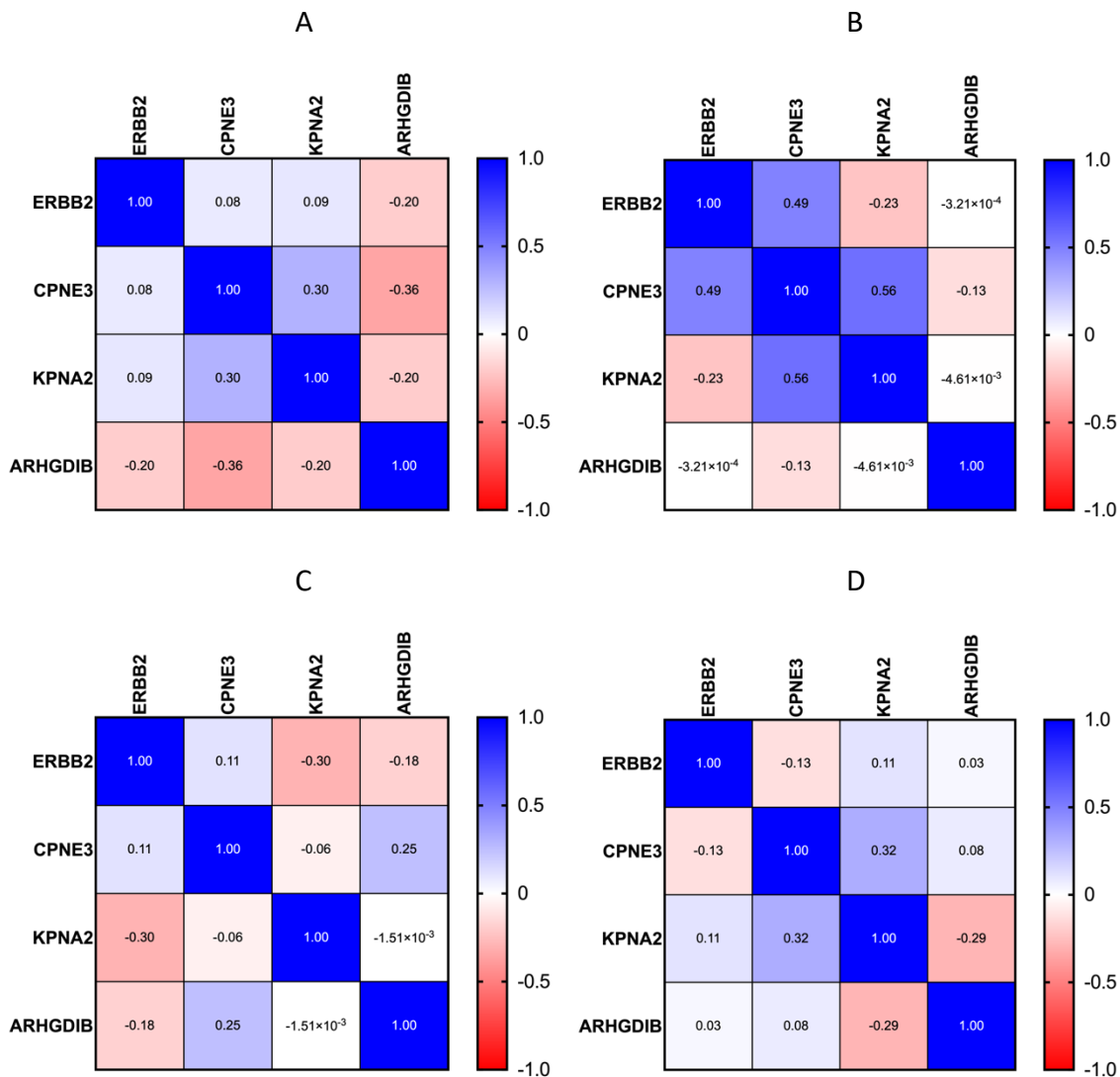
To confirm the expression of candidate biomarkers identified in the discovery cohort (Table 4.2.2) and proteins of interest from the 52 proteins identified as up/down regulated in response to CPNE3 knockdown in HMLECs (Appendix Table 3.4.1), a Pearson correlation test was implemented on the validation cohort for each of the proteins quantified in all patients. P-values were determined using a two tailed test and a 95% confidence interval. Proteins with quantified expression values in all 75 samples were compared across the four cancer subtypes, Basal-like, HER2-enriched (HER2 positive), Luminal A and Luminal B (Figure 4.3.3). Forty-five proteins were consistently quantified in all samples. Seven proteins, ALPP, and HIST1H3A, HMGA1, PNPLA7, PTGES, UBC and ZFP28, were omitted due to a high number of missing values for the samples or not detected at all during the quantification. The forty-five proteins quantified for all 75 patients were compared in the validation cohort and differences between the correlations of ERBB2 or CPNE3 with the other 45 proteins were determined for the four subtypes (Figure 4.3.3).

The HER2 positive subtype of the HER2-enriched subtype was characterised by three proteins, EIF5A, PYGM and HIST1H4A, that demonstrated a statistically significant correlation ( $p$ -value < 0.05). EIF5A and PYGM had a strong positive correlation to CPNE3 ( $r$ -value > 0.5) and HIST1H4A had a strong negative correlation to CPNE3 ( $r$ -value < -0.5) (Figure 4.3.3B). The Basal-like subtype was characterised by the absence of statistically significant correlations ( $p$ -value < 0.05) with CPNE3 (Figure 4.3.3BA). The Luminal A subtype was characterised by 2 proteins, KARS and PPP6R2, that had a statistically significant correlation ( $p$ -value < 0.05) with CPNE3 (Figure 4.3.3BC). The Luminal B subtype was found to yield 5 proteins, EIF5A, LRPPRC, PRDX3, RANBP1 and VDAC2, that demonstrated a statistically significant correlation ( $p$ -value < 0.05) with CPNE3 (Figure 4.3.3BD).

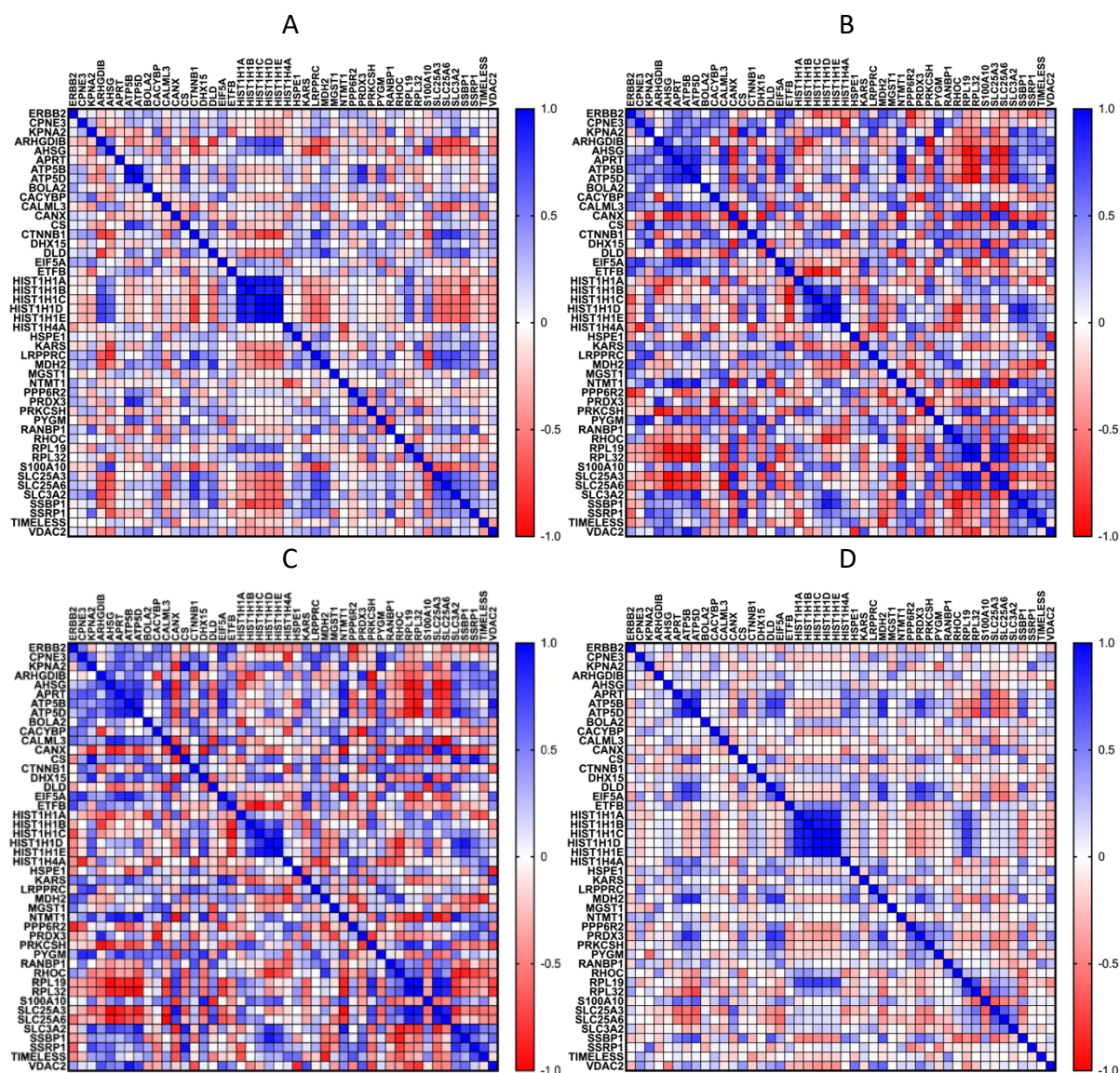
The vast majority of patient tumours evaluated for the HER2 positive subtype of the HER2-enriched subtype in the validation cohort (Figure 4.3.3B) were clinical Stage IIA (Table 4.2.1). Furthermore, the HER2 positive subtype had a moderate correlation between CPNE3 and ERBB2 ( $r$ -value ~ 0.5). As previously cited, the amplification of ERBB2/HER2 in breast cancer has long been known to correlate with disease progression, metastasis and poor therapeutic response (Ross & Fletcher, 1999; Slamon



et al., 1987) and has been shown to have a strong positive correlation with CPNE3 overexpression in a panel of HMLECs and breast tumour cell lines (Worthington, 2012).



**Figure 4.3.2** Heatmap of Pearson correlations for 4 protein candidates in the validation cohort. A) Basal-like B) HER2 positive C) Luminal A D) Luminal B



**Figure 4.3.3** Heatmap of subtype Pearson correlations for 45 proteins of interest in the validation cohort. A) Basal-like B) HER2 positive C) Luminal A D) Luminal B

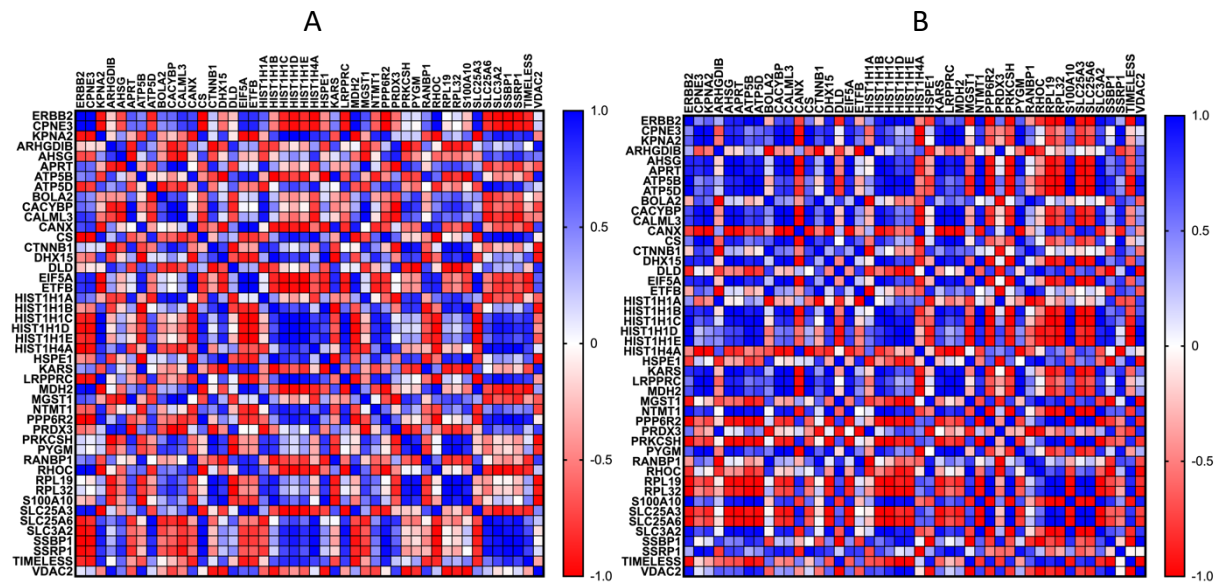
Thus, we would also expect proteins operating downstream or directly regulated by CPNE3 to have a strong correlation to both CPNE3 and ERBB2. The observed correlation pattern between ERBB2 and CPNE3 relative to proteins of interest for the HER2 positive subtype of the HER2-enriched subtype of the validation cohort suggests that the expected pattern is either absent in these patients or obscured by multistage grouping of patients. Therefore, we evaluated the correlation of ERBB2 and CPNE3 relative to our proteins of interest for the various TNM clinical stages of the HER2 positive subtype of the HER2-enriched subtype in the validation cohort (Figure 4.3.4).

The HER2 positive subtype of the HER2-enriched subtype was well represented in clinical stages IIA, IIB and IIIA (Table 4.2.1). However, Clinical stage IIB and IIIA had less than three patients for each stage and it was not possible to determine the correlations for any stage with less than 3 patient samples. Thus, the data for clinical stage IIB and IIIA was grouped and is evaluated as a unified later stage to present a juxtapose for the analysis of clinical stage IIA. The relevant data for each stage is shown as a heatmap of the respective correlation matrix (Figure 4.3.4). HER2 positive Stage IIA had 7 proteins, ERBB2, CS, HIST1H1D, MDH2, PPP6R2, RHOC and SSBP1, with statistically significant and very strong correlations with CPNE3 (r-value > 0.9) (Figure 4.3.4A). Stage IIA was particularly interesting because of a very strong positive correlation between CPNE3 and ERBB2 and very strong correlations for CS, HIST1H1D and RHOC with CPNE3 (r-value > 0.9). CS, HIST1H1D and RHOC were both linked to several of the top 10 enriched pathways found in the Panther and Reactome databases (Table 3.4.3 and 3.4.4). HER2 positive Stage IIB and IIIA had 2 proteins, CS and HIST1H4A, that demonstrated statistically significant and very strong correlations with CPNE3 (r-value > 0.9) (Figure 4.3.4B). Interestingly, CS was not unique to Stage IIB/IIIA, where it was found to have a strong positive correlation to CPNE3 (r-value > 0.9), but is also related to HER2 positive Stage IIA, where it was found to have a strong negative correlation to CPNE3 (r-value > 0.9). Thus, it would be an interesting consideration for a marker of HER2 positive status in breast cancer.

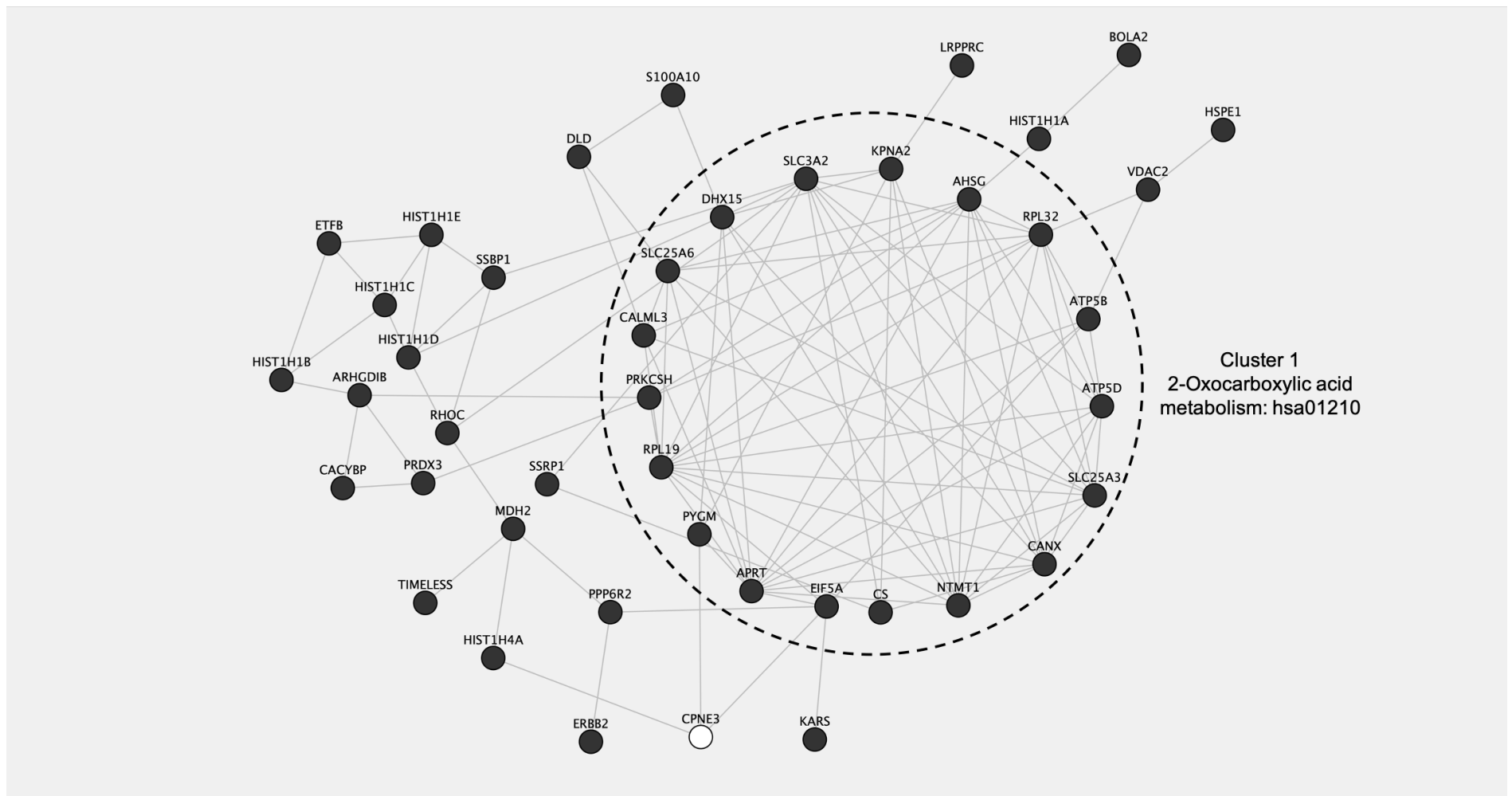
The Pearson correlation scores for the previously identified proteins which are shown to be differentially expressed or up/down regulated in response to CPNE3 knockdown in HMLECs (Appendix Table 3.4.1), were used to construct protein networks for the HER2 positive subtype (Figure 4.3.5), HER2 positive Stage IIA (Figure 4.3.6) and HER2 positive Stage IIB/IIIA (Figure 4.3.7) of the validation cohort. The protein networks were evaluated to identify the primary functional cluster containing either a set of highly interconnected nodes or CPNE3 respectively, and to identify the significant KEGG Pathways related to the cluster. The enrichment of glucose metabolism related terms and pathways was observed to be associated with each set of highly interconnected nodes or clusters containing CPNE3 for each analysis done on the validation cohort (Figure 4.3.5 to Figure 4.3.7).

The functional cluster (cluster 1) of the HER2 positive subtype of the validation cohort (Figure 4.3.5), predominantly contained molecules with enzyme or transporter function. The enrichment of glucose metabolism related terms and pathways was observed. For example, 2-Oxocarboxylic acid metabolism, glucagon signalling and insulin signalling pathways were enriched KEGG terms (Table 4.3.1). However, the cluster was observed to be more represented by glycoprotein binding and viral infection terms, such as the calnexin/calreticulin cycle and host interactions with Influenza factors pathway in a search of the Reactome database (Table 4.3.2). As shown in Table 4.3.7, most of the enriched genes observed in the KEGG (Table 4.3.1) and Reactome (Table 4.3.2) pathway databases are consistent with STRING interactions (Table 4.3.7). For instance, CANX, CS, KPNA2, RPL19, RPL32 and SLC25A6 were found to have protein to protein interactions with a high confidence score ( $\geq 0.90$ ) when mapped to the STRING interactions database.

The interaction shown in the constructed protein network from RPL19 to RPL32 (Figure 4.3.5), was found be consistent with the STRING interactions (Table 4.3.7). RPL19 and RPL32 have been reported to be associated with certain types of cancer and have been shown to have an oncogenic role in breast cancer. Previous studies have revealed that overexpression of RPL19 could sensitise breast cancer cells to endoplasmic reticulum stress-induced cell death by activating the unfolded protein response (M. Hong et al., 2014). A more recent study has revealed that RPL32 may decrease breast cancer cell migration and invasion by downregulating the expression of matrix metalloproteinase-2 (MMP-2) and matrix metalloproteinase-9 (MMP-9) (L. Xu et al., 2020). Considering these facts, the enrichment of KEGG terms and pathways related to RPL19 and RPL32 (Table 4.3.1), indicates that the knockdown of CPNE3 which reduced expression of both RPL19 and RPL32 in C3.6



**Figure 4.3.4** Heatmap of TNM staging Pearson correlations for 45 proteins of interest in the validation cohort.  
A) HER2 positive Stage IIA B) HER2 positive Stage IIB/IIIA



**Figure 4.3.5** Cytoscape network analysis of proteins associated with HER2 positive status in the validation cohort. Cluster 1 represents the primary functional cluster containing CPNE3 or a set of highly interconnected nodes and KEGG pathway with the highest enrichment ratio. Edges represent strong correlations ( $r$ -value > 0.5).

**Table 4.3.1** The table lists the top 10 enriched KEGG Pathways based on an ORA analysis of cluster 1 of the network of proteins associated with HER2 positive status in the validation cohort.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio
hsa04922	Glucagon signalling pathway	CALML3, PYGM	0.0114	12.09
hsa04910	Insulin signalling pathway	CALML3, PYGM	0.0195	9.09
hsa03010	Ribosome	RPL19, RPL32	0.0240	8.14
hsa04218	Cellular senescence	CALML3, SLC25A6	0.0261	7.78
hsa04217	Necroptosis	PYGM, SLC25A6	0.0268	7.68
hsa04022	cGMP-PKG signalling pathway	CALML3, SLC25A6	0.0271	7.64
hsa04141	Protein processing in endoplasmic reticulum	CANX, PRKCSH	0.0277	7.54
hsa01210	2-Oxocarboxylic acid metabolism	CS	0.0286	34.58
hsa05164	Influenza A	KPNA2, SLC25A6	0.0296	7.28
hsa04020	Calcium signalling pathway	CALML3, SLC25A6	0.0335	6.80

**Table 4.3.2** The table lists the top 10 enriched Reactome Pathways based on an ORA analysis of cluster 1 of the network of proteins associated with HER2 positive status in the validation cohort.

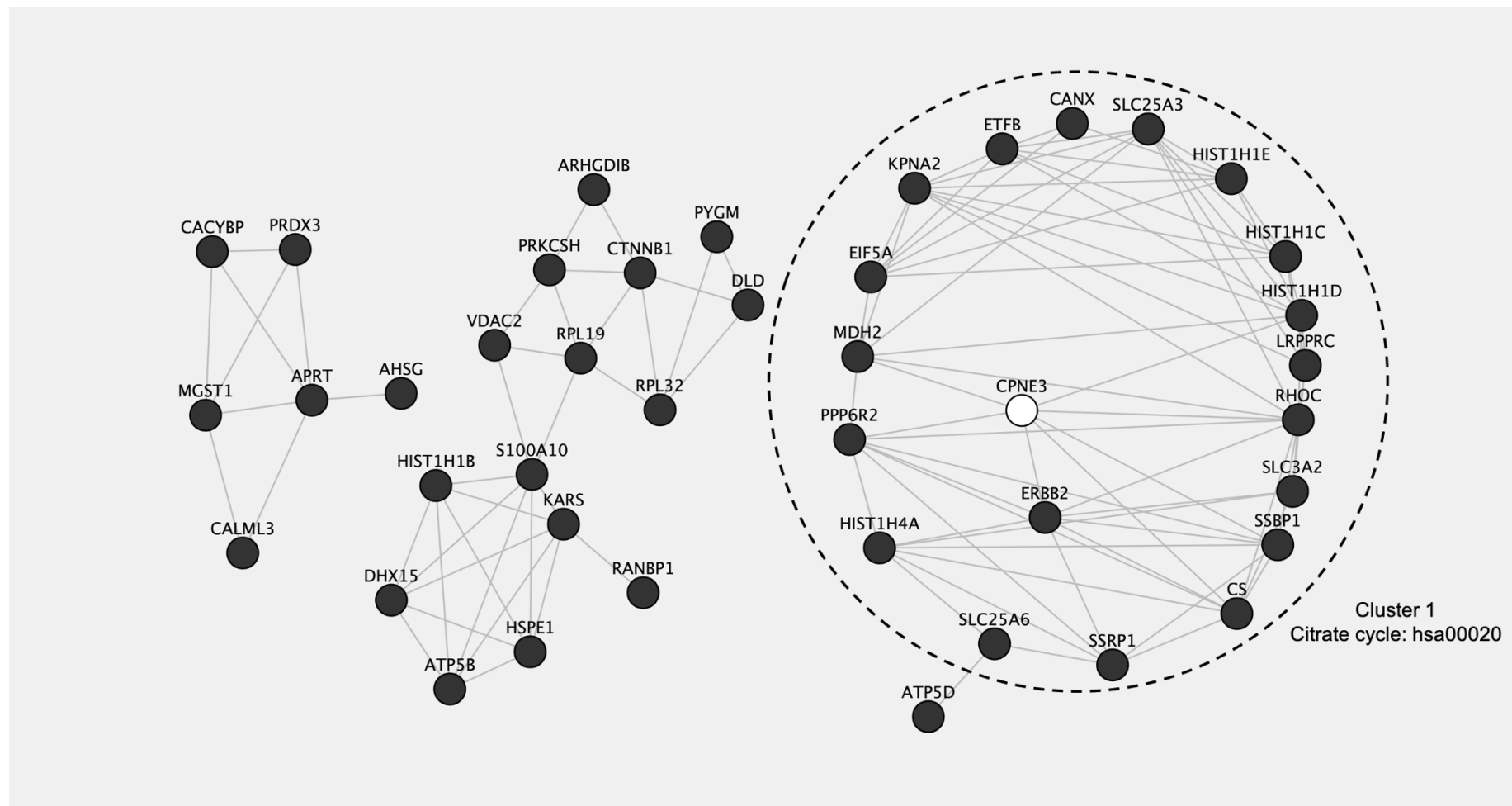
Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio
R-HSA-168254	Influenza Infection	CANX, KPNA2, RPL19, RPL32, SLC25A6	4.52E-07	28.56
R-HSA-5663205	Infectious disease	CANX, KPNA2, RPL19, RPL32, SLC25A6	3.88E-05	11.51
R-HSA-901042	Calnexin/calreticulin cycle	CANX, PRKCSH	3.79E-04	67.65
R-HSA-168255	Influenza Life Cycle	CANX, RPL19, RPL32	4.90E-04	18.45
R-HSA-532668	N-glycan trimming in the ER and Calnexin/Calreticulin cycle	CANX, PRKCSH	6.91E-04	50.26
R-HSA-168253	Host Interactions with Influenza Factors	KPNA2, SLC25A6	9.48E-04	42.90
R-HSA-1268020	Mitochondrial protein import	CS, SLC25A6	2.30E-03	27.48
R-HSA-1643685	Disease	CANX, KPNA2, RPL19, RPL32, SLC25A6	4.28E-03	4.17
R-HSA-156902	Peptide chain elongation	RPL19, RPL32	4.39E-03	19.76
R-HSA-192823	Viral mRNA Translation	RPL19, RPL32	4.39E-03	19.76

cells (Appendix Table 3.4.1), could play a role in mitigating endoplasmic reticulum stress-induced cell death and increase breast cancer migration by upregulating MMP expression.

The functional cluster (cluster 1) of the HER2 positive Stage IIA in the validation cohort (Figure 4.3.6), contained molecules with enzyme, kinase, transcription regulator, translation regulator and transporter function. The enrichment of glucose metabolism, DNA replication or viral infection related terms and pathways was observed. For example, citrate cycle (TCA cycle), glyoxylate and dicarboxylate metabolism, 2-Oxocarboxylic acid metabolism and pyruvate metabolism pathways were enriched KEGG terms related to glucose metabolism (Table 4.3.3). Likewise, the cluster was observed to be represented in glucose metabolism, DNA replication or viral infection related Reactome terms, such as apoptosis induced DNA fragmentation and citric acid cycle (TCA cycle) (Table 4.3.4). As shown in Table 4.3.7, most of the enriched genes observed in the KEGG (Table 4.3.3) and Reactome (Table 4.3.4) pathway databases are consistent with STRING interactions (Table 4.3.7). Moreover, CANX, CS, KPNA2, LRPPRC, MDH2, HIST1H1E, HIST1H4A and SLC25A6 were found to have protein to protein interactions with a high confidence score ( $\geq 0.90$ ) when mapped to the STRING interactions database.

The interactions shown in the constructed protein network from HIST1H4A to SSRP1, ERBB2 to RHOC and ERBB2 to CPNE3 (Figure 4.3.6), were found to be consistent with mapped STRING interactions (Table 4.3.7). The interaction between HIST1H4A and SSRP1 has been reported by several Affinity Capture-MS studies (Huttlin et al., 2017, 2021; Nakamura et al., 2019). Moreover, CPNE3 exhibits kinase activity, phosphorylates H1 histones and basic phospholipid proteins, activates downstream signalling pathways, and subsequently promotes tumour proliferation and metastasis (Mo et al., 2013; Thomas et al., 2008). This makes HIST1H4A an interesting candidate biomarker for ERBB2/HER2 overexpressing breast cancers and a plausible downstream interaction partner of CPNE3. Analysis of proteins related to CPNE3 knockdown in HER2 positive Stage IIA of the validation cohort reveals that HIST1H4A is among the genes associated with cellular senescence and DNA damage/telomere stress induced senescence terms and Reactome pathways.





**Figure 4.3.6** Cytoscape network analysis of proteins associated with HER2 positive Stage IIA in the validation cohort. Cluster 1 represents the primary functional cluster containing CPNE3 or a set of highly interconnected nodes and KEGG pathway with the highest enrichment ratio. Edges represent strong correlations ( $r$ -value  $> 0.5$ ).

**Table 4.3.3** The table lists the top 10 enriched KEGG Pathways based on an ORA analysis of cluster 1 of the network of proteins associated with HER2 positive Stage IIA in the validation cohort.

