

The University of Nottingham

Deciphering a molecular mechanism underlying lymphovascular invasion in breast cancer

Abrar Ibrahim Aljohani
B.Sc., M.Sc.

Thesis submitted to the University of Nottingham for the Degree of Doctorate of Philosophy

February 2022

Department of Pathology, School of Medicine

Abstract

Background: Despite advancements in the early detection of breast cancer (BC) and improved access to therapeutic options, the global mortality rate due to BC among women remains high. BC progression is a complex and multistep process. Lymphovascular invasion (LVI), one of the earliest stages of metastasis in BC, is controlled by specialised molecular pathways and is strongly associated with poor outcomes. LVI signifies the expansion of tumour cells towards the peritumoural region inside lymphatic spaces and blood vessels. A successful LVI event requires the involvement of a variety of cellular, biophysical, and biochemical pathways. Because the underlying molecular mechanisms of LVI are complex and involve connections with other oncogenic and metastatic pathways, no targeted treatments are currently available to prevent the progression of LVI and metastatic recurrence. Investigation into the molecular determinants of LVI is critical in elucidating the molecular mechanisms underlying LVI, discovering novel prognostic factors, and identifying possible treatment targets for invasive BC. The genes that play a role in the pathogenesis of LVI in invasive BC have received little attention; it was therefore hypothesised that by conducting bioinformatic analyses, key genes related to LVI in BC could be identified, some of which may have prognostic value, serve as potential therapeutic targets, and have roles in the various mechanisms regulating the development of LVI. This study aimed to identify genes associated with LVI, evaluate the clinicopathological and prognostic significance of the identified genes at the mRNA and protein levels, and investigate their functional roles.

Methodologies: Using artificial neural network (ANN) analysis and the weighted average difference (WAD) method on two transcriptome cohorts with well-characterised LVI status (Molecular Taxonomy of Breast Cancer International Consortium [METABRIC] and the Cancer Genome Atlas [TCGA]), key genes were identified. The candidate genes *MX1*, *IDH2*, *HMGB3*, *RRM2*, *CCNB1*, and *CCNB2* were each assessed at the transcriptome level using the METABRIC and TCGA data (n=1980 and

n=854, respectively) and at the proteome level via immunohistochemical (IHC) analysis and tissue microarray (TMA) sections derived from a large, well-characterised invasive BC cohort (n=3,173). The expression of these genes was evaluated with clinicopathological parameters that included LVI, tumour grade, lymph node (LN) stage, tumour size, patient outcome, BC molecular subtypes, and relevant biological markers. In a different approach, *in vitro* functional assays using a panel of BC cell lines were performed to investigate the effect of transient siRNA knockdown of *HMGB3* and *CCNB2* on cell proliferation, migration, cell cycle, and apoptosis. Furthermore, the effect of silencing *HMGB3* and *CCNB2* on the adherence and transmigration of BC cell lines across dermal lymphatic endothelial cells (DLECs) and human vascular endothelial cells (HUVECs) was evaluated to assess their role in LVI development.

Results: In this study, a 99-gene set associated with LVI status in invasive BC was generated using WAD analysis. In addition, ANN analysis produced a list of 100 genes of interest. From these gene lists, *MX1*, *IDH2*, *HMGB3*, *RRM2*, *CCNB1*, and *CCNB2* were selected for further investigation. All selected genes were significantly associated with LVI positivity at the mRNA level in both the METABRIC and TCGA cohorts, but *MX1*, *RRM2*, and *CCNB1* were not associated with LVI at the protein level. All biomarkers included in this study had a significant association with poor clinicopathological parameters, such as high tumour grade, poor NPI, hormonal receptor negativity, and HER2 positivity at the mRNA and/or protein levels. The results of the survival analysis varied among these biomarkers. High expression of these genes at the mRNA level was significantly associated with poor outcomes; however, only *IDH2*, *HMGB3*, and *CCNB2* expression at the protein level was associated with shorter BC-specific survival (BCSS). In multivariate analysis, the expression of MX1, IDH2, HMGB3, and CCNB2 were found to be factors that predicted patient survival independently from the other variables, including tumour size, histological grade, LN stage, and LVI. Additionally, all the selected biomarkers were significantly associated with EMT- and MMP-related markers, such as E-cadherin, N-cadherin, P-

cadherin, TWIST2, MMP1, MMP9, and MMP12 at the transcriptome and proteome levels.

In vitro assessment of *HMGB3* and *CCNB2* revealed the possible functional role of *HMGB3* in LVI development in BC. These preclinical investigations showed that silencing *HMGB3* and *CCNB2* in BC cell lines reduced cell proliferation and migration, enhanced apoptosis, and, importantly, reduced adherence and transmigration across endothelial cell lines. Silencing *HMGB3* arrested the cell cycle in G1 phase, while silencing *CCNB2* arrested the cell cycle in S phase.

Conclusion: Identifying key drivers of LVI and deciphering the underlying molecular mechanisms of this process is essential to preventing early spread of the disease and to developing a personalised approach to the management of patients with early-stage BC. This thesis demonstrated that the techniques of WAD and ANN could reveal key LVI genes with prognostic importance in invasive BC. The potential roles of these genes in the progression of BC were also explored. Furthermore, this study found that targeting *HMGB3* and *CCNB2* significantly reduced BC cell adhesion and transmigration across lymphatic and vascular endothelial cells, indicating the role of these genes in driving LVI. Targeting these genes may represent a novel approach to enhancing quality of life, health, and well-being in BC patients with aggressive metastatic disease. Exploration of the signalling pathways associated with LVI in invasive BC is fundamental to identifying the novel regulatory molecules controlling LVI and thereby developing effective therapeutic candidates and improved guidelines for metastatic cancer therapy. Further functional *in vitro* and *in vivo* investigations may provide new insights into the exact functional and molecular pathways of these biomarkers of LVI.

Publications

1. **ALJOHANI, A. I.**, JOSEPH, C., KUROZUMI, S., MOHAMMED, O. J., MILIGY, I. M., GREEN, A. R. & RAKHA, E. A. 2020. Myxovirus resistance 1 (MX1) is an independent predictor of poor outcome in invasive breast cancer. *Breast Cancer Res Treat,* 181**,** 541-551.

2. **ALJOHANI, A. I.**, TOSS, M. S., KUROZUMI, S., JOSEPH, C., ALESKANDARANY, M. A., MILIGY, I. M., ANSARI, R. E., MONGAN, N. P., ELLIS, I. O., GREEN, A. R. & RAKHA, E. A. 2020. The prognostic significance of wild-type isocitrate dehydrogenase 2 (IDH2) in breast cancer. *Breast Cancer Res Treat,* 179**,** 79-90.

3. **ALJOHANI, A. I.**, TOSS, M. S., El-SHARAWY, K., MIRZA, S., GRAHAM, B., GREEN, A. R. & RAKHA, E. A. 2022. Upregulation of cyclin B2 (*CCNB2*) in breast cancer contributes to the development of lymphovascular invasion. *Am J Cancer Res,* 12(2), 469-489.

4. **ALJOHANI, A. I.**, TOSS, M. S., GREEN, A. R. & RAKHA, E. A. The clinical significance of cyclin B1 (*CCNB1*) in invasive breast cancer with emphasis on its contribution to lymphovascular invasion development. ***(In preperation)***.

5. **ALJOHANI, A. I.**, TOSS, M. S., GREEN, A. R. & RAKHA, E. A. Clinicopathological significance of ribonucleotide reductase enzyme 2 (*RRM2*) expression in invasive breast cancer ***(In preperation)***.

6. **ALJOHANI, A. I.**, ALSAEED, S., TOSS, M. S., GREEN, A. R. & RAKHA, E. A. The prognostic significance and the potential pathogenetic role of HMGB3 in invasive breast cancer. ***(In preperation)***.

7. KARIRI, Y. A., ALSALEEM, M., JOSEPH, C., ALSAEED, S., **ALJOHANI, A.**, SHIINO, S., MOHAMMED, O. J., TOSS, M. S., GREEN, A. R. & RAKHA, E. A. 2021. The prognostic significance of interferon-stimulated gene 15 (ISG15) in invasive breast cancer. *Breast Cancer Res Treat,* 185**,** 293-305.

8. ELSHARAWY, K. A., ALTHOBITI, M., MOHAMMED, O. J., **ALJOHANI, A. I**., TOSS, M. S., GREEN, A. R. & RAKHA, E. A. 2021. Nucleolar protein 10 (NOP10) predicts poor prognosis in invasive breast cancer. *Breast Cancer Res Treat,* 185**,** 615-627.

9. JOSEPH, C., ALSALEEM, M. A., TOSS, M. S., KARIRI, Y. A., ALTHOBITI, M., ALSAEED, S., **ALJOHANI, A. I.**, NARASIMHA, P. L., MONGAN, N. P., GREEN, A. R. & RAKHA, E. A. 2020. The ITIM-containing receptor: leukocyte-associated immunoglobulin-like receptor-1 (LAIR-1) modulates immune response and confers poor prognosis in invasive breast carcinoma. *Cancers (Basel),* 13.

10. KUROZUMI, S., ALSAEED, S., ORAH, N., MILIGY, I. M., JOSEPH, C., **ALJOHANI, A**., TOSS, M. S., FUJII, T., SHIRABE, K., GREEN, A. R., ALESKANDARANY, M. A. & RAKHA, E. A. 2020. Clinicopathological significance of lipocalin 2 nuclear expression in invasive breast cancer. *Breast Cancer Res Treat,* 179**,** 557-564.

11. KUROZUMI, S., JOSEPH, C., SONBUL, S., ALSAEED, S., KARIRI, Y., **ALJOHANI, A**., RAAFAT, S., ALSALEEM, M., OGDEN, A., JOHNSTON, S. J., ALESKANDARANY, M. A., FUJII, T., SHIRABE, K., CALDAS, C., ASHANKYTY, I., DALTON, L., ELLIS, I. O., DESMEDT, C., GREEN, A. R., MONGAN, N. P. & RAKHA, E. A. 2019. A key genomic subtype associated with lymphovascular invasion in invasive breast cancer. *Br J Cancer,* 120**,** 1129-1136.

Conference presentations

As the main presenter:

Plenary oral presentation 'Cyclin B2 is a key gene in lymphovascular invasion in breast cancer.' 13th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, July 6–8, 2021, Manchester, United Kingdom.

Abrar Aljohani, Graham Ball, Andrew R. Green, Emad A. Rakha

Oral presentation 'Prognostic significance of isocitrate dehydrogenase 2 (IDH2): Associated with lymphovascular invasion in breast cancer.' 12th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, July 2–4, 2019, Leeds, United Kingdom.

Abrar Aljohani, Sasagu Kurozumi, Chitra Joseph, Mohammed A. Aleskandarany, Sami Al saeed, Yousif Kariri, Pavan Lakshmi Narasimha, Ian O. Ellis, Andrew R. Green, Emad A. Rakha

Poster presentation 'Potential pathogenesis role of high mobility group protein 3 (HMGB3) in lymphovascular invasion in breast cancer.' VIRTUAL 2022 4th Joint Winter Meeting of the Pathological Society and the Royal Society of Medicine, January 25–26, 2022, United Kingdom.

Abrar I Aljohani, Andrew R. Green, Emad A. Rakha

Poster presentation 'Prognostic significance of high mobility group protein 3 (HMGB3): A biomarker associated with lymphovascular invasion in breast cancer.' UK Interdisciplinary Breast Cancer Symposium 2020, January 27–28, 2020, ICC Birmingham, United Kingdom.

Abrar I Aljohani, Ian O. Ellis, Andrew R. Green, Emad A. Rakha

As a co-author:

Oral presentation 'Leukocyte-associated immunoglobulin-like receptor-1 confers poor prognosis in invasive breast carcinoma: Transcriptomic driven study.' UK Interdisciplinary Breast Cancer Symposium 2020, January 27–28, 2020, ICC Birmingham, United Kingdom.

Chitra Joseph, Micheal Toss, Sasagu Kurozumi, Mansour Alsaleem, Yousif Kariri, Pavan Lakshmi Narasimha, Sami Alsaeed, **Abrar Aljohani,** Nigel Mongan, Mohammed Aleskandarany, Andrew Green, Ian Ellis, Emad Rakha

Poster presentation 'Prognostic significance of cell division cycle associated 5 (CDCA5): A biomarker associated with lymphovascular invasion in breast cancer.' 12th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, July 2–4, 2019, Leeds, United Kingdom.

Yousif Ahmad Kariri, Chitra Joseph, Mansour Alsaleem, **Abrar Aljohani**, Sami Alsaeed, Pavan Lakshmi Narasimha, Nigel P. Mongan, Ian O. Ellis, Mohammed Aleskandarany, Andrew R. Green, Emad A. Rakha

Poster presentation 'Keratin 24 (KRT 24) confers poor patient outcome in invasive breast cancer.' 12th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, July 2–4, 2019, Leeds, United Kingdom.

Yousif Kariri, Chitra Joseph, Sasagu Kurozumi, Islam Miligy, Sami Alsaeed, **Abrar Aljohani**, Pavan Narasimha, Ian O. Ellis, Nigel Mongan, Mohammed Aleskandarany, Andrew R. Green, Emad A. Rakha

Acknowledgements

First, the most special thanks to Allah, the Almighty, for giving me the opportunity and ability to complete this project and earn this achievement.

I would like to express my heartfelt appreciation and gratitude to my supervisors, Professor Emad Rakha, Doctor Andy Green, and Professor Ian Ellis, for their valuable support, motivation, immense knowledge, crucial contributions, and sage advice throughout my PhD. Their truly scientific intuition has made them steadfast supporters with boundless ideas and passion for science. I am extremely appreciative in every possible way and hope to continue our collaboration in the future.

Moreover, I would like to thank Dr. Michael Toss, Dr. Sasagu Kurouzmi, Dr. Chitra Joseph, Dr. Omar Mohammed, Mr. Christopher Nolan, and Mrs. Holly Nicholls for their valuable technical support. I am also deeply grateful to my PhD colleagues and the entire breast pathology research group for being a part of all the joyful moments and achievements throughout my study.