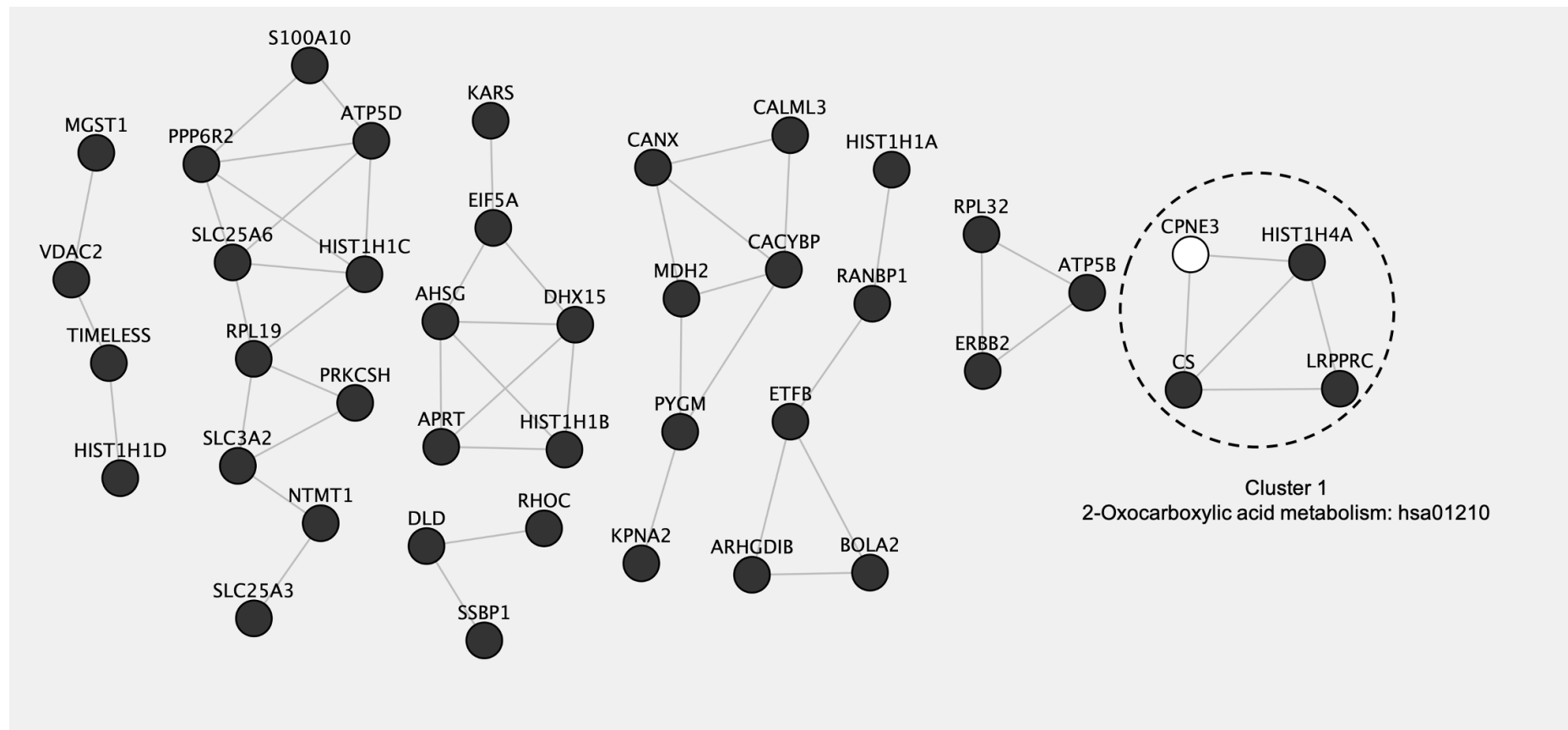
Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio
hsa00020	Citrate cycle (TCA cycle)	CS, MDH2	0.0004	62.24
hsa00630	Glyoxylate and dicarboxylate metabolism	CS, MDH2	0.0004	62.24
hsa01200	Carbon metabolism	CS, MDH2	0.0063	16.10
hsa05164	Influenza A	KPNA2, SLC25A6	0.0133	10.92
hsa01210	2-Oxocarboxylic acid metabolism	CS	0.0191	51.87
hsa03430	Mismatch repair	SSBP1	0.0244	40.59
hsa05166	Human T-cell leukemia virus 1 infection	CANX, SLC25A6	0.0284	7.32
hsa03030	DNA replication	SSBP1	0.0379	25.93
hsa00620	Pyruvate metabolism	MDH2	0.0410	23.94
hsa04216	Ferroptosis	SLC3A2	0.0421	23.34

**Table 4.3.4** The table lists the top 10 enriched Reactome Pathways based on an ORA analysis of cluster 1 of the network of proteins associated with HER2 positive Stage IIA in the validation cohort.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio
R-HSA-140342	Apoptosis induced DNA fragmentation	HIST1H1C, HIST1H1D, HIST1H1E	9.83E-07	143.27
R-HSA-211227	Activation of DNA fragmentation factor	HIST1H1C, HIST1H1D, HIST1H1E	9.83E-07	143.27
R-HSA-2559584	Formation of Senescence-Associated Heterochromatin Foci (SAHF)	HIST1H1C, HIST1H1D, HIST1H1E	1.92E-06	116.40
R-HSA-2559586	DNA Damage/Telomere Stress Induced Senescence	HIST1H1C, HIST1H1D, HIST1H1E, HIST1H4A	6.76E-06	31.04
R-HSA-75153	Apoptotic execution phase	HIST1H1C, HIST1H1D, HIST1H1E	7.31E-05	35.82
R-HSA-1428517	The citric acid (TCA) cycle and respiratory electron transport	CS, ETFB, LRPPRC, MDH2	1.44E-04	14.27
R-HSA-2559583	Cellular Senescence	HIST1H1C, HIST1H1D, HIST1H1E, HIST1H4A	2.23E-04	12.73
R-HSA-71403	Citric acid cycle (TCA cycle)	CS, MDH2	5.54E-04	56.44
R-HSA-168254	Influenza Infection	CANX, KPNA2, SLC25A6	1.78E-03	12.09
R-HSA-168253	Host Interactions with Influenza Factors	KPNA2, SLC25A6	1.93E-03	30.28

Considering these facts, CPNE3 may play a role in downstream phosphorylation of HIST1H4A, potentially suppressing cell senescence and DNA damage/telomere stress induced senescence pathways. Both ERBB2 and CPNE3 exhibit a consistent trend of negative correlations with HIST1H4A in most of the HER2 positive related correlation heatmaps for the discovery (Table 4.2.4 and 4.2.4) and the validation cohorts (Table 4.3.3 and 4.3.4). This is consistent with our findings from the TMT mass spectrometry analysis of global protein expression in response to CPNE3 knockdown in ERBB2 overexpressing C3.6 HMLECs (Appendix Table 3.4.1). Therefore, indicating that the abrogated expression of CPNE3 may contribute to cell senescence related pathways through one of the HIST1H4A linked pathways identified in Reactome (Table 4.3.4).

The functional cluster or CPNE3 containing cluster (cluster 1) of the HER2 positive Stage IIB/IIIA in the validation cohort (Figure 4.3.7), contained four molecules. The enrichment of glucose metabolism, DNA replication or viral infection related terms and pathways was observed. For example, 2-Oxocarboxylic acid metabolism, citrate cycle (TCA cycle) and glyoxylate and dicarboxylate metabolism were enriched KEGG terms related to glucose metabolism (Table 4.3.5). Likewise, the cluster was observed to be represented in glucose metabolism, DNA transcription or DNA modification related Reactome terms, such as RNA polymerase I promoter opening and the citric acid (TCA) cycle and respiratory electron transport (Table 4.3.6). All four genes observed in the KEGG (Table 4.3.5) and Reactome (Table 4.3.6) pathway results, CPNE3, CS, LRPPRC and HIST1H4A were found to have protein to protein interactions with a high confidence score ( $\geq 0.90$ ) when mapped to the STRING interactions database (Table 4.3.7). However, the interactions shown in the constructed protein network from CPNE3 to CS, CPNE3 to HIST1H4A, CS to HIST1H4A, LRPPRC to HIST1H4A and LRPPRC to CS (Figure 4.3.7), were not found among the mapped STRING interactions with a confidence score cut-off  $\geq 0.90$  (Table 4.3.7).



**Figure 4.3.7** Cytoscape network analysis of proteins associated with HER2 positive Stage IIB/IIIA in the validation cohort. Cluster 1 represents the primary functional cluster containing CPNE3 or a set of highly interconnected nodes and KEGG pathway with the highest enrichment ratio. Edges represent strong correlations ( $r$ -value  $> 0.5$ ).

**Table 4.3.5** The table lists the top 10 enriched KEGG Pathways based on an ORA analysis of cluster 1 of the network of proteins associated with HER2 positive Stage IIB/IIIA in the validation cohort.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio
hsa01210	2-Oxocarboxylic acid metabolism	CS	0.0048	207.47
hsa00020	Citrate cycle (TCA cycle)	CS	0.0080	124.48
hsa00630	Glyoxylate and dicarboxylate metabolism	CS	0.0080	124.48
hsa01230	Biosynthesis of amino acids	CS	0.0200	49.79
hsa01200	Carbon metabolism	CS	0.0308	32.19
hsa05322	Systemic lupus erythematosus	HIST1H4A	0.0353	28.08
hsa05034	Alcoholism	HIST1H4A	0.0476	20.75
hsa05203	Viral carcinogenesis	HIST1H4A	0.0531	18.58
hsa01100	Metabolic pathways	CS	0.3189	2.86

**Table 4.3.6** The table lists the top 10 enriched Reactome Pathways based on an ORA analysis of cluster 1 of the network of proteins associated with HER2 positive Stage IIB/IIIA in the validation cohort.

Gene Set	Pathway Description	Gene	P Value	Enrichment Ratio
R-HSA-1428517	The citric acid (TCA) cycle and respiratory electron transport	CS, LRPPRC	1.59E-03	30.33
R-HSA-71403	Citric acid cycle (TCA cycle)	CS	8.31E-03	119.93
R-HSA-3214842	HDMs demethylate histones	HIST1H4A	1.92E-02	51.74
R-HSA-171306	Packaging Of Telomere Ends	HIST1H4A	1.96E-02	50.74
R-HSA-71406	Pyruvate metabolism and Citric Acid (TCA) cycle	CS	2.07E-02	47.97
R-HSA-73728	RNA Polymerase I Promoter Opening	HIST1H4A	2.37E-02	41.88
R-HSA-1268020	Mitochondrial protein import	CS	2.40E-02	41.23
R-HSA-5334118	DNA methylation	HIST1H4A	2.44E-02	40.59
R-HSA-5625886	Activated PKN1 stimulates transcription of AR (androgen receptor) regulated genes KLK2 and KLK3	HIST1H4A	2.52E-02	39.38
R-HSA-427359	SIRT1 negatively regulates rRNA expression	HIST1H4A	2.55E-02	38.80

**Table 4.3.7** The table lists the STRING interactions with a confidence score cut-off  $\geq 0.90$  (high confidence) for significantly up/down - regulated proteins.

Name	Proximity on chromosome	Gene fusion	Phylogenetic cooccurrence	Homology	Coexpression	Experimentally determined interaction	Database annotated	Automated textmining	Combined score
ARHGDIB (interacts with) RHOC	0	0	0	0	0.082	0.645	0.6	0.564	0.935
ATP5B (interacts with) ATP5D	0.119	0	0	0	0.907	0.992	0.9	0.909	0.999
ATP5B (interacts with) SLC25A3	0	0	0	0	0.942	0	0	0.645	0.978
ATP5B (interacts with) MDH2	0.054	0	0	0	0.829	0.462	0	0.784	0.978
ATP5B (interacts with) CS	0	0	0	0	0.757	0.319	0.9	0.597	0.992
ATP5D (interacts with) MDH2	0.111	0	0	0	0.826	0.419	0	0.328	0.931
CACYBP (interacts with) CTNNB1	0	0	0	0	0.062	0.863	0.72	0.413	0.976
CANX (interacts with) PRKCSH	0	0	0	0	0.094	0.752	0.9	0.505	0.987
CPNE3 (interacts with) ERBB2	0	0	0	0	0.062	0.27	0	0.917	0.938
CS (interacts with) MDH2	0.075	0	0	0	0.596	0.843	0.9	0.871	0.999
CTNNB1 (interacts with) ERBB2	0	0	0	0	0	0.76	0.9	0.989	0.999
DLD (interacts with) MDH2	0.068	0	0	0	0.929	0.102	0	0.699	0.979
ERBB2 (interacts with) LRPPRC	0	0	0	0	0	0.129	0.9	0.044	0.909
ERBB2 (interacts with) RHOC	0	0	0	0	0.063	0.115	0.9	0.508	0.953
HIST1H1E (interacts with) HIST1H4A	0	0.003	0	0	0.215	0.814	0.36	0.497	0.946
HIST1H1E (interacts with) HIST1H3A	0	0.002	0	0	0.253	0.858	0.36	0.415	0.955
HIST1H3A (interacts with) SSRP1	0	0	0	0	0.063	0.947	0	0.238	0.959
HIST1H3A (interacts with) HIST1H4A	0	0.008	0	0	0.382	0.842	0.9	0.626	0.995
HIST1H4A (interacts with) SSRP1	0	0	0	0	0.062	0.868	0	0.34	0.911
KPNA2 (interacts with) RANBP1	0	0	0	0	0.345	0.253	0.9	0.617	0.978
RPL19 (interacts with) RPL32	0.295	0	0	0	0.978	0.997	0.9	0.742	0.999
SLC25A3 (interacts with) VDAC2	0	0	0	0	0.876	0.056	0	0.553	0.943
SLC25A6 (interacts with) VDAC2	0	0	0	0	0.145	0.116	0.8	0.577	0.927

#### 4.4. ERBB2-specific phosphosignalling in breast cancer patients

To define the possible biological mechanisms involved with CPNE3-mediated signal transduction in HER2 positive breast cancer, phosphoproteomic profiling data from the validation cohort was used to identify putative sites of phosphorylation for candidate biomarkers. Phosphopeptide enrichment was determined for protein interactions shown in the constructed protein networks (Figure 4.2.5 and Figure 4.3.5 to 4.3.7) and confirmed to have protein to protein interactions with a high confidence score ( $\geq 0.90$ ) when mapped to the STRING interactions. It is known that overexpression of ERBB2 leads to heterodimerization and activation of downstream signalling proteins (Graus-Porta et al., 1997; Tzahar et al., 1996). Thus, a cross-comparison of enriched phosphopeptides from HER2 positive patient samples and HER2-enriched patient samples was used to determine the changes in downstream phosphorylation events due to ERBB2/HER2 receptor signalling (Table 4.4.1).

Altered phosphopeptides with a  $\geq 1.5$ -fold change between HER2 positive (HER2+) and HER2-enriched patients were categorised depending on ERBB2 overexpression (Table 4.4.1). Most peptides (77%) were singly phosphorylated and a total of 159 changes for 73 unique phosphosites in 124 sequences (from 11 proteins) were identified. As expected for ERBB2 overexpressing cells, the ERBB2 protein ratio was greater than 2-fold and there were a greater number of phosphosites, or modified residues identified for ERBB2 compared to the other proteins in the evaluation. Interestingly, there was no significant difference in the CPNE3 protein ratio between HER2 positive (HER2+) and HER2-enriched patients. This observation is unsurprising as CPNE3 is known to correlate with ERBB2 expression in breast cancer. The positive correlation between CPNE3 and ERBB2 might be true of ERBB2 mRNA enriched cells such as HER2-enriched patient cells that do not overexpress the ERBB2/HER2 receptor. Moreover, the similarity in CPNE3 expression or near 1-fold protein ratio between HER2+ and HER2-enriched patients allows us to evaluate phosphopeptide alterations for previously identified proteins related to CPNE3 expression with only one upstream dependent variable, ERBB2/HER2 expression.

**Table 4.4.1** ERBB2 related phosphorylation in the validation cohort. 1pST, 2pST and 3pST refer to singly, doubly and triply phosphorylated peptide ratios for the indicated sites. Values in boldface display  $\geq 1.5$ -fold change in abundance as determined from ITRAQ ratios. Dark grey is up-regulated, and light grey is down-regulated.

			Ratio HER2+ vs. HER2-enriched		
Protein	Site	Ave Protein Ratio HER2+ vs. HER2-enriched	1pST	2pST	3pST
ATP5B	S160	1.08			
ATP5B	S465	1.08	<b>8.23</b>		
ATP5B	S415	1.08			
ATP5B	S106	1.08			
ATP5B	S433	1.08			
ATP5B	T213	1.08			
CANX	S583	<b>1.57</b>	<b>2.63</b>		
CANX	S564	<b>1.57</b>	0.95	1.24	
CANX	S554	<b>1.57</b>	0.73	1.24	
CANX	S562	<b>1.57</b>			
CPNE3	S159	1.14			
CPNE3	S243	1.14	1.15		
CPNE3	S242	1.14	<b>2.52</b>		
CPNE3	S240	1.14			
CPNE3	Y203	1.14			
CPNE3	S200	1.14			
CPNE3	S197	1.14	<b>2.01</b>		
CPNE3	S398	1.14			
CPNE3	T395	1.14			
CPNE3	Y392	1.14	0.84		
CPNE3	S260	1.14			
CPNE3	Y261	1.14			
CPNE3	S90	1.14			
CPNE3	S14	1.14	1.10		
CS	Y194	<b>1.70</b>			
CS	Y199	<b>1.70</b>			
ERBB2	S963	<b>2.18</b>			
ERBB2	T1166	<b>2.18</b>	<b>3.27</b>		
ERBB2	S819	<b>2.18</b>			
ERBB2	S1002	<b>2.18</b>			
ERBB2	S998	<b>2.18</b>	<b>3.07</b>		
ERBB2	S1107	<b>2.18</b>	<b>5.05</b>		
ERBB2	T1103	<b>2.18</b>	<b>6.26</b>		
ERBB2	S1100	<b>2.18</b>	<b>1.85</b>		
ERBB2	Y877	<b>2.18</b>			
ERBB2	S703	<b>2.18</b>			



ERBB2	T701	2.18	2.22	
ERBB2	S1083	2.18	1.46	5.73
ERBB2	S1078	2.18	2.91	5.73
ERBB2	S1073	2.18	13.42	1.54
ERBB2	S1066	2.18	6.42	5.30
ERBB2	S1174	2.18	3.81	
ERBB2	S1134	2.18	2.39	1.73
ERBB2	T1132	2.18		1.73
ERBB2	S1054	2.18	1.72	3.42
ERBB2	T1172	2.18		
HIST1H4A	Y52	1.71		
HIST1H4A	S47	1.71	3.87	
HIST1H4A	Y89	1.71		
PRKCSH	S126	1.42	0.81	
PRKCSH	S445	1.42		
PRKCSH	S442	1.42	0.17	
PRKCSH	S168	1.42	2.45	
RHOC	S152	0.91		
RPL19	S13	0.96		
RPL19	S12	0.96		
RPL19	S189	0.96	0.49	
RPL32	S25	1.37		
RPL32	S94	1.37		
SSRP1	S349	1.05		
SSRP1	S672	1.05		1.26
SSRP1	S673	1.05		1.72
SSRP1	S671	1.05	2.30	2.08
SSRP1	S668	1.05	12.17	4.96
SSRP1	S667	1.05		3.91
SSRP1	S444	1.05	1.16	1.36
SSRP1	Y441	1.05		
SSRP1	Y438	1.05		
SSRP1	S437	1.05		1.36
SSRP1	S120	1.05		
SSRP1	S531	1.05		
SSRP1	S657	1.05	0.98	
SSRP1	S627	1.05		

SSRP1 was found to have a comparatively high number of differential phosphorylation events compared to the other potential downstream interaction partners of CPNE3 (Table 4.4.1). Four amino acid residues exhibit statistically significant altered phosphopeptides with a  $\geq 1.5$ -fold change. The phosphopeptides with modified residues S668 and S671 were both singly and doubly phosphorylated, while S667 and S673 were only doubly phosphorylated. S668 is a putative site of *in vivo* kinases and has been shown to undergo phosphorylation by casein kinase II (CK2) which inhibits the DNA-binding activity of SSRP1 (Y. Li et al., 2005). Furthermore, SSRP1 is known to be a component of the histone chaperone FACT complex and plays a significant role in its involvement in chromatin-related processes of mammalian cells (Gurova et al., 2013; Prendergast et al., 2020). This is consistent with the high confidence score STRING interaction between HIST1H4A and SSRP1 (Table 4.3.7) and previously reported affinity capture studies (Huttlin et al., 2017, 2021; Nakamura et al., 2019). The high S688 phosphosite enrichment ratio of 12.17 indicates that the 2-fold ERBB2/HER2 protein ratio for HER2+ vs. HER2-enriched patients is associated with an increase in the putative phosphorylation of S688 by the serine/threonine protein kinase CK2. HIST14HA was found to be phosphorylated at amino acid residue S47 and this corresponds to a site an amino acid residue associated with upstream activity of the regulatory protein TBK1 and *in vivo* kinase PAK2.

CANX and PRKCSH, both previously shown to interact (Table 4.3.7) and to be enriched in pathways linked to Calnexin/calreticulin cycle, Protein processing in endoplasmic reticulum and Antigen processing and presentation (Table 4.2.3 and Table 4.2.4), had a higher expression in HER2+ vs. HER2-enriched patients based on the Average Protein Ratio value (Table 4.4.1). CANX (MHC class I Antigen-Binding Protein P88) was found to be singly phosphorylated at amino acid residue S583 and PRKCSH was found to be singly phosphorylated at amino acid residues S442 and S168. Albeit, the phosphorylation of amino acid residue S442 of PRKCSH was downregulated in HER2+ vs. HER2-enriched patients. The phosphorylation of amino acid residue S583 of CANX is a target site for the activity of *in vitro* and *in vivo* kinases, ERK1 and Cyclin dependent kinase 1 (CDK1). ATP5B which is also linked to MHC class I protein binding, had a high phosphosite enrichment ratio of 8.23 for the phosphorylation of amino acid residue S465. ATP5B is known to share a high confidence score STRING interaction between ATP5B and CS (Table 4.3.7). CS did not exhibit a statistically significant

enrichment of phosphopeptides; however, this indicates a potential link between the glucose metabolism related function of CS and MHC class I protein binding. Therefore, the proteins linked to CPNE3 knockdown and identified in cluster 1 of Figure 4.2.5, are linked to MHC class I related antigen processing and presentation, and CPNE3 could play a role as an upstream regulator of this process.

#### 4.5. Developing predictive models for HER2-positive breast cancer

To determine if marker combinations showed superior sensitivity and specificity compared to single markers, we perform logistic regression using 2 models. Model 1 was built using a combination of six proteins previously identified as potential clinical biomarkers for HER2 breast cancer in experimental and *in silico* studies: ERBB2, ARHGDIB, CACYBP, CPNE3, KPNA2 and EIF5A and is a predictive model for HER2 status. The training and testing were implemented using a 70-30 train-test split of the discovery and validation cohort datasets. Model 2 was built using a combination of six proteins discovered to have a significant correlation to CPNE3 and HER2 positive status (Table 4.5.1) in the validation cohort: CPNE3, CS, EIF5A, ERBB2, HIST1H4A and PYGM and is a predictive model of HER2 positive status. The training and testing were implemented using a 70-30 train-test split of the discovery and validation cohort datasets. Both models were evaluated using the AUC of a ROC curve with confidence intervals computed for a given score function based on labels and predictions using a bootstrapping method. Leave one out cross validation (LOOCV) and the root mean squared error (RMSE) were used to measure how well the predictions made by the model match the observed data. The Youden J index was used to capture the performance of each dichotomous diagnostic test and to account for the imbalanced class distribution in the two datasets (Table 4.2.2).

##### 4.5.1. Model 1: A predictive model of HER2 positive status

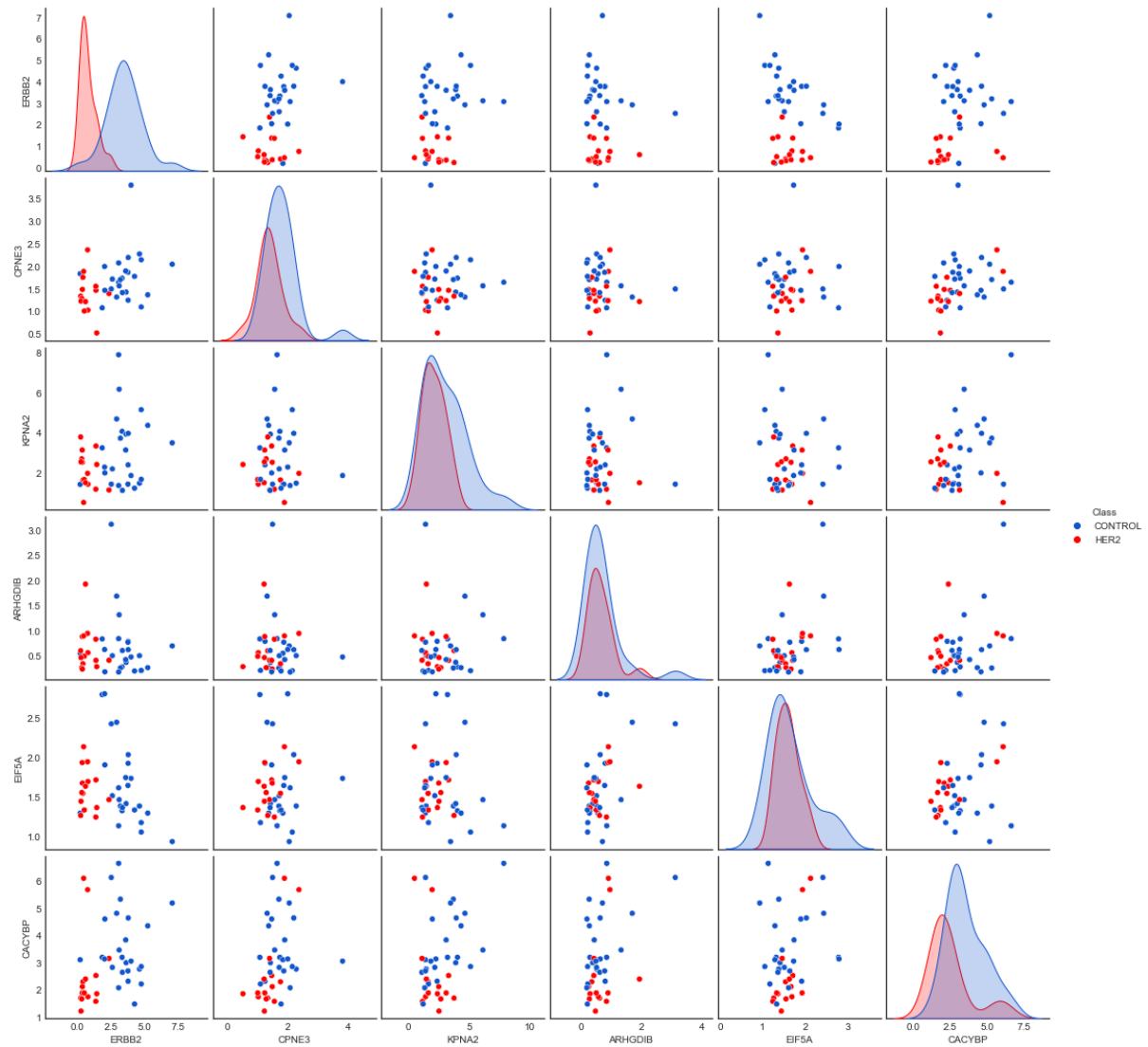
Logistic regression model 1 was trained on a combination of independent variables derived from a set of proteins previously identified as potential biomarkers of breast cancer: ERBB2, ARHGDIB, CACYBP, CPNE3, KPNA2 and EIF5A. A total of 6 predictor combinations were selected in line with the experimental hypothesis that CPNE3 operates downstream of ERBB2 and upstream of

ARHGDIB, CACYBP, EIF5A and KPNA2. The diagnostic performance of predictor combinations on the discovery cohort was initially explored using a pair plot from the Python Seaborn package (Figure 4.5.1).

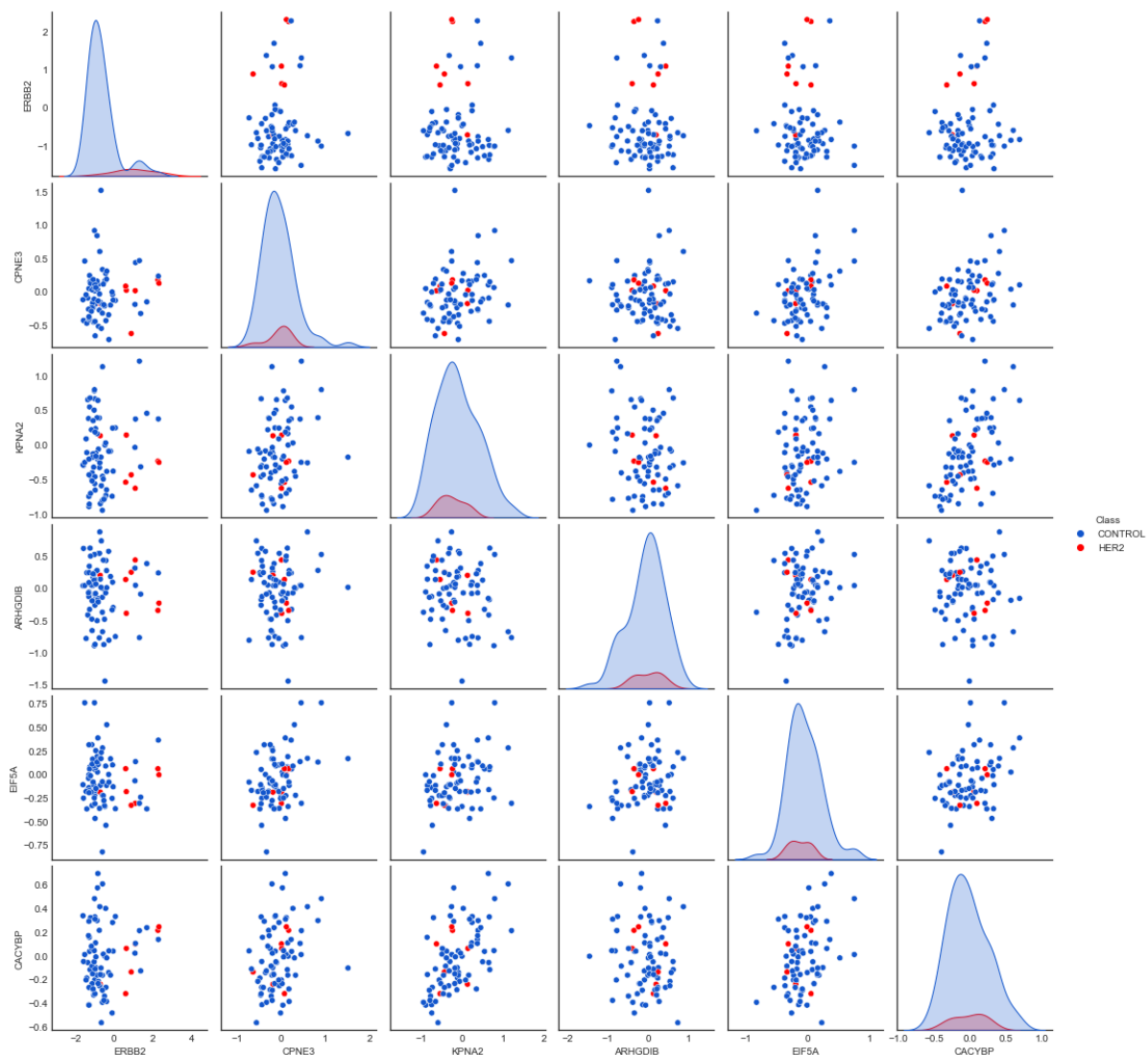
ERBB2 and CACYBP were found to be the most effective patient diagnosis pair for HER2 positive and HER2 negative subtypes in the discovery cohort. Conferring a slightly denser clustering of the HER2 patients with a wider distribution than either ARHGDIB, CPNE3, KPNA2 and EIF5A (Figure 4.5.1). The same pattern was observed in the validation cohort and CACYBP conferred a denser clustering of both HER2 positive and negative patient groups for the ERBB2 and CACYBP pairwise plot (Figure 4.5.3).

Using all six proteins, the resulting performance scores for model 1 on the discovery cohort were a sensitivity of 86%, specificity of 100% and an AUC of 0.96 (95% CI: 0.79, 1.00) (Figure 4.5.3A) (Table 4.5.2). This result was similar to the performance of the predictor pair CPNE3 & CACYBP (Table 4.5.2). However, the predictor pair ERBB2 & CPNE3 outperformed all other predictor combinations resulting in a sensitivity of 100%, specificity of 100% and an AUC of 1.0 (95% CI: 1.00, 1.00). CPNE3 & ARHGDIB, CPNE3 & EIF5A and CPNE3 & KPNA2 performed poorly, with an equal sensitivity of 14%, specificity of 100% and an AUC of 0.57 (95% CI: 0.50, 0.75). Overall, three predictor combinations, ERBB2 & CPNE3, CPNE3 & CACYBP and all six predictors, resulted in an acceptable diagnostic performance with a Youden's J index > 0.5 and RMSE of 0.274 for the LOOCV on the discovery cohort (Table 4.5.2).

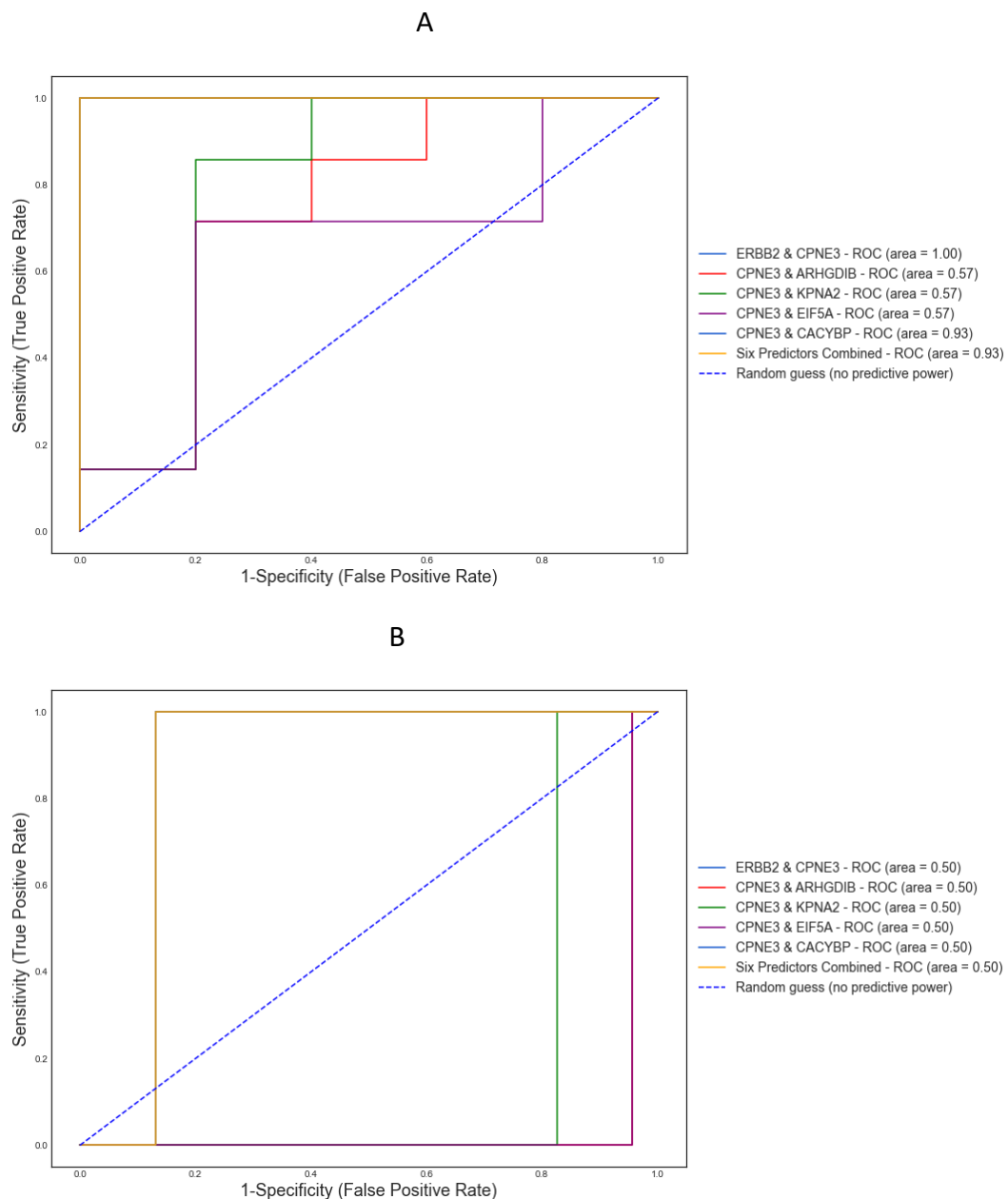
Model 1 and its respective predictors performed less favourably on the validation cohort. Using all six proteins, the resulting performance scores for model 1 on the validation cohort were a sensitivity of 50%, specificity of 92% and an AUC of 0.50 (95% CI: 0.50, 0.50) (Figure 4.5.3B) (Table 4.5.2). The combined markers outperformed all other predictor pairs on the validation cohort. However, all predictor pairs evaluated had an AUC of 0.50 (95% CI: 0.50, 0.50) and Youden's J index < 0.5 (Table 4.5.2). Therefore, none of the predictor pairs resulted in a meaningful diagnostic performance on the validation cohort.



**Figure 4.5.1** Pair plot of candidate biomarker combinations correlated with HER2 positive status in the discovery cohort (n=40)



**Figure 4.5.2** Pair plot of candidate biomarker combinations correlated with HER2 positive status in the validation cohort (n=75)



**Figure 4.5.3** Receiver operator curve analysis of multivariate logistic regression model 1 with individual or combined proteins as predictor variables. A) Discovery cohort (n=40) B) Validation cohort (n=75)

#### 4.5.2. Model 2: A predictive model of HER2-positive and HER2 enriched status

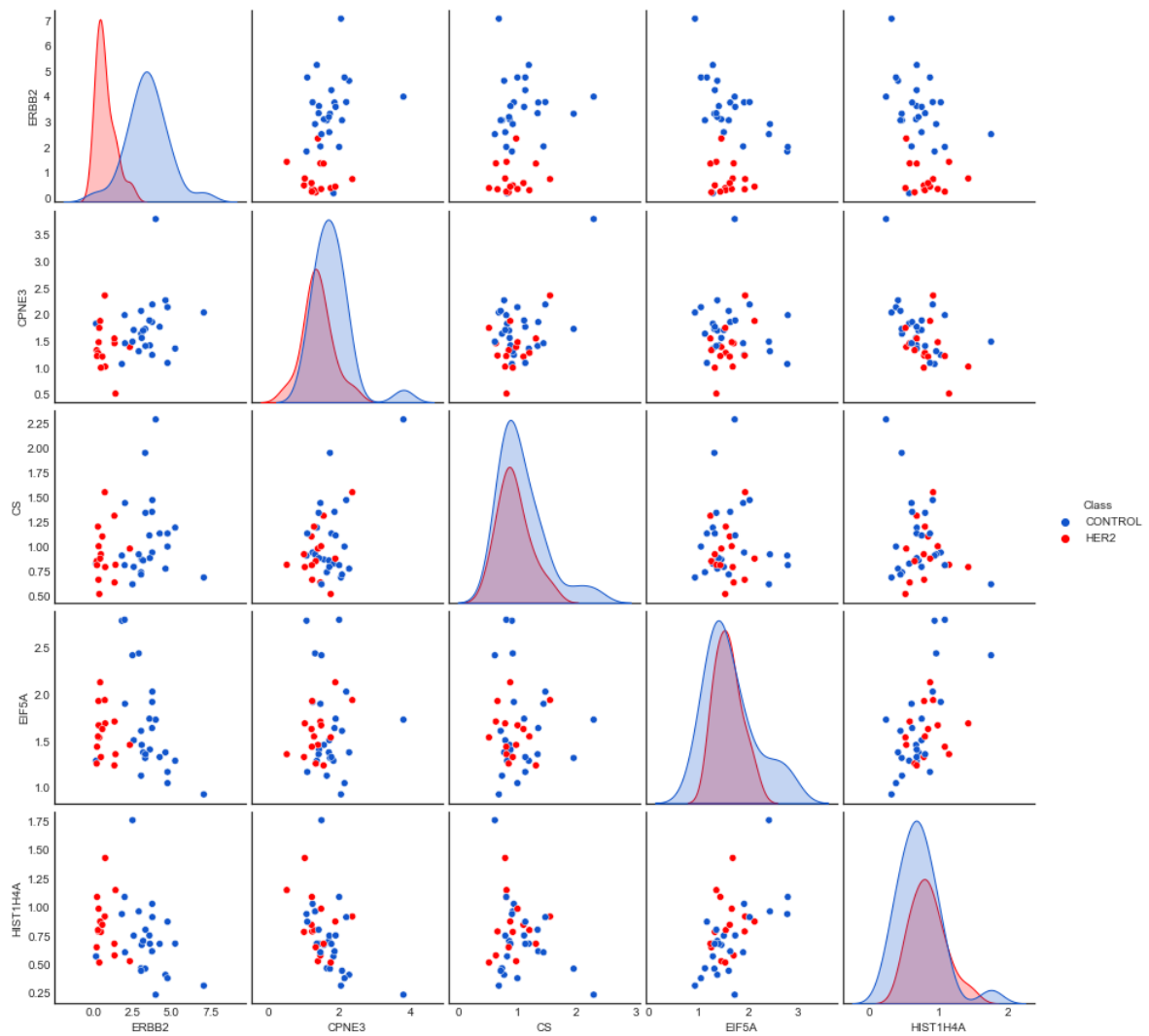
Logistic regression model 2 was trained on a combination of independent variables derived from a set of proteins discovered to correlate with HER2 positive status among the HER2-enriched subtype of our validation cohort: CPNE3, CS, EIF5A, ERBB2, HIST1H4A and PYGM. The proteins were selected on the basis of their predictive potential for HER2 positive status and a correlation with the predominant HER2-enriched status (Table 4.5.1).

**Table 4.5.1** The table lists candidate biomarkers correlated with HER2 positive/HER2 enriched status.

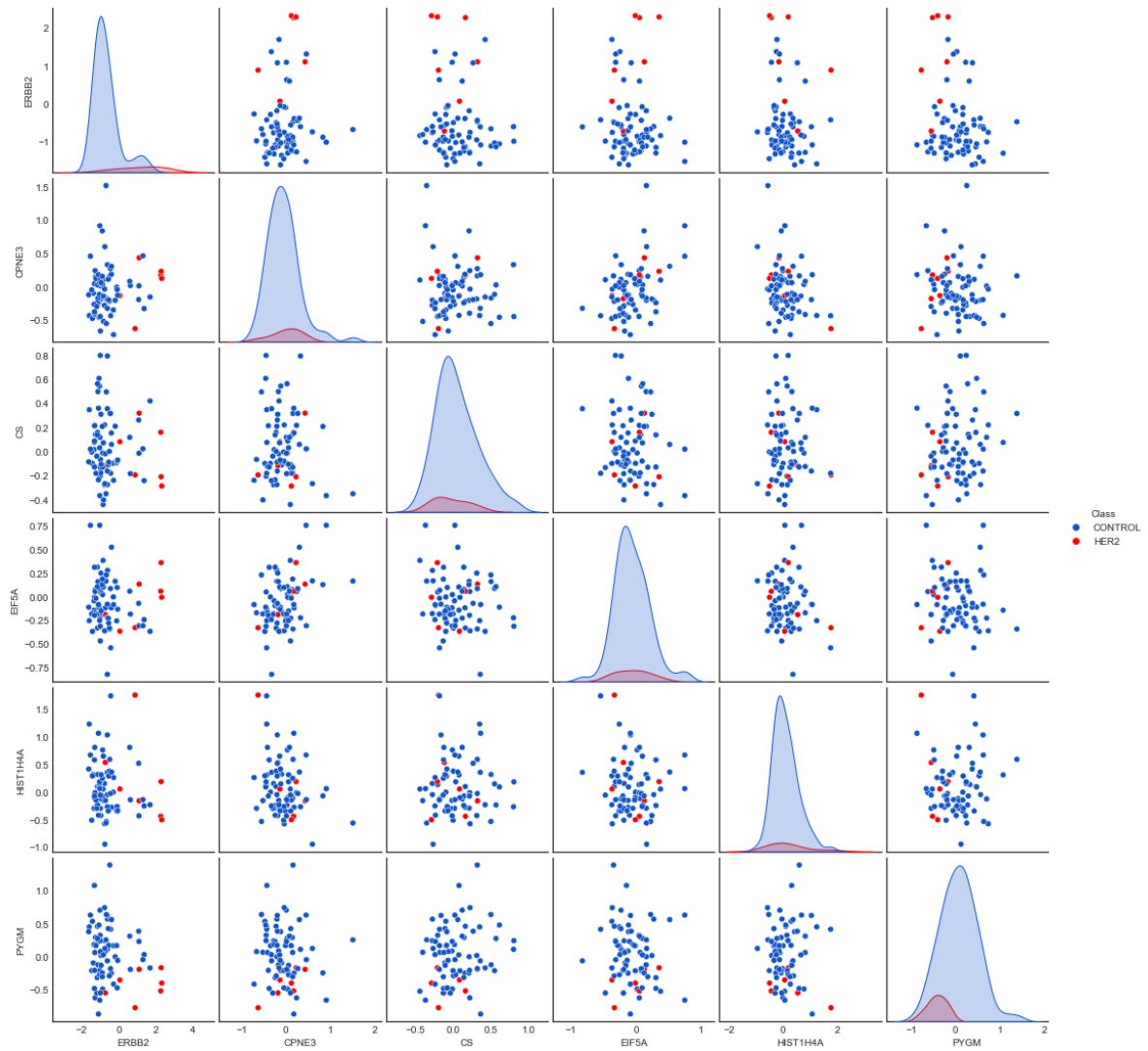
Gene Symbol	Molecule Type	Potential Biomarker Application
CS	Enzyme	HER2 diagnosis
EIF5A	Translation regulator	HER2 diagnosis
HIST1H4A	Other	HER2 diagnosis
PYGM	Enzyme	HER2 diagnosis

A total of 6 predictor combinations were selected in line with the experimental hypothesis that CPNE3 operates downstream of ERBB2 and potentially upstream of CS, EIF5A, HIST1H4A and PYGM. The HER2 diagnostic performance of predictor combinations was explored for the validation cohort using a pair plot from the Python Seaborn package (Figure 4.5.4). It was unclear from a pairwise plot which candidate biomarkers correlated with HER2-enriched status were the most effective for patient diagnosis in the validation cohort (Figure 4.5.4).

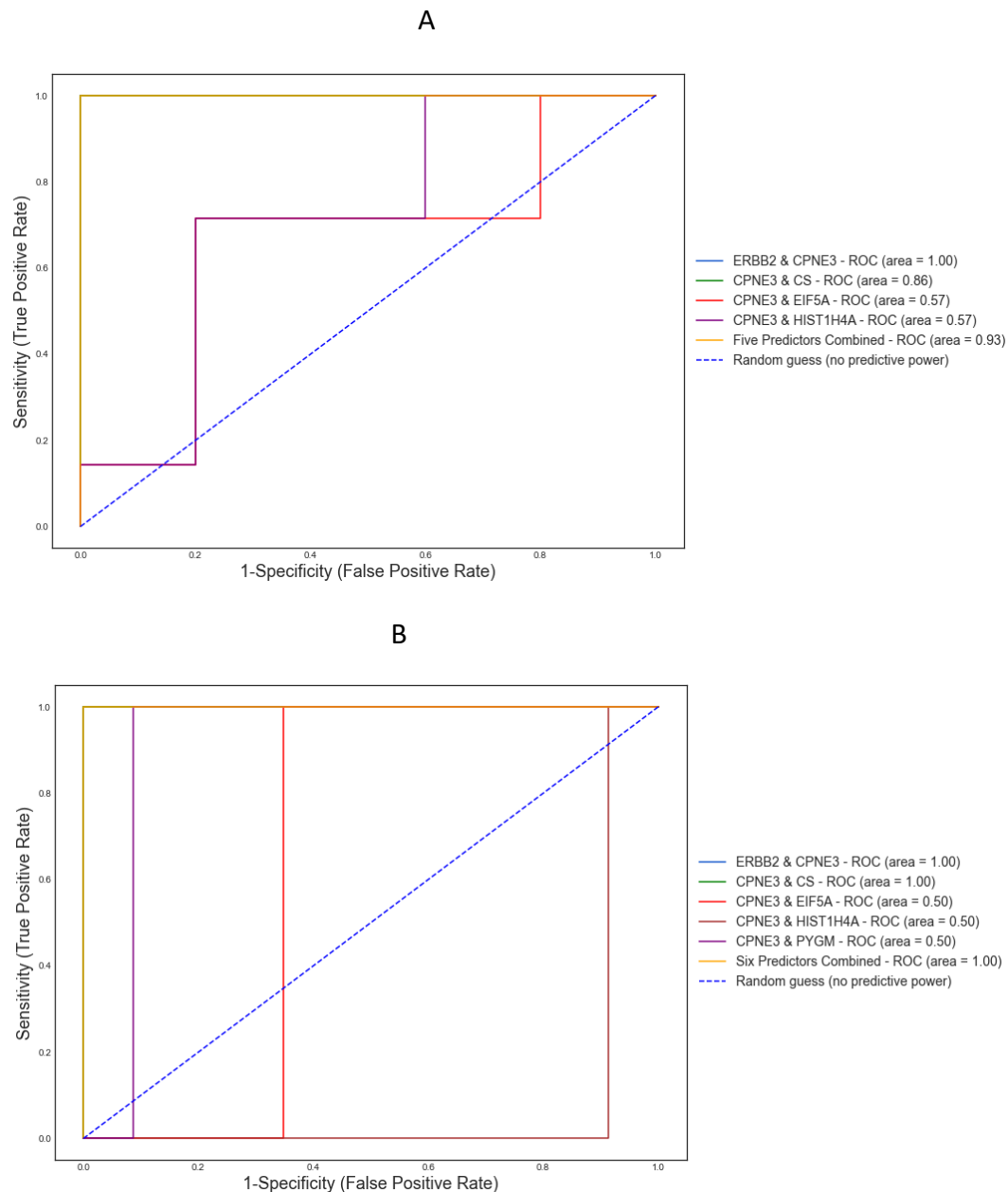




**Figure 4.5.4** Pair plot of candidate biomarkers correlated with both HER2 positive and HER2-enriched status in the discovery cohort (n=40)



**Figure 4.5.5** Pair plot of candidate biomarker combinations correlated with both HER2 positive and HER2-enriched status in the validation cohort (n=75)



**Figure 4.5.6** Receiver operator curve analysis of multivariate logistic regression model 2 with individual or combined proteins as predictor variables. A) Discovery cohort (n=40) B) Validation cohort (n=75)

Using five proteins as predictor variables, the performance scores for model 2 on the discovery cohort were a sensitivity of 100%, specificity of 100% and an AUC of 0.93 (95% CI: 0.75, 1.00) (Figure 4.5.6A) (Table 4.5.2). PYGM was not quantified in all samples of the discovery cohort and it was not possible to evaluate its predictive performance on the discovery cohort and only five predictors were used to evaluate the model. The predictor pair ERBB2 & CPNE3 outperformed the five combined predictors resulting in a sensitivity of 100%, specificity of 100% and an AUC of 1.0 (95% CI: 1.00, 1.00). Interestingly, the CPNE3 & CS predictor pair selected from an evaluation of TNM clinical stages

of the HER2 positive subtype of the HER2-enriched subtype of the validation cohort performed well on the discovery cohort, with a sensitivity of 71%, specificity of 100% and an AUC of 0.86 (95% CI: 0.67, 1.00). CPNE3 & EIF5A and CPNE3 & HIST1H4A performed poorly, with an equal sensitivity of 14%, specificity of 100% and an AUC of 0.57 (95% CI: 0.50, 0.75). Overall, three predictor combinations, ERBB2 & CPNE3, CPNE3 & CS and the five predictors combined, resulted in an acceptable diagnostic performance with a Youden's J index  $> 0.5$  and RMSE of  $\leq 0.274$  for the LOOCV on the discovery cohort (Table 4.5.2).

Using all six proteins, the resulting performance scores for model 2 on the validation cohort were a sensitivity of 100%, specificity of 100% and an AUC of 1.0 (95% CI: 1.00, 1.00) (Figure 4.5.6B) (Table 4.5.2). This result was similar to the performance of the predictor pairs ERBB2 & CPNE3 and CPNE3 & CS (Table 4.5.2). However, the predictor pairs CPNE3 & EIF5A, CPNE3 & HIST1H4A and CPNE3 & PYGM performed very poorly, with an equal sensitivity of 0%, specificity of 100% and an AUC of 0.50 (95% CI: 0.50, 0.50). Overall, three predictor combinations, ERBB2 & CPNE3, CPNE3 & CS and the six predictors combined, resulted in an acceptable diagnostic performance with a Youden's J index  $> 0.5$  and RMSE of 0.253 for the LOOCV on the validation cohort (Table 4.5.2).

The logistic regression models have confirmed that CACYBP and CS are much better predictors of HER2 status in clinical patients than the previously proposed candidate proteins, ARHGDIB and KPNA2 (Table 4.5.2). These candidate proteins discovered through comparative proteomic profiling of C3.6 cells, an ERBB2 overexpressing HMLEC and SKBR3 cells, a metastatic ERBB2 overexpressing breast cancer cell line, are shown to be ineffective at patient diagnosis and determination of clinical HER2 patient status in two patient cohorts. The expression of ARHGDIB and KPNA2 is sporadic and non-specific to the HER2 subtype of patients in the clinical samples (Figure 4.2.2 & Figure 4.3.2). Correlations between CPNE3 and the candidate markers CACYBP and CS (Figure 4.3.4), present a much more favourable HER2 status prediction performance (Table 4.2.2). Furthermore, CACYBP is already known to play a role in clinical progression of breast cancer (N. Wang et al., 2010). The data presented herein supports the findings from the literature and confirms a predictive correlation between CACYBP and CPNE3 in HER2 patients.

**Table 4.5.2** Summary of logistic regression values for candidate biomarkers predicting HER2 status.

<b>Model</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>AUC</b>	<b>95% CI</b>	<b>Youden Index</b>	<b>LOOCV (RMSE)</b>
<b>Model 1 - Discovery Cohort</b>						
ERBB2 & CPNE3	1.00	1.00	1.00	1.00 - 1.00	1.00	0.274
CPNE3 & ARHGDIB	0.14	1.00	0.57	0.50 - 0.75	0.14	0.592
CPNE3 & KPNA2	0.14	1.00	0.57	0.50 - 0.75	0.14	0.592
CPNE3 & EIF5A	0.14	1.00	0.57	0.50 - 0.75	0.14	0.592
CPNE3 & CACYBP	0.86	1.00	0.93	0.79 - 1.00	0.86	0.274
Six Predictors Combined	0.86	1.00	0.93	0.79 - 1.00	0.86	0.274
<b>Model 1 - Validation Cohort</b>						
ERBB2 & CPNE3	0.50	0.89	0.50	0.50 - 0.50	0.39	0.300
CPNE3 & ARHGDIB	0.00	1.00	0.50	0.50 - 0.50	0.00	0.300
CPNE3 & KPNA2	0.00	1.00	0.50	0.50 - 0.50	0.00	0.300
CPNE3 & EIF5A	0.00	1.00	0.50	0.50 - 0.50	0.00	0.300
CPNE3 & CACYBP	0.50	0.89	0.50	0.50 - 0.50	0.39	0.277
Six Predictors Combined	0.50	0.92	0.50	0.50 - 0.50	0.42	0.300
<b>Model 2 - Discovery Cohort</b>						
ERBB2 & CPNE3	1.00	1.00	1.00	1.00 - 1.00	1.00	0.274
CPNE3 & CS	0.71	1.00	0.86	0.67 - 1.00	0.71	0.224
CPNE3 & EIF5A	0.14	1.00	0.57	0.50 - 0.75	0.14	0.632
CPNE3 & HIST1H4A	0.14	1.00	0.57	0.50 - 0.75	0.14	0.632
CPNE3 & PYGM	nil	nil	nil	nil	nil	nil
Five Predictors Combined	1.00	1.00	0.93	0.75 - 1.00	1.00	0.224
<b>Model 2 - Validation Cohort</b>						
ERBB2 & CPNE3	1.00	1.00	1.00	1.00 - 1.00	1.00	0.253
CPNE3 & CS	1.00	1.00	1.00	1.00 - 1.00	1.00	0.253
CPNE3 & EIF5A	0.00	1.00	0.50	0.50 - 0.50	0.00	0.253
CPNE3 & HIST1H4A	0.00	1.00	0.50	0.50 - 0.50	0.00	0.300
CPNE3 & PYGM	0.00	1.00	0.50	0.50 - 0.50	0.00	0.300
Six Predictors Combined	1.00	1.00	1.00	1.00 - 1.00	1.00	0.253

## 4.6. Chapter Conclusion

In conclusion, our findings have important implications for the use of proteomic markers to diagnose HER2 status in breast cancer patients. Our results contrast with previous studies that used cell lines for comparative proteomic profiling, which did not identify CPNE3 as a specific marker for HER2-positive breast cancer (Heinrich et al., 2010; Kulkarni et al., 2010; Worthington et al., 2010, 2017). Our study has shown that CPNE3 expression is variable between different breast cancer subtypes, suggesting that the tumour microenvironment plays a significant role in CPNE3 expression. Further investigation is necessary to elucidate how the tumour microenvironment affects the expression of CPNE3 and its downstream candidates, such as ARHGDIB and KPNA2, which have previously been implicated as potential biomarkers for HER2-positive breast cancer. Additionally, treatment with growth factors has been shown to increase CPNE3 expression in cells that do not overexpress ERBB2 (Bertani, 2005; Gharbi et al., 2002; White et al., 2004), indicating that further research is needed to determine how this may affect downstream candidates linked to CPNE3 expression.

Our analysis has also identified novel candidates that have a unique correlation to CPNE3 expression in HER2-positive breast cancer patients. We confirmed the specificity of these correlations using two patient cohorts and known STRING protein-to-protein interactions. These candidates have a functional role in the TCA cycle and other glucose metabolism related pathways and were previously unknown as biomarkers for HER2-positive breast cancer. Our analysis of the phosphoproteome of these candidates revealed interesting functional associations, particularly in DNA binding and antigen processing and presentation. Furthermore, our observation of a phosphopeptide enrichment pattern suggests that CPNE3 could play an upstream regulatory role in MHC class I protein binding and antigen processing and presentation via the MHC I pathway. These findings warrant further investigation into the role of CPNE3 in MAPK3 kinase signaling via ERK1 or cell cycle regulation via CDK1. Future research should focus on validating the candidates linked to these pathways as biomarkers for HER2-positive breast cancer and investigating their potential role in the development and progression of this subtype of breast cancer.

## Chapter 5

### 5. Discussion

#### 5.1. Chapter Introduction

Copine 3 (CPNE3) expression is correlation with the amplification of tyrosine kinase receptor ERBB2/Her2 in human mammary luminal epithelial cells (HMLECs) and breast tumour tissues. ERBB2/Her2 amplification in breast cancer is implicated as a key contributor to aggressive disease phenotype, metastasis and poor prognosis, and CPNE3 is one of several gene products that are hypothesised to bring about these effects via ERBB2-dependent signalling (Bertani, 2005; Durán et al., 2008; Gharbi et al., 2002; Mackay et al., 2003; White et al., 2004; Worthington et al., 2010). Despite previous experiments to establish the role of CPNE3 with respect to ERBB2 over-expression and/or growth factor-dependent modulation, the candidate protein has hitherto remained poorly characterised and lacked a definitive role in ERBB2-dependent tumourigenicity. The functional characterisation of CPNE3 and identification of potential interaction partners or proteins that correlate with CPNE3 expression is therefore essential to unravel its role in tumourigenicity.

The principal aims of this study were to validate previous findings related to the expression of CPNE3 and cell adhesion-related proteins and to establish a functional role in cell adhesion for CPNE3 in the context of ERBB2 overexpression. Previous findings from gene expression profiling studies, including the downregulation of cell adhesion proteins ITGA6 and ITGB4 in response to ERBB2 overexpression were validated at the protein level in a HMLEC model cell system. CPNE3 was functionally characterised by evaluating the effect of siRNA (small interfering RNA)-mediated gene silencing on cellular adhesion and the spread of ERBB2 over-expressing HMLECs on an adherent surface. Further aims were to elucidate possible molecular networks involved in producing the observed siRNA-mediated phenotypes by global proteomic profiling of CPNE3 knockdown in an ERBB2 over-expressing HMLEC (see Chapter 3). Proteins differentially regulated by siRNA-mediated knockdown of CPNE3 were identified using two bottom-up global protein expression profiling methods on two

separate mass spectrometry instrument platforms. To resolve the dissimilarity in the two datasets generated by the two methods, principal component analysis and k-means clustering were applied to extract an intrinsic dimension (id) of each data set that describes the proteins with the highest expression similarity to CPNE3.