With my most warm and heartfelt appreciation, I would like to dedicate this work to my mother, Mrs. Mona Abbas, whose dreams for me brought about this accomplishment. Without her love, her praying for me all the time, and sending me her blessings, I would not be where I am today. My special gratitude to my late father, Mr. Ibrahim Aljohani. I wish you could see this achievement. I would also like to thank my brothers for caring and praying for me, which is what has sustained me thus far.

Words are not enough to thank the biggest sources of my power: my husband, Dr. Turki Alqurashi, and my son, Omar, for helping me in the difficult moments, standing by me, encouraging me on to follow my dreams of obtaining this degree, and supporting me throughout my whole life. Thank you from the bottom of my heart.

Finally, my sincerest appreciation to my sponsor in the Kingdom of Saudi Arabia, Taif University, and the Saudi Arabian Cultural Bureau in the United Kingdom, for the financial support for my PhD study.

Thank you.

List of abbreviations

αKG	Alpha-ketoglutarate
µg	Microgram
µl	Microlitre
µM	Micromole
2-HG	2-hydroxyglutarate
A	
ANN	Artificial neural network
ASCO	American Society of Clinical Oncology
B	
BC	Breast cancer
BCSS	Breast cancer-specific survival
BSA	Bovine serum albumin
BVI	Blood vascular invasion
C	
CAFs	Cancer-associated fibroblasts
CCL21	C-C motif chemokine 21
CCNB1	Cyclin B1
CCNB2	Cyclin B2
CCNE1	Cyclin-E1
CCR7	C-C chemokine receptor type 7
CDC2	Cell division control
CDKs	Cyclin-dependent kinases
CK	Cytokeratin
CNA	Copy number aberration
CO_2	Carbon dioxide
COX2	Cyclooxygenase 2
CSF-1	Colony-stimulating factor 1
CTNNB1	Catenin beta 1
D	
DAB	3,3’-diaminobenzidine
DEGs	Differentially expressed genes
DFS	Disease-free survival
DLECs	Dermal lymphatic endothelial cells
DMFS	Distant metastasis-free survival
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E	
E-cadherin	Epithelial cadherin
ECM	Extracellular matrix
EGFR	Epidermal growth factor receptor
EMT	Epithelial-mesenchymal transition
ER	Oestrogen receptor

ESCC	Oesophageal squamous cell carcinoma
F	
FBS	Foetal bovine serum
FFPE	Formalin-fixed paraffin-embedded
G	
GSK3B	Glycogen synthase kinase 3 beta
GTP	Guanosine triphosphate
H	
H&E	Haematoxylin and eosin
HER2	Human epidermal growth factor receptor 2
HMG	High mobility group protein
HMGB	High mobility group box
HMGB3	High mobility group box 3
HR	Hormone receptor
HR-	Hormone receptor-negative
HR+	Hormone receptor-positive
HRP	Horseradish peroxidase
H-score	Histological scoring
HUVECs	Human umbilical vein endothelial cells
I	
ICC	Intra-class correlation coefficient
IDC	Invasive ductal carcinoma
IDC of NST	IDC of no specific type
IDH2	Isocitrate dehydrogenase 2
IFN	Interferon
IgG	Immunoglobulin G
IHC	Immunohistochemistry
ILC	Invasive lobular carcinoma
IMPC	Invasive micropapillary carcinoma
IMS	Industrial methylated spirits
IntClust	Integrative clusters
ISGs	Interferon-stimulated genes
K	
kDa	Kilodalton
L	
LN	Lymph node
LVI	Lymphovascular invasion
M	
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
miRNA	microRNA
ml	Millilitre
ML	Machine learning
mm	Millimetre

MMP	Metalloproteinase
mRNA	Messenger ribonucleic acid
MT1	Membrane type 1
MX1	Myxovirus resistance 1
N	
NADH	Nicotinamide adenine dinucleotide hydrogen
NADPH	Nicotinamide adenine dinucleotide phosphate hydrogen
N-cadherin	Neural cadherin
NF-kB	Nuclear factor kappa light chain-enhancer of activated B cells
NGS	Next-generation sequence
NHSB	Nottingham Health Science Biobank
nm	Nanometre
NPC	Nasopharyngeal carcinoma
NPI	Nottingham prognostic index
NPS	Nottingham primary series
O	
OD	Optical density
OS	Overall survival
P	
p53	Protein 53
PAM50	Prediction Analysis of Microarray 50
PBS	Phosphate-buffered saline
P-cadherin	Placental cadherin
PD-L1	Programmed cell death ligand 1
PE	Plating efficiency
pERK1/2	Phosphorylated extracellular-regulated kinase 1/2
PI	Propidium iodide
pmol	Picomole
PR	Progesterone receptor
PS	Phosphatidylserine
R	
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNaseA	Ribonuclease A
RNR	Ribonucleotide reductase
RPMI	Roswell Park Memorial Institute
RRM2	Ribonucleotide reductase enzyme 2
RT-qPCR	Quantitative reverse transcription-polymerase chain reaction
S	
SF	Surviving fraction
siRNA	Small double-stranded interfering RNA
SPSS	Statistical Package for the Social Sciences
SUMO	Small ubiquitin-like modifier protein

T	
TBS	Tris-buffered saline
TBS-T	TBS containing 0.1% Tween 20
TCA	Tricarboxylic acid
TCGA	The Cancer Genome Atlas
TGFβ1	Transforming growth factor beta 1
TILs	Tumour-infiltrating lymphocytes
TMA	Tissue microarray
TME	Tumour microenvironment
TN	Triple-negative
TNF-α	Tumour necrosis factor α
TTDM	Time to distant metastasis
TWIST2	Twist-related protein 2
U	
uPA	Urokinase plasminogen activator
V	
VEGF	Vascular endothelial growth factor
VEGFR3	Vascular endothelial growth factor receptor 3
VI	Vascular invasion
VM	Vascular mimicry
W	
WAD	Weighted average difference
WB	Western blotting
WebGestalt	WEB-based GEne SeT AnaLysis Toolkit
Z	
ZEB1	Zinc finger E-box binding homeobox 1

Table of contents

Chapter 1: Introduction 1
1. General introduction to the breast and breast cancer 2
1.1 Anatomical and histological structure of normal breast tissue 2
1.2 Development of invasive BC 2
1.3 BC incidence and mortality 4
1.4 Risk factors for BC 5
1.4.1 Gender 5
1.4.2 Age 5
1.4.3 Oestrogen exposure 5
1.4.4 Family history 6
1.4.5 Obesity, alcohol consumption, and smoking 6
1.4.6 Benign breast disorders 6
1.5 Prognostic and predictive factors of BC 6
1.5.1 Prognostic factors 7
1.5.1.1 Histological type 7
1.5.1.2 Tumour grade 7
1.5.1.3 Tumour size 8
1.5.1.4 Lymph node (LN) stage 9
1.5.1.5 NPI 9
1.5.1.6 LVI 10
1.5.2 Predictive factors 12
1.5.2.1 Hormone receptors: ER and PR 12
1.5.2.2 HER2 12
1.6 Molecular classification of BC 13
1.7 Molecular mechanisms underlying LVI in invasive BC 16
1.7.1 Introduction to LVI 16
1.7.2 Morphological assessment of LVI 17
1.7.3 Tumour histological characteristics and LVI 19
1.7.4 Vascular or lymphatic invasion 20
1.7.5 Tumour invasiveness, tumour microenvironment (TME), and LVI 21
1.7.6 Gene signatures, gene expression profiling, and LVI 25
1.7.7 Proteomics and LVI 27
1.7.8 Unified mechanism of LVI 29
1.7.9 Conclusions 30
1.8 Hypothesis and aims 30
1.8.1 Hypothesis 30

1.8.2 Aims 31
1.8.3 Significance of the project 31
Chapter 2: Material and methods 32
2.1 Study cohorts 33
2.1.1 Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort 33
2.1.2 Nottingham subset from METABRIC cohort 34
2.1.3 The Cancer Genome Atlas (TCGA) cohort 36
2.1.4 Nottingham Primary Series (NPS) cohort of invasive BC 37
2.1.5 Biomarker data 40
2.2 Determination of binding specificity for selected antibodies with Western blotting (WB) 41
2.2.1 Overview 41
2.2.2 Cell line lysates 41
2.2.3 Cell culture and preparation of cell lysates 42
2.2.4 Cell lysate protein quantification (Pierce™ BCA protein assay) 43
2.2.5 Preparation of WB reagents and tools 43
2.2.6 Sample preparation 44
2.2.7 Loading of samples and separation of the denatured protein mixture 45
2.2.8 Protein transfer onto nitrocellulose membranes 45
2.2.9 Blocking and antibody probing 46
2.2.10 Visualisation of protein bands 46
2.3 Tissue handling and processing 47
2.3.1 FFPE block retrieval from the Nottingham Health Science Biobank (NHSB). 47
2.3.2 TMA construction 47
2.3.2.1 Sectioning in the microtome 47
2.3.2.2 Preparation of donor blocks and TMA construction 48
2.4 Immunohistochemical staining 49
2.4.1 Overview 49
2.4.2 IHC tools, chemicals, and procedures 50
2.4.2.1 Labelling of slides 50
2.4.2.2 Deparaffinisation and rehydration of tissue sections 50
2.4.2.3 Heat-induced epitope retrieval (HIER) using citrate buffer at pH 6... 51
2.4.2.4 Preparation of sodium citrate buffer at pH 6 51
2.4.2.5 Blocking of endogenous peroxidases 52
2.4.2.6 Blocking non-specific reacting proteins 53
2.4.2.7 Applying the optimised primary antibody 53

2.4.2.8 Application of the post-primary antibody 53
2.4.2.9 Application of NovolinkTM polymer 54
2.4.2.10 Detection with DAB 54
2.4.2.11 Counterstaining of tissue slides with haematoxylin and tissue dehydration 54
2.4.3 Optimisation of primary antibodies 55
2.4.4 Negative and positive controls 55
2.5 Imaging of tissue sections and the histological scoring (H-score) system 57
2.5.1 Whole-slide imaging 57
2.5.2 H-score 57
2.5.3 Identification of the cut-off points of IHC staining 58
2.6 Cell culture (*in vitro*) 59
2.6.1 Cell lines, general growth conditions, and media 59
2.6.1.1 BC cell lines 59
2.6.1.2 Endothelial cell lines 60
2.6.1.3 Subculturing of adherent cell lines 61
2.6.1.4 Cryopreservation and thawing of cell lines 61
2.6.2 Small interfering RNA (siRNA) transfection of HMGB3 and CCNB2 62
2.6.2.1 Overview 62
2.6.2.2 Cell line selection 63
2.6.2.3 Subculture of the selected cell line and cell plating 63
2.6.2.4 Preparation of transfection reagents 64
2.6.2.5 Adding transfection reagents 65
2.6.2.6 Cell lysate collection and protein quantification for transfected cells 65
2.6.2.7 Analysing the transfection results 66
2.6.3 CellTitre 96 Aqueous One Solution Cell Proliferation Assay (MTS assay) 66
2.6.3.1 Overview and principles 66
2.6.3.2 MTS assay procedure 66
2.6.4 Clonogenic survival assay 67
2.6.4.1 Overview 67
2.6.4.2 Clonogenic assay procedure 67
2.6.5 Wound-healing assay 68
2.6.5.1 Overview 68
2.6.5.2 Migration assay procedure 69
2.6.6 Cell cycle assay 70
2.6.6.1 Overview and principles 70

2.6.6.2 Cell cycle assay procedure 70
2.6.7 Apoptosis assay 71
2.6.7.1 Overview and principles 71
2.6.7.2 Apoptosis assay procedure 72
2.6.8 *In vitro* models of tumour–endothelium interactions 73
2.6.8.1 Overview 73
2.6.8.2 Adhesion assay 73
2.6.8.2.1 Adding a coating agent 73
2.6.8.2.2 Seeding endothelial cells 73
2.6.8.2.3 Labelling and seeding BC tumour cells 74
2.6.8.3 Transmigration assay 75
2.6.8.3.1 Adding a coating agent 75
2.6.8.3.2 Seeding endothelial cells 75
2.6.8.3.3 Labelling and seeding BC tumour cells 76
2.7 Statistical analysis 77
2.7.1 Introduction 77
2.7.2 Descriptive statistics 77
2.7.3 Intra-class correlation coefficient (ICC) 78
2.7.4 Univariate and correlation analyses 78
2.7.5 Survival analysis 79
2.8 Ethical approval 79
Chapter 3: Bioinformatic analysis for gene selection 80
3.1 Introduction 81
3.2 WAD analysis for ranking the differential gene expression of transcriptomic cohorts 82
3.3 ANN analysis of the transcriptomic cohorts 82
3.4 Hypothesis 83
3.5 Aims 83
3.6 Methods 83
3.6.1 Study cohorts 83
3.6.1.1 METABRIC cohort 83
3.6.1.2 TCGA cohort 83
3.6.1.3 WAD bioinformatic analysis for DEGs 84
3.6.1.4 ML for biomarker set enrichment 85
3.6.1.5 Pathway analysis using an accessible online resource 87
3.7 Results 88
3.7.1 Genes associated with LVI from WAD analysis 88
3.7.2 ANN analysis of transcriptomic cohorts to discover genes associated with LVI 89