In this study, we further evaluated the expression pattern of the differentially regulated proteins identified using one of the global protein expression profiling methods (TMT-LC-MS/MS) by applying biostatistical approaches to retrospective LC-MS/MS data from two breast cancer patient cohorts to gain further insight into the possible molecular networks involved in producing the observed siRNA-mediated phenotypes or alternative molecular functions of CPNE3 as a candidate in downstream signalling of ERBB2 positive breast. This approach coupled with logistic regression provided *in silico* validation of the ERBB2/Her2 subtype specificity of candidate biomarkers differentially regulated by CPNE3 expression. Bioinformatics analysis of data was used in an attempt to define the functional consequences of candidate gene silencing and to link changes with cellular phenotype or alternative molecular mechanisms. The phosphoproteomic enrichment of proteins with STRING interactions of high confidence ( $\geq 0.90$ ) and specific to the ERBB2/Her2 subtype following biostatistical analysis was evaluated in an attempt to establish potential links between proteins differentially regulated by CPNE3 and the ERBB signalling network.

Despite extensive research efforts to establish the functional role of CPNE3 in several cancers, the full mechanistic role of CPNE3 in relation to ERBB2 over-expression and/or growth factor-dependent modulation remains ambiguous. The present study aimed to evaluate the effects of ERBB2 amplification on CPNE3 expression and proteins differentially regulated by CPNE3 in early signal transduction events in a HMLEC model cell system and ERBB2/Her2 overexpressing patients (see Chapters 3 and 4). Downstream ERBB2 signalling targets differentially regulated by CPNE3 and their respective putative sites of phosphorylation were determined using phosphopeptide enrichment analysis of CPTAC LC-MS/MS data for ERBB2/Her2 overexpressing vs Her2-enriched patients. Finally, a binary classification model was used to evaluate the potential diagnostic potential of proteins determined to share a unique pairwise correlation with CPNE3 in ERBB2/Her2 patients. The



following section presents the major findings from the present study with a more detailed discussion provided within each chapter.

## 5.2. Validating the role of CPNE3 in cell adhesion

Previous expression profiling studies of a HMLEC model cell system and breast tumour tissues have identified the downregulation of cell adhesion-related proteins ITGA6 and ITGB4 in response to ERBB2 overexpression. In addition, CPNE3 knockdown in ERBB2 overexpressing breast tumour tissues was found to confer a slight reduction in cell adhesion. The present study aimed to validate the previous findings and to evaluate the effect of siRNA-mediated knockdown of CPNE3 on cellular adhesion and spread in ERBB2 over-expressing HMLECs (see Chapter 3). The present study verified previously observed differences in ITGA6 and ITGB4 expression between the model cell lines C3.6 and HB4a. ITGA6 and ITGB4 were confirmed to be down-regulated in ERBB2 over-expressing clones relative to parental HMLECs. In contrast, the siRNA-mediated knockdown of ERBB2 in ERBB2 over-expressing clones did not induce an up-regulation of ITGA6 and ITGB4 at the protein level. The present study did however confirm that the down-regulation of ITGA6 and ITGB4 also correlates with the overexpression of CPNE3 in ERBB2 over-expressing clones. However, like ERBB2, the siRNA-mediated knockdown of CPNE3 in ERBB2 over-expressing clones did not induce an up-regulation of ITGA6 and ITGB4 at the protein level. While regulation of ITGA6 and ITGB4 expression isn't apparent at the protein level following siRNA-mediated knockdown of CPNE3 and ERBB2, knockdown of these proteins may still elicit a long-term adaptive regulation of expression that cannot be reversed by acute knockdown of ERBB2 in the absence of growth factor-induced receptor triggering. Indeed, previous studies have shown that treatment with growth factors can induce a moderate increase in CPNE3 expression in the HB4a cells, which do not overexpress ERBB2 (Bertani, 2005; Gharbi et al., 2002; White et al., 2004). Further experimentation is therefore required to determine whether ERBB receptor triggering with growth factors would yield an upregulation of ITGA6 and ITGB4 following the siRNA-mediated knockdown of either CPNE3 or ERBB2.

The present study also identified that ERBB2 over-expressing clones detach and adhere to an adherent surface more readily than parental HMLECs. This further supports the hypothesis that

ERBB2 signalling networks interact to modulate cell adhesion. In addition, the siRNA-mediated knockdown of CPNE3 and ERBB2 confers a slight shift in the rate of cell spread on an adherent surface from that of ERBB2 over-expressing clones towards one similar to parental HMLECs. Although not statistically significant, this highlights the subtlety of the mechanical dynamics involved and perhaps suggests that a more discernible and statistically significant difference could be achieved in response to ERBB receptor triggering with growth factors. The present data is therefore inconclusive and supports the need for a follow-up study with triggering of ERBB signalling networks by growth factor treatment to address this in the future.

Prospective interaction partners with potential roles in producing the observed phenotype in a CPNE3-dependent manner were predicted by evaluating the effect of candidate knockdown on global protein expression of ERBB2 over-expressing clones using TMT LC-MS/MS and Label-Free LC-MS/IMS/MS (see Chapter 3). The TMT LC-MS/MS proteomic profiling experiments identified a total of 52 significantly up/ down-regulated ( $>1.5$  fold) changes in protein expression. This included 28 proteins that have previously been reported to have a causal relevance to breast cancer or the cell adhesion phenotype based on a Qiagen Ingenuity Pathway Analysis. The 52 proteins were compared to prior data generated through CPNE3 knockdown of ERBB2 over-expressing SKBR3 cells using SILAC-based LC-MS/MS to identify 2 proteins, ARGHDIB and KPNA2, differentially regulated by CPNE3 in a similar manner in both cell lines. The list of 52 significantly up/ down-regulated proteins was mostly enriched for Reactome pathways associated with DNA fragmentation and DNA-induced cell senescence, with  $p$ -value  $< 0.05$  and FDR  $< 0.05$ . The Label-Free LC-MS/IMS/MS proteomic profiling experiments identified a total of 40 significantly up/ down-regulated ( $>1.5$  fold) changes in protein expression. This included 17 proteins that have previously been reported to have a causal relevance to breast cancer or the cell adhesion phenotype based on a Qiagen Ingenuity Pathway Analysis. The 40 proteins were compared to prior data generated through CPNE3 knockdown of ERBB2 over-expressing SKBR3 cells using SILAC-based LC-MS/MS but identified no proteins differentially regulated by CPNE3 in a similar manner in both cell lines. The list of 40 significantly up/ down-regulated proteins was mostly enriched for Reactome pathways related to DNA replication. However, the results were not statistically significant FDR  $> 0.05$ . At first glance, the two datasets lack

any shared patterns in altered protein expression. However, Qiagen Ingenuity Pathway Analysis identified several proteins with a causal relevance to breast cancer and a known role in cell adhesion including the formation, turnover, size and quantity of focal adhesions. This suggests that CPNE3 may play a role in regulating focal adhesions in ERBB2 over-expressing clones during cell adhesion.

To reveal functional similarities between the two mass spectrometry datasets, we repeated our analysis using a cluster-based dimensionality reduction method to identify the intrinsic dimension of each dataset. The dimensionality reduction was intended to provide additional evidence of a relationship between CPNE3 and cell adhesion by extracting a minimal dataset of proteins with a close expression similarity to CPNE3 using K-means clustering and PCA. According to our results, the clusters generated by analysing the data from both mass spectrometry methods demonstrate reduced levels of ribonucleoprotein and ribonucleoprotein-associated proteins in response to CPNE3 knockdown in HMLEC cells. This shared protein expression pattern is not apparent in the initial analysis and reveals functional similarities between the two datasets. Ribonucleoproteins are widely known to interact with nuclear actin in the nucleus where cytoskeletal components have been temporally implicated in gene expression (Percipalle et al., 2003). However, an association also occurs in a cell structure that exists only in the early stages of cell spreading known as a spreading initiation centre (SIC). SICs are ribonucleoprotein complexes that appear to be surrounded by an actin sheath, contain focal adhesion markers and play a role in the initiation of cell spreading (de Hoog et al., 2004). This suggests that CPNE3 expression may affect the formation of SICs through reduced expression of ribonucleoprotein and ribonucleoprotein-associated proteins, SNRPB, HNRNPU, HNRNPD, SNRPD3 and XRN2, revealed by the analysis of TMT-LC-MS/MS data and SYNCRIP, HNRNPU, HNRNPH1 and HNRNPR, revealed by Label Free-LC-MS/IMS/MS data.

Upregulation of several ribonucleoproteins discovered by TMT-LC-MS/MS and Label Free-LC-MS/IMS/MS is associated with an aggressive phenotype in BRCA patients. HNRNPD has been shown to be upregulated in BRCA tissues compared to normal tissues and is associated with high cancer stages (J. Zhou et al., 2021). HNRNPU and SYNCRIP have previously been shown to be upregulated in BRCA tissues compared to normal tissues and are also associated with high cancer stages (J. Zhou et al., 2021). HNRNPH1 is known to function as an oncogene in BRCA progression

and is known to result in the production of HER2 splice variants in BRCA (J. Zhou et al., 2021). HNRNPR is upregulated in invasive BRCA tissue and has a significant association with high cancer stages and lymph node metastases (J. Zhou et al., 2021) and depletion of SNRPD3 has been shown to inhibit the proliferation of TNBC cell lines (Koedoot et al., 2021). The aggressive phenotype associated with the upregulation of these ribonucleoproteins in BRCA tissues may be reversible through the knockdown of CPNE3. The subsequent reduction in ribonucleoprotein expression demonstrated by analysis of both data sets would in turn increase cell spreading by interfering with the formation of SICs (de Hoog et al., 2004). This is consistent with the lower duration of cell spread demonstrated by the CPNE3 knockdown of ERBB2 over-expressing clones compared to parental HMLECs during real-time cell adhesion assays. However, it must be noted that the real-time cell adhesion assay results had p-value > 0.05 according to a Welch's unpaired t-test. With this background in mind, it is quite clear that these ribonucleoprotein and ribonucleoprotein-associated proteins play a role in disease progression and are good candidate biomarkers for HER2-positive breast cancer.

### 5.3. Characterisation of CPNE3 mediated cell signalling

The role of CPNE3-related signalling mechanisms in ERBB2-mediated transformation and breast tumour progression is currently poorly understood. The present study aimed to elucidate potential interaction partners of CPNE3, gain further insight into the possible molecular networks and identify putative sites of phosphorylation using data from two patient cohorts comprised of SILAC LC-MS/MS data and CPTAC LC-MS/MS data for ERBB2/Her2 overexpressing vs Her2-enriched patients (see Chapter 4). The present biostatistical analysis was run using the 52 proteins identified from the TMT LC-MS/MS global protein expression profiling to establish pairwise correlations on SILAC LC-MS/MS clinical data from a 40-patient discovery cohort. 10 proteins, ATP5B, BOLA2, CACYBP, CANX, EIF5A, HSPE1, LRPPRC, PRKCSH, RPL19 and SLC25A3 were found to have statistically significant (p-value < 0.05) and strong positive correlations with CPNE3 (r-value > 0.5) for the Her2 positive subgroup. Most of these proteins were also identified in a cluster containing 13 highly interconnected proteins, ATP5B, ATP5D, BOLA2, CACYBP, CANX, DLD, DHX15, EIF5A, LRPPRC, PRKCSH, RHOC, RPL19 and SLC25A3 identified by constructing a biological network with strong correlations

( $r$ -value > 0.5) as edges or interconnecting lines in Cytoscape. This included 6 proteins, CANX, DLD, LRPPRC, PRKCSH, RHOC and RPL19 observed in all KEGG and Reactome pathways and confirmed to have STRING interactions of high confidence score ( $\geq 0.90$ ). 3 proteins CANX, PRKCSH and RPL19 were found to be either singly or doubly phosphorylated with a total of 7 putative sites of phosphorylation identified for ERBB2/Her2 overexpressing vs Her2-enriched patient phosphosite enrichment data. Notably, amino acid residue S583 of CANX was found to be 2.63-fold singly phosphorylated and has been identified as a target site for the activity of *in vitro* and *in vivo* kinases, ERK1 and Cyclin-dependent kinase 1 (CDK1). The upstream role of ERK1 remains unconfirmed. However, the role of CDK1 has been shown by covalent capture of kinase-specific phosphopeptides to reveal Cdk1-cyclin B substrates (Blethrow et al., 2008). The interaction of ERBB2 with cyclin-dependent kinase activity has been investigated using *in vivo* murine mammary tumour models to demonstrate that the function of cyclin D1 mediates transformation by ERBB2 (C. Yang et al., 2004). CANX which is a downstream target of CDK1 is upregulated 1.57-fold in ERBB2/Her2 overexpressing vs Her2-enriched tumours and 3.129-fold in response to CPNE3 knockdown.

CANX is an endoplasmic reticulum chaperone that controls the flux of  $\text{Ca}^{2+}$  ions into mitochondria to either increase or decrease the energetic output of the oxidative phosphorylation (OXPHOS) pathway and thereby control the energy balance between OXPHOS and glycolysis in cells (Gutiérrez et al., 2020). Studies into CDK1 and its role in mitochondrial OXPHOS have demonstrated that cyclin B1/Cdk1 proteins relocate to the matrix of mitochondria and elevate mitochondrial bioenergetics in G2/M transition (Z. Wang et al., 2014). These findings coupled with the downregulation of PYGM, glycogen phosphorylase and upregulation of CS, a TCA cycle enzyme suggest that the abrogation of CPNE3 expression increases the energetic output of OXPHOS and glycolysis. Albeit, suppressing glycogen metabolism via glycogenolysis. This is significant because breast cancer cells have been reported to undergo an increase in glycogen stores in response to hypoxia and are known to utilise these hypoxic glycogen stores via glycogen phosphorylases to promote metastatic phenotypes (Altemus et al., 2019). The present study, therefore, suggests that the knockdown of CPNE3 is a plausible method to reverse metastatic phenotypes in breast cancer under hypoxic or post-hypoxic conditions.

The present study confirmed the constitutive correlations of gene products previously identified by TMT LC-MS/MS global protein expression profiling following CPNE3 knockdown by secondary biostatistical analysis to establish pairwise correlations on a 75-patient validation cohort of CPTAC LC-MS/MS clinical data. 3 proteins, EIF5A, PYGM and HIST1H4A were found to have statistically significant ( $p\text{-value} < 0.05$ ) and strong correlations with CPNE3 ( $r\text{-value} > 0.5$ ) for the Her2 positive subgroup of the Her2-enriched subtype. 2 of these proteins were also identified in a cluster containing 17 highly interconnected proteins, AHSG, APRT, ATP5B, ATP5D, CALML3, CANX, CS, DHX15, KPNA2, NTMT1, PRKCSH, PYGM, RPL19, RPL32, SLC25A3, SLC25A6 and SLC3A2 identified by constructing a biological network with strong correlations ( $r\text{-value} > 0.5$ ) as edges or interconnecting lines in Cytoscape. This included 1 protein, PYGM observed in 3 KEGG pathways but lacked STRING interactions of high confidence score ( $\geq 0.90$ ).

The lack of any plausible proteins identified prompted an evaluation of the TNM stages of the CPTAC LC-MS/MS patient group. 7 proteins, ERBB2, CS, HIST1H1D, MDH2, PPP6R2, RHOC and SSBP1, were found to have statistically significant ( $p\text{-value} < 0.05$ ) and very strong correlations with CPNE3 ( $r\text{-value} > 0.9$ ) for the Her2 positive Stage IIA subgroup of the Her2-enriched subtype. All were identified in a cluster containing 20 highly interconnected proteins, ATP5D, CANX, CS, EIF5A, ERBB2, ETFB, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H4A, KPNA2, LRPPRC, MDH2, PPP6R2, RHOC, SLC25A3, SLC25A6, SLC3A2 SSBP1 and SSRP1 identified by constructing a biological network with strong correlations ( $r\text{-value} > 0.5$ ) as edges or interconnecting lines in Cytoscape. This included 2 proteins, CS and MDH2, observed in several KEGG and Reactome pathways and confirmed to have STRING interactions of high confidence score ( $\geq 0.90$ ). HER2 positive Stage IIB and IIIA subgroups of the Her2-enriched subtype had 2 proteins, CS and HIST1H4A, that demonstrated statistically significant ( $p\text{-value} < 0.05$ ) and very strong correlations with CPNE3 ( $r\text{-value} > 0.9$ ). Both proteins were identified in a cluster containing 4 highly interconnected proteins, CPNE3, CS, HIST1H4 and LRPPRC, identified by constructing a biological network with strong correlations ( $r\text{-value} > 0.5$ ) as edges or interconnecting lines in Cytoscape. This included 2 proteins, CS and HIST1H4A, observed in all KEGG and Reactome pathways and confirmed to have STRING

interactions of high confidence score ( $\geq 0.90$ ). Taken together the present data found 1 protein HIST1H4A to be either singly or doubly phosphorylated with a total of 1 putative site of phosphorylation identified for ERBB2/Her2 overexpressing vs Her2-enriched patient phosphosite enrichment data. Amino acid residue S47 of HIST1H4A was found to be 3.87-fold singly phosphorylated and has been identified as a target site for the activity of in vitro and in vivo kinases.

HIST1H4A is known to be phosphorylated on S47 in vivo and in vitro in a PAK2-dependent manner. Furthermore, PAK2 is able to phosphorylate single H4 or the H3-H4 tetramer but not nucleosomal H4. The phosphorylation of H4 at S47 has not been directly associated with transcription regulation. However, it is known to be preferentially enriched in H3.3-containing nucleosomes that are typically deposited into chromatin in a transcription-dependant manner and this demonstrates a novel mechanism by which a histone PTM can influence the incorporation of histone variants into chromatin resulting in the specialisation of chromatin domains (Kang et al., 2011). Notably, PAK2 does not localise to the nucleus upon stimulation of ERBB2/Her2 overexpressing cells (Arias-Romero et al., 2013). A recent study that investigated E-cadherin mechanosignaling pathways has demonstrated that PAK2 is required for glucose uptake and reinforcement of the actin cytoskeleton. This increase in mechanosignaling stimulates glucose uptake into the cell which drives OXPHOS to generate ATP and provide enough energy to support the actin cytoskeletal rearrangements necessary to resist the forces applied to the cell (Campbell et al., 2019). Nevertheless, the present study can only link CPNE3 knockdown to an upregulation of H4 and a 3.87-fold higher rate of S47 phosphorylation in ERBB2/Her2 overexpressing cells. It remains unknown what the mechanistic implications of the H4 upregulation in ERBB2/Her2 overexpressing cells could be. However, since S47 is a downstream target of PAK2, the mechanistic role might involve glucose uptake and actin remodelling. Therefore, linking CPNE3 expression to the regulation of several proteins such as CANX, CS and PYGM that are downstream effectors of OXPHOS, and HIST1H4A, a downstream target of OXPHOS implicated phosphosignalling.

#### 5.4. Combining multiple biomarkers in logistic regression models

Many potential prognostic and diagnostic biomarkers have been proposed to tackle the problem of early detection of cancer and prediction of patient outcomes. However, due to the heterogeneity of proteomic datasets generated by different platforms or laboratories, the biomarkers inferred from one single patient often lack reproducibility and the intersection of the inferences is small. This motivated us to evaluate the diagnostic potential of proteins identified as unique to the HER2-positive subtype and correlating to CPNE3 expression by statistical inference. We applied two logistic regression models characterised by 6 biomarker combinations on data from the two patient cohorts used for the biostatistical inference. The findings of this study suggest that some multiple biomarker combinations of CPNE3 and other proteins can perform with high sensitivity and specificity for the detection of HER2-positive breast cancer. This is particularly the case for the combination of CPNE3 and CS or CPNE3 and CACYBP. This could be caused by the fact that CPNE3 and either one of these proteins has a central role in HER2-positive breast cancer progression. The main influence on the performance outcomes of these biomarker combinations that perform as well as ERBB2/Her2 appears to favour proteins that were upregulated in response to CPNE3 knockdown in ERBB2 overexpressing HMLECs (Chapter 3).

A problem that was encountered was a class imbalance in the two datasets. For a binary classification model such as logistic regression, the class distribution will often reflect on the probabilities of certain predictions made by the model whether true positive, true negative, false positive or false negative. This in turn affects the sensitivity or true positive rate (TPR) and the specificity or true negative rate (TNR) which are calculated using the true positive, true negative, false positive and false negative values. This problem is greatly evident in the validation cohort which has  $n = 6$  sample size for cases vs  $n = 69$  for controls. Under such circumstances, a model that predicts all cases to be negative or controls, yields an accuracy of 92% and the classifier will tend to predict the majority class regardless of the variables or biomarkers used in a multinomial logistic regression. The high-performance scores observed for the CPNE3 and CS combination were right on the border of significance, as the observed accuracy values of 100% were high enough to demonstrate that the



classifier does not simply predict the majority class. This is further demonstrated by an acceptable Youden index and RMSE for the cross-validation.

## 5.5. Future prospects

The present study verified previously observed differences in cell adhesion-related gene expression in ERBB2 over-expressing HMLECs. Further studies should determine whether ERBB receptor triggering with growth factors would yield an upregulation of ITGA6 and ITGB4 following the siRNA-mediated knockdown of either CPNE3 or ERBB2. Moreover, the detachment and adherence of ERBB2 over-expressing clones to an adherent surface should further be evaluated following ERBB triggering with growth factors and siRNA-mediated knockdown of CPNE3. Real-time cell adhesion assays offer a high degree of sensitivity and a distinct advantage over endpoint methods such as MTT assays, which is essential for an assay of this kind. The rate of cell adhesion and spread for each of the siRNA treatment conditions evaluated during this study could be determined at various concentrations of growth factor treatment using real-time cell adhesion assays.

The present findings propose that CPNE3 expression regulates selected candidates with a role in glucose homeostasis in breast cancer and identified downstream interaction partners potentially involved in a mechanism of metabolic reprogramming in HER2-positive breast cancer. Future work should first validate the differential regulation of these potential downstream effectors by western blotting in the knockdown cell lysates. Correlations between potential interaction partners and CPNE3 expression should be evaluated by immunoblotting in a panel of HMLEC and breast cancer cell lines. Work should next focus on the mechanisms of CPNE3 regulation in ERBB2 over-expressing breast cancer cell lines and how this can elicit a metabolic reprogramming phenotype. This would involve the use of specific signalling (e.g. PI3K and MAPK), proteasomal and translational inhibitors (e.g. CDK1), with an assessment of mRNA and protein changes by quantitative real-time polymerase chain reaction (qRT-PCR) and immunoblotting, respectively. Integrin  $\alpha 6 \beta 4$  is known to amplify downstream pathways such as PI3K, AKT, MAPK, and the Rho family small GTPases by cooperating with growth factor receptors including EGFR, ERBB-2, and c-Met (Stewart & O'Connor, 2015). Since upregulation

of CPNE3 expression correlates with reduced expression of integrin  $\alpha 6\beta 4$  and upregulation of CANX (Chapter 3), this work would establish whether the siRNA-mediated knockdown of CPNE3 leads to amplified MAPK/ERK1 signalling particularly targeting CANX at S583 and a potential mechanosensing role involving integrin  $\alpha 6\beta 4$  amplification of MAPK/ERK1. The integrin chain  $\alpha 6$  has already been shown to associate with the chaperone CANX prior to integrin assembly (Lenter & Vestweber, 1994) and CPNE3 expression could regulate or serve as a marker for integrin assembly, particularly at the SIC.

## 5.6. Conclusions

In conclusion, the work presented in this thesis validated previously observed differences in integrin  $\alpha 6\beta 4$  expression between a parental control HMLEC and its ERBB2 over-expressing derivatives and established that acute knockdown of ERBB2 or CPNE3 does not reverse this expression pattern. Initially, end point cell adhesion assays were used to evaluate differences in cell adhesion due to differences in the expression of ERBB2 or CPNE3 in HMLECs. However, the present findings go further to demonstrate that end point cell adhesion assays are inadequate to evaluate a dynamic multistage process such as passive in vitro cell adhesion and implements real time cell adhesion assays to assess cell adhesion and cell spreading on an adherent surface. The work presented also combines two mass spectrometry based targeted proteomics methods, tandem mass tagging (TMT) LC-MS/MS on a Thermo Orbitrap LTQ and label-free time of flight liquid chromatography- ion mobility tandem mass spectrometry (TOF/LC-MS/IMS/MS) on a Waters SYNAPT G2, to quantify changes in global protein expression due to siRNA mediated knockdown of CPNE3 in ERBB2/HER2 overexpressing HMLECs. Furthermore, candidate biomarkers regulated by CPNE3 expression and likely participant in CPNE3 function were identified and their validity supported using statistical analysis of candidate biomarker expression in two breast cancer patient cohorts to select the most plausible candidates for functional enrichment analysis and phosphopeptide enrichment analysis.

Taken together, real-time cell adhesion or spreading assays, mass spectrometry based proteomic workflows, statistical analysis, network construction, causal and functional enrichment analysis have

revealed an interesting link between CPNE3 expression and subsets of genes which play an important role in metabolism of breast cancer cells and are downstream effectors or targets of OXPHOS. Mass spectrometry based proteomic studies of ERBB2 overexpressing HMLECs, and breast tumour lines have already provided a wealth of information on the association between ERBB2 overexpression, CPNE3 expression and genes related to cell adhesion. The results presented here provide more insight into the specific role of CPNE3 and link the previously known gene expression in ERBB2 overexpressing HMLECs and breast tumour lines to an adaptive mechanosensing related metabolic reprogramming and suggest a role for CPNE3 in glucose homeostasis of breast cancer.

## Chapter 6

### 6. References

- Adra, C. N., Ko, J., Leonard, D., Wirth, L. J., Cerione, R. A., & Lim, B. (1993). Identification of a novel protein with GDP dissociation inhibitor activity for the ras-like proteins CDC42Hs and rac 1. *Genes, Chromosomes & Cancer*, 8(4), 253–261. <https://doi.org/10.1002/gcc.2870080408>
- Alam, A., Taye, N., Patel, S., Thube, M., Mullick, J., Shah, V. K., Pant, R., Roychowdhury, T., Banerjee, N., Chatterjee, S., Bhattacharya, R., Roy, R., Mukhopadhyay, A., Mogare, D., & Chattopadhyay, S. (2019). SMAR1 favors immunosurveillance of cancer cells by modulating calnexin and MHC I expression. *Neoplasia (New York, N.Y.)*, 21(10), 945–962. <https://doi.org/10.1016/j.neo.2019.07.002>
- Alfarouk, K. O. (2016). Tumor metabolism, cancer cell transporters, and microenvironmental resistance. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 31(6), 859–866. <https://doi.org/10.3109/14756366.2016.1140753>
- Alonso, A., Sasin, J., Bottini, N., Friedberg, I., Friedberg, I., Osterman, A., Godzik, A., Hunter, T., Dixon, J., & Mustelin, T. (2004). Protein tyrosine phosphatases in the human genome. *Cell*, 117(6), 699–711. <https://doi.org/10.1016/j.cell.2004.05.018>
- Altamir, M. A., Goo, L. E., Little, A. C., Yates, J. A., Cheriyan, H. G., Wu, Z. F., & Merajver, S. D. (2019). Breast cancers utilize hypoxic glycogen stores via PYGB, the brain isoform of glycogen phosphorylase, to promote metastatic phenotypes. *PloS One*, 14(9), e0220973. <https://doi.org/10.1371/journal.pone.0220973>
- Anders, C., & Carey, L. A. (2008). Understanding and treating triple-negative breast cancer. *Oncology (Williston Park, N.Y.)*, 22(11), 1233–1239; discussion 1239-40, 1243.
- Anderson, N. M., & Simon, M. C. (2020). The tumor microenvironment. *Current Biology*, 30(16), R921–R925. <https://doi.org/10.1016/j.cub.2020.06.081>
- Andrechek, E. R., White, D., & Muller, W. J. (2005). Targeted disruption of ErbB2/Neu in the mammary epithelium results in impaired ductal outgrowth. *Oncogene*, 24(5), 932–937. <https://doi.org/10.1038/sj.onc.1208230>
- Aragay, A. M., Diaz, P., & Daban, J. R. (1988). Association of nucleosome core particle DNA with different histone oligomers. Transfer of histones between DNA-(H2A,H2B) and DNA-(H3,H4) complexes. *Journal of Molecular Biology*, 204(1), 141–154. [https://doi.org/10.1016/0022-2836\(88\)90605-5](https://doi.org/10.1016/0022-2836(88)90605-5)
- Arias-Romero, L. E., Villamar-Cruz, O., Huang, M., Hoeflich, K. P., & Chernoff, J. (2013). Pak1 kinase links ErbB2 to  $\beta$ -catenin in transformation of breast epithelial cells. *Cancer Research*, 73(12), 3671–3682. <https://doi.org/10.1158/0008-5472.CAN-12-4453>

- Arnold, M., Cavalcanti-Adam, E. A., Glass, R., Blümmel, J., Eck, W., Kantlehner, M., Kessler, H., & Spatz, J. P. (2004). Activation of Integrin Function by Nanopatterned Adhesive Interfaces. *ChemPhysChem*, 5(3), 383–388. <https://doi.org/10.1002/cphc.200301014>
- Aroian, R. V., Koga, M., Mendel, J. E., Ohshima, Y., & Sternberg, P. W. (1990). The let-23 gene necessary for *Caenorhabditis elegans* vulval induction encodes a tyrosine kinase of the EGF receptor subfamily. *Nature*, 348(6303), 693–699. <https://doi.org/10.1038/348693a0>
- Attwood, K., Fleyshman, D., Prendergast, L., Paszkiewicz, G., Omilian, A. R., Bshara, W., & Gurova, K. (2017). Prognostic value of histone chaperone FACT subunits expression in breast cancer. *Breast Cancer (Dove Medical Press)*, 9, 301–311. <https://doi.org/10.2147/BCTT.S126390>
- Bachelder, R. E., Ribick, M. J., Marchetti, A., Falcioni, R., Soddu, S., Davis, K. R., & Mercurio, A. M. (1999). p53 inhibits alpha 6 beta 4 integrin survival signaling by promoting the caspase 3-dependent cleavage of AKT/PKB. *The Journal of Cell Biology*, 147(5), 1063–1072. <https://doi.org/10.1083/jcb.147.5.1063>
- Baghban, R., Roshangar, L., Jahanban-Esfahlan, R., Seidi, K., Ebrahimi-Kalan, A., Jaymand, M., Kolahian, S., Javaheri, T., & Zare, P. (2020). Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Communication and Signaling*, 18(1), 59. <https://doi.org/10.1186/s12964-020-0530-4>
- Baker, P. R. S., Armando, A. M., Campbell, J. L., Quehenberger, O., & Dennis, E. A. (2014). Three-dimensional enhanced lipidomics analysis combining UPLC, differential ion mobility spectrometry, and mass spectrometric separation strategies. *Journal of Lipid Research*, 55(11), 2432–2442. <https://doi.org/10.1194/jlr.D051581>
- Ball, L. E., Agana, B. A., Comte-Walters, S., Bethard, J. R., & Burnette, B. B. (2023). An Introduction to Mass Spectrometry-Based Proteomics. In *Encyclopedia of Cell Biology* (pp. 132–140). Elsevier. <https://doi.org/10.1016/B978-0-12-821618-7.00143-7>
- Bastien, R. R., Rodríguez-Lescure, Á., Ebbert, M. T., Prat, A., Munárriz, B., Rowe, L., Miller, P., Ruiz-Borrego, M., Anderson, D., Lyons, B., Álvarez, I., Dowell, T., Wall, D., Seguí, M. Á., Barley, L., Boucher, K. M., Alba, E., Pappas, L., Davis, C. A., ... Martín, M. (2012). PAM50 Breast Cancer Subtyping by RT-qPCR and Concordance with Standard Clinical Molecular Markers. *BMC Medical Genomics*, 5(1), 44. <https://doi.org/10.1186/1755-8794-5-44>
- Baum, B., & Georgiou, M. (2011). Dynamics of adherens junctions in epithelial establishment, maintenance, and remodeling. *Journal of Cell Biology*, 192(6), 907–917. <https://doi.org/10.1083/jcb.201009141>
- Bazley, L. A., & Gullick, W. J. (2005). The epidermal growth factor receptor family. *Endocrine-Related Cancer*, 12 Suppl 1, S17-27. <https://doi.org/10.1677/erc.1.01032>
- Bellard, M., Oudet, P., Germond, J. E., & Chambon, P. (1976). Subunit structure of simian-virus-40 minichromosome. *European Journal of Biochemistry*, 70(2), 543–553. <https://doi.org/10.1111/j.1432-1033.1976.tb11046.x>
- Bertani, M. F. (2005). *Analysis of ErbB2 and growth factor dependent gene expression in breast cancer*.

- Blethrow, J. D., Glavy, J. S., Morgan, D. O., & Shokat, K. M. (2008). Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proceedings of the National Academy of Sciences of the United States of America*, 105(5), 1442–1447. <https://doi.org/10.1073/pnas.0708966105>
- Borradori, L., & Sonnenberg, A. (1999). Structure and function of hemidesmosomes: more than simple adhesion complexes. *The Journal of Investigative Dermatology*, 112(4), 411–418. <https://doi.org/10.1046/j.1523-1747.1999.00546.x>
- Bouyain, S., Longo, P. A., Li, S., Ferguson, K. M., & Leahy, D. J. (2005). The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand. *Proceedings of the National Academy of Sciences of the United States of America*, 102(42), 15024–15029. <https://doi.org/10.1073/pnas.0507591102>
- Brabletz, T., Jung, A., Reu, S., Porzner, M., Hlubek, F., Kunz-Schughart, L. A., Knuechel, R., & Kirchner, T. (2001). Variable  $\beta$ -catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proceedings of the National Academy of Sciences*, 98(18), 10356–10361. <https://doi.org/10.1073/pnas.171610498>
- Britt, H. M., Cragolini, T., Khatun, S., Hatimy, A., James, J., Page, N., Williams, J. P., Hughes, C., Denny, R., Thalassinou, K., & Vissers, J. P. C. (2022). Evaluation of acquisition modes for semi-quantitative analysis by targeted and untargeted mass spectrometry. *Rapid Communications in Mass Spectrometry*, 36(13). <https://doi.org/10.1002/rcm.9308>
- Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., & Reiter, L. (2015). Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Molecular & Cellular Proteomics*, 14(5), 1400–1410. <https://doi.org/10.1074/mcp.M114.044305>
- Burnett, G., & Kennedy, E. P. (1954). The enzymatic phosphorylation of proteins. *The Journal of Biological Chemistry*, 211(2), 969–980.
- Cai, Q., Zhao, M., Liu, X., Wang, X., Nie, Y., Li, P., Liu, T., Ge, R., & Han, F. (2017). Reduced expression of citrate synthase leads to excessive superoxide formation and cell apoptosis. *Biochemical and Biophysical Research Communications*, 485(2), 388–394. <https://doi.org/10.1016/j.bbrc.2017.02.067>
- Calvo, E., Camafeita, E., Fernández-Gutiérrez, B., & López, J. A. (2011). Applying selected reaction monitoring to targeted proteomics. *Expert Review of Proteomics*, 8(2), 165–173. <https://doi.org/10.1586/epr.11.11>
- Campbell, H. K., Salvi, A. M., O'Brien, T., Superfine, R., & DeMali, K. A. (2019). PAK2 links cell survival to mechanotransduction and metabolism. *The Journal of Cell Biology*, 218(6), 1958–1971. <https://doi.org/10.1083/jcb.201807152>
- Carmichael, A. R., & Bates, T. (2004). Obesity and breast cancer: a review of the literature. *Breast (Edinburgh, Scotland)*, 13(2), 85–92. <https://doi.org/10.1016/j.breast.2003.03.001>

- Carmona, F. J., Montemurro, F., Kannan, S., Rossi, V., Verma, C., Baselga, J., & Scaltriti, M. (2016). AKT signaling in ERBB2-amplified breast cancer. *Pharmacology & Therapeutics*, 158, 63–70. <https://doi.org/10.1016/j.pharmthera.2015.11.013>
- Carraway, K. L., & Sweeney, C. (2006). Co-opted integrin signaling in ErbB2-induced mammary tumor progression. *Cancer Cell*, 10(2), 93–95. <https://doi.org/10.1016/j.ccr.2006.07.015>
- Caudell, E. G., Caudell, J. J., Tang, C. H., Yu, T. K., Frederick, M. J., & Grimm, E. A. (2000). Characterization of human copine III as a phosphoprotein with associated kinase activity. *Biochemistry*, 39(42), 13034–13043. <https://doi.org/10.1021/bi001250v>
- Chakraborty, A., Chen, W., & Gebler, J. (2009). *Analysis of PEGs by ESI-IMS-TOF MS Coupled with Ion Molecule Reactions*. <https://www.waters.com/webassets/cms/library/docs/720002907en.pdf>
- Chao, C., Lotz, M. M., Clarke, A. C., & Mercurio, A. M. (1996). A function for the integrin  $\alpha 6 \beta 4$  in the invasive properties of colorectal carcinoma cells. *Cancer Research*, 56(20), 4811–4819.
- Chen, W. Y. (2008). Exogenous and endogenous hormones and breast cancer. *Best Practice & Research. Clinical Endocrinology & Metabolism*, 22(4), 573–585. <https://doi.org/10.1016/j.beem.2008.08.001>
- Chi, F., Wu, R., Jin, X., Jiang, M., & Zhu, X. (2016). HER2 induces cell proliferation and invasion of non-small-cell lung cancer by upregulating COX-2 expression via MEK/ERK signaling pathway. *OncoTargets and Therapy*, 9, 2709–2716. <https://doi.org/10.2147/OTT.S96197>
- Cho, H.-S., & Leahy, D. J. (2002). Structure of the extracellular region of HER3 reveals an interdomain tether. *Science (New York, N. Y.)*, 297(5585), 1330–1333. <https://doi.org/10.1126/science.1074611>
- Cho, H.-S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W., & Leahy, D. J. (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature*, 421(6924), 756–760. <https://doi.org/10.1038/nature01392>
- Choi, H. Y., Park, N., Na, J. B., Ko, E. S., Park, J.-Y., & Yoo, J. C. (2016). Direct binding of Copine3 with Jab1 activates downstream ErbB2 signaling and motility in SKBr3 breast cancer cells. *Oncology Reports*, 35(2), 1147–1152. <https://doi.org/10.3892/or.2015.4472>
- Chook, Y. M., & Blobel, G. (2001). Karyopherins and nuclear import. *Current Opinion in Structural Biology*, 11(6), 703–715. [https://doi.org/10.1016/s0959-440x\(01\)00264-0](https://doi.org/10.1016/s0959-440x(01)00264-0)
- Claude, E., Towers, M., Neeson, K., & Vissers, J. (2013). MALDI Imaging HDMSE: A Novel Data Independent Technique for the Visualisation and Identification of Lipids Directly from a Single Tissue Section. *Journal of Biomolecular Techniques: JBT*, 24((Suppl)), S53–S53.
- Colangelo, D., Guo, H. Y., Connors, K. M., Silvestro, L., & Hoffman, R. M. (1992). Noncolorimetric measurement of cell activity in three-dimensional histoculture using the tetrazolium dye 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide: the pixel image analysis of formazan crystals. *Analytical Biochemistry*, 205(1), 8–13. [https://doi.org/10.1016/0003-2697\(92\)90571-n](https://doi.org/10.1016/0003-2697(92)90571-n)
- Colognato, H., Baron, W., Avellana-Adalid, V., Relvas, J. B., Baron-Van Evercooren, A., Georges-Labouesse, E., & French-Constant, C. (2002). CNS integrins switch growth factor signalling to

- promote target-dependent survival. *Nature Cell Biology*, 4(11), 833–841.  
<https://doi.org/10.1038/ncb865>
- Cox E, & Bonner J T. (2001). The Advantages of Togetherness. *Science*, 292(5516), 448.
- Creutz, C. E., Tomsig, J. L., Snyder, S. L., Gautier, M. C., Skouri, F., Beisson, J., & Cohen, J. (1998). The copines, a novel class of C2 domain-containing, calcium-dependent, phospholipid-binding proteins conserved from Paramecium to humans. *The Journal of Biological Chemistry*, 273(3), 1393–1402. <https://doi.org/10.1074/jbc.273.3.1393>
- Cui, X., Wang, H., Wu, X., Huo, K., & Jing, X. (2021). Increased expression of KPNA2 predicts unfavorable prognosis in ovarian cancer patients, possibly by targeting KIF4A signaling. *Journal of Ovarian Research*, 14(1), 71. <https://doi.org/10.1186/s13048-021-00818-9>
- Daemen, A., & Manning, G. (2018). HER2 is not a cancer subtype but rather a pan-cancer event and is highly enriched in AR-driven breast tumors. *Breast Cancer Research*, 20(1), 8.  
<https://doi.org/10.1186/s13058-018-0933-y>
- Darcy, K. M., Wohlhueter, A. L., Zangani, D., Vaughan, M. M., Russell, J. A., Masso-Welch, P. A., Varela, L. M., Shoemaker, S. F., Horn, E., Lee, P. P., Huang, R. Y., & Ip, M. M. (1999). Selective changes in EGF receptor expression and function during the proliferation, differentiation and apoptosis of mammary epithelial cells. *European Journal of Cell Biology*, 78(7), 511–523.  
[https://doi.org/10.1016/S0171-9335\(99\)80077-6](https://doi.org/10.1016/S0171-9335(99)80077-6)
- Darcy, K. M., Zangani, D., Wohlhueter, A. L., Huang, R. Y., Vaughan, M. M., Russell, J. A., & Ip, M. M. (2000). Changes in ErbB2 (her-2/neu), ErbB3, and ErbB4 during growth, differentiation, and apoptosis of normal rat mammary epithelial cells. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society*, 48(1), 63–80.  
<https://doi.org/10.1177/002215540004800107>
- De Arcangelis, A., Mark, M., Kreidberg, J., Sorokin, L., & Georges-Labouesse, E. (1999). Synergistic activities of alpha3 and alpha6 integrins are required during apical ectodermal ridge formation and organogenesis in the mouse. *Development (Cambridge, England)*, 126(17), 3957–3968.  
<https://doi.org/10.1242/dev.126.17.3957>
- de Hoog, C. L., Foster, L. J., & Mann, M. (2004). RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers. *Cell*, 117(5), 649–662.  
[https://doi.org/10.1016/s0092-8674\(04\)00456-8](https://doi.org/10.1016/s0092-8674(04)00456-8)
- Diaz, G., Melis, M., Musin, A., Piludu, M., Piras, M., & Falchi, A. M. (n.d.). Localization of MTT formazan in lipid droplets. An alternative hypothesis about the nature of formazan granules and aggregates. *European Journal of Histochemistry: EJH*, 51(3), 213–218.
- Diaz, L. K., Cristofanilli, M., Zhou, X., Welch, K. L., Smith, T. L., Yang, Y., Sneige, N., Sahin, A. A., & Gilcrease, M. Z. (2005). Beta4 integrin subunit gene expression correlates with tumor size and nuclear grade in early breast cancer. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, 18(9), 1165–1175.  
<https://doi.org/10.1038/modpathol.3800411>



- Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., & Alice, M. B. (1968). Molecular Beams of Macroions. *The Journal of Chemical Physics*, 49(5), 2240–2249.  
<https://doi.org/10.1063/1.1670391>
- Donato, R. (2001). S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *The International Journal of Biochemistry & Cell Biology*, 33(7), 637–668. [https://doi.org/10.1016/S1357-2725\(01\)00046-2](https://doi.org/10.1016/S1357-2725(01)00046-2)
- Duan, M., Hu, F., Li, D., Wu, S., & Peng, N. (2020). Silencing KPNA2 inhibits IL-6-induced breast cancer exacerbation by blocking NF- $\kappa$ B signaling and c-Myc nuclear translocation in vitro. *Life Sciences*, 253, 117736. <https://doi.org/10.1016/j.lfs.2020.117736>
- Durán, M. C., Vega, F., Moreno-Bueno, G., Artiga, M. J., Sanchez, L., Palacios, J., Ridley, A., & Timms, J. F. (2008). Characterisation of tumoral markers correlated with ErbB2 (HER2/Neu) overexpression and metastasis in breast cancer. *PROTEOMICS - CLINICAL APPLICATIONS*, 2(9), 1313–1326. <https://doi.org/10.1002/prca.200780020>
- Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., MacLean, B. X., Ting, Y. S., Canterbury, J. D., Marsh, D. M., Kellmann, M., Zabrouskov, V., Wu, C. C., & MacCoss, M. J. (2013). Multiplexed MS/MS for improved data-independent acquisition. *Nature Methods*, 10(8), 744–746. <https://doi.org/10.1038/nmeth.2528>
- Ellis, M. J., Gillette, M., Carr, S. A., Paulovich, A. G., Smith, R. D., Rodland, K. K., Townsend, R. R., Kinsinger, C., Mesri, M., Rodriguez, H., & Liebler, D. C. (2013). Connecting Genomic Alterations to Cancer Biology with Proteomics: The NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discovery*, 3(10), 1108–1112. <https://doi.org/10.1158/2159-8290.CD-13-0219>
- Elmore, J. G., Armstrong, K., Lehman, C. D., & Fletcher, S. W. (2005). Screening for breast cancer. *JAMA*, 293(10), 1245–1256. <https://doi.org/10.1001/jama.293.10.1245>
- Fan, J., Cai, B., Zeng, M., Hao, Y., Giancotti, F. G., & Fu, B. M. (2011). Integrin  $\beta$ 4 signaling promotes mammary tumor cell adhesion to brain microvascular endothelium by inducing ErbB2-mediated secretion of VEGF. *Annals of Biomedical Engineering*, 39(8), 2223–2241.  
<https://doi.org/10.1007/s10439-011-0321-6>
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N. Y.)*, 246(4926), 64–71.  
<https://doi.org/10.1126/science.2675315>
- Ferguson, K. M. (2008). Structure-based view of epidermal growth factor receptor regulation. *Annual Review of Biophysics*, 37, 353–373. <https://doi.org/10.1146/annurev.biophys.37.032807.125829>
- Ferguson, K. M., Berger, M. B., Mendrola, J. M., Cho, H. S., Leahy, D. J., & Lemmon, M. A. (2003). EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Molecular Cell*, 11(2), 507–517. [https://doi.org/10.1016/s1097-2765\(03\)00047-9](https://doi.org/10.1016/s1097-2765(03)00047-9)
- Frame, M. C. (2004). Newest findings on the oldest oncogene; how activated src does it. *Journal of Cell Science*, 117(Pt 7), 989–998. <https://doi.org/10.1242/jcs.01111>
- Freedman, V. (1974). Cellular tumorigenicity in nude mice: Correlation with cell growth in semi-solid medium. *Cell*, 3(4), 355–359. [https://doi.org/10.1016/0092-8674\(74\)90050-6](https://doi.org/10.1016/0092-8674(74)90050-6)

- FRIEDMAN, D. L., & LARNER, J. (1963). STUDIES ON UDPG-ALPHA-GLUCAN TRANSGLUCOSYLASE. III. INTERCONVERSION OF TWO FORMS OF MUSCLE UDPG-ALPHA-GLUCAN TRANSGLUCOSYLASE BY A PHOSPHORYLATION-DEPHOSPHORYLATION REACTION SEQUENCE. *Biochemistry*, 2, 669–675.  
<https://doi.org/10.1021/bi00904a009>
- Friedrichs, K., Ruiz, P., Franke, F., Gille, I., Terpe, H. J., & Imhof, B. A. (1995). High expression level of alpha 6 integrin in human breast carcinoma is correlated with reduced survival. *Cancer Research*, 55(4), 901–906.
- Gallien, S., Duriez, E., & Domon, B. (2011). Selected reaction monitoring applied to proteomics. *Journal of Mass Spectrometry*, 46(3), 298–312. <https://doi.org/10.1002/jms.1895>
- Galvão, F. C., Rossi, D., Silveira, W. da S., Valentini, S. R., & Zanelli, C. F. (2013). The deoxyhypusine synthase mutant dys1-1 reveals the association of eIF5A and Asc1 with cell wall integrity. *PloS One*, 8(4), e60140. <https://doi.org/10.1371/journal.pone.0060140>
- Garcia, H., Miecznikowski, J. C., Safina, A., Commane, M., Ruusulehto, A., Kilpinen, S., Leach, R. W., Attwood, K., Li, Y., Degan, S., Omilian, A. R., Guryanova, O., Papantonopoulou, O., Wang, J., Buck, M., Liu, S., Morrison, C., & Gurova, K. V. (2013). Facilitates Chromatin Transcription Complex Is an “Accelerator” of Tumor Transformation and Potential Marker and Target of Aggressive Cancers. *Cell Reports*, 4(1), 159–173. <https://doi.org/10.1016/j.celrep.2013.06.013>
- Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Kofler, M., Jorissen, R. N., Nice, E. C., Burgess, A. W., & Ward, C. W. (2003). The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. *Molecular Cell*, 11(2), 495–505. [https://doi.org/10.1016/s1097-2765\(03\)00048-0](https://doi.org/10.1016/s1097-2765(03)00048-0)
- Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H.-J., Walker, F., Frenkel, M. J., Hoyne, P. A., Jorissen, R. N., Nice, E. C., Burgess, A. W., & Ward, C. W. (2002). Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell*, 110(6), 763–773. [https://doi.org/10.1016/s0092-8674\(02\)00940-6](https://doi.org/10.1016/s0092-8674(02)00940-6)
- Geiger, T., Cox, J., & Mann, M. (2010). Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-ion Fragmentation. *Molecular & Cellular Proteomics*, 9(10), 2252–2261.  
<https://doi.org/10.1074/mcp.M110.001537>
- Georges-Labouesse, E., Mark, M., Messaddeq, N., & Gansmüller, A. (1998). Essential role of alpha 6 integrins in cortical and retinal lamination. *Current Biology: CB*, 8(17), 983–986.  
[https://doi.org/10.1016/s0960-9822\(98\)70402-6](https://doi.org/10.1016/s0960-9822(98)70402-6)
- Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M. J., Cramer, R., Waterfield, M. D., & Timms, J. F. (2002). Evaluation of Two-dimensional Differential Gel Electrophoresis for Proteomic Expression Analysis of a Model Breast Cancer Cell System. *Molecular & Cellular Proteomics*, 1(2), 91–98.  
<https://doi.org/10.1074/mcp.T100007-MCP200>

- Ghasemi, M., Turnbull, T., Sebastian, S., & Kempson, I. (2021). The MTT Assay: Utility, Limitations, Pitfalls, and Interpretation in Bulk and Single-Cell Analysis. *International Journal of Molecular Sciences*, 22(23), 12827. <https://doi.org/10.3390/ijms222312827>
- Ghazali, R., Mehta, K. J., Bligh, S. A., Tewfik, I., Clemens, D., & Patel, V. B. (2020). High omega arachidonic acid/docosahexaenoic acid ratio induces mitochondrial dysfunction and altered lipid metabolism in human hepatoma cells. *World Journal of Hepatology*, 12(3), 84–98. <https://doi.org/10.4254/wjh.v12.i3.84>
- Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., & Aebersold, R. (2012). Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics*, 11(6), O111.016717. <https://doi.org/10.1074/mcp.O111.016717>
- Gil-Solsona, R., Sancho, J. V., Gassner, A., Weyermann, C., Hernández, F., Delémont, O., & Bijlsma, L. (2021). Use of ion mobility-high resolution mass spectrometry in metabolomics studies to provide near MS/MS quality data in a single injection. *Journal of Mass Spectrometry*, 56(5). <https://doi.org/10.1002/jms.4718>
- Git, A., Spiteri, I., Blenkiron, C., Dunning, M. J., Pole, J. C., Chin, S.-F., Wang, Y., Smith, J., Livesey, F. J., & Caldas, C. (2008). PMC42, a breast progenitor cancer cell line, has normal-like mRNA and microRNA transcriptomes. *Breast Cancer Research*, 10(3), R54. <https://doi.org/10.1186/bcr2109>
- Glamann, J., & Hansen, A. J. (2006). Dynamic detection of natural killer cell-mediated cytotoxicity and cell adhesion by electrical impedance measurements. *Assay and Drug Development Technologies*, 4(5), 555–563. <https://doi.org/10.1089/adt.2006.4.555>
- Gokduman, K., & Gok, A. (2020). In Vitro Investigation of Therapeutic Potential of Bare Magnetite (Fe<sub>3</sub>O<sub>4</sub>) Nanoparticles (≤100 ppm) on Hepatocellular Carcinoma Cells. *Journal of Nanoscience and Nanotechnology*, 20(3), 1391–1400. <https://doi.org/10.1166/jnn.2020.17152>
- Goldfarb, D. S., Corbett, A. H., Mason, D. A., Harreman, M. T., & Adam, S. A. (2004). Importin alpha: a multipurpose nuclear-transport receptor. *Trends in Cell Biology*, 14(9), 505–514. <https://doi.org/10.1016/j.tcb.2004.07.016>
- Görlich, D. (1998). Transport into and out of the cell nucleus. *The EMBO Journal*, 17(10), 2721–2727. <https://doi.org/10.1093/emboj/17.10.2721>
- Graus-Porta, D., Beerli, R. R., Daly, J. M., & Hynes, N. E. (1997). ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *The EMBO Journal*, 16(7), 1647–1655. <https://doi.org/10.1093/emboj/16.7.1647>
- Griffin, T. J., Lock, C. M., Li, X., Patel, A., Chervetsova, I., Lee, H., Wright, M. E., Ranish, J. A., Chen, S. S., & Aebersold, R. (2003). Abundance ratio-dependent proteomic analysis by mass spectrometry. *Analytical Chemistry*, 75(4), 867–874. <https://doi.org/10.1021/ac026127j>
- Gschwind, A., Fischer, O. M., & Ullrich, A. (2004). The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nature Reviews. Cancer*, 4(5), 361–370. <https://doi.org/10.1038/nrc1360>

- Gupta, G. P., & Massagué, J. (2006). Cancer Metastasis: Building a Framework. *Cell*, 127(4), 679–695. <https://doi.org/10.1016/j.cell.2006.11.001>
- Gupta, P., & Srivastava, S. K. (2014). Inhibition of Integrin-HER2 signaling by Cucurbitacin B leads to in vitro and in vivo breast tumor growth suppression. *Oncotarget*, 5(7), 1812–1828. <https://doi.org/10.18632/oncotarget.1743>
- Gurova, K. V., Garcia, H., Miecznikowski, J., Omilian, A. R., & Morrison, C. (2013). Level of SSRP1 in Cancer as a Prognostic Marker of Aggressive Disease. *American Journal of Clinical Pathology*, 140(suppl 1), A152–A152. <https://doi.org/10.1093/ajcp/140.suppl1.152>
- Gutiérrez, T., Qi, H., Yap, M. C., Tahbaz, N., Milburn, L. A., Lucchinetti, E., Lou, P.-H., Zaugg, M., LaPointe, P. G., Mercier, P., Overduin, M., Bischof, H., Burgstaller, S., Malli, R., Ballanyi, K., Shuai, J., & Simmen, T. (2020). The ER chaperone calnexin controls mitochondrial positioning and respiration. *Science Signaling*, 13(638). <https://doi.org/10.1126/scisignal.aax6660>
- Guy, P. M., Platko, J. V., Cantley, L. C., Cerione, R. A., & Carraway, K. L. (1994). Insect cell-expressed p180erbB3 possesses an impaired tyrosine kinase activity. *Proceedings of the National Academy of Sciences of the United States of America*, 91(17), 8132–8136. <https://doi.org/10.1073/pnas.91.17.8132>
- Gygi, S. P., & Aebersold, R. (2000). Mass spectrometry and proteomics. *Current Opinion in Chemical Biology*, 4(5), 489–494. [https://doi.org/10.1016/s1367-5931\(00\)00121-6](https://doi.org/10.1016/s1367-5931(00)00121-6)
- Hahn, W. C., & Weinberg, R. A. (2002). Modelling the molecular circuitry of cancer. *Nature Reviews Cancer*, 2(5), 331–341. <https://doi.org/10.1038/nrc795>
- Hamid, Z., Zimmerman, K. D., Guillen-Ahlers, H., Li, C., Nathanielsz, P., Cox, L. A., & Olivier, M. (2022). Assessment of label-free quantification and missing value imputation for proteomics in non-human primates. *BMC Genomics*, 23(1), 496. <https://doi.org/10.1186/s12864-022-08723-1>
- Harris, R. A., Eichholtz, T. J., Hiles, I. D., Page, M. J., & O'Hare, M. J. (1999). New model of ErbB-2 over-expression in human mammary luminal epithelial cells. *International Journal of Cancer*, 80(3), 477–484. [https://doi.org/10.1002/\(sici\)1097-0215\(19990129\)80:3<477::aid-ijc23>3.0.co;2-w](https://doi.org/10.1002/(sici)1097-0215(19990129)80:3<477::aid-ijc23>3.0.co;2-w)
- Harris, R. C., Chung, E., & Coffey, R. J. (2003). EGF receptor ligands. *Experimental Cell Research*, 284(1), 2–13. [https://doi.org/10.1016/s0014-4827\(02\)00105-2](https://doi.org/10.1016/s0014-4827(02)00105-2)
- Hazan, R., Margolis, B., Dombalagian, M., Ullrich, A., Zilberstein, A., & Schlessinger, J. (1990). Identification of autophosphorylation sites of HER2/neu. *Cell Growth & Differentiation: The Molecular Biology Journal of the American Association for Cancer Research*, 1(1), 3–7.
- Heinrich, C., Keller, C., Boulay, A., Vecchi, M., Bianchi, M., Sack, R., Lienhard, S., Duss, S., Hofsteenge, J., & Hynes, N. E. (2010). Copine-III interacts with ErbB2 and promotes tumor cell migration. *Oncogene*, 29(11), 1598–1610. <https://doi.org/10.1038/onc.2009.456>
- Hemler, M. E., Crouse, C., & Sonnenberg, A. (1989). Association of the VLA alpha 6 subunit with a novel protein. A possible alternative to the common VLA beta 1 subunit on certain cell lines. *The Journal of Biological Chemistry*, 264(11), 6529–6535.