3.8 Validation of identified gene signatures from the bioinformatic analysis 91
3.8.1 Validation of identified gene signatures from the WAD bioinformatic analysis. ... 91
3.8.2 Validation of gene signatures identified from ANN bioinformatic analysis. ... 92
3.9 Pathway analysis results ... 94
3.10 Gene selection criteria for protein expression studies ... 95
3.11 Discussion ... 96
3.12 Conclusion ... 98
Chapter 4: The prognostic significance of MX1 and IDH2 in invasive breast cancer ... 100
4.1 Introduction ... 101
4.1.1 MX1 ... 102
4.1.2 IDH2 ... 103
4.2 Hypothesis ... 104
4.3 Aims ... 104
4.4 Methods ... 105
4.4.1 Study cohort ... 105
4.4.1.1 METABRIC cohort ... 105
4.4.1.2 TCGA cohort ... 105
4.4.1.3 Nottingham BC cohort ... 105
4.4.2 WB ... 106
4.4.3 IHC ... 106
4.5 Results ... 107
4.5.1 Validation of MX1 and IDH2 antibodies by WB ... 107
4.5.2 Expression of MX1 and IDH2 in BC ... 107
4.5.3 Association of *MX1* and *IDH2* with clinicopathological parameters ... 111
4.5.3.1 MX1 and IDH2 mRNA expression ... 111
4.5.3.2 MX1 and IDH2 protein expression ... 111
4.5.4 *MX1* and *IDH2* expression in molecular BC subtypes ... 112
4.5.5 *MX1* and *IDH2* expression and other associated biomarkers ... 119
4.5.5.1 Associations with MX1 ... 119
4.5.5.2 Associations with IDH2 ... 119
4.5.5.3 Associations between MX1 and IDH2 ... 120
4.5.6 Association of *MX1* and *IDH2* with patient outcome ... 123
4.6 Discussion ... 133
4.6.1 MX1 ... 133
4.6.2 IDH2 ... 135
4.7 Conclusion ... 138

Chapter 5: The prognostic significance and potential pathogenetic role of HMGB3 in invasive BC 139
5.1 Introduction 140
5.2 Hypothesis 141
5.3 Aims 141
5.4 Methods 141
5.4.1 Study cohort 141
5.4.1.1 METABRIC cohort 141
5.4.1.2 TCGA cohort 142
5.4.1.3 Nottingham BC cohort 142
5.4.2 WB 142
5.4.3 IHC 142
5.4.4 *In vitro* studies of *HMGB3* 143
5.4.4.1 BC tumour cells 143
5.4.4.2 Endothelial cells 143
5.4.4.3 Silencing HMGB3 using siRNA 143
5.4.4.4 MTS assay 144
5.4.4.5 Colony formation assay 144
5.4.4.6 Wound-healing assay 144
5.4.4.7 Flow cytometric analysis of the cell cycle 144
5.4.4.8 Annexin V-based flow cytometric analysis of apoptosis 144
5.4.4.9 In vitro assessment of lymphatic and blood vessel invasion . 145
5.4.4.9.1 Static adhesion assay 145
5.4.4.9.2 Transmigration assay 145
5.5 Results 145
5.5.1 Validation of HMGB3 antibody by WB 145
5.5.2 Expression of *HMGB3* in BC 146
5.5.3 Association of *HMGB3* with clinicopathological parameters 149
5.5.3.1 HMGB3 mRNA expression 149
5.5.3.2 HMGB3 protein expression 149
5.5.4 *HMGB3* expression in molecular BC subtypes 149
5.5.5 *HMGB3* expression and other associated biomarkers 152
5.5.6 Association of *HMGB3* with patient outcomes 154
5.5.7 Preclinical studies (*in vitro* investigation of the mechanistic role of *HMGB3* in LVI-related biological processes) 159
5.5.7.1 Silencing of HMGB3 expression using siRNA 159
5.5.7.2 HMGB3 silencing suppresses cell proliferation and cell growth among the selected BC cell lines 161

5.5.7.3 HMGB3 silencing suppresses cell migration among the selected BC cell lines 161
5.5.7.4 HMGB3 silencing arrests cells in the G1 phase 164
5.5.7.5 HMGB3 silencing enhances cell apoptosis 164
5.5.7.6 The role of HMGB3 in the adhesion of BC cells to endothelial cells..... 167
5.5.8.7 The role of HMGB3 in the transmigration of BC cells through endothelial cells 167
5.6 Discussion 170
Chapter 6: RRM2, CCNB1, and CCNB2 as key drivers of LVI in invasive breast cancer: prognostic significance and *in vitro* investigation 174
6.1 Introduction 175
6.1.1 RRM2 175
6.1.2 CCNB1 176
6.1.3 CCNB2 177
6.2 Hypothesis 177
6.3 Aims 178
6.4 Methods 178
6.4.1 Study cohort 178
6.4.1.1 METABRIC cohort 178
6.4.1.2 TCGA cohort 178
6.4.1.3 Nottingham BC cohort 179
6.4.2 WB 179
6.4.3 IHC 179
6.4.4 *In vitro* assessment of *CCNB2* 180
6.4.4.1 BC tumour cells 180
6.4.4.2 Endothelial cells 180
6.4.4.3 Transfection of CCNB2 using siRNA 181
6.4.4.4 Proliferation assay 181
6.4.4.5 Clonogenic assay 181
6.4.4.6 Wound-healing assay 181
6.4.4.7 Cell cycle assay 182
6.4.4.8 Apoptosis assay 182
6.4.4.9 Static adhesion assay 182
6.4.4.10 Transmigration assay 182
6.5 Results 183
6.5.1 Validation of RRM2, CCNB1, and CCNB2 antibodies by WB 183
6.5.2 Expression of *RRM2*, *CCNB1*, and *CCNB2* in BC 183
6.5.3 Association of *RRM2*, *CCNB1*, and *CCNB2* with clinicopathological parameters 189

6.5.3.1 RRM2, CCNB1, and CCNB2 mRNA expression 189
6.5.3.2 RRM2, CCNB1, and CCNB2 protein expression 189
6.5.4 *RRM2*, *CCNB1*, and *CCNB2* expression in molecular BC subtypes 199
6.5.5 *RRM2*, *CCNB1*, and *CCNB2* expression and other associated biomarkers 200
6.5.5.1 Associations with RRM2 200
6.5.5.2 Associations with CCNB1 202
6.5.5.3 Associations with CCNB2 202
6.5.6 Associations between *RRM2*, *CCNB1*, and *CCNB2* 203
6.5.7 Association between *RRM2*, *CCNB1*, and *CCNB2* and patient outcome 205
6.5.8 Preclinical studies (*in vitro* investigation of the role of *CCNB2* in LVI-related biological processes) 220
6.5.8.1 Suppression of CCNB2 expression using siRNA 220
6.5.8.2 Downregulation of CCNB2 suppresses cell proliferation and cell growth in the selected BC cell lines 222
6.5.8.3 Downregulation of CCNB2 suppresses cell migration in the selected BC cell lines 222
6.5.8.4 Downregulation of CCNB2 arrests cells in S phase 225
6.5.8.5 Downregulation of CCNB2 enhances cell apoptosis 225
6.5.8.6 CCNB2 enhances BC cell adhesion to endothelial cells 228
6.5.8.7 CCNB2 enhances BC cell transmigration to endothelial cells . 228
6.6 Discussion 231
6.6.1 RRM2 231
6.6.2 CCNB1 234
6.6.3 CCNB2 237
6.7 Conclusion 240
Chapter 7: General discussion and conclusion 242
7.1 General discussion 243
7.2 Major findings of the study 244
7.3 Conclusion 247
7.4 Strengths of the study 248
7.5 Limitations of the study 248
7.6 Future work 249
7.7 Impact of COVID-19 250
References 251
Appendices 272

List of tables

Table 2.1 Frequency of clinical characteristics of the METABRIC cohort (n=1,980) and Nottingham subsets from the METABRIC cohort (n=288). 35
Table 2.2 Frequency of clinical features of the TCGA cohort (n=854). 37
Table 2.3 Clinicopathological features of the Nottingham Primary Series of invasive BC (n=3,173). 39
Table 2.4 Antibody clones, suppliers, and cut-off values for previously examined biomarkers in the Nottingham cohort. 41
Table 2.5 Details of the primary antibodies used in this study..... 56
Table 2.6 Characteristics of BC cell lines used in *in vitro* experiments. 60
Table 3.1 List of 99 genes correlated with LVI status in the METABRIC cohort.....88
Table 3.2 Top 100 concordant genes from the METABRIC and TCGA cohorts associated with LVI, based on ANN analysis. 90
Table 3.3 Validation of the association between 42 upregulated genes identified via WAD analysis and the presence of LVI. 92
Table 3.4 Validation of the association between the 100 genes identified via ANN analysis and LVI status. 93
Table 3.5 Pathway analysis results from WebGestalt, used to identify differentially regulated pathways in 42 upregulated genes and the top 100 concordant genes (adapted from http://www.webgestalt.org/).....95
Table 4.1 Statistical associations between *MX1* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets.....114
Table 4.2 Statistical associations between *IDH2* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets 115
Table 4.3 Statistical associations between MX1 protein expression and clinicopathological factors in the Nottingham BC cohort (n=901). 116
Table 4.4 Statistical associations between IDH2 protein expression and clinicopathological factors in the Nottingham BC cohort (n=859). 118
Table 4.5 Correlations between *MX1* expression and mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes. 121
Table 4.6 Correlations between *IDH2* expression and mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes. 122

Table 4.7 Correlations between *MX1* and *IDH2* in mRNA and protein expression levels. 122
Table 4.8 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *MX1* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort. 131
Table 4.9 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *IDH2* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort. 132
Table 5.1 Statistical associations between *HMGB3* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets..... 150
Table 5.2 Statistical associations between HMGB3 protein expression and clinicopathological factors in the Nottingham BC cohort (n=1,647). 151
Table 5.3 Correlations between HMGB3 expression and mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes. 153
Table 5.4 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *HMGB3* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort. 158
Table 6.1 Statistical associations between *RRM2* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets..... 191
Table 6.2 Statistical associations between *CCNB1* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets. 192
Table 6.3 Statistical associations between *CCNB2* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets. 193
Table 6.4 Statistical associations between RRM2 protein expression and clinicopathological factors in the Nottingham BC cohort (n=1,940). 194
Table 6.5 Statistical associations between CCNB1 protein expression and clinicopathological factors in the Nottingham BC cohort (n=2,480). 196
Table 6.6 Statistical associations between CCNB2 protein expression and clinicopathological factors in the Nottingham BC cohort (n=3,178). 198
Table 6.7 Correlations of RRM2 expression with mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes. 201

Table 6.8 Correlations of CCNB1 expression with mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes. 203
Table 6.9 Correlations of CCNB2 expression with mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes. 204
Table 6.10 Correlations between *RRM2*, *CCNB1*, and *CCNB2* mRNA and protein expression levels..... 205
Table 6.11 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *RRM2* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort. 217
Table 6.12 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *CCNB1* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort. 218
Table 6.13 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *CCNB2* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort. 219

List of figures

Figure 1.1 True lymphovascular invasion (LVI) as an illustration, and on a histological specimen with haematoxylin and eosin (H&E stain ×20), CD31, and CD34 staining. 11
Figure 2.1 Experimental workflow for a wound-healing assay using the ibidi Culture-Insert 2 Well.....70
Figure 2.2 Illustration of staining with Annexin V and propidium iodide (PI) by flow cytometry.. 72
Figure 2.3 Illustration of the transwell migration model.. 77
Figure 3.1 Schematic diagram represents WAD bioinformatic analysis for DEGs.....85
Figure 3.2 Workflow of machine learning strategy grounded in an artificial neural network. Based on the genomic expression data obtained from mRNA datasets, ANN-based neural data mining was used to identify an enriched set of concordant biomarkers associated with LVI.....87
Figure 3.3 Overlapping differentially expressed genes (DEGs) from the weighted average difference (WAD) and artificial neural network (ANN) analysis. 94
Figure 4.1 Western blot (WB) reflecting antibody specificity of MX1 and IDH2.....107
Figure 4.2 The distribution of the expression of *MX1* and *IDH2* mRNA and protein.... 109
Figure 4.3 Cytoplasmic expression of MX1 and IDH2 protein in invasive breast cancer. 110
Figure 4.4 Kaplan–Meier survival plots showing the association between *MX1* mRNA and protein expression and breast cancer-specific survival (BCSS) and total time to distant metastasis (TTDM). 127
Figure 4.5 Kaplan–Meier survival plots showing the association between *IDH2* mRNA and protein expression and breast cancer-specific survival (BCSS) and total time to distant metastasis (TTDM). 130
Figure 5.1 Western blot (WB) reflecting the antibody specificity of HMGB3.....146
Figure 5.2 The distribution of the expression of *HMGB3* mRNA and protein.....147
Figure 5.3 Nuclear expression of HMGB3 protein in invasive breast cancer. 148
Figure 5.4 Kaplan–Meier survival plots showing the association between *HMGB3* mRNA and protein expression and breast cancer-specific survival (BCSS) and total time to distant metastasis (TTDM).....157
Figure 5.5 HMGB3 protein expression in breast cancer (BC) cell lines 160
Figure 5.6 The effect of *HMGB3* knockdown (KD) by siRNA on cell proliferation and the colony formation assays.....162
Figure 5.7 The effect of *HMGB3* knockdown (KD) by siRNA on cell migration 163
Figure 5.8 The effect of siRNA knockdown (KD) of *HMGB3* on the cell cycle..... 165

Figure 5.9 The effect of siRNA knockdown (KD) of *HMGB3* on apoptosis.....166
Figure 5.10 Representative photomicrographs of tumour cell adhesion across vascular and lymphatic endothelial cells (HUVECs and DLECs). 168
Figure 5.11 Representative photomicrographs of tumour cell (MCF-7, ZR-75-1, and SK-BR-3) transmigration across HUVECs and DLECs.. 169
Figure 6.1 Western blot (WB) reflecting the antibody specificity of RRM2, CCNB1, and CCNB2.....183
Figure 6.2 The distribution of the expression of *RRM2*, *CCNB1,* and *CCNB2* mRNA and protein..... 185
Figure 6.3 Cytoplasmic expression of RRM2, CCNB1, and CCNB2 protein in invasive breast cancer..... 188
Figure 6.4 Kaplan–Meier survival plots showing the association between *RRM2* mRNA and protein expression and breast cancer-specific survival and total time to distant metastasis (TTDM) 210
Figure 6.5 Kaplan–Meier survival plots showing the association between *CCNB1* mRNA and protein expression and breast cancer-specific survival (BCSS) and total time to distant metastasis (TTDM)..... 213
Figure 6.6 Kaplan–Meier survival plots showing the association between *CCNB2* mRNA and protein expression and breast cancer-specific survival (BCSS) and total time to distant metastasis (TTDM).....216
Figure 6.7 CCNB2 protein expression in breast cancer (BC) cell lines..... 221
Figure 6.8 The effect of knockdown (KD) *CCNB2* by siRNA on cell proliferation and colony formation assays.....223
Figure 6. 9 The effect of knockdown (KD) CCNB2 by siRNA on cell migration..... 224
Figure 6. 10 The effect of siRNA knockdown (KD) of *CCNB2* on the cell cycle 226
Figure 6.11 The effect of siRNA knockdown (KD) of *CCNB2* on apoptosis 227
Figure 6.12 Representative photomicrographs of tumour cell adhesion across vascular and lymphatic endothelial cells (HUVECs and DLECs). 229
Figure 6.13 Representative photomicrographs of tumour cell (MCF-7, ZR-75-1, and SK-BR-3) transmigration across HUVECs and DLECs. 230