- Hill, R. J., & Sternberg, P. W. (1992). The gene *lin-3* encodes an inductive signal for vulval development in *C. elegans*. *Nature*, 358(6386), 470–476. <https://doi.org/10.1038/358470a0>
- Hines, K. M., & Xu, L. (2019). Lipidomic consequences of phospholipid synthesis defects in *Escherichia coli* revealed by HILIC-ion mobility-mass spectrometry. *Chemistry and Physics of Lipids*, 219, 15–22. <https://doi.org/10.1016/j.chemphyslip.2019.01.007>
- Hoffman, G. R., Nassar, N., & Cerione, R. A. (2000). Structure of the Rho Family GTP-Binding Protein Cdc42 in Complex with the Multifunctional Regulator RhoGDI. *Cell*, 100(3), 345–356. [https://doi.org/10.1016/S0092-8674\(00\)80670-4](https://doi.org/10.1016/S0092-8674(00)80670-4)
- Hofmann, W., Reichart, B., Ewald, A., Müller, E., Schmitt, I., Stauber, R. H., Lottspeich, F., Jockusch, B. M., Scheer, U., Hauber, J., & Dabauvalle, M. C. (2001). Cofactor requirements for nuclear export of Rev response element (RRE)- and constitutive transport element (CTE)-containing retroviral RNAs. An unexpected role for actin. *The Journal of Cell Biology*, 152(5), 895–910. <https://doi.org/10.1083/jcb.152.5.895>
- Hong, M., Kim, H., & Kim, I. (2014). Ribosomal protein L19 overexpression activates the unfolded protein response and sensitizes MCF7 breast cancer cells to endoplasmic reticulum stress-induced cell death. *Biochemical and Biophysical Research Communications*, 450(1), 673–678. <https://doi.org/10.1016/j.bbrc.2014.06.036>
- Hong, S., Ergezen, E., Lec, R., & Barbee, K. A. (2006). Real-time analysis of cell-surface adhesive interactions using thickness shear mode resonator. *Biomaterials*, 27(34), 5813–5820. <https://doi.org/10.1016/j.biomaterials.2006.07.031>
- Hoque, M., Park, J. Y., Chang, Y.-J., Luchessi, A. D., Cambiaghi, T. D., Shamanna, R., Hanauske-Abel, H. M., Holland, B., Pe'ery, T., Tian, B., & Mathews, M. B. (2017). Regulation of gene expression by translation factor eIF5A: Hypusine-modified eIF5A enhances nonsense-mediated mRNA decay in human cells. *Translation (Austin, Tex.)*, 5(2), e1366294. <https://doi.org/10.1080/21690731.2017.1366294>
- Hu, A., Noble, W. S., & Wolf-Yadlin, A. (2016). Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research*, 5, 419. <https://doi.org/10.12688/f1000research.7042.1>
- Huang, B., Luo, H., & Guo, X. B. (n.d.). High expression of KPNA2 promotes colorectal cancer development by activating the PI3K/AKT pathway. *Journal of Biological Regulators and Homeostatic Agents*, 34(6), 2317–2324. <https://doi.org/10.23812/20-581-L>
- Huang, W., Anvari, B., Torres, J. H., LeBaron, R. G., & Athanasiou, K. A. (2003). Temporal effects of cell adhesion on mechanical characteristics of the single chondrocyte. *Journal of Orthopaedic Research: Official Publication of the Orthopaedic Research Society*, 21(1), 88–95. [https://doi.org/10.1016/S0736-0266\(02\)00130-4](https://doi.org/10.1016/S0736-0266(02)00130-4)
- Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., & Hauer, C. R. (1986). Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 83(17), 6233–6237. <https://doi.org/10.1073/pnas.83.17.6233>

- Hutchinson-Bunch, C., Sanford, J. A., Hansen, J. R., Gritsenko, M. A., Rodland, K. D., Piehowski, P. D., Qian, W.-J., & Adkins, J. N. (2021). Assessment of TMT Labeling Efficiency in Large-Scale Quantitative Proteomics: The Critical Effect of Sample pH. *ACS Omega*, 6(19), 12660–12666. <https://doi.org/10.1021/acsomega.1c00776>
- Huttlin, E. L., Bruckner, R. J., Navarrete-Perea, J., Cannon, J. R., Baltier, K., Gebreab, F., Gygi, M. P., Thornock, A., Zarraga, G., Tam, S., Szpyt, J., Gassaway, B. M., Panov, A., Parzen, H., Fu, S., Golbazi, A., Maenpaa, E., Stricker, K., Guha Thakurta, S., ... Gygi, S. P. (2021). Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell*, 184(11), 3022–3040.e28. <https://doi.org/10.1016/j.cell.2021.04.011>
- Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M. P., Parzen, H., Szpyt, J., Tam, S., Zarraga, G., Pontano-Vaites, L., Swarup, S., White, A. E., Schweppe, D. K., Rad, R., Erickson, B. K., ... Harper, J. W. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature*, 545(7655), 505–509. <https://doi.org/10.1038/nature22366>
- Hynes, R. O. (1987). Integrins: a family of cell surface receptors. *Cell*, 48(4), 549–554. [https://doi.org/10.1016/0092-8674\(87\)90233-9](https://doi.org/10.1016/0092-8674(87)90233-9)
- Hynes, R. O. (1992). Integrins: versatility, modulation, and signaling in cell adhesion. *Cell*, 69(1), 11–25. [https://doi.org/10.1016/0092-8674\(92\)90115-s](https://doi.org/10.1016/0092-8674(92)90115-s)
- Ieguchi, K., Fujita, M., Ma, Z., Davari, P., Taniguchi, Y., Sekiguchi, K., Wang, B., Takada, Y. K., & Takada, Y. (2010). Direct binding of the EGF-like domain of neuregulin-1 to integrins ( $\alpha_3\beta_3$  and  $\alpha_6\beta_4$ ) is involved in neuregulin-1/ErbB signaling. *The Journal of Biological Chemistry*, 285(41), 31388–31398. <https://doi.org/10.1074/jbc.M110.113878>
- Iqbal, N., & Iqbal, N. (2014). Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. *Molecular Biology International*, 2014, 1–9. <https://doi.org/10.1155/2014/852748>
- Iribarne, J. V. (1976). On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics*, 64(6), 2287. <https://doi.org/10.1063/1.432536>
- Iselt, M., Holtei, W., & Hilgard, P. (1989). The tetrazolium dye assay for rapid in vitro assessment of cytotoxicity. *Arzneimittel-Forschung*, 39(7), 747–749.
- Jackson, J. G., St Clair, P., Sliwkowski, M. X., & Brattain, M. G. (2004). Blockade of epidermal growth factor- or heregulin-dependent ErbB2 activation with the anti-ErbB2 monoclonal antibody 2C4 has divergent downstream signaling and growth effects. *Cancer Research*, 64(7), 2601–2609. <https://doi.org/10.1158/0008-5472.can-03-3106>
- Jauliac, S., López-Rodríguez, C., Shaw, L. M., Brown, L. F., Rao, A., & Toker, A. (2002). The role of NFAT transcription factors in integrin-mediated carcinoma invasion. *Nature Cell Biology*, 4(7), 540–544. <https://doi.org/10.1038/ncb816>
- Jiao, H., Soejima, Y., Ohe, Y., Miura, K., Tamura, T., & Saijo, N. (1992). Differential macrophage-mediated cytotoxicity to P388 leukemia cells and its drug-resistant cells examined by a new MTT assay. *Leukemia Research*, 16(12), 1175–1180. [https://doi.org/10.1016/0145-2126\(92\)90115-N](https://doi.org/10.1016/0145-2126(92)90115-N)

- JOHNS, E. W., & BUTLER, J. A. (1962). Further fractionations of histones from calf thymus. *The Biochemical Journal*, 82, 15–18. <https://doi.org/10.1042/bj0820015>
- Jones, F. E., & Stern, D. F. (1999). Expression of dominant-negative ErbB2 in the mammary gland of transgenic mice reveals a role in lobuloalveolar development and lactation. *Oncogene*, 18(23), 3481–3490. <https://doi.org/10.1038/sj.onc.1202698>
- Jones, F. E., Welte, T., Fu, X. Y., & Stern, D. F. (1999). ErbB4 signaling in the mammary gland is required for lobuloalveolar development and Stat5 activation during lactation. *The Journal of Cell Biology*, 147(1), 77–88. <https://doi.org/10.1083/jcb.147.1.77>
- Jones, J. L., Royall, J. E., Critchley, D. R., & Walker, R. A. (1997). Modulation of myoepithelial-associated alpha6beta4 integrin in a breast cancer cell line alters invasive potential. *Experimental Cell Research*, 235(2), 325–333. <https://doi.org/10.1006/excr.1997.3662>
- Jones, J. T., Akita, R. W., & Sliwkowski, M. X. (1999). Binding specificities and affinities of egf domains for ErbB receptors. *FEBS Letters*, 447(2–3), 227–231. [https://doi.org/10.1016/s0014-5793\(99\)00283-5](https://doi.org/10.1016/s0014-5793(99)00283-5)
- Kabir, K. M. M., & Donald, W. A. (2017). Microscale differential ion mobility spectrometry for field deployable chemical analysis. *TrAC Trends in Analytical Chemistry*, 97, 399–427. <https://doi.org/10.1016/j.trac.2017.10.011>
- Kallioniemi, O. P., Holli, K., Visakorpi, T., Koivula, T., Helin, H. H., & Isola, J. J. (1991). Association of c-erbB-2 protein over-expression with high rate of cell proliferation, increased risk of visceral metastasis and poor long-term survival in breast cancer. *International Journal of Cancer*, 49(5), 650–655. <https://doi.org/10.1002/ijc.2910490504>
- Kang, B., Pu, M., Hu, G., Wen, W., Dong, Z., Zhao, K., Stillman, B., & Zhang, Z. (2011). Phosphorylation of H4 Ser 47 promotes HIRA-mediated nucleosome assembly. *Genes & Development*, 25(13), 1359–1364. <https://doi.org/10.1101/gad.2055511>
- Kang, H. A., & Hershey, J. W. (1994). Effect of initiation factor eIF-5A depletion on protein synthesis and proliferation of *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry*, 269(6), 3934–3940.
- Karas, Michael., & Hillenkamp, Franz. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, 60(20), 2299–2301. <https://doi.org/10.1021/ac00171a028>
- Khalili, A., & Ahmad, M. (2015). A Review of Cell Adhesion Studies for Biomedical and Biological Applications. *International Journal of Molecular Sciences*, 16(8), 18149–18184. <https://doi.org/10.3390/ijms160818149>
- Kho, D., MacDonald, C., Johnson, R., Unsworth, C. P., O'Carroll, S. J., du Mez, E., Angel, C. E., & Graham, E. S. (2015). Application of xCELLigence RTCA Biosensor Technology for Revealing the Profile and Window of Drug Responsiveness in Real Time. *Biosensors*, 5(2), 199–222. <https://doi.org/10.3390/bios5020199>
- Klapper, L. N., Glathe, S., Vaisman, N., Hynes, N. E., Andrews, G. C., Sela, M., & Yarden, Y. (1999). The ErbB-2/HER2 oncoprotein of human carcinomas may function solely as a shared coreceptor

- for multiple stroma-derived growth factors. *Proceedings of the National Academy of Sciences of the United States of America*, 96(9), 4995–5000. <https://doi.org/10.1073/pnas.96.9.4995>
- Klebe, M., Fremd, C., Kriegsmann, M., Kriegsmann, K., Albrecht, T., Thewes, V., Kirchner, M., Charoentong, P., Volk, N., Haag, J., Wirtz, R., Oskarsson, T., Schulz, A., Heil, J., Schneeweiss, A., Winter, H., & Sinn, P. (2020). Frequent Molecular Subtype Switching and Gene Expression Alterations in Lung and Pleural Metastasis From Luminal A-Type Breast Cancer. *JCO Precision Oncology*, 4, 848–859. <https://doi.org/10.1200/PO.19.00337>
- Kleizen, B., & Braakman, I. (2004). Protein folding and quality control in the endoplasmic reticulum. *Current Opinion in Cell Biology*, 16(4), 343–349. <https://doi.org/10.1016/j.ceb.2004.06.012>
- Koedoot, E., van Steijn, E., Vermeer, M., González-Prieto, R., Vertegaal, A. C. O., Martens, J. W. M., Le Dévédec, S. E., & van de Water, B. (2021). Splicing factors control triple-negative breast cancer cell mitosis through SUN2 interaction and sororin intron retention. *Journal of Experimental & Clinical Cancer Research*, 40(1), 82. <https://doi.org/10.1186/s13046-021-01863-4>
- Kok, D., Peeters, C. M. M., Mardina, Z., Oterdoom, D. L. M., Bulstra, S. K., Veldhuizen, A. G., Kuijer, R., & Wapstra, F. H. (2019). Is remaining intervertebral disc tissue interfering with bone generation during fusion of two vertebrae? *PloS One*, 14(4), e0215536. <https://doi.org/10.1371/journal.pone.0215536>
- Kondaveeti, Y., Guttilla Reed, I. K., & White, B. A. (2015). Epithelial–mesenchymal transition induces similar metabolic alterations in two independent breast cancer cell lines. *Cancer Letters*, 364(1), 44–58. <https://doi.org/10.1016/j.canlet.2015.04.025>
- Konecny, G., Pauletti, G., Pegram, M., Untch, M., Dandekar, S., Aguilar, Z., Wilson, C., Rong, H.-M., Bauerfeind, I., Felber, M., Wang, H.-J., Beryt, M., Seshadri, R., Hepp, H., & Slamon, D. J. (2003). Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer. *Journal of the National Cancer Institute*, 95(2), 142–153. <https://doi.org/10.1093/jnci/95.2.142>
- Krämer, A., Green, J., Pollard, J., & Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*, 30(4), 523–530. <https://doi.org/10.1093/bioinformatics/btt703>
- Krebs, E. G. (1983). Historical perspectives on protein phosphorylation and a classification system for protein kinases. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 302(1108), 3–11. <https://doi.org/10.1098/rstb.1983.0033>
- KREBS, E. G., & FISCHER, E. H. (1956). The phosphorylase b to a converting enzyme of rabbit skeletal muscle. *Biochimica et Biophysica Acta*, 20(1), 150–157. [https://doi.org/10.1016/0006-3002\(56\)90273-6](https://doi.org/10.1016/0006-3002(56)90273-6)
- Kulkarni, Y. M., Suarez, V., & Klink, D. J. (2010). Inferring predominant pathways in cellular models of breast cancer using limited sample proteomic profiling. *BMC Cancer*, 10(1), 291. <https://doi.org/10.1186/1471-2407-10-291>



- Lacroix, M., & Leclercq, G. (2004). Relevance of breast cancer cell lines as models for breast tumours: an update. *Breast Cancer Research and Treatment*, 83(3), 249–289. <https://doi.org/10.1023/B:BREA.0000014042.54925.cc>
- Lange, A., Mills, R. E., Lange, C. J., Stewart, M., Devine, S. E., & Corbett, A. H. (2007). Classical nuclear localization signals: definition, function, and interaction with importin alpha. *The Journal of Biological Chemistry*, 282(8), 5101–5105. <https://doi.org/10.1074/jbc.R600026200>
- Lasfargues, E. Y., Coutinho, W. G., & Redfield, E. S. (1978). Isolation of two human tumor epithelial cell lines from solid breast carcinomas. *Journal of the National Cancer Institute*, 61(4), 967–978.
- LeBaron, R. G., & Athanasiou, K. A. (2000). Ex vivo synthesis of articular cartilage. *Biomaterials*, 21(24), 2575–2587. [https://doi.org/10.1016/s0142-9612\(00\)00125-3](https://doi.org/10.1016/s0142-9612(00)00125-3)
- Leckband, D. E., & de Rooij, J. (2014). Cadherin adhesion and mechanotransduction. *Annual Review of Cell and Developmental Biology*, 30, 291–315. <https://doi.org/10.1146/annurev-cellbio-100913-013212>
- Lee, S. I., Kim, D. K., Seo, E. J., Choi, E. J., Kwon, Y. W., Jang, I. H., Lee, J. C., Kim, H. Y., Shong, M., Kim, J. H., & Kim, S.-J. (2017). Role of Krüppel-Like Factor 4 in the Maintenance of Chemoresistance of Anaplastic Thyroid Cancer. *Thyroid: Official Journal of the American Thyroid Association*, 27(11), 1424–1432. <https://doi.org/10.1089/thy.2016.0414>
- Leffers, H., Nielsen, M. S., Andersen, A. H., Honoré, B., Madsen, P., Vandekerckhove, J., & Celis, J. E. (1993). Identification of two human Rho GDP dissociation inhibitor proteins whose overexpression leads to disruption of the actin cytoskeleton. *Experimental Cell Research*, 209(2), 165–174. <https://doi.org/10.1006/excr.1993.1298>
- Lenter, M., & Vestweber, D. (1994). The integrin chains beta 1 and alpha 6 associate with the chaperone calnexin prior to integrin assembly. *The Journal of Biological Chemistry*, 269(16), 12263–12268.
- Levin, Y., & Bahn, S. (2010). Quantification of proteins by label-free LC-MS/MS. *Methods in Molecular Biology (Clifton, N.J.)*, 658, 217–231. [https://doi.org/10.1007/978-1-60761-780-8\\_13](https://doi.org/10.1007/978-1-60761-780-8_13)
- Levkowitz, G., Waterman, H., Zamir, E., Kam, Z., Oved, S., Langdon, W. Y., Beguinot, L., Geiger, B., & Yarden, Y. (1998). c-Cbl/Sli-1 regulates endocytic sorting and ubiquitination of the epidermal growth factor receptor. *Genes & Development*, 12(23), 3663–3674. <https://doi.org/10.1101/gad.12.23.3663>
- Li, F., Mandal, M., Barnes, C. J., Vadlamudi, R. K., & Kumar, R. (2001). Growth factor regulation of the molecular chaperone calnexin. *Biochemical and Biophysical Research Communications*, 289(3), 725–732. <https://doi.org/10.1006/bbrc.2001.6001>
- Li, J., Smith, L. S., & Zhu, H.-J. (2021). Data-independent acquisition (DIA): An emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters. *Drug Discovery Today: Technologies*, 39, 49–56. <https://doi.org/10.1016/j.ddtec.2021.06.006>
- Li, Y., Keller, D. M., Scott, J. D., & Lu, H. (2005). CK2 phosphorylates SSRP1 and inhibits its DNA-binding activity. *The Journal of Biological Chemistry*, 280(12), 11869–11875. <https://doi.org/10.1074/jbc.M413944200>

- Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., & Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research*, 47(W1), W199–W205. <https://doi.org/10.1093/nar/gkz401>
- Lilja, J., Zacharchenko, T., Georgiadou, M., Jacquemet, G., Franceschi, N. D., Peuhu, E., Hamidi, H., Pouwels, J., Martens, V., Nia, F. H., Beifuss, M., Boeckers, T., Kreienkamp, H.-J., Barsukov, I. L., & Ivaska, J. (2017). SHANK proteins limit integrin activation by directly interacting with Rap1 and R-Ras. *Nature Cell Biology*, 19(4), 292–305. <https://doi.org/10.1038/ncb3487>
- Lin, C.-C., Cheng, T.-L., Tsai, W.-H., Tsai, H.-J., Hu, K.-H., Chang, H.-C., Yeh, C.-W., Chen, Y.-C., Liao, C.-C., & Chang, W.-T. (2012). Loss of the respiratory enzyme citrate synthase directly links the Warburg effect to tumor malignancy. *Scientific Reports*, 2, 785. <https://doi.org/10.1038/srep00785>
- Lin, H., Zhang, F., Geng, Q., Yu, T., Cui, Y., Liu, X., Li, J., Yan, M., Liu, L., He, X., Li, J., & Yao, M. (2013). Quantitative Proteomic Analysis Identifies CPNE3 as a Novel Metastasis-promoting Gene in NSCLC. *Journal of Proteome Research*, 12(7), 3423–3433. <https://doi.org/10.1021/pr400273z>
- Lin, S. Y., Xia, W., Wang, J. C., Kwong, K. Y., Spohn, B., Wen, Y., Pestell, R. G., & Hung, M. C. (2000). Beta-catenin, a novel prognostic marker for breast cancer: its roles in cyclin D1 expression and cancer progression. *Proceedings of the National Academy of Sciences of the United States of America*, 97(8), 4262–4266. <https://doi.org/10.1073/pnas.060025397>
- Lipscomb, E. A., Simpson, K. J., Lyle, S. R., Ring, J. E., Dugan, A. S., & Mercurio, A. M. (2005). The alpha6beta4 integrin maintains the survival of human breast carcinoma cells in vivo. *Cancer Research*, 65(23), 10970–10976. <https://doi.org/10.1158/0008-5472.CAN-05-2327>
- Liu, J., Wang, P., Zhang, P., Zhang, X., Du, H., Liu, Q., Huang, B., Qian, C., Zhang, S., Zhu, W., Yang, X., Xiao, Y., Liu, Z., & Luo, D. (2019). An integrative bioinformatics analysis identified miR-375 as a candidate key regulator of malignant breast cancer. *Journal of Applied Genetics*, 60(3–4), 335–346. <https://doi.org/10.1007/s13353-019-00507-w>
- Liu, Y., Du, F., Chen, W., Yao, M., Lv, K., & Fu, P. (2015). EIF5A2 is a novel chemoresistance gene in breast cancer. *Breast Cancer (Tokyo, Japan)*, 22(6), 602–607. <https://doi.org/10.1007/s12282-014-0526-2>
- Liu, Y., Komohara, Y., Domenick, N., Ohno, M., Ikeura, M., Hamilton, R. L., Horbinski, C., Wang, X., Ferrone, S., & Okada, H. (2012). Expression of antigen processing and presenting molecules in brain metastasis of breast cancer. *Cancer Immunology, Immunotherapy: CII*, 61(6), 789–801. <https://doi.org/10.1007/s00262-011-1137-9>
- Liu, Y., Peterson, D. A., Kimura, H., & Schubert, D. (1997). Mechanism of cellular 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) reduction. *Journal of Neurochemistry*, 69(2), 581–593. <https://doi.org/10.1046/j.1471-4159.1997.69020581.x>
- Mackay, A., Jones, C., Dexter, T., Silva, R. L. A., Bulmer, K., Jones, A., Simpson, P., Harris, R. A., Jat, P. S., Neville, A. M., Reis, L. F. L., Lakhani, S. R., & O'Hare, M. J. (2003). cDNA microarray

- analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. *Oncogene*, 22(17), 2680–2688. <https://doi.org/10.1038/sj.onc.1206349>
- Małaczewska, J., Kaczorek-Łukowska, E., & Kazuń, B. (2021). High cytotoxicity of betulin towards fish and murine fibroblasts: Is betulin safe for nonneoplastic cells? *BMC Veterinary Research*, 17(1), 198. <https://doi.org/10.1186/s12917-021-02905-x>
- Mann, M., & Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology*, 21(3), 255–261. <https://doi.org/10.1038/nbt0303-255>
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science (New York, N.Y.)*, 298(5600), 1912–1934. <https://doi.org/10.1126/science.1075762>
- Marchiò, C., Annaratone, L., Marques, A., Casorzo, L., Berrino, E., & Sapino, A. (2021). Evolving concepts in HER2 evaluation in breast cancer: Heterogeneity, HER2-low carcinomas and beyond. *Seminars in Cancer Biology*, 72, 123–135. <https://doi.org/10.1016/j.semcancer.2020.02.016>
- Margolis, B. L., Lax, I., Kris, R., Dombalagian, M., Honegger, A. M., Howk, R., Givol, D., Ullrich, A., & Schlessinger, J. (1989). All autophosphorylation sites of epidermal growth factor (EGF) receptor and HER2/neu are located in their carboxyl-terminal tails. Identification of a novel site in EGF receptor. *The Journal of Biological Chemistry*, 264(18), 10667–10671.
- Martinez-Serra, J., Gutierrez, A., Muñoz-Capó, S., Navarro-Palou, M., Ros, T., Amat, J. C., Lopez, B., Marcus, T. F., Fueyo, L., Suquia, A. G., Gines, J., Rubio, F., Ramos, R., & Besalduch, J. (2014). xCELLigence system for real-time label-free monitoring of growth and viability of cell lines from hematological malignancies. *OncoTargets and Therapy*, 7, 985–994. <https://doi.org/10.2147/OTT.S62887>
- Marzluff, W. F., & McCarty, K. S. (1970). Two classes of histone acetylation in developing mouse mammary gland. *The Journal of Biological Chemistry*, 245(21), 5635–5642.
- McIlwain, S., Mathews, M., Bereman, M. S., Rubel, E. W., MacCoss, M. J., & Noble, W. S. (2012). Estimating relative abundances of proteins from shotgun proteomics data. *BMC Bioinformatics*, 13, 308. <https://doi.org/10.1186/1471-2105-13-308>
- Mercurio, A. M. (1995). Laminin receptors: achieving specificity through cooperation. *Trends in Cell Biology*, 5(11), 419–423. [https://doi.org/10.1016/s0962-8924\(00\)89100-x](https://doi.org/10.1016/s0962-8924(00)89100-x)
- 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., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J. T., Gatza, M. L., Wilkerson, M., Perou, C. M., Yellapantula, V., ... NCI CPTAC. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*, 534(7605), 55–62. <https://doi.org/10.1038/nature18003>
- Miki, I., Ishihara, N., Otoshi, M., & Kase, H. (1993). Simple colorimetric cell-cell adhesion assay using MTT-stained leukemia cells. *Journal of Immunological Methods*, 164(2), 255–261. [https://doi.org/10.1016/0022-1759\(93\)90318-2](https://doi.org/10.1016/0022-1759(93)90318-2)

- Mishra, S. K., Mandal, M., Mazumdar, A., & Kumar, R. (2001). Dynamic chromatin remodeling on the HER2 promoter in human breast cancer cells. *FEBS Letters*, 507(1), 88–94.  
[https://doi.org/10.1016/s0014-5793\(01\)02951-9](https://doi.org/10.1016/s0014-5793(01)02951-9)
- Mitteer, D. R., & Greer, B. D. (2022). Using GraphPad Prism's Heat Maps for Efficient, Fine-Grained Analyses of Single-Case Data. *Behavior Analysis in Practice*, 15(2), 505–514.  
<https://doi.org/10.1007/s40617-021-00664-7>
- Mo, W., Zhang, J., Li, X., Meng, D., Gao, Y., Yang, S., Wan, X., Zhou, C., Guo, F., Huang, Y., Amente, S., Avvedimento, E. V., Xie, Y., & Li, Y. (2013). Identification of novel AR-targeted microRNAs mediating androgen signalling through critical pathways to regulate cell viability in prostate cancer. *PloS One*, 8(2), e56592. <https://doi.org/10.1371/journal.pone.0056592>
- Moore, S. C., Rice, P., Iskandar, M., & Ausió, J. (1997). Reconstitution of native-like nucleosome core particles from reversed-phase-HPLC-fractionated histones. *The Biochemical Journal*, 328 ( Pt 2), 409–414. <https://doi.org/10.1042/bj3280409>
- Mor-Yossef Moldovan, L., Kislev, N., Lustig, M., Pomeranec, L., & Benayahu, D. (2020). Biomechanical stimulation effects on the metabolism of adipocyte. *Journal of Cellular Physiology*, 235(11), 8702–8713. <https://doi.org/10.1002/jcp.29714>
- Mosmann, T. (1983a). Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *Journal of Immunological Methods*, 65(1–2), 55–63.  
[https://doi.org/10.1016/0022-1759\(83\)90303-4](https://doi.org/10.1016/0022-1759(83)90303-4)
- Mosmann, T. (1983b). Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *Journal of Immunological Methods*, 65(1–2), 55–63.  
[https://doi.org/10.1016/0022-1759\(83\)90303-4](https://doi.org/10.1016/0022-1759(83)90303-4)
- Mukhopadhyay, R., Theriault, R. L., & Price, J. E. (1999). Increased levels of alpha6 integrins are associated with the metastatic phenotype of human breast cancer cells. *Clinical & Experimental Metastasis*, 17(4), 325–332. <https://doi.org/10.1023/a:1006659230585>
- Nakamura, K., Saredi, G., Becker, J. R., Foster, B. M., Nguyen, N. V., Beyer, T. E., Cesa, L. C., Faull, P. A., Lukauskas, S., Frimurer, T., Chapman, J. R., Bartke, T., & Groth, A. (2019). H4K20me0 recognition by BRCA1–BARD1 directs homologous recombination to sister chromatids. *Nature Cell Biology*, 21(3), 311–318. <https://doi.org/10.1038/s41556-019-0282-9>
- Nesvizhskii, A. I., & Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Molecular & Cellular Proteomics: MCP*, 4(10), 1419–1440.  
<https://doi.org/10.1074/mcp.R500012-MCP200>
- Nie, F., Yu, X.-L., Wang, X.-G., Tang, Y.-F., Wang, L.-L., & Ma, L. (2010). Down-regulation of CacyBP is associated with poor prognosis and the effects on COX-2 expression in breast cancer. *International Journal of Oncology*, 37(5), 1261–1269. [https://doi.org/10.3892/ijo\\_00000777](https://doi.org/10.3892/ijo_00000777)
- Nikoloff, N., Carranza Martin, A. C., Fabra, M. C., & Furnus, C. C. (2021). Amitraz induced cytotoxic effect on bovine cumulus cells and impaired oocyte maturation. *Environmental Science and Pollution Research International*, 28(23), 29188–29199. <https://doi.org/10.1007/s11356-021-12670-x>

- Ning, L., Wang, L., Zhang, H., Jiao, X., & Chen, D. (2020). Eukaryotic translation initiation factor 5A in the pathogenesis of cancers. *Oncology Letters*, 20(4), 81. <https://doi.org/10.3892/ol.2020.11942>
- O'Connor, K. L., Shaw, L. M., & Mercurio, A. M. (1998). Release of cAMP gating by the  $\alpha 6 \beta 4$  integrin stimulates lamellae formation and the chemotactic migration of invasive carcinoma cells. *The Journal of Cell Biology*, 143(6), 1749–1760. <https://doi.org/10.1083/jcb.143.6.1749>
- Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.-H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., & Yokoyama, S. (2002). Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell*, 110(6), 775–787. [https://doi.org/10.1016/s0092-8674\(02\)00963-7](https://doi.org/10.1016/s0092-8674(02)00963-7)
- Okazaki, Y., Ohno, H., Takase, K., Ochiai, T., & Saito, T. (2000). Cell surface expression of calnexin, a molecular chaperone in the endoplasmic reticulum. *The Journal of Biological Chemistry*, 275(46), 35751–35758. <https://doi.org/10.1074/jbc.M007476200>
- Olsen, D. T., Peng, L., Træholt, S. D., Duus, K., Højrup, P., & Houen, G. (2013). Purification and characterization of a soluble calnexin from human placenta. *Protein Expression and Purification*, 92(1), 105–111. <https://doi.org/10.1016/j.pep.2013.09.006>
- Ostman, A., Hellberg, C., & Böhmer, F. D. (2006). Protein-tyrosine phosphatases and cancer. *Nature Reviews. Cancer*, 6(4), 307–320. <https://doi.org/10.1038/nrc1837>
- Paglia, G., Angel, P., Williams, J. P., Richardson, K., Olivos, H. J., Thompson, J. W., Menikarachchi, L., Lai, S., Walsh, C., Moseley, A., Plumb, R. S., Grant, D. F., Palsson, B. O., Langridge, J., Geromanos, S., & Astarita, G. (2015). Ion mobility-derived collision cross section as an additional measure for lipid fingerprinting and identification. *Analytical Chemistry*, 87(2), 1137–1144. <https://doi.org/10.1021/ac503715v>
- Paglia, G., Smith, A. J., & Astarita, G. (2021). Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics. *Mass Spectrometry Reviews*. <https://doi.org/10.1002/mas.21686>
- Park, M. H., Lee, Y. B., & Joe, Y. A. (1997). Hypusine is essential for eukaryotic cell proliferation. *Biological Signals*, 6(3), 115–123. <https://doi.org/10.1159/000109117>
- Park, M. H., Wolff, E. C., & Folk, J. E. (1993). Hypusine: its post-translational formation in eukaryotic initiation factor 5A and its potential role in cellular regulation. *BioFactors (Oxford, England)*, 4(2), 95–104.
- Patra, T., & Gupta, M. K. (2020). Evaluation of sodium alginate for encapsulation-vitrification of testicular Leydig cells. *International Journal of Biological Macromolecules*, 153, 128–137. <https://doi.org/10.1016/j.ijbiomac.2020.02.233>
- Peng, M., Yang, D., Hou, Y., Liu, S., Zhao, M., Qin, Y., Chen, R., Teng, Y., & Liu, M. (2019). Intracellular citrate accumulation by oxidized ATM-mediated metabolism reprogramming via PFKP and CS enhances hypoxic breast cancer cell invasion and metastasis. *Cell Death & Disease*, 10(3), 228. <https://doi.org/10.1038/s41419-019-1475-7>
- Percipalle, P., Fomproix, N., Kylberg, K., Miralles, F., Bjorkroth, B., Daneholt, B., & Visa, N. (2003). An actin-ribonucleoprotein interaction is involved in transcription by RNA polymerase II.

- Proceedings of the National Academy of Sciences of the United States of America*, 100(11), 6475–6480. <https://doi.org/10.1073/pnas.1131933100>
- Pick, E., Kluger, Y., Giltneane, J. M., Moeder, C., Camp, R. L., Rimm, D. L., & Kluger, H. M. (2007). High HSP90 expression is associated with decreased survival in breast cancer. *Cancer Research*, 67(7), 2932–2937. <https://doi.org/10.1158/0008-5472.CAN-06-4511>
- Pieters, R., Huismans, D. R., Leyva, A., & Veerman, A. J. (1988). Adaptation of the rapid automated tetrazolium dye based (MTT) assay for chemosensitivity testing in childhood leukemia. *Cancer Letters*, 41(3), 323–332. [https://doi.org/10.1016/0304-3835\(88\)90294-7](https://doi.org/10.1016/0304-3835(88)90294-7)
- Pinkstaff, J. K., Detterich, J., Lynch, G., & Gall, C. (1999). Integrin subunit gene expression is regionally differentiated in adult brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 19(5), 1541–1556.
- Prendergast, L., Hong, E., Safina, A., Poe, D., & Gurova, K. (2020). Histone chaperone FACT is essential to overcome replication stress in mammalian cells. *Oncogene*, 39(28), 5124–5137. <https://doi.org/10.1038/s41388-020-1346-9>
- Qu, S., Rinehart, C., Wu, H.-H., Wang, S. E., Carter, B., Xin, H., Kotlikoff, M., & Arteaga, C. L. (2006). Gene targeting of ErbB3 using a Cre-mediated unidirectional DNA inversion strategy. *Genesis (New York, N.Y. : 2000)*, 44(10), 477–486. <https://doi.org/10.1002/dvg.20243>
- Rabinovitz, I., & Mercurio, A. M. (1997). The integrin alpha6beta4 functions in carcinoma cell migration on laminin-1 by mediating the formation and stabilization of actin-containing motility structures. *The Journal of Cell Biology*, 139(7), 1873–1884. <https://doi.org/10.1083/jcb.139.7.1873>
- Reid, G. E., & McLuckey, S. A. (2002). “Top down” protein characterization via tandem mass spectrometry. *Journal of Mass Spectrometry: JMS*, 37(7), 663–675. <https://doi.org/10.1002/jms.346>
- Ridley, A. J. (1995). Rho-related proteins: actin cytoskeleton and cell cycle. *Current Opinion in Genetics & Development*, 5(1), 24–30. [https://doi.org/10.1016/s0959-437x\(95\)90049-7](https://doi.org/10.1016/s0959-437x(95)90049-7)
- Rigbolt, K. T. G., & Blagoev, B. (2010). Proteome-wide quantitation by SILAC. *Methods in Molecular Biology (Clifton, N.J.)*, 658, 187–204. [https://doi.org/10.1007/978-1-60761-780-8\\_11](https://doi.org/10.1007/978-1-60761-780-8_11)
- Robinson, D. R., Wu, Y. M., & Lin, S. F. (2000). The protein tyrosine kinase family of the human genome. *Oncogene*, 19(49), 5548–5557. <https://doi.org/10.1038/sj.onc.1203957>
- Roshan Moniri, M., Young, A., Reinheimer, K., Rayat, J., Dai, L.-J., & Warnock, G. L. (2015). Dynamic assessment of cell viability, proliferation and migration using real time cell analyzer system (RTCA). *Cytotechnology*, 67(2), 379–386. <https://doi.org/10.1007/s10616-014-9692-5>
- Roskoski, R. (2004). The ErbB/HER receptor protein-tyrosine kinases and cancer. *Biochemical and Biophysical Research Communications*, 319(1), 1–11. <https://doi.org/10.1016/j.bbrc.2004.04.150>
- Ross, J. S., & Fletcher, J. A. (1999). HER-2/neu (c-erb-B2) gene and protein in breast cancer. *American Journal of Clinical Pathology*, 112(1 Suppl 1), S53-67.

- Rozanova, S., Barkovits, K., Nikolov, M., Schmidt, C., Urlaub, H., & Marcus, K. (2021). *Quantitative Mass Spectrometry-Based Proteomics: An Overview* (pp. 85–116). [https://doi.org/10.1007/978-1-0716-1024-4\\_8](https://doi.org/10.1007/978-1-0716-1024-4_8)
- Rubin, I., & Yarden, Y. (2001). The basic biology of HER2. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 12 Suppl 1, S3-8. [https://doi.org/10.1093/annonc/12.suppl\\_1.s3](https://doi.org/10.1093/annonc/12.suppl_1.s3)
- Ruiz-Saenz, A., Dreyer, C., Campbell, M. R., Steri, V., Gulizia, N., & Moasser, M. M. (2018). HER2 Amplification in Tumors Activates PI3K/Akt Signaling Independent of HER3. *Cancer Research*, 78(13), 3645–3658. <https://doi.org/10.1158/0008-5472.CAN-18-0430>
- Santamaria-Kisiel, L., Rintala-Dempsey, A. C., & Shaw, G. S. (2006). Calcium-dependent and -independent interactions of the S100 protein family. *The Biochemical Journal*, 396(2), 201–214. <https://doi.org/10.1042/BJ20060195>
- Santoro, M. M., Gaudino, G., & Marchisio, P. C. (2003). The MSP receptor regulates alpha6beta4 and alpha3beta1 integrins via 14-3-3 proteins in keratinocyte migration. *Developmental Cell*, 5(2), 257–271. [https://doi.org/10.1016/s1534-5807\(03\)00201-6](https://doi.org/10.1016/s1534-5807(03)00201-6)
- Schaeffer, L., Duclert, N., Huchet-Dymanus, M., & Changeux, J. P. (1998). Implication of a multisubunit Ets-related transcription factor in synaptic expression of the nicotinic acetylcholine receptor. *The EMBO Journal*, 17(11), 3078–3090. <https://doi.org/10.1093/emboj/17.11.3078>
- Scherle, P., Behrens, T., & Staudt, L. M. (1993). Ly-GDI, a GDP-dissociation inhibitor of the RhoA GTP-binding protein, is expressed preferentially in lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America*, 90(16), 7568–7572. <https://doi.org/10.1073/pnas.90.16.7568>
- Schlessinger, J. (1994). SH2/SH3 signaling proteins. *Current Opinion in Genetics & Development*, 4(1), 25–30. [https://doi.org/10.1016/0959-437x\(94\)90087-6](https://doi.org/10.1016/0959-437x(94)90087-6)
- Schneider, G., Nieznanski, K., Kilanczyk, E., Bieganski, P., Kuznicki, J., & Filipek, A. (2007). CacyBP/SIP interacts with tubulin in neuroblastoma NB2a cells and induces formation of globular tubulin assemblies. *Biochimica et Biophysica Acta*, 1773(11), 1628–1636. <https://doi.org/10.1016/j.bbamcr.2007.07.013>
- Schroeder, J. A., & Lee, D. C. (1998). Dynamic expression and activation of ERBB receptors in the developing mouse mammary gland. *Cell Growth & Differentiation: The Molecular Biology Journal of the American Association for Cancer Research*, 9(6), 451–464.
- Sebastian, J., Richards, R. G., Walker, M. P., Wiesen, J. F., Werb, Z., Derynck, R., Hom, Y. K., Cunha, G. R., & DiAugustine, R. P. (1998). Activation and function of the epidermal growth factor receptor and erbB-2 during mammary gland morphogenesis. *Cell Growth & Differentiation: The Molecular Biology Journal of the American Association for Cancer Research*, 9(9), 777–785.
- Selevsek, N., Chang, C.-Y., Gillet, L. C., Navarro, P., Bernhardt, O. M., Reiter, L., Cheng, L.-Y., Vitek, O., & Aebersold, R. (2015). Reproducible and Consistent Quantification of the *Saccharomyces*

- cerevisiae Proteome by SWATH-mass spectrometry \*. *Molecular & Cellular Proteomics*, 14(3), 739–749. <https://doi.org/10.1074/mcp.M113.035550>
- Shah, A. A., Ito, A., Nakata, A., & Yoshida, M. (2016). Identification of a Selective SIRT2 Inhibitor and Its Anti-breast Cancer Activity. *Biological & Pharmaceutical Bulletin*, 39(10), 1739–1742. <https://doi.org/10.1248/bpb.b16-00520>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Shaw, L. M., Rabinovitz, I., Wang, H. H., Toker, A., & Mercurio, A. M. (1997). Activation of phosphoinositide 3-OH kinase by the alpha6beta4 integrin promotes carcinoma invasion. *Cell*, 91(7), 949–960. [https://doi.org/10.1016/s0092-8674\(00\)80486-9](https://doi.org/10.1016/s0092-8674(00)80486-9)
- Shi, T., Song, E., Nie, S., Rodland, K. D., Liu, T., Qian, W., & Smith, R. D. (2016). Advances in targeted proteomics and applications to biomedical research. *PROTEOMICS*, 16(15–16), 2160–2182. <https://doi.org/10.1002/pmic.201500449>
- Shliaha, P. V., Bond, N. J., Gatto, L., & Lilley, K. S. (2013). Effects of Traveling Wave Ion Mobility Separation on Data Independent Acquisition in Proteomics Studies. *Journal of Proteome Research*, 12(6), 2323–2339. <https://doi.org/10.1021/pr300775k>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539. <https://doi.org/10.1038/msb.2011.75>
- Simpson, R. T. (1976). Histones H3 and H4 interact with the ends of nucleosome DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 73(12), 4400–4404. <https://doi.org/10.1073/pnas.73.12.4400>
- Sinclair, J., & Timms, J. F. (2011). Quantitative profiling of serum samples using TMT protein labelling, fractionation and LC-MS/MS. *Methods (San Diego, Calif.)*, 54(4), 361–369. <https://doi.org/10.1016/j.ymeth.2011.03.004>
- Singletary, K. W., & Gapstur, S. M. (2001). Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. *JAMA*, 286(17), 2143–2151. <https://doi.org/10.1001/jama.286.17.2143>
- Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., & McGuire, W. L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science (New York, N.Y.)*, 235(4785), 177–182. <https://doi.org/10.1126/science.3798106>
- Smeltzer, S., Quadri, Z., Miller, A., Zamudio, F., Hunter, J., Stewart, N. J. F., Saji, S., Lee, D. C., Chaput, D., & Selenica, M.-L. B. (2021). Hypusination of Eif5a regulates cytoplasmic TDP-43 aggregation and accumulation in a stress-induced cellular model. *Biochimica et Biophysica Acta. Molecular Basis of Disease*, 1867(1), 165939. <https://doi.org/10.1016/j.bbadis.2020.165939>



- Stamps, A. C., Davies, S. C., Burman, J., & O'Hare, M. J. (1994). Analysis of proviral integration in human mammary epithelial cell lines immortalized by retroviral infection with a temperature-sensitive SV40 T-antigen construct. *International Journal of Cancer*, 57(6), 865–874. <https://doi.org/10.1002/ijc.2910570616>
- Stefanowicz-Hajduk, J., & Ochocka, J. R. (2020). Real-time cell analysis system in cytotoxicity applications: Usefulness and comparison with tetrazolium salt assays. *Toxicology Reports*, 7, 335–344. <https://doi.org/10.1016/j.toxrep.2020.02.002>
- Stein, R. A., & Staros, J. V. (2000). Evolutionary analysis of the ErbB receptor and ligand families. *Journal of Molecular Evolution*, 50(5), 397–412. <https://doi.org/10.1007/s002390010043>
- Stein, R. A., & Staros, J. V. (2006). Insights into the evolution of the ErbB receptor family and their ligands from sequence analysis. *BMC Evolutionary Biology*, 6, 79. <https://doi.org/10.1186/1471-2148-6-79>
- Stewart, R. L., & O'Connor, K. L. (2015). Clinical significance of the integrin  $\alpha 6 \beta 4$  in human malignancies. *Laboratory Investigation; a Journal of Technical Methods and Pathology*, 95(9), 976–986. <https://doi.org/10.1038/labinvest.2015.82>
- Stockert, J. C., Blázquez-Castro, A., Cañete, M., Horobin, R. W., & Villanueva, A. (2012). MTT assay for cell viability: Intracellular localization of the formazan product is in lipid droplets. *Acta Histochemica*, 114(8), 785–796. <https://doi.org/10.1016/j.acthis.2012.01.006>
- Stockert, J. C., Horobin, R. W., Colombo, L. L., & Blázquez-Castro, A. (2018). Tetrazolium salts and formazan products in Cell Biology: Viability assessment, fluorescence imaging, and labeling perspectives. *Acta Histochemica*, 120(3), 159–167. <https://doi.org/10.1016/j.acthis.2018.02.005>
- Surin, A. M., Sharipov, R. R., Krasil'nikova, I. A., Boyarkin, D. P., Lisina, O. Y., Gorbacheva, L. R., Avetisyan, A. V., & Pinelis, V. G. (2017). Disruption of Functional Activity of Mitochondria during MTT Assay of Viability of Cultured Neurons. *Biochemistry. Biokhimiia*, 82(6), 737–749. <https://doi.org/10.1134/S0006297917060104>
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, 49(D1), D605–D612. <https://doi.org/10.1093/nar/gkaa1074>
- Tan, M., Li, P., Klos, K. S., Lu, J., Lan, K.-H., Nagata, Y., Fang, D., Jing, T., & Yu, D. (2005). ErbB2 promotes Src synthesis and stability: novel mechanisms of Src activation that confer breast cancer metastasis. *Cancer Research*, 65(5), 1858–1867. <https://doi.org/10.1158/0008-5472.CAN-04-2353>
- Tan, M., & Yu, D. (2007). *Molecular Mechanisms of ErbB2-Mediated Breast Cancer Chemoresistance* (pp. 119–129). [https://doi.org/10.1007/978-0-387-74039-3\\_9](https://doi.org/10.1007/978-0-387-74039-3_9)
- Tarcic, G., Boguslavsky, S. K., Wakim, J., Kiuchi, T., Liu, A., Reinitz, F., Nathanson, D., Takahashi, T., Mischel, P. S., Ng, T., & Yarden, Y. (2009). An unbiased screen identifies DEP-1 tumor

- suppressor as a phosphatase controlling EGFR endocytosis. *Current Biology: CB*, 19(21), 1788–1798. <https://doi.org/10.1016/j.cub.2009.09.048>
- Taubenberger, A., Cisneros, D. A., Friedrichs, J., Puech, P.-H., Muller, D. J., & Franz, C. M. (2007). Revealing early steps of alpha2beta1 integrin-mediated adhesion to collagen type I by using single-cell force spectroscopy. *Molecular Biology of the Cell*, 18(5), 1634–1644. <https://doi.org/10.1091/mbc.e06-09-0777>
- Taylor, G. I. (1964). Disintegration of water drops in an electric field. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 280(1382), 383–397. <https://doi.org/10.1098/rspa.1964.0151>
- Teng, L., Zheng, Y., & Wang, H. (2008). BRCA1/2 associated hereditary breast cancer. *Journal of Zhejiang University. Science. B*, 9(2), 85–89. <https://doi.org/10.1631/jzus.B0710617>
- Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., Crenshaw, A., Cancel-Tassin, G., Staats, B. J., Wang, Z., Gonzalez-Bosquet, J., Fang, J., Deng, X., Berndt, S. I., Calle, E. E., ... Chanock, S. J. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nature Genetics*, 40(3), 310–315. <https://doi.org/10.1038/ng.91>
- Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K. A., & Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry*, 75(8), 1895–1904. <https://doi.org/10.1021/ac0262560>
- Tidcombe, H., Jackson-Fisher, A., Mathers, K., Stern, D. F., Gassmann, M., & Golding, J. P. (2003). Neural and mammary gland defects in ErbB4 knockout mice genetically rescued from embryonic lethality. *Proceedings of the National Academy of Sciences of the United States of America*, 100(14), 8281–8286. <https://doi.org/10.1073/pnas.1436402100>
- Timms, J. F., White, S. L., O'Hare, M. J., & Waterfield, M. D. (2002). Effects of ErbB-2 overexpression on mitogenic signalling and cell cycle progression in human breast luminal epithelial cells. *Oncogene*, 21(43), 6573–6586. <https://doi.org/10.1038/sj.onc.1205847>
- Trempe, G. L. (1976). Human breast cancer in culture. *Recent Results in Cancer Research. Fortschritte Der Krebsforschung. Progres Dans Les Recherches Sur Le Cancer*, 57, 33–41. [https://doi.org/10.1007/978-3-642-81043-5\\_5](https://doi.org/10.1007/978-3-642-81043-5_5)
- Tyanova, S., Albrechtsen, R., Kronqvist, P., Cox, J., Mann, M., & Geiger, T. (2016). Proteomic maps of breast cancer subtypes. *Nature Communications*, 7, 10259. <https://doi.org/10.1038/ncomms10259>
- Tzahar, E., Waterman, H., Chen, X., Levkowitz, G., Karunakaran, D., Lavi, S., Ratzkin, B. J., & Yarden, Y. (1996). A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor. *Molecular and Cellular Biology*, 16(10), 5276–5287. <https://doi.org/10.1128/MCB.16.10.5276>
- Unwin, R. D. (2010). Quantification of proteins by iTRAQ. *Methods in Molecular Biology (Clifton, N.J.)*, 658, 205–215. [https://doi.org/10.1007/978-1-60761-780-8\\_12](https://doi.org/10.1007/978-1-60761-780-8_12)

- Usher, K. C., Remington, S. J., Martin, D. P., & Drueckhammer, D. G. (1994). A very short hydrogen bond provides only moderate stabilization of an enzyme-inhibitor complex of citrate synthase. *Biochemistry*, 33(25), 7753–7759. <https://doi.org/10.1021/bi00191a002>
- Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A., & Yates, J. R. (2004). Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods*, 1(1), 39–45. <https://doi.org/10.1038/nmeth705>
- Venerando, A., Cesaro, L., & Pinna, L. A. (2017). From phosphoproteins to phosphoproteomes: a historical account. *The FEBS Journal*, 284(13), 1936–1951. <https://doi.org/10.1111/febs.14014>
- Verbeek, B. S., Adriaansen-Slot, S. S., Rijksen, G., & Vroom, T. M. (1997). Grb2 overexpression in nuclei and cytoplasm of human breast cells: a histochemical and biochemical study of normal and neoplastic mammary tissue specimens. *The Journal of Pathology*, 183(2), 195–203. [https://doi.org/10.1002/\(SICI\)1096-9896\(199710\)183:2<195::AID-PATH901>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1096-9896(199710)183:2<195::AID-PATH901>3.0.CO;2-Y)
- VerBerkmoes, N. C., Bundy, J. L., Hauser, L., Asano, K. G., Razumovskaya, J., Larimer, F., Hettich, R. L., & Stephenson, J. L. (2002). Integrating 'top-down' and 'bottom-up' mass spectrometric approaches for proteomic analysis of *Shewanella oneidensis*. *Journal of Proteome Research*, 1(3), 239–252. <https://doi.org/10.1021/pr025508a>
- Wang, H., & Hanash, S. (2015). Mass spectrometry based proteomics for absolute quantification of proteins from tumor cells. *Methods*, 81, 34–40. <https://doi.org/10.1016/j.ymeth.2015.03.007>
- Wang, H., Leavitt, L., Ramaswamy, R., & Rapraeger, A. C. (2010). Interaction of syndecan and alpha6beta4 integrin cytoplasmic domains: regulation of ErbB2-mediated integrin activation. *The Journal of Biological Chemistry*, 285(18), 13569–13579. <https://doi.org/10.1074/jbc.M110.102137>
- Wang, J., Duncan, D., Shi, Z., & Zhang, B. (2013). WEB-based GEne SeT Analysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Research*, 41(W1), W77–W83. <https://doi.org/10.1093/nar/gkt439>
- Wang, M., You, J., Bemis, K. G., Tegeler, T. J., & Brown, D. P. G. (2008). Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. *Briefings in Functional Genomics and Proteomics*, 7(5), 329–339. <https://doi.org/10.1093/bfpg/eln031>
- Wang, N., Ma, Q., Wang, Y., Ma, G., & Zhai, H. (2010). CacyBP/SIP expression is involved in the clinical progression of breast cancer. *World Journal of Surgery*, 34(11), 2545–2552. <https://doi.org/10.1007/s00268-010-0690-2>
- Wang, S. E., Xiang, B., Zent, R., Quaranta, V., Pozzi, A., & Arteaga, C. L. (2009). Transforming growth factor beta induces clustering of HER2 and integrins by activating Src-focal adhesion kinase and receptor association to the cytoskeleton. *Cancer Research*, 69(2), 475–482. <https://doi.org/10.1158/0008-5472.CAN-08-2649>
- Wang, X., Bi, X., Huang, X., Wang, B., Guo, Q., & Wu, Z. (2020). Systematic investigation of biomarker-like role of ARHGDI1 in breast cancer. *Cancer Biomarkers: Section A of Disease Markers*, 28(1), 101–110. <https://doi.org/10.3233/CBM-190562>

- Wang, Z., Fan, M., Candas, D., Zhang, T.-Q., Qin, L., Eldridge, A., Wachsmann-Hogiu, S., Ahmed, K. M., Chromy, B. A., Nantajit, D., Duru, N., He, F., Chen, M., Finkel, T., Weinstein, L. S., & Li, J. J. (2014). Cyclin B1/Cdk1 coordinates mitochondrial respiration for cell-cycle G2/M progression. *Developmental Cell*, 29(2), 217–232. <https://doi.org/10.1016/j.devcel.2014.03.012>
- Warren, C. M., & Landgraf, R. (2006). Signaling through ERBB receptors: multiple layers of diversity and control. *Cellular Signalling*, 18(7), 923–933. <https://doi.org/10.1016/j.cellsig.2005.12.007>
- Washburn, M. P., Wolters, D., & Yates, J. R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, 19(3), 242–247. <https://doi.org/10.1038/85686>
- White, S. L., Gharbi, S., Bertani, M. F., Chan, H.-L., Waterfield, M. D., & Timms, J. F. (2004). Cellular responses to ErbB-2 overexpression in human mammary luminal epithelial cells: comparison of mRNA and protein expression. *British Journal of Cancer*, 90(1), 173–181. <https://doi.org/10.1038/sj.bjc.6601458>
- Wiesen, J. F., Young, P., Werb, Z., & Cunha, G. R. (1999). Signaling through the stromal epidermal growth factor receptor is necessary for mammary ductal development. *Development (Cambridge, England)*, 126(2), 335–344. <https://doi.org/10.1242/dev.126.2.335>
- Wiest, D. L., Burgess, W. H., McKean, D., Kears, K. P., & Singer, A. (1995). The molecular chaperone calnexin is expressed on the surface of immature thymocytes in association with clonotype-independent CD3 complexes. *The EMBO Journal*, 14(14), 3425–3433. <https://doi.org/10.1002/j.1460-2075.1995.tb07348.x>
- Wilhelm, J. A., & McCarty, K. S. (1970). Partial characterization of the histones and histone acetylation in cell cultures. *Cancer Research*, 30(2), 409–417.
- Wilhelmsen, K., Litjens, S. H. M., & Sonnenberg, A. (2006). Multiple functions of the integrin alpha6beta4 in epidermal homeostasis and tumorigenesis. *Molecular and Cellular Biology*, 26(8), 2877–2886. <https://doi.org/10.1128/MCB.26.8.2877-2886.2006>
- Wolff, M. M., & Stephens, W. E. (1953). A Pulsed Mass Spectrometer with Time Dispersion. *Review of Scientific Instruments*, 24(8), 616–617. <https://doi.org/10.1063/1.1770801>
- Worthington, J. (2012). *Functional characterisation of targets of ErbB2-dependent signalling in breast cancer*.
- Worthington, J., Bertani, M., Chan, H.-L., Gerrits, B., & Timms, J. F. (2010). Transcriptional profiling of ErbB signalling in mammary luminal epithelial cells - interplay of ErbB and IGF1 signalling through IGFBP3 regulation. *BMC Cancer*, 10(1), 490. <https://doi.org/10.1186/1471-2407-10-490>
- Worthington, J., Spain, G., & Timms, J. F. (2017). Effects of ErbB2 Overexpression on the Proteome and ErbB Ligand-specific Phosphosignaling in Mammary Luminal Epithelial Cells. *Molecular & Cellular Proteomics: MCP*, 16(4), 608–621. <https://doi.org/10.1074/mcp.M116.061267>
- Wu, J., Tan, X., Peng, X., Yuan, J., & Qiang, B. (2003). Translocation and phosphorylation of calcyclin binding protein during retinoic acid-induced neuronal differentiation of neuroblastoma SH-SY5Y cells. *Journal of Biochemistry and Molecular Biology*, 36(4), 354–358. <https://doi.org/10.5483/bmbrep.2003.36.4.354>

- Wu, Y., Moissoglu, K., Wang, H., Wang, X., Frierson, H. F., Schwartz, M. A., & Theodorescu, D. (2009). Src phosphorylation of RhoGDI2 regulates its metastasis suppressor function. *Proceedings of the National Academy of Sciences of the United States of America*, 106(14), 5807–5812. <https://doi.org/10.1073/pnas.0810094106>
- Xie, W., Paterson, A. J., Chin, E., Nabell, L. M., & Kudlow, J. E. (1997). Targeted Expression of a Dominant Negative Epidermal Growth Factor Receptor in the Mammary Gland of Transgenic Mice Inhibits Pubertal Mammary Duct Development. *Molecular Endocrinology*, 11(12), 1766–1781. <https://doi.org/10.1210/mend.11.12.0019>
- Xu, L., Wang, L., Jiang, C., Zhu, Q., Chen, R., Wang, J., & Wang, S. (2020). Biological effect of ribosomal protein L32 on human breast cancer cell behavior. *Molecular Medicine Reports*, 22(3), 2478–2486. <https://doi.org/10.3892/mmr.2020.11302>
- Xu, W., Yuan, X., Beebe, K., Xiang, Z., & Neckers, L. (2007). Loss of Hsp90 association up-regulates Src-dependent ErbB2 activity. *Molecular and Cellular Biology*, 27(1), 220–228. <https://doi.org/10.1128/MCB.00899-06>
- Xu, Y. J., Hu, Y. M., Qin, C., Wang, F., Cao, W., Yu, Y. W., Zhao, L., Li, J., Chen, W. Q., Li, N., & He, J. (2021). [CacyBP promotes the proliferation and invasion of non-small cell lung cancer]. *Zhonghua Zhong Liu Za Zhi [Chinese Journal of Oncology]*, 43(9), 924–931. <https://doi.org/10.3760/cma.j.cn112152-20210421-00329>
- Yang, C., Ionescu-Tiba, V., Burns, K., Gadd, M., Zukerberg, L., Louis, D. N., Sgroi, D., & Schmidt, E. V. (2004). The role of the cyclin D1-dependent kinases in ErbB2-mediated breast cancer. *The American Journal of Pathology*, 164(3), 1031–1038. [https://doi.org/10.1016/S0002-9440\(10\)63190-2](https://doi.org/10.1016/S0002-9440(10)63190-2)
- Yang, X. H., Flores, L. M., Li, Q., Zhou, P., Xu, F., Krop, I. E., & Hemler, M. E. (2010). Disruption of laminin-integrin-CD151-focal adhesion kinase axis sensitizes breast cancer cells to ErbB2 antagonists. *Cancer Research*, 70(6), 2256–2263. <https://doi.org/10.1158/0008-5472.CAN-09-4032>
- Yap, A. S., Duszyc, K., & Viasnoff, V. (2018). Mechanosensing and Mechanotransduction at Cell-Cell Junctions. *Cold Spring Harbor Perspectives in Biology*, 10(8). <https://doi.org/10.1101/cshperspect.a028761>
- Yarden, Y., & Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nature Reviews. Molecular Cell Biology*, 2(2), 127–137. <https://doi.org/10.1038/35052073>
- Yip, S.-C., Saha, S., & Chernoff, J. (2010). PTP1B: a double agent in metabolism and oncogenesis. *Trends in Biochemical Sciences*, 35(8), 442–449. <https://doi.org/10.1016/j.tibs.2010.03.004>
- Yu, D., & Hung, M.-C. (2000). Overexpression of ErbB2 in cancer and ErbB2-targeting strategies. *Oncogene*, 19(53), 6115–6121. <https://doi.org/10.1038/sj.onc.1203972>
- Zecha, J., Satpathy, S., Kanashova, T., Avanesian, S. C., Kane, M. H., Clauser, K. R., Mertins, P., Carr, S. A., & Kuster, B. (2019). TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Molecular & Cellular Proteomics*, 18(7), 1468–1478. <https://doi.org/10.1074/mcp.TIR119.001385>

- Zhang, B., Kirov, S., & Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Research*, 33(Web Server issue), W741-8. <https://doi.org/10.1093/nar/gki475>
- Zhou, B., & Hung MC. (2003). Dysregulation of cellular signaling by HER2 in breast cancer. *Seminars in Oncology*, 30, 38–48. <https://doi.org/10.1053/j.seminoncol.2003.08.006>
- Zhou, J., Guo, Y., Huo, Z., Xing, Y., Fang, J., Ma, G., Han, Q., Wang, M., & Xu, Q. (2021). Identification of therapeutic targets and prognostic biomarkers from the hnRNP family in invasive breast carcinoma. *Aging*, 13(3), 4503–4521. <https://doi.org/10.18632/aging.202411>

## **Chapter 7**

### **7. Appendix**

#### **7.1. Mass spectrometry data tables**

**Appendix Table 3.4.1. 52 proteins identified as up/down regulated in response to CPNE3 knockdown.** The table shows a list of up/ down-regulated (>1.5 fold) proteins and their molecular functions. The identified proteins were filtered by selecting peptides with a score >20 and below the Mascot significance threshold filter of  $p = 0.05$  were included and single peptide identifications required a score equal to or above the Mascot identity threshold (Sinclair & Timms, 2011). 25 proteins highlighted in yellow were commonly differentially expressed in response to CPNE3 knockdown in both C3.6 and SKBR3 cell lines.