Chapter 1: Introduction

1. General introduction to the breast and breast cancer

1.1 Anatomical and histological structure of normal breast tissue

The understanding of normal, healthy breast structure and its ductal lobar anatomy provides insight into the origins of breast cancer (BC) [1]. Terminal duct lobular units (TDLUs) are lactating functional units of origin for epithelial neoplasms. The structure of the breast is composed of three components: subcutaneous tissue, skin, and breast tissue (parenchyma and stroma). The parenchyma is compartmentalised into ~20 lobes converging radially at the nipple. The lobes have ducts that converge into ~10 collecting ducts that open at the nipple area [1]; these ducts form the glandular tissue and are supported by stromal tissue. The breasts are suspended by the ligaments of Cooper, supporting ducts, and glandular tissue, and span from the second to the sixth ribs, and spreading to the anterior axillary line from the edge of the sternum [2]. The ligaments of Cooper are hollow, cone-shaped, fibrous tissue projections that contain breast tissue. The areola is composed of smooth muscles organised in concentric rings and outwardly in subcutaneous tissue [2]. Thick skin with corrugations encloses the nipple and close to its apex are the openings of the lactiferous ducts. An inner layer of cuboidal or columnar epithelial cells and an outer layer of myoepithelial cells line the ducts and ductules, which are all anchored on a basement membrane layer. The interlobular and intralobular regions contain fibro-adipose connective tissue with various cellular components, including fibroblasts, myofibroblasts, adipocytes, histiocytes, and occasional lymphocytes [3, 4]. Breast lymphatics drain mainly into the internal breast and axillary lymph nodes [5].

1.2 Development of invasive BC

Breast tissue is capable of producing a wide variety of benign and malignant lesions. Although these lesions are most frequently found in the ducts (70–80%), they can also occur in the lobules (10–15%) or in the tissue between the ducts and lobules [6]. Uncontrolled cell division or hyperplasia can

result in typical ductal or lobular hyperplasia that can lead to pre-invasive lesions, such as ductal carcinoma *in situ* (DCIS) or lobular carcinoma *in situ* (LCIS) [7, 8]. Lesions *in situ* have a high risk of breaching the myoepithelial layer and basement membrane, ultimately resulting in invasive BC. Malignant transformation is guided by highly sophisticated systems that result in significant cytological or architectural dysplasia and stimulate cells to evade mechanisms that regulate their survival, proliferation, and migration. This malignant transformation entails a series of multistep processes referred to as 'cancer hallmarks', which include evasion of tumour suppressors, proliferation of signalling pathways, activation of replication immortality, reversal of apoptosis, metastasis, metabolic reprogramming, and suppression of immunity [9]. Neoplastic lesions in BC are characterised by exhibiting different degrees of differentiation and clinical prognosis. The lesions initially reach a pre-invasive stage that progresses in association with tumour grading of biological, molecular, and genetic pathways. While progressing from pre-invasive to invasive BC, molecular modifications occur in myoepithelial and tumour stromal cells [10]. The majority of BCs are of the invasive ductal subtype, where tumour growth occurs in the epithelium from the nipple ends of the lactiferous ducts, spreading to the tubules at the terminal units [11]. The different BC types develop because the tumours initiate from different epithelial cell subtypes (details will be discussed in Section 1.5.1.1). Some BCs can be aggressive, rapidly metastasising to other parts of the body, while others grow slowly and respond differently to treatments. Cellular heterogeneity in tumours occurs when neoplastic events develop alongside genetic modifications, epigenetic changes, and the microenvironment [11, 12].

Some reports attribute the invasiveness of BC to specific molecular subtypes that play a role in hormonal signalling pathways [10]. It has been shown that luminal B, which is an aggressive subtype of invasive BC, is more common in younger patients [13]. Another report stated that, in younger patients, invasive BC of the luminal B subtype showed a strong

association with lymphovascular invasion (LVI) [14]. In patients with invasive BC, the presence of LVI is an unfavourable event that has an impact on patient survival (details will be discussed in Section 1.5.1.6). The exact nature of LVI is not yet understood, as BC is a disease with varying molecular and biological features. Therefore, LVI is the focus of this thesis, and the molecular mechanisms controlling LVI are discussed later in this chapter.

1.3 BC incidence and mortality

BC is the most common non-cutaneous malignant disease in women worldwide. Approximately 13% of all women will be diagnosed with BC in their lifetime [15]. BC has a much higher incidence in women than other cancers in both transitioned (55.9 per 100,000) and transitioning (29.7 per 100,000) countries. BC among women surpassed lung cancer as the main cause of cancer-related deaths globally in 2020, with an estimated 2.3 million new cases accounting for 11.7% of all cancer-related deaths [16]. In the UK, BC affects approximately 55,000 women and 370 men each year, which amounts to the diagnosis of approximately 150 cases daily. Each year, more than 11,500 women and 85 men in the UK die of BC, representing nearly 1,000 deaths per month and 31 deaths per day [17]. By 2025, it is expected that the incidence of BC in women will rise, and 20 million cases could be reported, especially in less developed countries with low incomes [18].

Based on the 2020 Global Cancer Statistics, BC was the most common cancer in developed and developing countries, with 794,000 and 883,000 cases reported in these countries, respectively. In developing countries, BC was the leading cause of death, with 324,000 deaths; in developed countries, it was the second most prevalent cause of cancer-related deaths among women, with 198,000 deaths [16, 18]. Early detection, avoidance of risk factors and increasing awareness of BC could reduce the number of new cases and improve treatment options.

1.4 Risk factors for BC

Predisposing factors for BC vary from person to person and depend on environmental or hereditary factors [15]. Some of these factors are controllable, while others, such as gender, genetics and age, are not. The risk factors for BC are discussed below.

1.4.1 Gender

Gender is the most influential predisposing factor for BC. Although men can develop BC, it is almost 100 times more prevalent in women. This is usually because women's mammary tissue is more exposed to hormones, such as oestrogen, that promote abnormal cell differentiation [15].

1.4.2 Age

Age is another important risk factor associated with an increased risk of BC. Approximately 85% of women who develop BC are over 50 years old, while around 5% of cases occur in women below the age of 40 years [19]. However, aggressive phenotypes of invasive BC, such as human epidermal growth factor receptor 2 (HER2)-enriched and triple-negative (TN), are more common in younger patients and are strongly associated with poor outcomes [20].

1.4.3 Oestrogen exposure

The initiation and progression of BC are strongly associated with hormonal factors. Oestrogen is the main hormone that regulates many biological functions in women. The risk of developing carcinoma of the breast increases with increased exposure to oestrogen from oral contraceptives or menopausal hormonal therapies [21, 22]. A form of progesterone, which is a combination of oestrogen and medroxyprogesterone (MPA), regulates ovulation and has also been shown to increase the risk of invasive BC in menopausal women [23].

1.4.4 Family history

Women with a family history of ovarian or breast carcinoma are at a higher risk of developing BC. One of the major causes of BC is the hereditary germline mutation of genes, commonly those encoding breast cancer type 1 susceptibility protein (*BRCA1*) and breast cancer type 2 susceptibility protein (*BRCA2*). Other genetic variations may also influence the risk of developing BC. Family history-related risks account for about 5–10% of all cases of BC [19].

1.4.5 Obesity, alcohol consumption, and smoking

Obesity is another risk factor for BC in postmenopausal women. This link could be due to the production of aromatase-driven oestrogen in adipose tissue and the increased secretion of inflammatory cytokines, such as insulin-like growth factor I and circulating insulin, which are associated with BC development [24, 25]. Furthermore, women who consume alcohol or smoke cigarettes have a higher risk of developing BC than other women. The risk of developing BC increases with the amount of alcohol consumed and with long-term smoking [26, 27].

1.4.6 Benign breast disorders

Women who have undergone a biopsy showing ductal cell hyperplasia have a higher risk of developing BC, especially if the cells appeared atypical. These benign breast disorders include benign growth, mastitis, and fat necrosis [15].

1.5 Prognostic and predictive factors of BC

Studies have defined various prognostic factors for the identification of aggressive tumours. Any tumour feature that can predict BC patient outcome, such as BC-specific survival (BCSS), disease-free survival (DFS), and distant metastasis-free survival (DMFS), is defined as a prognostic factor. Any clinicopathological BC characteristic that can estimate the potential for a response to a particular therapy is called a predictive factor

[28]. Identifying precise predictive and prognostic factors can be helpful in the selection of treatment methods and can improve patient outcomes. The following sub-sections introduce the common prognostic and predictive factors of BC.

1.5.1 Prognostic factors

1.5.1.1 Histological type

BC has a variety of histological subtypes, each with a unique clinical course and outcome. The most frequent type of BC is invasive ductal carcinoma (IDC), which comprises a variety of tumours with different morphologies. Some have distinct cytoarchitectural characteristics and are categorised as 'special types', such as tubular, medullary-like, invasive cribriform, and mucinous; however, the majority of tumours (75%) are classified as IDCs of no specific type (NST) [29]. The second most frequent type of BC, invasive lobular carcinoma (ILC), accounts for 5–15% of all breast carcinomas and has several histological variations [30]. Mucinous, tubular, tubulolobular, and invasive cribriform subtypes have a good prognosis, with survival rates of more than 80% at 10 years. In contrast, IDCs of NST, solid lobular carcinomas, and mixed ductal and lobular carcinomas have a poor prognosis, with survival rates of less than 50% at 10 years [31].

1.5.1.2 Tumour grade

BC and normal breast cells are described based on the extent of the resemblance between them via a three-tier grading scheme. A prognostic indicator for such an assessment is tumour grading, which describes how BC cells in a cancerous lesion are organised. The tumour grading system categorises BC into three different types [32]:

1. Grade I: BC cells are well differentiated, with slow growth, and appear similar to normal cells.
2. Grade II: BC cells are well differentiated, with fast growth, and do not appear similar to normal cells.

3. Grade III: BC cells are poorly differentiated, with fast growth, and the appearance of these cells differs markedly from that of normal breast cells [32].

Tumour grading is calculated based on three components: tubule or acinus formation is the primary component, nuclear pleomorphism is the second component, and mitotic count is the third component. The scoring measures are as follows [32, 33]:

1. Tubule or acinus formation: score = 1 (more than 75% of the tumour), score = 2 (moderate degree, 10–75% of the tumour), score = 3 (less than 10% of the tumour).
2. Nuclear pleomorphism: score = 1 (presence of small or regular uniform cells), score = 2 (variability and moderate size), score = 3 (marked variability).
3. Mitotic count: score = 1 (less than 9), score = 2 (between 10 and 19), score = 3 (more than 20).

Light microscopy with a field diameter of 0.58 mm and 10 high-power fields is a suitable method for the analysis of tumour cell mitosis. When digitally visualising a tumour section, the image size should be 512 × 512 μm [34]. A tumour grade is assigned by adding the values of each component of a case: an excellent prognosis is represented by a score of 3, 4, or 5, and is assigned tumour grade I; moderate prognosis is represented by a score of 6 or 7 and is assigned tumour grade II; and poor prognosis is represented by a score of 8 or 9 and is assigned tumour grade III [32, 33]. For ease of evaluation and understanding of the BC level, the tumour grading system defined above was integrated with the Nottingham prognostic index (NPI) [34].

1.5.1.3 Tumour size

Tumour size is another essential prognostic factor for predicting the behaviour of the BC and the patient's survival status. It is an important factor for studying invasive BC as there is a direct correlation between tumour size and increased mortality in patients with invasive BC, with

patients with smaller tumour size having a better prognosis than those with larger tumour size [35-37]. Tumour size is also included in the NPI [34].

1.5.1.4 Lymph node (LN) stage

LN stage is also a strong prognostic factor in invasive BC. BCSS and DFS are firmly correlated with LN staging during the early stages of BC [38, 39]. This study classifies tumours into three groups based on LN staging: when no LN involvement is present, BC is considered LN stage I; LN stage II is considered if there are one, two, or three LNs positive with invasive tumour cells; and the presence of more than three positive LNs is classified as LN stage III [40, 41]. There is a direct relationship between increased mortality in invasive BC patients and their LN stage status, with lower LN stage being associated with better survival and *vice versa*. Over a 5-year period, it was reported that approximately 83% of BC patients with LN stage I survived [39]. The LN staging system is also included in the NPI [34].

However, according to recent studies, if patients are followed up for a sufficiently long period, approximately 15–20% of LN-negative patients will eventually develop distant metastases. [42]. Therefore, a need for diagnostic and prognostic markers for such patients is of crucial importance, as conventional markers (such as NPI) cannot group such patients. Recently, numerous studies have highlighted that the evaluation of LVI may influence the classification of such LN-negative patients [43, 44].

1.5.1.5 NPI

The NPI is a survival and prognostic stratification tool for patients with invasive BC. The NPI also serves as a common platform for various prognostic factors [40], including the tumour grading system, tumour size, and LN staging. In addition to providing a strong platform for patients, NPI values are calculated through a consensus formula based on clinical data and the values of three significant prognostic factors for the NPI:

NPI = 0.2 × size of tumour (cm) + LN (stage 1–3) + tumour grade (grade 1–3)

Based on the NPI value, patients are categorised into three classes: 1) good-prognosis NPI, with an NPI value less than 3.4; 2) moderate-prognosis NPI, with an NPI value between 3.4 and 5.4; and 3) poor-prognosis NPI, with an NPI value greater than 5.4 [40]. An improved NPI grouping system has also been proposed that further divides the three aforementioned NPI prognostic groups into six groups: excellent-prognostic group (EPG; NPI value 2.08–2.40), good-prognostic group (GPG; NPI value 2.41–3.40), moderate-prognostic group I (MPG-I; NPI value 3.41–4.40), moderate-prognostic group II (MPG-II; NPI value 4.41–5.40), poor-prognostic group (PPG; NPI value 5.41–6.40), and very poor-prognostic group (VPG; NPI value >6.40) [45].