Protein name	Gene name	Function	Average ratio (CPNE3kd vs. Ctrl (127+129+131)/(126+128+130))
Histone H3.1	HIST1H3A	DNA binding, cadherin binding, histone binding, protein heterodimerization activity	4.842
Histone H4	HIST1H4A	DNA binding, histone binding, protein heterodimerization activity, protein domain specific binding, RNA binding	4.074
Prostaglandin E synthase	PTGES	glutathione binding, prostaglandin-E synthase activity	3.397
Microsomal glutathione S-transferase 1	MGST1	glutathione binding, glutathione peroxidase activity, glutathione transferase activity, protein homodimerization activity	3.140
Calnexin	CANX	apolipoprotein binding, calcium ion binding, carbohydrate binding, glycoprotein binding, ionotropic glutamate receptor binding, RNA binding, unfolded protein binding	3.129
ADP/ATP translocase 3	SLC25A6	adenine transmembrane transporter activity, ATP:ADP antiporter activity	2.995
Citrate synthase, mitochondrial	CS	citrate (si)- synthetase activity, RNA binding	2.478
Thioredoxin-dependent peroxide reductase, mitochondrial	PRDX3	cysteine-type endopeptidase inhibitor activity involved in apoptotic process, alkyl hydroperoxide reductase activity, protein kinase binding, thioredoxin peroxidase activity, identical protein binding	2.443
Phosphate carrier protein, mitochondrial	SLC25A3	phosphate ion carrier activity, protein complex binding, symporter activity	2.221
Single-stranded DNA-binding protein, mitochondrial	SSBP1	chromatin binding, RNA binding, single-stranded DNA binding	1.992
Voltage-dependent anion-selective channel protein 2	VDAC2	nucleotide binding, porin activity, voltage-gated anion channel activity	1.990
Rho-related GTP-binding protein RhoC	RHOC	GTPase activity, GTP binding, signal transduce activity	1.973
Dihydrolipoyl dehydrogenase, mitochondrial	DLD	dihydrolipoyl dehydrogenase activity, flavin adenine dinucleotide binding, lipoamide binding, NAD binding	1.967
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	beta-tubulin binding, endonuclease activity, microtubule binding, RNA binding, single-stranded DNA binding, ubiquitin protein ligase binding	1.960
60 kDa heat shock protein, mitochondrial	CACYBP	apolipoprotein binding, ATPase activity, ATP binding, chaperone binding, DNA replication origin binding, double stranded RNA binding, unfolded protein binding, protease binding	1.892
10 kDa heat shock protein, mitochondrial	HSPE1	ATP binding, chaperone binding, metal ion binding, RNA binding, unfolded protein binding	1.713
4F2 cell-surface antigen heavy chain	SLC3A2	cadherin binding, calcium: sodium antiporter activity, catalytic activity, double-stranded RNA binding, neutral amino acid transmembrane transport activity, RNA binding	1.710



Ran-specific GTPase-activating protein	RANBP1	cadherin binding, GDP-dissociation inhibitor activity, GTPase activator activity, Ran GTPase binding	1.677
ATP synthase subunit delta, mitochondrial	ATP5D	ATPase activity, proton-transport ATP synthase activity, transmembrane transport activity, transport activity	1.669
Malate dehydrogenase, mitochondrial	MDH2	L-malate dehydrogenase activity, malate dehydrogenase (NADP+) activity, protein self-association, RNA binding	1.664
Polyubiquitin-C	UBC	protease binding, RNA binding	1.650
High mobility group protein HMG-I/HMG-Y	HMGA1	5'-deoxyribose-5-phosphate lyase activity, AT DNA binding, chromatin binding, DNA-(apurinic or apyrimidinic site) lyase activity, DNA binding, enhancer sequence-specific DNA binding, enzyme binding, transcription factor binding	1.638
Calmodulin-like protein 3	CALML3	calcium ion binding	1.630
Alkaline phosphatase, placental type	ALPP	alkaline phosphatase activity, magnesium ion binding, zinc ion binding	1.607
Histone H1.5	HIST1H1B	chromatin binding, histone deacetylase binding, RNA binding	1.599
Electron transfer flavoprotein subunit beta	ETFB	electron carrier activity	1.596
Catenin beta-1	CTNNB1	alpha-catenin binding, androgen receptor binding, cadherin binding, chromatin binding, estrogen receptor binding, ion channel binding, nuclear hormone receptor binding, protein kinase binding, protein phosphokinase binding, I-SMAD binding, RNA polymerase II activating transcription factor binding, signal transducer activity, SMAD binding, transcription factor binding.	1.586
Histone H1.1	HIST1H1A	chromatin DNA binding, heparin binding	1.583
Histone H1.4	HIST1H1E	ADP binding, AMP binding, ATP binding, calcium ion binding, chromatin DNA binding, dATP binding, double-stranded DNA binding, GTP binding, RNA binding	1.567
Histone H1.3	HIST1H1D	chromatin DNA binding, RNA binding	1.567
Histone H1.2	HIST1H1C	chromatin DNA binding, RNA binding	1.567
Protein S100-A10	S100A10	calcium ion binding, ion channel binding, protein homodimerization activity	1.552
Protein timeless homolog	TIMELESS	protein heterodimerization activity, protein homodimerization activity	1.550
Eukaryotic translation initiation factor 5A-1	EIF5A	protein N-terminus binding, ribosome binding, RNA binding, translation elongation factor activity, U6 snRNA binding	1.540
ATP synthase subunit beta, mitochondrial	ATP5B	angiotensin binding, ATP binding, MHC class I protein binding, proton-transporting ATPase activity, proton-transport ATP synthase activity, transmembrane transporter activity, transporter activity	1.540
Rho GDP-dissociation inhibitor 2	ARHGD1B	GTPase activator activity, GTPase activity, Rac GTPase binding, Rho GDP-dissociation inhibitor activity	1.539
Bola-like protein 2	BOLA2	Acts as a cytosolic iron-sulfur (Fe-S) cluster assembly factor that facilitates [2Fe-2S] cluster insertion into a subset of cytosolic proteins	1.538
Glucosidase 2 subunit beta	PRKCSH	calcium ion binding, ion channel binding, phosphoprotein binding, protein kinase C binding	1.510
Glycogen phosphorylase, muscle form	PYGM	glycogen phosphorylase activity, nucleotide binding, pyridoxal phosphate binding	0.661
Lysine--tRNA ligase	KARS	amino acid binding, ATP binding, lysine-tRNA ligase activity, metal ion binding, tRNA binding	0.650
60S ribosomal protein L32	RPL32	RNA binding, structural constituent of ribosome	0.640
N-terminal Xaa-Pro-Lys N-methyltransferase 1	NTMT1	histone methyltransferase activity, N-terminal protein N-methyltransferase activity, protein methyltransferase activity	0.630
60S ribosomal protein L19	RPL19	5.8S rRNA binding, large ribosomal subunit rRNA binding, RNA binding, structural constituent of ribosome	0.618

Serine/threonine-protein phosphatase 6 regulatory subunit 2	PPP6R2	regulatory subunit of protein phosphatase 6 (PP6)	0.618
Zinc finger protein 28 homolog	ZFP28	DNA binding, metal ion binding, transcription factor activity, sequence-specific DNA binding	0.598
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15	ATP binding, ATP-dependent RNA helicase activity, double-stranded RNA binding, RNA binding, RNA helicase activity	0.593
Copine-3	CPNE3	calcium-dependent phospholipid binding, calcium-dependent protein binding, protein serine/threonine kinase activity, receptor tyrosine kinase binding, RNA binding, transporter activity	0.588
Importin subunit alpha-1	KPNA2	histone deacetylase binding, nuclear localization sequence binding, protein transporter activity, RNA binding	0.550
Alpha-2-HS-glycoprotein	AHSG	cysteine-type endopeptidase inhibitor activity, endopeptidase inhibitor activity, kinase inhibitor activity	0.543
Adenine phosphoribosyltransferase	APRT	adenine binding, adenine phosphoribosyltransferase activity, AMP binding	0.543
Patatin-like phospholipase domain-containing protein 7	PNPLA7	lysophospholipase activity	0.488
FACT complex subunit SSRP1	SSRP1	chromatin binding, DNA binding, RNA binding	0.467

**Appendix Table 3.4.6. 40 proteins identified as up/down regulated in response to CPNE3 knockdown.** The table shows a list of up/ down-regulated (>1.5-fold) proteins and their molecular functions.

Protein name	Gene name	Function	Expr Fold Change
Ribosomal protein S21	RPS21	protein N-terminus binding, RNA binding, structural constituent of ribosome	1.723
Ribosomal protein S15	RPS15	DNA binding, MDM2/MDM4 family protein binding, RNA binding, structural constituent of ribosome, ubiquitin ligase inhibitor activity	0.664
Actin related protein 2	ACTR2	actin binding, ATP binding, structural constituent of cytoskeleton	0.664
Archain 1	ARCN1	RNA binding	0.661
Dihydropyrimidinase like 2	DPYSL2	dihydropyrimidinase activity, identical protein binding, microtubule binding	0.659
Hypoxia up-regulated 1	HYOU1	ATPase activity, ATP binding, chaperone binding, unfolded protein binding	0.651
FKBP prolyl isomerase 1A	FKBP1A	activin binding, FK506 binding, ion channel binding, macrolide binding, peptidyl-prolyl cis-trans isomerase activity, SMAD binding, transforming growth factor beta receptor binding, type I transforming growth factor beta receptor binding	0.650
Prolyl endopeptidase	PREP	endopeptidase activity, oligopeptidase activity, serine-type endopeptidase activity, serine-type peptidase activity	0.638
Proteasome 20S subunit alpha 4	PSMA4	endopeptidase activity, threonine-type endopeptidase activity	0.634
Minichromosome maintenance complex component 3	MCM3	ATP binding, DNA binding, DNA helicase activity, DNA replication origin binding, hydrolase activity, single-stranded DNA binding	0.634
Sorting nexin 6	SNX6	dynactin binding, phosphatidylinositol binding, protein homodimerization activity	0.633
Tyrosyl-tRNA synthetase 1	YARS1	ATP binding, interleukin-8 receptor binding, RNA binding, tRNA binding, tyrosine-tRNA ligase activity	0.621
Sosondowah ankyrin repeat domain family member A	SOWAHA	n/a	0.616
Lactate dehydrogenase B	LDHB	identical protein binding, L-lactate dehydrogenase activity	0.613
Chloride nucleotide-sensitive channel 1A	CLNS1A	RNA binding	0.611
Staphylococcal nuclease and tudor domain containing 1	SND1	cadherin binding, endonuclease activity, endoribonuclease activity, nuclease activity, RISC complex binding, RNA binding, transcription coregulator activity	0.611
EH domain containing 1	EHD1	ATP binding, cadherin binding, calcium ion binding, GTP binding, identical protein binding, small GTPase binding	0.607
Non-POU domain containing octamer binding	NONO	chromatin binding, identical protein binding, nucleic acid binding, RNA binding, transcription regulatory region sequence-specific DNA binding	0.607
Far upstream element binding protein 3	FUBP3	mRNA binding, RNA binding, single-stranded DNA binding	0.604
Arrestin domain containing 3	ARRDC3	beta-3 adrenergic receptor binding	0.603

Ornithine aminotransferase	OAT	identical protein binding, ornithine(lysine) transaminase activity, ornithine-oxo-acid transaminase activity, pyridoxal phosphate binding	0.600
N-acetylated alpha-linked acidic dipeptidase 2	NAALAD2	carboxypeptidase activity, dipeptidase activity, dipeptidyl-peptidase activity, metal ion binding, metallocarboxypeptidase activity, serine-type peptidase activity	0.599
X-ray repair cross complementing 5	XRCC5	ATPase activity, acting on DNA, ATP binding, damaged DNA binding, DNA binding, DNA end binding, DNA helicase activity, double-stranded DNA binding, enzyme activator activity, protein-containing complex binding, protein C-terminus binding, RNA binding, telomeric DNA binding, transcription regulatory region sequence-specific DNA binding, U3 snoRNA binding, ubiquitin protein ligase binding	0.580
Dynein cytoplasmic 1 light intermediate chain 2	DYNC1LI2	ATP binding, dynein heavy chain binding, identical protein binding	0.580
Proteasome 20S subunit beta 6	PSMB6	cadherin binding, endopeptidase activity, threonine-type endopeptidase activity	0.576
Creatine kinase, mitochondrial 1B	CKMT1A	ATP binding, creatine kinase activity, kinase activity	0.571
Phosphofructokinase, muscle	PFKM	6-phosphofructokinase activity, AMP binding, ATP binding, fructose-6-phosphate binding, fructose binding, identical protein binding, kinase binding, metal ion binding, monosaccharide binding, protein C-terminus binding	0.564
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase non-catalytic subunit	DDOST	enzyme activator activity	0.558
Serpin family A member 2 (gene/pseudogene)	SERPINA2	serine-type endopeptidase inhibitor activity	0.557
RAB39B, member RAS oncogene family	RAB39B	GTPase activity, GTP binding, myosin V binding	0.557
Hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit beta	HADHB	3-hydroxyacyl-CoA dehydrogenase activity, acetyl-CoA C-acyltransferase activity, acetyl-CoA C-myristoyltransferase activity, enoyl-CoA hydratase activity, RNA binding	0.556
Replication protein A3	RPA3	damaged DNA binding, single-stranded DNA binding	0.540
Ubiquinol-cytochrome c reductase core protein 2	UQCRC2	metal ion binding, protein-containing complex binding	0.537
Dpy-30 histone methyltransferase complex regulatory subunit	DPY30	identical protein binding, protein homodimerization activity	0.535
Polypyrimidine tract binding protein 1	PTBP1	mRNA binding, poly-pyrimidine tract binding, pre-mRNA binding, RNA binding	0.535
Cullin associated and neddylation dissociated 1	CAND1	TBP-class protein binding	0.534
Phosphoribosyl pyrophosphate synthetase 1	PRPS1	ATP binding, identical protein binding, kinase activity, magnesium ion binding, protein homodimerization activity, ribose phosphate diphosphokinase activity	0.478
Ubiquitin C-terminal hydrolase L3	UCHL3	peptidase activity, thiol-dependent deubiquitinase, ubiquitin binding	0.437
Actinin alpha 2	ACTN2	actin filament binding, calcium ion binding, cytoskeletal protein binding, FATZ binding, identical protein binding, integrin binding, ion channel binding, LIM domain binding, nuclear receptor coactivator activity, phosphatidylinositol-4,5-bisphosphate binding, protein domain specific binding, structural constituent of muscle, titin binding, titin Z domain binding	0.421
Copine 3	CPNE3	calcium-dependent phospholipid binding, calcium-dependent protein binding, metal ion binding, protein serine/threonine kinase activity, receptor tyrosine kinase binding, RNA binding	0.319

**Appendix Table 3.5.1. 52 proteins identified by PCA and K-means clustering of TMT LC-MS/MS data.** The table shows a list of proteins with a high expression similarity score to CPNE3 in cluster 4 and their molecular functions.

Protein name	Gene name	Function	Expr Fold Change
Glutamine--tRNA ligase	QARS	ATP binding; glutamate-tRNA ligase activity; GTPase binding; identical protein binding; proline-tRNA ligase activity; protein homodimerization activity; RNA stem-loop binding; zinc ion binding	0.813
Podocan	PODN	collagen binding	0.778
Splicing factor, proline- and glutamine-rich	SFPQ	chromatin binding; DNA binding; histone deacetylase binding; protein homodimerization activity; RNA binding; transcription cis-regulatory region binding	0.774
Mitogen-activated protein kinase kinase kinase 4	MAP4K4	ATP binding; creatine kinase activity; microtubule binding; protein serine/threonine/tyrosine kinase activity; protein serine/threonine kinase activity; protein serine kinase activity	0.762
Macrophage migration inhibitory factor	MIF	chemoattractant activity; cytokine activity; cytokine receptor binding; dopachrome isomerase activity; identical protein binding; phenylpyruvate tautomerase activity; protease binding	0.758
AP-1 complex subunit beta-1	AP1B1	clathrin binding; protein kinase binding	0.758
Eukaryotic initiation factor 4A-III	EIF4A3	ATP binding; ATP hydrolysis activity; mRNA binding; poly(A) binding; ribonucleoprotein complex binding; RNA binding; RNA helicase activity; RNA stem-loop binding; selenocysteine insertion sequence binding; translation regulator activity	0.757
Eukaryotic translation initiation factor 4H	EIF4H	cadherin binding; ribosomal small subunit binding; RNA binding; RNA strand annealing activity; RNA strand-exchange activity; translation factor activity, RNA binding; translation initiation factor activity	0.757
Putative ribosomal RNA methyltransferase NOP2	NOP2	RNA binding; rRNA (cytosine-C5-)-methyltransferase activity	0.756
DNA-directed RNA polymerase II subunit RPB2	POLR2B	chromatin binding; DNA binding; DNA-directed 5'-3' RNA polymerase activity; metal ion binding; ribonucleoside binding; RNA binding	0.754
Clathrin light chain A	CLTA	clathrin heavy chain binding; GTPase binding; peptide binding; protein-containing complex binding; structural molecule activity	0.754
Splicing factor 3B subunit 2	SF3B2	RNA binding	0.751
60S ribosomal protein L7	RPL7	DNA binding; identical protein binding; mRNA binding; RNA binding; structural constituent of ribosome	0.748
60S ribosomal protein L6	RPL6	cadherin binding; DNA binding; RNA binding; structural constituent of ribosome	0.746
Threonine--tRNA ligase, cytoplasmic	TARS	ATP binding; identical protein binding; threonine-tRNA ligase activity; tRNA binding; zinc ion binding	0.741
Glucose-6-phosphate 1-dehydrogenase	G6PD	glucose-6-phosphate dehydrogenase activity; glucose binding; identical protein binding; NADP binding; protein homodimerization activity	0.735
60S ribosomal protein L27	RPL27	RNA binding; structural constituent of ribosome	0.732

Histidine triad nucleotide-binding protein 1	HINT1	adenosine 5'-monophosphoramidase activity; hydrolase activity; nucleotide binding; protein kinase C binding; SUMO-specific isopeptidase activity	0.731
DNA replication licensing factor MCM7	MCM7	ATP binding; ATP hydrolysis activity; DNA helicase activity; single-stranded DNA binding	0.729
Small nuclear ribonucleoprotein-associated proteins B and B'	SNRPB	histone pre-mRNA DCP binding; RNA binding; telomerase RNA binding; U1 snRNP binding; U2 snRNP binding	0.722
Serine/arginine-rich splicing factor 3	SRSF3	phospholipase binding; RNA binding	0.719
Histidine--tRNA ligase, cytoplasmic	HARS	ATP binding; histidine-tRNA ligase activity; identical protein binding; protein homodimerization activity	0.719
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	actin binding; ATP binding; chromatin binding; chromatin DNA binding; DNA binding; double-stranded DNA binding; double-stranded RNA binding; identical protein binding; mRNA 3'-UTR binding; poly(A) binding; poly(C) RNA binding; poly(G) binding; pre-mRNA binding; promoter-specific chromatin binding; protein-containing complex binding; ribonucleoprotein complex binding; RNA binding; RNA polymerase II cis-regulatory region sequence-specific DNA binding; RNA polymerase II complex binding; RNA polymerase II C-terminal domain binding; sequence-specific double-stranded DNA binding; single-stranded DNA binding; single-stranded RNA binding; snRNA binding; telomerase RNA binding; TFIIH-class transcription factor complex binding; transcription corepressor activity	0.703
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	chromatin binding; histone deacetylase binding; minor groove of adenine-thymine-rich DNA binding; mRNA 3'-UTR AU-rich region binding; RNA binding; telomeric DNA binding	0.698
Small nuclear ribonucleoprotein Sm D3	SNRPD3	enzyme binding; histone pre-mRNA DCP binding; RNA binding; telomerase RNA binding; U7 snRNA binding	0.692
Serine/threonine-protein phosphatase 2A activator	PPP2R4	ATP binding; peptidyl-prolyl cis-trans isomerase activity; protein homodimerization activity; protein phosphatase 2A binding; protein phosphatase regulator activity; protein tyrosine phosphatase activator activity; signaling receptor binding	0.692
SUMO-activating enzyme subunit 2	UBA2	ATP binding; magnesium ion binding; protein heterodimerization activity; small protein activating enzyme binding; SUMO activating enzyme activity; SUMO binding; transferase activity; ubiquitin-like protein conjugating enzyme binding	0.691
Gem-associated protein 4	GEMIN4	ribonucleoprotein complex binding	0.688
Vacuolar protein sorting-associated protein 26A	VPS26A	unknown	0.688
60S ribosomal protein L28	RPL28	RNA binding; structural constituent of ribosome	0.685
DNA replication licensing factor MCM3	MCM3	ATP binding; ATP hydrolysis activity; DNA binding; DNA helicase activity; single-stranded DNA binding	0.684
Xaa-Pro dipeptidase	PEPD	manganese ion binding; metalloaminopeptidase activity; metallocarboxypeptidase activity; peptidase activity; proline dipeptidase activity	0.681
Serpin H1	SERPINH1	collagen binding; RNA binding; serine-type endopeptidase inhibitor activity; unfolded protein binding	0.677
Eukaryotic translation initiation factor 4 gamma 1	EIF4G1	ATP binding; eukaryotic initiation factor 4E binding; identical protein binding; molecular adaptor activity; mRNA binding; RNA binding; translation factor activity, RNA binding; translation initiation factor activity; translation initiation factor binding	0.674
UMP-CMP kinase	CMPK1	ATP binding; CMP kinase activity; cytidylate kinase activity; dCMP kinase activity; nucleoside diphosphate kinase activity; nucleoside monophosphate kinase activity; UMP kinase activity; uridine kinase activity	0.670

Glycogen phosphorylase, muscle form	PYGM	glycogen phosphorylase activity; linear malto-oligosaccharide phosphorylase activity; nucleotide binding; pyridoxal phosphate binding; SHG alpha-glucan phosphorylase activity	0.661
Lysine--tRNA ligase	KARS	amino acid binding; ATP adenyltransferase activity; ATP binding; identical protein binding; lysine-tRNA ligase activity; protein homodimerization activity; tRNA binding	0.650
60S ribosomal protein L32	RPL32	RNA binding; structural constituent of ribosome	0.640
Ubiquitin carboxyl-terminal hydrolase 5	USP5	cysteine-type endopeptidase activity; thiol-dependent deubiquitinase; ubiquitin binding; zinc ion binding	0.633
N-terminal Xaa-Pro-Lys N-methyltransferase 1	NTMT1	histone methyltransferase activity; methyltransferase activity; N-terminal protein N-methyltransferase activity; protein methyltransferase activity	0.630
Chromatin assembly factor 1 subunit A	CHAF1A	chromatin binding; chromo shadow domain binding; identical protein binding; unfolded protein binding	0.614
U6 snRNA-associated Sm-like protein LSM5	LSM5	protein heterodimerization activity; RNA binding	0.601
Argininosuccinate synthase	ASS1	amino acid binding; argininosuccinate synthase activity; ATP binding; identical protein binding; RNA binding; toxic substance binding	0.595
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15	ATP binding; ATP hydrolysis activity; double-stranded RNA binding; RNA binding; RNA helicase activity	0.593
Copine-3	CPNE3	calcium-dependent phospholipid binding; calcium-dependent protein binding; metal ion binding; protein serine/threonine kinase activity; receptor tyrosine kinase binding; RNA binding	0.588
Olfactory receptor 11A1	OR11A1	G protein-coupled receptor activity; olfactory receptor activity	0.57
Importin subunit alpha-1	KPNA2	histone deacetylase binding; nuclear import signal receptor activity; nuclear localization sequence binding; RNA binding	0.55
Alpha-2-HS-glycoprotein	AHSG	cysteine-type endopeptidase inhibitor activity; endopeptidase inhibitor activity; kinase inhibitor activity	0.543
Adenine phosphoribosyltransferase	APRT	adenine binding; adenine phosphoribosyltransferase activity; AMP binding	0.543
5'-3' exoribonuclease 2	XRN2	3'-5'-exoribonuclease activity; 5'-3' exonuclease activity; 5'-3' exoribonuclease activity; identical protein binding; metal ion binding; nuclease activity; RNA binding; transcription termination site sequence-specific DNA binding	0.531
FACT complex subunit SSRP1	SSRP1	DNA binding; histone binding; nucleosome binding; RNA binding	0.467
Eukaryotic translation initiation factor 3 subunit F	EIF3F	identical protein binding; isopeptidase activity; metal-dependent deubiquitinase activity; thiol-dependent deubiquitinase; translation initiation factor activity; translation initiation factor binding	0.451

**Appendix Table 3.5.4. 44 proteins identified by PCA and K-means clustering of Label Free LC-MS/MS data.** The table shows a list of proteins with a high expression similarity score to CPNE3 in cluster 1 and their molecular functions.

Protein name	Gene name	Function	Expr Fold Change
Copine-3	CPNE3	calcium-dependent phospholipid binding; calcium-dependent protein binding; metal ion binding; protein serine/threonine kinase activity; receptor tyrosine kinase binding; RNA binding	0.319
Alpha-actinin-2	ACTN2	actin filament binding; calcium ion binding; cytoskeletal protein binding; FAT2 binding; identical protein binding; integrin binding; LIM domain binding; nuclear receptor coactivator activity; phosphatidylinositol-4,5-bisphosphate binding; protein domain specific binding; structural constituent of muscle; titin binding; titin Z domain binding; transmembrane transporter binding	0.421
26S proteasome regulatory subunit 10B	PSMC6	ATP binding; ATP hydrolysis activity; identical protein binding; proteasome-activating activity; protein-macromolecule adaptor activity	0.266
Ubiquitin carboxyl-terminal hydrolase isozyme L3	UCHL3	NEDD8-specific protease activity; peptidase activity; thiol-dependent deubiquitinase; ubiquitin binding	0.437
ATP-dependent RNA helicase A	DHX9	3'-5' DNA/RNA helicase activity; 3'-5' DNA helicase activity; 3'-5' RNA helicase activity; ATP binding; ATP hydrolysis activity; chromatin DNA binding; DNA binding; DNA helicase activity; DNA replication origin binding; double-stranded DNA binding; double-stranded RNA binding; importin-alpha family protein binding; metal ion binding; mRNA binding; nucleoside-triphosphatase activity; nucleoside-triphosphate diphosphatase activity; polysome binding; promoter-specific chromatin binding; regulatory region RNA binding; RISC complex binding; RNA binding; RNA helicase activity; RNA polymerase binding; RNA polymerase II cis-regulatory region sequence-specific DNA binding; RNA polymerase II complex binding; RNA polymerase II-specific DNA-binding transcription factor binding; RNA stem-loop binding; sequence-specific mRNA binding; single-stranded 3'-5' DNA helicase activity; single-stranded DNA binding; single-stranded RNA binding; siRNA binding; transcription coactivator activity; transcription coregulator activity; triplex DNA binding	0.438
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	mRNA 5'-UTR binding; mRNA binding; RNA binding	0.543
Catenin alpha-1	CTNNA1	actin filament binding; beta-catenin binding; cadherin binding; gamma-catenin binding; identical protein binding; RNA binding; structural molecule activity; vinculin binding	0.387
Putative heat shock protein HSP 90-alpha A5	HSP90AA5P	ATP binding; ATP hydrolysis activity; unfolded protein binding	0.518
Keratin_type I cytoskeletal 16	KRT16	structural constituent of cytoskeleton	0.396
Eukaryotic translation initiation factor 2 subunit 1	EIF2S1	ribosome binding; RNA binding; translation initiation factor activity	0.362
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	actin binding; ATP binding; chromatin binding; chromatin DNA binding; DNA binding; double-stranded DNA binding; double-stranded RNA binding; identical protein binding; mRNA 3'-UTR binding; poly(A) binding; poly(C) RNA binding; poly(G) binding; pre-mRNA binding; promoter-specific chromatin binding; protein-containing complex binding; ribonucleoprotein complex binding; RNA binding; RNA polymerase II cis-regulatory region sequence-specific DNA binding; RNA polymerase II complex binding; RNA polymerase II C-terminal domain binding; sequence-specific double-stranded DNA binding; single-stranded DNA binding; single-stranded RNA binding; snRNA binding; telomerase RNA binding; TFIIF-class transcription factor complex binding; transcription corepressor activity	0.483
Protein Niban 2	NIBAN2	cadherin binding; transcription coactivator activity	0.350
Elongation factor 2	EEF2	cadherin binding; GTPase activity; GTP binding; protein kinase binding; ribosome binding; RNA binding; translation elongation factor activity	0.437



Anterior gradient protein 3	AGR3	dystroglycan binding	0.515
NEDD8	NEDD8	protein tag; ubiquitin protein ligase binding	0.903
Methionine aminopeptidase 2	METAP2	aminopeptidase activity; metal ion binding; metalloaminopeptidase activity; metalloexopeptidase activity; RNA binding	0.576
A-kinase anchor protein 12	AKAP12	adenylate cyclase binding; calmodulin binding; protein kinase A binding	0.554
Heterogeneous nuclear ribonucleoprotein H	HNRNPH1	identical protein binding; poly(U) RNA binding; RNA binding	0.480
RNA-binding protein 3	RBM3	ribosomal large subunit binding; RNA binding	0.422
RE1-silencing transcription factor	REST	chromatin binding; DNA-binding transcription factor activity; DNA-binding transcription repressor activity, RNA polymerase II-specific; identical protein binding; metal ion binding; RNA polymerase II cis-regulatory region sequence-specific DNA binding; RNA polymerase II core promoter sequence-specific DNA binding; RNA polymerase II-specific DNA-binding transcription factor binding; transcription cis-regulatory region binding	0.343
Putative UPF0607 protein ENSP00000383783	3 SV	unknown	0.432
Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2	endopeptidase inhibitor activity; hyaluronic acid binding; serine-type endopeptidase inhibitor activity	0.507
Formin-1	FMN1	actin binding; microtubule binding; SH3 domain binding	0.470
Adenylate kinase 2_ mitochondrial	AK2	adenylate kinase activity; ATP binding	0.482
Putative protein FAM10A4	ST13P4	heat shock protein binding; protein dimerization activity	0.438
Armadillo repeat-containing protein 8	ARMC8	proteasomal degradation of the transcription factor HBP1	0.549
Platelet-activating factor acetylhydrolase IB subunit alpha2	PAFAH1B2	1-alkyl-2-acetyl glycerophosphocholine esterase activity; platelet-activating factor acetyltransferase activity; protein-containing complex binding; protein heterodimerization activity; protein homodimerization activity	0.524
Acetyl-CoA acetyltransferase_ mitochondrial	ACAT1	acetyl-CoA C-acetyltransferase activity; C-acetyltransferase activity; coenzyme A binding; enzyme binding; identical protein binding; potassium ion binding	0.555
Glycine N-acyltransferase-like protein 3	GLYATL3	glycine N-acyltransferase activity; N-acyltransferase activity	0.553
Ankyrin repeat and SOCS box protein 2	ASB2	cullin family protein binding	0.577
Dihydrolipoyl dehydrogenase_ mitochondrial	DLD	dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding	0.578
Protein Atg16L2	ATG16L2	may play a role in regulating epithelial homeostasis in an ATG16L1-dependent manner.	0.567
S-methyl-5'-thioadenosine phosphorylase	MTAP	1,4-alpha-oligoglucan phosphorylase activity; S-methyl-5-thioadenosine phosphorylase activity	0.524
Interleukin-18	IL18	cytokine activity; interleukin-18 receptor binding	0.641
Endoplasmic reticulum resident protein 44	ERP44	protein disulfide isomerase activity	0.638
DnaJ homolog subfamily A member 2	DNAJA2	ATPase activator activity; ATP binding; chaperone binding; Hsp70 protein binding; metal ion binding; unfolded protein binding	0.591

Jupiter microtubule associated homolog 1	JPT1	plays a role in the regulation of cell cycle and cell adhesion	0.647
Protein-L-isoaspartate(D-aspartate) O-methyltransferase	PCMT1	cadherin binding; protein-L-isoaspartate (D-aspartate) O-methyltransferase activity	0.624
Cadherin-5	CDH5	beta-catenin binding; BMP receptor binding; cadherin binding; calcium ion binding; fibrinogen binding; protein phosphatase binding; protein tyrosine kinase binding; signaling receptor binding; transmembrane transporter binding; vascular endothelial growth factor receptor 2 binding	0.609
Heterogeneous nuclear ribonucleoprotein R	HNRNPR	mRNA binding; RNA binding	0.631
Thioredoxin-like protein 1	TXNL1	disulfide oxidoreductase activity	0.620
Ubiquitin carboxyl-terminal hydrolase 13	USP13	BAT3 complex binding; chaperone binding; cysteine-type endopeptidase activity; Lys48-specific deubiquitinase activity; proteasome binding; thiol-dependent deubiquitinase; ubiquitin binding; ubiquitin-like protein ligase binding; ubiquitin protein ligase binding; zinc ion binding	0.636
Vitamin D-binding protein	GC	actin binding; calcidiol binding; vitamin D binding; vitamin transmembrane transporter activity	0.692
Alpha-enolase	ENO1	cadherin binding; DNA-binding transcription repressor activity, RNA polymerase II-specific; GTPase binding; magnesium ion binding; phosphopyruvate hydratase activity; protein homodimerization activity; RNA binding; RNA polymerase II transcription regulatory region sequence-specific DNA binding; transcription corepressor activity; transcription corepressor binding	0.739

## 7.2. Supplementary data tables & code files

All supplementary data has been provided separately in the respective data formats.