1.5.1.6 LVI

In patients with invasive BC, LVI is an independent prognostic factor that is strongly associated with BCSS and DMFS [46]. LVI signifies the expansion of tumour cells towards the peritumoural region within lymphatic spaces and/or blood vessels. LVI can be identified by the microscopic examination of primary tumour morphology for the presence or absence of specific endothelial markers (Figure 1.1A). Previous research identified LVI as an independent prognostic factor and a sign of metastatic potential [46-48]. Vascular invasion (VI) can be identified by the assessment of tumour cells within vascular spaces lined by endothelial cells in tumour sections stained with haematoxylin and eosin (H&E) (Figure 1.1B). LVI status can also be assessed through immunohistochemistry (IHC) by utilising CD31, CD34, and D2-40 biomarkers [49] (Figure 1.1B).

Nevertheless, false detection of VI during assessment can occur due to retraction artefacts or VI may be missed if tumour cells are packed tightly within a vessel. A study by Ellis et al. highlighted a low level of concordance (kappa=0.3) between two observers in the identification of VI by

conventional means, and a very high level of concordance (kappa=0.9) between observers in the implementation of immunohistochemical markers, such as D2-40 [50, 51].

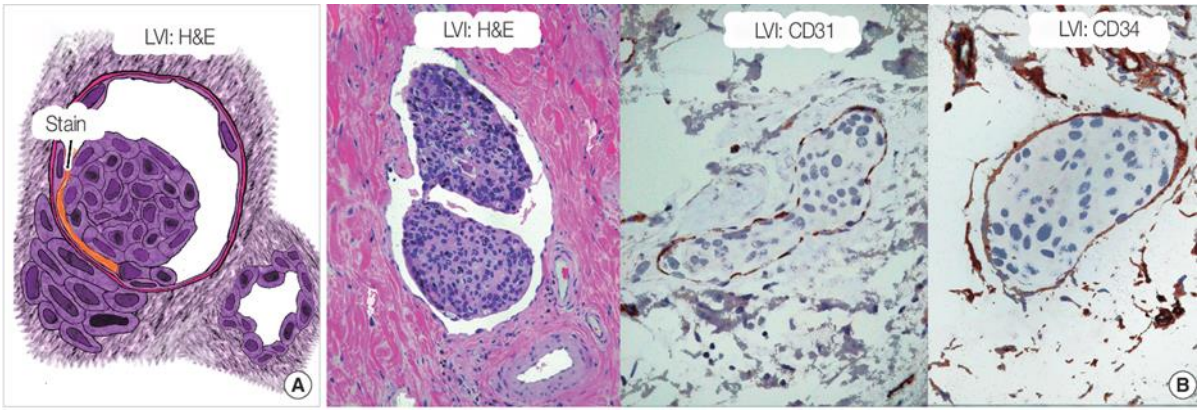


Figure 1.1 'True lymphovascular invasion (LVI) as (A) an illustration, and (B) on a histological specimen stained with haematoxylin and eosin (H&E; ×20), CD31, and CD34 staining (both at ×40 magnification, with detail emphasising the lymphatic space involved).' (adapted from Zaorsky et al. [2012]) [49].

*** This figure was reused according to the agreement between Mrs. Abrar Aljohani and Mr. Seong Ouk Lee (Editorial Assistant of Journal of Breast Cancer). See Appendix A.**

Mohammed et al. found that 96.4% of 177 well-identified primary invasive BC specimens showed lymphatic vessel involvement, while only 3.5% showed blood vessel invasion [52]. The presence of lymphatic vessel invasion in the abovementioned research was significantly associated with the presence of LN metastasis, the occurrence of distant metastases, regional or partial recurrence, shorter disease-free interval, and a decreased overall survival (OS). These findings are consistent with other studies that found an association between lymphatic vessel invasion and larger tumour size, lymph node metastasis, and distant metastasis. Moreover, lymphatic vessel invasion is a poor independent prognostic factor for BC [53, 54]. More details about the molecular mechanisms underlying LVI are discussed in depth later in this chapter.

1.5.2 Predictive factors

Predictive factors can help to predict the prognosis and response of the BC to a certain therapy. Predictive factors are critical for the management and stratification of BC patients, particularly when paired with the clinicopathological data and prognostic criteria outlined previously. The principle predictive features of a tumour that are involved in discerning therapy-related response are oestrogen receptor (ER), progesterone receptor (PR), and HER2 [55].

1.5.2.1 Hormone receptors: ER and PR

The College of American Pathologists (CAP) and the American Society of Clinical Oncology (ASCO) recommend a positive detection threshold of 1% of tumour nuclei for ER and PR in invasive BC patients [56]. ER status is a powerful predictor of long-term response to adjuvant endocrine treatment; however, the duration of a hormonal therapy regimen for ER^+ patients is still debated [57]. The prognostic significance of ER positivity may be more beneficial in postmenopausal BC patients, whereas younger patients are more likely to have aggressive BC phenotypes. Nevertheless, ER status might mitigate the influence of age as an independent prognostic factor and allow these patients to benefit from chemotherapy [58, 59]. Clinical co-expression of the PR in ER^+ cases has been used to predict a patient's response to adjuvant hormonal treatments. ER^+PR^- cases have a 50% response rate to hormonal therapy, but this rises to 80% if the receptors are co-expressed [58]. Hormone receptors (ER and PR) could be correlated with the pathways of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kB) and the tumour necrosis factor α (TNF-α); these pathways are associated with invasion and metastasis in BC [60].

1.5.2.2 HER2

The HER2 protein, encoded by the *ERBB2* gene, is a receptor tyrosine-kinase protein erbB-2. Overexpression of HER2 is found in approximately 15% of BC cases and is associated with invasive BC with poor prognosis

[61]. $HER2^+$ cases show noncohesive BC cells undergoing epithelial-mesenchymal transition (EMT) and demonstrating tumorigenic potential [62]. An analysis showed a strong correlation between $HER2^+$ and LVI in invasive ER^+ BC cases [63]. In invasive BC, HER2 positivity is correlated with aggressive features, such as invasion, abnormal cellular polarity, tumour growth, and poor outcome [64]. A monoclonal antibody, trastuzumab, has previously been used to treat patients with $HER2^+$ BC; however, effective therapeutic outcomes are observed only in early-stage $HER2^+$ BC patients [65]. According to a recent study, LVI exhibits excellent correlation with 30% of a large cohort of patients with invasive BC treated with neoadjuvant chemotherapy. LVI is correlated with poor DMFS, BCSS, and DFS in all invasive BC patients; remarkably, LVI has a hazard ratio of 6.21 in $HER2^+$ patients [66].

1.6 Molecular classification of BC

BC is a heterogeneous disease in which various phenotypes and genotypes can have clinical relevance. BC gene expression patterns have been known to the scientific community since 2001 [67]. Deoxyribonucleic acid (DNA) microarray and hierarchical clustering analysis have determined two different sets of hormone receptor tumours: hormone receptor-positive (HR^+) and hormone receptor-negative (HR^-). The receptors under consideration are ER and PR. Post-clustering gene expression profiles of HR^+ BC revealed two subgroups similar to those observed for luminal epithelial cells from mammary glands [67]. HR^- clustering analysis produced three tumour classes: amplified (HER2), normal-like BC, and basal-like BC.

Microarray-based gene expression profiling (GEP) has provided invaluable new perspectives on the molecular heterogeneity of BC. Breast tumours have been categorised into different subgroups with distinct clinical relevance based on changes in their gene expression patterns. The Prediction Analysis of Microarray 50 (PAM50), a 50-gene classifier that divides breast tumours into five intrinsic subtypes, is the best-known gene

expression classification method. This approach was utilised by Perou et al. to classify breast carcinomas into four subtypes: luminal, HER2-enriched, basal-like, and normal breast-like tumours [68]. After further clustering analysis, the luminal tumour subtype was further divided into luminal A and luminal B, yielding a total of five molecular subtypes [67, 69, 70]:

1. Luminal A subtype. This subtype has low levels of the proliferation marker Ki67 and is ER^+ and/or PR^+ and $HER2^-$. It is the most prevalent subtype, accounting for 50–60% of all BCs. Tumours in this group are low grade, slow growing, and have good prognosis and survival rates when compared with other subtypes [71].
2. Luminal B subtype. This subtype accounts for 10–20% of all BCs and is defined by ER^+, PR^+, and/or $HER2^+$. PR levels are lower in luminal B than in luminal A tumours (79% and 94%, respectively) [72]. In addition, luminal B tumours are more aggressive than luminal A tumours, with a higher histological grade, higher expression of proliferative markers, worse prognosis, and poorer outcome [69, 71, 73]. Changes in Ki67 expression levels constitute an important distinction between the two luminal subtypes. Using an IHC assay, Cheang et al. defined luminal A tumours as ER^+ and/or PR^+, $HER2^-$, with low Ki67, and luminal B tumours as ER^+ and/or PR^+, $HER2^-$, with high Ki67 [74].
3. HER2-enriched subtype. This subtype accounts for 15–20% of all BCs and is defined as ER^-, PR^-, and $HER2^+$, with high proliferative gene expression and low expression of luminal and basal-like group genes. This subtype grows more rapidly than luminal subtypes and has a poor prognosis [69]. While microarray-based approaches identified roughly 70% of clinical $HER2^+$ BC as HER2-enriched, the remaining 30% were mostly luminal B subtype [70, 75].
4. TN/basal-like subtype. This subtype accounts for about 10–20% of all BCs and is characterised as ER^-, PR^-, and $HER2^-$. Positive expression of the basal markers cytokeratin 5 and 6 (CK5/6) and/or

epidermal growth factor receptor (EGFR) distinguishes basal-like BCs from non-basal TN tumours [76]. TN tumours, in general, have a higher histological grade and are more aggressive than other BC subtypes [77].

5. Normal breast-like subtype. This group has not been described in a consistent manner; while some studies defined it as luminal A tumour, others classified it as a TN (ER^-, PR^-, and $HER2^-$) tumour with an absence of basal-like gene expression [70, 78-80]. Normal breast-like tumours have been revealed to be technical artefacts resulting from high contamination of samples with normal tissue [77]. These tumours have a favourable prognosis in general, but they have a worse prognosis than luminal A tumours [80].

A group of 10 integrative clusters (IntClust) from samples of 2,000 BC cases was studied through comprehensive genomic and transcriptomic profiling. The clusters had different genomic landscapes, molecular and clinical characteristics, and unique expression profiles in terms of ER, PR, and HER2 [81]. IntClust 3, 4, 7, and 8 demonstrated good prognosis until 10 years post-treatment in BCSS. IntClust 1, 6, and 9 demonstrated moderate outcomes, while IntClust 2, 5, and 10 were associated with poor outcomes [82]. IntClust possesses distinct molecular features at the chromosome level, as listed below:

1. IntClust-1: amplification of 17q23
2. IntClust-2: amplification of 11q13/14
3. IntClust-3: inadequacy in copy number changes
4. IntClust-4: lack of copy number
5. IntClust-5: amplification of HER2
6. IntClust-6: amplification of 8p12
7. IntClust-7: gain in 16p, loss in 17q, and amplification in 8q

8. IntClust-8: gain in 1q, loss of 16q

9. IntClust-9: gain of 8q, amplification of 20q

10. IntClust-10: loss of 5q, gain of 8q, gain of 10p, and gain of 12p [82].

1.7 Molecular mechanisms underlying LVI in invasive BC

1.7.1 Introduction to LVI

While the minority (<10%) of BC patients have metastatic disease, 20–30% of patients diagnosed with early-stage disease will develop distant metastases later in their lives [71, 83, 84]. For a tumour cell to metastasise, a series of consecutive steps must occur, collectively called the 'metastatic cascade': (a) detachment from the primary tumour mass, (b) invasion and migration through the extracellular matrix (ECM), (c) penetration into the local blood and/or lymphatic vessels, (d) evasion of the host immune response, (e) extravasation into the new organ, and (f) adaption into a new tumour site and division to form the new tumour [86]. Distant metastases, identified as poor prognostic events in cancers, progress and develop via balanced and highly selective biological processes and related pathways that integrate many connected mechanisms, including cell proliferation, interaction with the cell matrix, stromal invasion, migration, and LVI [85-87].

In a study involving 3,812 BC cases [46], it was revealed that LVI could be used not only as an independent prognostic variable in all the BC cases, but also in different prognostic subgroups involving LN-negative cases. Within this subgroup, LVI is a high risk factor with a survival disadvantage equivalent to the involvement of one or two lymph nodes and to that provided by one higher size category (pT1 to pT2) [46]. At the molecular level, tumour cells can use several biological interactions to facilitate their metastatic development via vascular and lymphatic channels [47, 88-90]. Identification of these signalling pathways is important for discovering the molecules driving the vital steps of LVI.

The hallmarks of cancer have been thoroughly studied, and they indicate that a tumour cell's genes are controlled by its intrinsic and extrinsic characteristics [9]. A high level of protein expression, resistance to apoptosis, high proliferation rate, adhesion, invasion, and cellular motility resulting in cellular migration are features that characterise malignant cells [91]. Although there is no direct association between these features and the occurrence of LVI—as they can be found in both LVI^- and LVI^+ cancers—they are important features of its development and of the molecular mechanisms that characterise these processes. The aforementioned factors must be active, at least in some cellular clones within the carcinogenic pathway, from the initiation to the development of vascular metastasis. This section focuses on the clinicopathological features of LVI in BC and the features of molecular modification in invasion and LVI.

1.7.2 Morphological assessment of LVI

Histomorphological characteristics have been used to distinguish LVI from blood vascular invasion (BVI). BVI is distinguished from LVI by the presence of tumour emboli, including fibrin clots and/or erythrocytes, inside the endothelium-lined region [44]. Several vascular markers have been found and successfully used to identify VI and distinguish between BVI and LVI; however, their sensitivity and specificity vary. Published research evaluated VI in 177 specimens using four vascular markers: podoplanin, D2-40, CD31, and CD34. D2-40/podoplanin has been demonstrated to be lymphatic-specific and blood vessel-negative. CD34 expression was found to be high in blood vessels but low in lymphatic vessels [44, 92]. CD31 expression was observed in a fraction of lymphocytes in some tumours and blood vessels, but its expression varied in lymph vessels. LVI was differentiated from BVI by evaluating the differential expression of these four markers [44, 93].

Morphological evaluation of LVI has been conducted in peritumoural areas. This achievement is mostly due to the simplicity with which LVI can be detected outside the tumorous tissue, hence avoiding the controversy

around the existence of intratumoural lymphatics, the existence of tumour retraction artefacts, the prognostic importance of peritumoural over intratumoural LVI, and the presence of a rare, pure intratumoural LVI [44, 94-96]. The significance of peritumoural LVI in BC prognosis is that it has been reported to facilitate the spread of tumour cells (LN metastasis) compared with intratumoural LVI, which has been found to be associated with tumour growth [44]. Therefore, both peritumoural and intratumoural LVI are predictive of poor prognosis in BC. In our laboratory, LVI was detected in both peritumoural and intratumoural vessels [44]. In an evaluation of LVI-positive cases, 75% had peritumoural LVI and 25% had intratumoural LVI; furthermore, only a few (4%) of the intratumoural LVI-positive cases had LVI in the intratumoural compartment alone. Focusing only on the cases in which intratumoural LVI is present with a positive LN status, it is possible that the presence of intratumoural LVI could explain the metastatic cascade. When such node involvement is not evident, however, caution is warranted in disregarding the value of intratumoural LVI. The pragmatic approach posits that the foundational mechanism of tumour cell progression is in the vasculature; as a result, clinical vigilance is warranted. Based on several pieces of evidence, it is possible that the occurrence of LVI in breast tumours lacks the morphological features of LVI, such as the presence of circulating tumour cells [97], bone marrow micro-metastases [98], and LN positivity [99]. This may have an indirect impact on our ability to understand the molecular mechanisms underlying LVI in BC, as determining metastatic potential in patients without LVI is challenging. The challenge of understanding the molecular mechanisms driving LVI and their association with the metastatic process of invasion and the lack of a model to identify and scrutinise intravasation in BC complicate the investigation of driver molecules, biological and molecular pathways, and potential treatment targets correlated with LVI in BC.

1.7.3 Tumour histological characteristics and LVI

There is a direct correlation between tumour morphological characteristics and the occurrence of LVI, including tumour type, grade, and size. An example of such a tumour type is invasive micropapillary carcinoma (IMPC), which is characterised by the frequent development of LVI [100]. Genomic aberrations in several protein-encoding genes could play a role in cytoskeletal organisation, abnormal cellular polarity, and primary cilium biogenesis involving formin 2 (*FMN2*), dynein heavy chain 9 (*DNAH9*), translocation protein SEC63 (*SEC63*), and zinc finger FYVE domain-containing protein 26 (*ZFYVE26*) [101]. This discovery was possible due to a comparison between the exome and RNA sequencing of IMPC compared with IDC. Additionally, it is important to note that IMPCs show a high frequency of LVI despite the number of micropapillary elements in an invasive tumour [102], which suggests the presence of an essential biological characteristic of this tumour type that promotes the development of LVI. On the contrary, ILC has been shown to demonstrate a low frequency of LVI, even when the LVI was identified using lymphatic endothelium-specific biomarkers [103], despite having a similar frequency of nodal involvement and the tendency to spread to more nodes than grade-matched IDC [104]. There is an imbalance between the metastatic rate and the frequency of morphologically identified LVI in ILC; this can be explained by the discohesive single-cell infiltration of ILC with the inherited loss of E-cadherin, leading to intratumoural or peritumoural intravasation of uniform small cells rather than the easily detected cohesive clusters of IDC cells. In the rare case of a pleomorphic variant of ILC, LVI can be easily identified, although it shares similar molecular alterations with classical ILC [105-107]. LVI can be observed in benign papillary lesions of the breast, which are a unique type of benign breast tumour. This phenomenon may be related to their inherent friability, which leads to epithelial intravascular displacement after needle instrumentation procedures, rather than the biological characteristics correlated with their metastatic potential [108].

Furthermore, there is an association between LVI and both tumour grade and size. For example, LVI is rarely found in grade I tumours; however, up to one-half of grade III tumours are correlated with LVI [46, 109, 110], and similar correlations are often observed with tumour size [111]. Despite its rarity, the presence of LVI in small BC is an independent predictor of node metastasis and short survival [111, 112], which may suggest that these tumours can be aggressive.

There is also an association between LVI and the BC molecular subtype. There are many molecular subtypes, such as TN or basal-like BC, that are less correlated with LVI despite their poor prognosis and high tendency for local recurrence [113]. On the contrary, luminal and $HER2^+$ molecular subtypes demonstrate a greater incidence of LVI [114]. An explanation for this difference is a lower level of expression of tight junction proteins, claudins, and higher levels of proteins that are essential for transforming the malignant epithelial cells to mesenchymal cells in TN and basal-like tumours compared with luminal and $HER2^+$ BC subtypes [115, 116]. Therefore, studies regarding the molecular determinants of LVI should endorse the best choice of cohorts and control samples for more robust identification of drivers controlling LVI.

1.7.4 Vascular or lymphatic invasion

Invasive BC can spread via blood vessels or lymph vessels. Evidence from current studies reveals that lymph vessels, rather than blood vessels, are critical to the initial spread of BC due to the intrinsic tropism of malignant cells to lymphatic channels [44]. According to one recent study, BVI, rather than LVI, was significantly correlated with the recurrence of endometrial cancer [117]. Based on the structural differences between lymphatics and blood vessels, it is possible to determine the initial route of spread. Lymphatic capillaries do not have the sufficient number of tight junctions typically present between endothelial cells lining blood capillaries [118]; moreover, this lack of tight junctions renders lymphatics more dilated and leakier than blood vessels, thereby facilitating tumour cell intravasation.

Therefore, peritumoural lymphatic vessels may represent the primary route of intravasation for invasive BC cells that cannot overcome the physical barriers of vascular endothelial cells [119]. It has also been highlighted that interstitial fluid pressure controls the directional motility of metastatic tumour cells in the direction of the lymphatics at the periphery of a tumour, passively increasing the likelihood of their intravasation into lymphatic channels [120]. Another plausible explanation involves active molecular mechanisms that attract malignant cells to either type of vessel [121]. Lymphangiogenic factor expression may exert an influence on the route of metastatic spread via lymphatics because of their specific structure, and this may enhance tumour growth [118, 122]. For example, the expression of vascular endothelial growth factor receptor 3 (VEGFR3) exclusively on lymphatic endothelial cells and its ligand vascular endothelial growth factor C (VEGFC), which is overexpressed in tumour cells, can influence lymphangiogenesis. However, based on some studies, there is a lack of correlation between VEGFC and LVI or node metastasis; this can be explained by the potential role of numerous factors in addition to VEGFC and vascular endothelial growth factor D (VEGFD), and the possibility of heterogeneity in the study samples [123]. Despite the nature of the vessels involved—whether blood, lymphatic, native, or newly formed—it is of crucial importance to understand that neither angiogenesis nor lymphangiogenesis is equivalent to LVI, despite their strong correlation. Angiogenesis is the process of neovasculature development that supports tumour growth and invasion, while LVI involves the process by which tumour emboli become lodged in vascular or lymphatic channels in the initial stages of the metastatic cascade.

1.7.5 Tumour invasiveness, tumour microenvironment (TME), and LVI

Before BC cells reach the vascular lumen, initial detachment from the tumour cell mass must occur accompanied by invasion and migration through the ECM to the vascular walls; these processes are considered

prerequisites for the development of LVI. Infiltration into the surrounding ECM occurs as single malignant cells or as groups of cells. BC tumour cells upregulate many genes, such as *SNAIL*, zinc finger E-box binding homeobox 1 (*ZEB1*), *SLUG*, and other transcriptional factors that facilitate the potential evolution of malignant epithelial cells to the mesenchymal phase via suppression of E-cadherin and other cell–cell adhesion molecules, increasing cell motility [124, 125]. The mechanobiological interactions between stromal components and tumour cells can be controlled by the TME [126]. Through genetic and epigenetic regulation studies, the interactions between tumour cells and the components of the TME have been shown to alter adhesion and enhance invasion and migration of tumour cells through the adjacent ECM, leading to an increase in their penetration into vascular channels [9].

Proteolytic degradation occurs with the assistance of stromal cells via matrix metalloproteinases (MMPs) and it is caused by malignant cells and their epigenetic regulation. LVI is strongly correlated with the expression of MMP1 and MMP9 [127-129]. Degradation of the vascular basement membrane and cancer cell intravasation can be induced by BC cells and their overexpression of cancer cell-related membrane type 1 (MT1)-MMP or MMP14 [130]. The expression of MMP14 in cancer cells induces its expression in vascular endothelial cells and results in remodelling of the vascular basement membrane and cancer cell invasion into blood vessels, but not into lymphatic vessels. MMP14 can also act as a predictive factor and therapeutic target for haematogenous metastasis in BC patients with the TN subtype [131]. Podoplanin, a glycoprotein encoded by the *PDPN* gene, is expressed in the lymphatic endothelium. In combination with other markers, such as LYVE-1 and Prox-1, podoplanin is considered a specific lymphatic endothelial marker. Cancer-associated fibroblasts (CAFs) also express podoplanin, which has been shown to have a role in increasing cellular motility and reducing cellular adhesion. The expression of podoplanin by CAFs in BC is associated with aggressive characteristics,

including node metastasis, LVI, and short survival [132]; podoplanin expression is also associated with the expression of VEGFC in BC cells and the number of intratumoural microvessels [133]. Interestingly, increased podoplanin expression at the invasive front of BC plays an essential role in invasion and metastasis by facilitating the development of invasive protrusions and filopodia, but not by activating endothelial-mesenchymal transition [132]. This phenomenon could represent a powerful synergy between metastasis-related paracrine signalling from CAFs and epigenetic regulation by malignant cells to enhance lymphatic vascular invasion.

C-C chemokine receptor type 7 (CCR7), a lymphatic-homing chemokine receptor, demonstrates the concert between tumour cells, dendritic cells, and the vascular endothelium in LVI. CCR7 and VEGFC expression by tumour cells has been reported to promote the development of LVI [134]. Transcellular lymphatic drainage that occurs as a result of increased interstitial flow of lymphatic fluids may also polarise tumour cells towards the adjacent lymphatic vessel via the autologous release of CCR7 [120]. The ligand of CCR7 (C-C motif chemokine 21 [CCL21]) is located on the vascular endothelial surface, and its expression by the vascular endothelium is increased due to VEGFC secreted by malignant cells and due to the upstream signalling of its receptor VEGFR3. Consequently, $CCR7^{+}$ tumour cells will migrate to lymphatic vessels and the possibility of developing LVI will increase. This phenomenon demonstrates how tumour cells can increase their invasiveness and how they use the chemokinetics of lymphangiogenesis to direct themselves to their preferred invasion target.

The existence of macrophages in primary tumours is correlated with the development of LVI and the tumour's metastatic potential; however, the exact mechanism is unknown. A study reported that tumour cell and macrophage interactions play a role in the process of migration [135]. The expression of EGF and/or colony-stimulating factor 1 (CSF-1) can enhance the accumulation of tumour cells and macrophages into microneedles. The

inhibition of either EGF- or CSF-1-stimulated signalling decreases migration of these cell types [135]. EGFR overexpression by mammary malignant cells renders cells chemotactic to EGF secreted by the macrophages that line blood vessels [135]. These findings provide evidence that a synergistic interaction between macrophages and malignant cells in the process of cell migration *in vivo* can partially explain the rise in the tendency of basal-like BCs—which increase the expression of the EGFR—to be more node negative and demonstrate haematogenous dissemination to viscera and bone rather than lymphatic dissemination [136, 137]. Macrophages are known to release interleukin 10 and angiogenic indicators [138]. Under hypoxic conditions, there is upregulation of angiogenesis and lymphangiogenesis stimulated by the growth factors released by tumour cells [139], which act as chemo-attractants for macrophages. Another study demonstrated a correlation between the presence of perivascular macrophages and malignant cell intravasation [140]. Invasion and metastasis are promoted by recruited macrophages via a dual-direction reaction loop of paracrine signalling with tumour epithelial cells; therefore, intravasation into vascular channels and migration occur through the vessel lumen [141]. Accordingly, immunomodulatory drug therapy may inhibit tumour-induced immunosuppression processes and may provide beneficial adjuncts regarding non-specific cytotoxic drugs [142].

The term ‘vasculogenic mimicry’ or ‘vascular mimicry (VM)’ describes the formation of new microvascular channels without the involvement of endothelial cells and is independent from angiogenesis. Initially, VM was demonstrated *in vitro* in highly aggressive and metastatic melanoma cells [143]. The development of these highly patterned microvascular channels by aggressive, genetically dysregulated tumour cells occurs independently of endothelial cells and fibroblasts [144]. The findings were further confirmed in a xenograft inflammatory BC (IBC) model and cell lines [145, 146]. A study showed that the presence of VM in tumour tissues was correlated with node metastasis and poor outcome [147]. The poor

prognosis associated with VM was found to be related to the possibility of aggressive tumour cells surviving longer and the potential for metastatic tumour cells to reach the bloodstream and distant organs [148]. Various molecular biomarkers are overexpressed in VM in BC cells, such as VEGF, twist-related protein 2 (TWIST2), and cyclooxygenase 2 (COX2), as well as molecules conferring the cancer stem cell phenotype and those involved in the process of EMT [149].

1.7.6 Gene signatures, gene expression profiling, and LVI

To identify key genes and pathways promoting LVI, genes that are differentially expressed between LVI-positive and LVI-negative BC must be identified. Through the advances in high-throughput molecular and bioinformatic techniques, analysing genes associated with LVI on a large scale becomes possible [150]. One of the challenges in identifying the genes associated with LVI in genome-wide profiling studies is the nature of the cohorts under exploration (as described above) and the use of archival frozen tissue with inaccurate or incomplete LVI data. According to a previous study of endometrial carcinoma with an LVI gene signature, 18 genes were correlated with LVI [151] and were further demonstrated in BC using 11 open-access gene expression datasets focusing on data collected from more than 2,400 BC patients [152]. An association was found between the 18-gene signature and morphologically detected LVI and other poor prognostic characteristics of BC, such as advanced stage, high grade, HER2 receptor positivity, hormone receptor negativity, and the basal-like phenotype [152]. Candidate genes comprised 11 downregulated genes and 7 upregulated genes, both groups including a diverse array of genes controlling ECM proteolysis, ECM collagens, angiopoietin, interleukins, tumour necrosis index, and fibroblast growth factor receptors. Although this gene signature is correlated with LVI in BC, it reflects features of aggressive tumour cell behaviour and interactions with the surrounding microenvironment rather than specific molecular mechanisms controlling the development of LVI. Our efforts to study LVI mechanisms in invasive

BC have found that Myxovirus resistance 1 (*MX1*), isocitrate dehydrogenase 2 (*IDH2*), high mobility group box 3 (*HMGB3*), cyclin B1 (*CCNB1*), cyclin B2 (*CCNB2*), and ribonucleotide reductase enzyme 2 (*RRM2*) are novel LVI-associated genes. The rationale for selecting these genes will be discussed in Chapter 3.

Another study aiming to develop the gene signature analysed 214 hepatocellular carcinomas that predicted the presence of LVI among patients who were surgically treated for hepatocellular carcinoma [153]. It is necessary to mention that 14 genes were identified as upregulated and 21 genes as downregulated in correlation with LVI, but no overlap was found with the 18-gene signature discussed above [152], although this could be due to the different cancer types. Some genes, such as *CDKN3, CD24*, *PIK3R1*, and *UBE2C*, are important in BC development and progression [153]. In one study, array comparative genomic hybridisation, quantitative reverse transcription-polymerase chain reaction (RT-qPCR), and complementary DNA (cDNA) microarray were used to identify DNA copy number aberrations (CNAs) and the gene expression profile associated with LVI in a cohort of invasive BCs [129]. Although the authors did not find a significant correlation between LVI and increased genomic instability (even taking into consideration the total number of chromosomal changes), they found that there was an increase in the frequency of certain alterations that correlated with LVI, including gains at 5p, 16p, 17q12, and 19, and losses at 8p, 11q, 18q, and 21. They also found three novel small-scale rearrangements in LVI-positive grade III tumours that integrated putative BC genes involving amplicons at 11p11.2, harbouring G1/S-specific cyclin-E1 (*CCNE1*), zinc finger protein 217 (*ZNF217*), 19q12, very-long-chain 3-oxoacyl-CoA reductase (*HSD17B12*), 20q13.2, and 4q13.3, and covering a disinterring and metalloproteinase with thrombospondin motifs 3 (*ADAMTS3*). In addition to these findings, a deletion at 12p12.3 was detected that integrated Ras-related and oestrogen-regulated growth inhibitor-like protein (*RERGL*) and phosphatidylinositol 4-phosphate 3-

kinase C2 domain-containing subunit gamma (*PIK3C2G*). Gene expression analysis revealed networks outlining interstitial collagenase gene *MMP1*, protein S100-A8 (*S100A8*), and an intermediary of RNA polymerase II transcript subunit 1 (*MED1*) as potential genes involved in LVI-positive grade III tumours [129]. Several genes demonstrated a concordant pattern of copy number gains and an upregulation of gene expression in the LVI-positive samples, including uromodulin (*UMOD*), MYCBP-associated protein (*MYCBPAP*), arylsulfatase G (*ARSG*), and *MED1* [129]. Although the results of this study were promising, it is important to note that the BC cohort used was small (n=57) and frozen tissue was utilised with no further confirmation of LVI status.

Despite the controversy surrounding microRNAs (miRNAs) and their role in BC, some researchers have shown a correlation between LVI and levels of miRNAs, such as miR-9-3 [154], miR-10b [155], and miRNA-205 [156]. Applying bioinformatic analytical approaches to finding target genes in a well-defined BC cohort, classified according to stringent histopathological criteria for LVI, may represent a promising strategy. Ideally, such tools should be considered in the discovery and validation of exclusively sets of cases with definite LVI and, as a control, the LVI-negative cases should have no morphological or biological patterns associated with LVI (e.g., positive LN or circulating tumour cells). Candidate genes require specific analysis based not only on their differential expression but also on their biological features and the involvement of certain molecular pathways. This approach can help to narrow down the number of possible genes for further testing using *in vivo* models and *in vitro* functional studies.

1.7.7 Proteomics and LVI

To date, the association of LVI with proteins has been indirect or incidental. An IHC-based analysis of 177 stained and paraffin-embedded BC tissue sections with CD34, podoplanin, VEGFA, VEGFC, VEGFD was performed. High expression of VEGFA and VEGFC, but not VEGFD, was correlated with higher microvessel and lymph vessel density, the occurrence of both nodal

and distant metastases, and poor outcomes [51]. This finding suggests that high levels of VEGFA and VEGFC expression predicts a poor prognosis, due to increased lymphangiogenesis and angiogenesis, and is associated with LVI. In an IHC-based study that focused on 77 patients, the expression of VEGFC and VEGFD in BC was correlated significantly with peritumoural lymph vessel density, which was associated with the development of LVI and nodal metastases [157]. Another study indicated that stained IHC sections of 70 BCs showing high expression of VEGFC and COX2 were correlated with higher lymph vessel density and D2-40-positive LVI [158]. Therefore, despite the existence of other studies with more controversial findings [123], there is enough evidence for the significance of the VEGF pathways in the initial stages of the metastatic cascade and LVI. VEGF and its autocrine functions include tumour cell adhesion, migration, invasion, and survival [159]. Deconstruction of the pathways and their members would help to elucidate the pathomechanisms of LVI. The attempts of some studies to discover proteins related to LVI have led to further gene analyses [160].

Another approach using the artificial neural network (ANN) assessed gene expression arrays of 91 BC patients and detected 89 transcripts related to LVI. Low expression levels of one of the transcripts, an inhibitor of the calpain proteases called calpastatin, was correlated with intratumoural LVI on PCR and was detected by IHC platforms. Calpain-1 and calpain-2, however, which have a role in tumourigenesis and the proteolysis of substrates, such as inhibitors of NF-kB and focal adhesion proteins, were found to be unrelated to LVI [161]. Therefore, calpastatin is the main regulator of LVI in this pathway. Other proteins related to LVI have been discovered as incidental findings while exploring the roles of novel proteins in BC. For example, a study that identified phosphorylated extracellular-regulated kinase 1/2 (pERK1/2) indicated that members of the mitogen-activated protein kinase family in BC could be correlated with LVI negativity [162]. This association is also viable for the ER^+ subgroup in this series.

LVI has been revealed to have an inverse correlation with hormone receptor expression (ER and PR) [54], and molecules, such as pERK1/2 may provide evidence of the biological processes underlying LVI. The analysis of DNA repair proteins in BC has also demonstrated an association between the proteins and LVI. High expression of X-ray repair cross-complementing protein-6/X-ray repair cross-complementing protein-5 (KU70/KU80) was highly correlated with LVI positivity [163]. A member of the family of small ubiquitin-like modifier proteins (SUMO) also demonstrated an association with LVI. Despite positive correlations between LVI and loss of SUMO nuclear expression, LVI was positively correlated with the cytoplasmic expression of E3 SUMO-protein ligase PIAS1 (*PIAS1*), E3 SUMO-protein ligase PIAS4 (*PIAS4*), and SUMO-conjugating enzyme UBC9 (*UBC9*) [164]. This finding demonstrates the significance of SUMO pathways in LVI. Poly(ADP-ribose) polymerase-1 (PARP1), the main facilitator of DNA repair, was not correlated with LVI in either cleaved or non-cleaved forms [165], nor was the serine/threonine-protein kinase Chk1 (CHEK1), the main member of the DNA damage-activated checkpoint signalling responses [166].

Based on several innovative studies, the identification of more proteins is needed to understand the molecular mechanisms controlling LVI. Therefore, this study utilised weighted average difference (WAD) and ANN approaches to elucidate the proteins controlling LVI in BC. Details on the WAD and ANN analyses and the selection criteria for the genes is discussed in Chapter 3.

1.7.8 Unified mechanism of LVI

The various important mechanisms underlying LVI cooperate under the influence of tumour genetic, transcriptomic, and proteomic profiles, and the resulting interactions between these biological processes guarantee tumour cells' invasion of vascular spaces and metastasis. ECM is the medium for vibrant interactions between tumour cells and stromal non-tumoural cells in LVI. When tumour cells induce molecular modifications in the ECM, they

secure their progressive pathways of invasion and thus ensure metastasis. Synthesis, crosslinking, alignment, and degradation of the ECM represent remodelling activities, and tumour cells will select a combination of remodelling activities to establish the ideal ECM microenvironment and conditions for invading the target vessels and metastasising to other body sites [167]. Endothelial cells, macrophages, fibroblasts, and other immune cells are perceived as LVI mediators due to their various contributions to LVI mechanisms. Angiogenesis and lymphangiogenesis are important biological processes managed by many chemokines and cytokines in the TME. High interstitial fluid pressure, growing tumour mass, and hypoxia all contribute to the formation of new lymphatic and blood vessels; these vessels are not only required to maintain access to nourishment and to exchange damaged intratumoural vessels, but also to increase the contact surface area and subsequent likelihood of developing LVI.

1.7.9 Conclusions

Many molecular factors controlling the development of LVI have been elucidated, and other key contributors remain on the verge of discovery. To date, there are questions regarding the drivers of molecular mechanisms that cause BC tumour cells to invade vascular spaces and metastasise. Large-scale genomic and transcriptomic profiling of histologically confirmed LVI in BC may result in a high output of genes and pathways containing LVI master regulators. The candidate genes comprise these signatures and pathways, which could be further investigated by utilising various research approaches to identify new therapeutic targets.

1.8 Hypothesis and aims

1.8.1 Hypothesis

Using bioinformatic analyses, key genes of LVI-positive and LVI-negative settings in BC can be identified, and some of the identified genes show prognostic value, are potential therapeutic targets, and play a role in various mechanisms involved in the development of LVI.

1.8.2 Aims

The aims of this thesis are as follows:

1. To identify genes associated with LVI in patients with early-stage BC using bioinformatic analyses, including ANN and WAD approaches.
2. Evaluate the clinicopathological and prognostic significance of the genes at the messenger ribonucleic acid (mRNA) and protein levels using large, well-characterised invasive BC cohorts with long-term follow-up.
3. Evaluate and confirm the general functions of the selected genes using *in vitro* experimental studies (evaluate the impact of gene knockdown on tumour cell proliferation, migration, cell cycle, and apoptosis using a panel of relevant BC cell lines).
4. Assess the functional roles of the selected genes in LVI by conducting an *in vitro* investigation of BC cell lines with lymphatic and vascular endothelial cell models to mimic *in vivo* interactions and assess BC cell adhesion and transmigration with endothelial cells.

1.8.3 Significance of the project

Despite the use of conventional methods for the assessment of LVI in routine clinical practice, the identification of more factors is needed to determine whether cancer cells will potentially invade vascular or lymphatic channels and thereby spread to distant organs. The lack of targeted therapy to inhibit LVI and prevent metastasis typically stems from the complex underlying molecular mechanisms of LVI, which interact with other carcinogenic pathways. Deciphering the mechanisms of LVI and identifying key genes is likely to improve patient outcomes and could produce potential therapeutic targets. This thesis provides a unique insight into the key proteins involved in the development of LVI in invasive BC patients by assessing the roles of the main cell cycle regulators and other pathway regulators; these regulators will provide valuable insight into which molecular pathway could represent a potential molecular mechanism controlling LVI in BC.

Chapter 2: Material and methods

2.1 Study cohorts

2.1.1 Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort

The METABRIC study cohort consisted of 1,980 samples from invasive BC patients, and information regarding the validated clinicopathological, genomic, and transcriptomic data was readily available [81]. Primary fresh-frozen BC specimens were collected from tumour repositories in the United Kingdom and Canada. For copy number and genotype analysis and transcriptional profiling, DNA and RNA were extracted from the primary tumours of patients. The integrity of the extracted RNA was tested using the Agilent 2100 Bioanalyser Nanochip and the quality of the isolated DNA was assessed using agarose gel electrophoresis (Agilent Technologies, Wokingham, UK). AROS Applied Biotechnology (Aarhus, Denmark) hybridised DNA probes to Affymetrix SNP 6.0 arrays (Affymetrix, Santa Clara, CA). The Illumina Totalprep RNA amplification kit (Ambion, Warrington, UK) was used to generate biotin-labelled cRNA from total RNA, which was then hybridised on Illumina Human HT-12 v3 platforms to evaluate mRNA expression. As previously indicated, the raw gene expression data were pre-processed and normalised [81]. Differential expression analysis was performed using the linear models of microarray data (LIMMA) Bioconductor software package. LVI status had previously been assessed for 1,565 patients, including the Nottingham subset (n=288) from the METABRIC cohort, through histological analysis of H&E-stained paraffin-embedded tissues. LVI status was also defined in the Nottingham subset through the use of IHC staining for CD31, CD34, and D2-40 [44].

The median age of the patients at diagnosis was 61 years, with a range of 74 years. Patients with grade I tumours accounted for 9.0% of the whole cohort, whereas patients with grade II and grade III tumours represented 40.7% and 50.3% of the cohort, respectively. Positive regional LN metastasis was observed in 47.5% of the patients, with LN involvement in stage II and stage III tumours being 31.5% and 16.0%, respectively. LVI

status was histologically evaluated on slides stained with H&E and was positive in 635 of the 1,565 patients (41.0%) and negative in 930 patients (59.0%). The BCSS for each patient was calculated in months (up to a maximum of 300 months) from the date of the primary surgical treatment to the time of death due to BC. Follow-up status revealed that 505 (25.5%) of the patients in the cohort died of BC, while 1,071 (54.1%) survived the disease. The overall distribution of intrinsic BC subtypes in this cohort was assessed via prediction analysis of 50 genes using the PAM50 method, a technique based on RT-qPCR. The normal subtype accounted for 10.4% of the sample cohort, with the luminal A subtype present in 36.3%, luminal B subtype present in 24.6%, HER2-enriched subtype present in 12.1%, and the basal-like subtype present in 16.6% of the sample cohort. The clinical characteristics of the METABRIC cohort are presented in Table 2.1.

2.1.2 Nottingham subset from METABRIC cohort

The Nottingham subset (n=288) of the METABRIC cohort was used in this study to assess the correlation between mRNA and protein expression. As mentioned above, LVI status for this subset was assessed histologically by reviewing H&E-stained slides and through immunohistochemical biomarkers including CD31, CD34, and D2-40. The clinicopathological features of the Nottingham subset are presented in Table 2.1.

Table 2.1 Frequency of clinical characteristics of the METABRIC cohort (n=1,980) and Nottingham subsets from the METABRIC cohort (n=288).

Clinicopathological features	METABRIC cohort Number (%)	Nottingham subset cohort Number (%)
Age		
<50 years	424 (21.4)	94 (32.6)
≥50 years	1556 (78.6)	194 (67.4)
Tumour size		
<2 cm	622 (31.7)	131 (45.5)
≥2 cm	1338 (68.3)	157 (54.5)
Grade		
I	171 (9.0)	36 (12.7)
II	770 (40.7)	101 (35.7)
III	952 (50.3)	146 (51.6)
Nottingham prognostic index (NPI)		
Good prognostic group (GPG)	680 (34.3)	103 (35.8)
Moderate prognostic group (MPG)	1101 (55.6)	166 (57.6)
Poor prognostic group (PPG)	199 (10.1)	19 (6.6)
Lymph node stage		
I	1035 (52.5)	164 (56.9)
II	622 (31.5)	97 (33.7)
III	316 (16.0)	27 (9.4)
Lymphovascular invasion (LVI) (H&E/IHC: CD31, CD34, and D2-40)		
Negative	930 (59.0)	168 (59.1)
Positive	635 (41.0)	117 (40.9)
Follow-up status		
Alive	1071 (54.1)	172 (59.9)
Died from breast cancer	505 (25.5)	72 (25.1)
Died from other causes/unknown	384 (19.4)	25 (8.7)
Lost to follow-up	20 (1.0)	18 (6.3)
PAM50		
Normal-like	205 (10.4)	18 (6.3)
Luminal-A	718 (36.3)	108 (37.6)
Luminal-B	488 (24.6)	80 (27.9)
HER2-enriched	240 (12.1)	30 (10.4)

Basal-like	329 (16.6)	51 (17.8)
Oestrogen receptor (ER) (IHC)		
Negative	474 (23.9)	71 (24.6)
Positive	1506 (76.1)	217 (75.4)
Progesterone receptor (PR) (IHC)		
Negative	940 (47.5)	133 (46.2)
Positive	1040 (52.5)	155 (53.8)
Human epidermal growth factor receptor 2 (HER2) (IHC)		
Negative	1733 (87.5)	247 (85.8)
Positive	247 (12.5)	41 (14.2)

2.1.3 The Cancer Genome Atlas (TCGA) cohort

The Genomic Data Commons Data Portal and cBioPortal Websites [168, 169] were used to extract the TCGA cohort comprising female BC patients (n=854) to further evaluate the mRNA expression of all biomarkers included in this study [170]. The cohort was then accessed for RNASeqV2 derived mRNA expression data. De-identified clinical patient information was also accessed for certain clinicopathological features and their outcomes. cBioPortal was used to extract digital H&E-stained slides for the TCGA breast carcinoma cohort. The median age of the patients at diagnosis was 59 years. Among the entire cohort, 10.9% of the patients had grade I tumours, 45.9% had grade II tumours, and 43.1% had grade III tumours. Approximately 72% of the cases had a large tumour size (≥2 cm). In the TCGA cohort, LVI was present in 34.5% of the cases, and 65.5% of the cases were LVI negative. The cases were histologically evaluated using slides stained with H&E. The BCSS for each patient was calculated in months (up to 120 months). Survival data were also available, which revealed that 89.7% of the patients were alive at the time the data were accessed and 10.3% had died. Patients with HER2 positivity accounted for 19.0% of the cases, while 81.0% were HER2-negative. Hormone receptor positivity, including ER and PR positivity, was present in 77.5% and 66.9% of cases, respectively. Patients with ER^- tumours made up 22.5%, and PR^-

tumours made up 33.1% of the cohort. The clinical characteristics for the TCGA cohort are shown in Table 2.2.

Table 2.2 Frequency of clinical features of the TCGA cohort (n=854).

Clinicopathological features	Number (%)
Age	
<50 years	231 (27.1)
≥50 years	623 (72.9)
Tumour size	
<2 cm	239 (28.0)
≥2 cm	615 (72.0)
Grade	
I	89 (10.9)
II	375 (45.9)
III	352 (43.1)
Lymphovascular invasion (LVI) (H&E)	
Negative	559 (65.5)
Positive	295 (34.5)
Follow-up status	
Alive	765 (89.7)
Died from breast cancer	88 (10.3)
Oestrogen receptor (ER) (IHC)	
Negative	185 (22.5)
Positive	639 (77.5)
Progesterone receptor (PR) (IHC)	
Negative	271 (33.1)
Positive	547 (66.9)
Human epidermal growth factor receptor 2 (HER2) (IHC)	
Negative	567 (81.0)
Positive	133 (19.0)

2.1.4 Nottingham Primary Series (NPS) cohort of invasive BC

Tissue samples in this cohort were derived from the well-characterised NPS cohort, which consisted of 3,173 unselected, operable, invasive primary BCs from female patients. The NPS patients had previously undergone surgery at Nottingham City Hospital between 1986 and 2006. The biopsies

from the patients' samples in this cohort were previously used to establish tissue microarray (TMA) formalin-fixed paraffin-embedded (FFPE) blocks. Robust clinicopathological profiles for each patient were readily available that consisted of the following: age at diagnosis, primary tumour size, lymph node stage, histological grade, NPI, and LVI status. Data on BC molecular subtypes were also available. The BC subtypes of the Nottingham cohort were identified using immunohistochemical profiling and the Elston–Ellis mitotic score [33]. Four molecular subtypes were identified: $ER^+/HER2^-$ low proliferation (mitotic score 1); $ER^+/HER2^-$ high proliferation (mitotic score 2 or 3); $HER2^+$, including $HER2^+$ irrespective of ER status and triple-negative subtype (ER^-, PR^-, and $HER2^-$) [171]. The time to distant metastasis (TTDM), from surgery to the occurrence of the first distant metastasis, and BCSS for each patient were recorded for up to 10 years (median 120 months). The total follow-up time for this study was 10 years. Details of the clinicopathological factors are shown in Table 2.3.

Patient management was uniform and based on tumour features as determined by NPI and hormone receptor status. Patients with an NPI score ≤3.4, representing the excellent prognostic group, underwent no adjuvant therapy, while patients with an NPI score >3.4 who were ER^+ were offered tamoxifen (with or without goserelin [Zoladex] in premenopausal patients). Patients who were ER^- and fit enough to receive chemotherapy received cyclophosphamide, methotrexate, and 5-flurouracil (CMF). In this study, no patients received neoadjuvant therapy or Herceptin.

Table 2.3 Clinicopathological features of the Nottingham Primary Series of invasive BC (n=3,173).

Clinicopathological features	Number (%)
Age	
<50 years	1042 (32.8)
≥50 years	2131 (67.2)
Tumour size	
<2 cm	1768 (55.8)
≥2 cm	1403 (44.2)
Grade	
I	543 (17.1)
II	1131 (35.7)
III	1496 (47.2)
Nottingham prognostic index (NPI)	
Good prognostic group (GPG)	1041 (32.8)
Moderate prognostic group (MPG)	1632 (51.5)
Poor prognostic group (PPG)	496 (15.7)
Stage	
I	1989 (62.8)
II	903 (28.5)
III	277 (8.7)
Lymphovascular invasion (LVI) (H&E/IHC: CD31, CD34, and D2-40)	
Negative	2205 (69.8)
Positive	954 (30.2)
Oestrogen receptor (ER) (IHC)	
Negative	704 (22.3)
Positive	2453 (77.7)
Progesterone receptor (PR) (IHC)	
Negative	1256 (40.9)
Positive	1813 (59.1)
Human epidermal growth factor receptor 2 (HER2) (IHC)	
Negative	2677 (86.9)
Positive	402 (13.1)
Triple-negative (IHC)	
No	2630 (84.5)
Yes	483 (15.5)
Immunohistochemistry subtypes	

$ER^+/HER2^-$ low proliferation	1051 (36.9)
$ER^+/HER2^-$ high proliferation	906 (31.9)
Triple-negative	483 (17)
$HER2^+$	402 (14.2)

2.1.5 Biomarker data

A broad panel of IHC data of biomarkers related to BC that had previously been investigated in the Nottingham cohort was used in this study to elucidate the molecular interactions of the genes of interest. The following biomarkers were assessed in this cohort: HR (ER and PR) [172]; HER2 [173]; and EMT-related markers comprising epithelial cadherin (E-cadherin), placental cadherin (P-cadherin), neural cadherin (N-cadherin), transforming growth factor beta 1 (TGFβ1), and TWIST2, as previously published [76, 172, 174, 175]. All biomarkers were immunohistochemically evaluated via TMA. To determine HER2 immunoreactivity, the standard HercepTest (Dako) protocols were followed. In brief, a score of 0 or 1+ indicated that HER2 staining was negative, a score of 3+ indicated that HER2 staining was positive, and a score of 2+ indicated that HER2 staining was equivocal. HER2 fluorescence in situ hybridisation (FISH) pharmDxTM and HER2 chromogenic in situ hybridisation (CISH) pharmDxTM kit (Dako) were used to quantify HER2 gene amplification in equivocal cases and the results were evaluated according to ASCO standards [176]. Table 2.4 lists the antibody clones, suppliers, and cut-off values used to determine the expression levels of all biomarkers.

Table 2.4 Antibody clones, suppliers, and cut-off values for previously examined biomarkers in the Nottingham cohort.

Biomarker	Antibody clone	Supplier	Cut-off value
ER	1D5	DAKO	≥1%
PgR	PgR 636	DAKO	≥1%
HER2	cerbB-2	DAKO	3+ and 2+ with amplification
E-cadherin	NCH-38	DAKO	>100 H-score
P-cadherin	56	BD	≥5%
N-cadherin	C3865	Sigma	>100 H-score
TGFβ1	TB21	Abcam	>100 H-score
TWIST2	Ab57997	Abcam	>160 H-score

2.2 Determination of binding specificity for selected antibodies with Western blotting (WB)

2.2.1 Overview

WB was performed to evaluate the binding specificity of the selected antibodies to the target protein prior staining the tissues via IHC. The qualitative detection of a single protein or an epitope in several BC cell lysates was used to determine the binding specificity of the selected antibodies. The detection of multiple bands in cell lysates might be due to non-specific binding of the antibody or recognition of different isoforms of a particular protein by the same primary antibody. The term 'blotting' denotes the transfer of an isolated protein using a separation gel medium onto the surface of a non-reactive membrane, most often a nitrocellulose membrane. This technique was first introduced for transferring proteins from polyacrylamide gels to nitrocellulose membranes [177].

2.2.2 Cell line lysates

The cell lysates that were subjected to WB were obtained from different BC cell lines (MCF-7, ZR-75-1, SK-BR-3, and MDA-MB-231) or lysates from another cell line (HEK-293) to validate the specificity of the antibodies used in IHC analysis. The specificity was based on the presence of a specific protein band.

2.2.3 Cell culture and preparation of cell lysates

MCF-7, ZR-75-1, SK-BR-3, and MDA-MB-231 cell lines were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). MCF-7, ZR-75-1, and MDA-MB-231 cells were cultured in Roswell Park Memorial Institute (RPMI 1640) medium with L-glutamine (Cytiva, SH30027.01, UK) supplemented with 10% foetal bovine serum (FBS) (Sigma, F9665, UK); the SK-BR-3 cell line was grown in McCoy's 5A medium modified with L-glutamine and sodium bicarbonate liquid (Sigma, M9309, UK) supplemented with 10% FBS. Cells were cultured at 37°C in a humidified incubator containing 5% carbon dioxide (CO_2). The cells were grown to about 80–90% confluency, at which point the growth medium was removed and the cells washed with sterile Dulbecco's phosphate-buffered saline (DPBS) (Sigma, D8537, UK) to eliminate the medium residue. The cells were then trypsinised by adding trypsin ethylenediamine tetraacetic acid (EDTA) solution (Sigma, T3924, UK) and incubated at 37°C for less than 5 min to aid in the detachment of the cells, which were then examined under a microscope to confirm their detachment. Trypsin was neutralised by adding fresh medium to the detached cells and centrifuged at 1,000 rpm for 5 min. The cell pellet was then resuspended in fresh medium and counted to a concentration of $1x10^6$ cells/ml. A Neuber haemocytometer was used to count viable cells, and a mixture was prepared by adding trypan blue (a substance that stains non-viable cells blue) to the cell suspension at a 1:1 ratio. On the haemocytometer, 20 µl of the trypan blue/cell mixture was loaded beneath a coverslip and only viable cells were counted in the central (5 × 5) grid using light microscopy. All counts were performed in duplicate and the average cell count was multiplied by the dilution factor 2×10^4 to obtain the number of cells/ml. Following the determination of the cell count, the cell suspension was centrifuged again for 5 min and the pellet was placed on ice. A total of 1 ml of RIPA buffer for each 10^6-cell pellet (on ice) (25 mM Tris, 150 mM sodium chloride, 1% Nonidet P-40, 1% sodium deoxycholate, and 0.1% sodium dodecyl sulphate

[SDS]) (Thermo Fisher Scientific, 89900) and protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific, 04693124001, UK) were added and the pellet was resuspended. The lysed cells were incubated for an additional 20 min on ice with frequent gentle shaking and then centrifuged at 4°C at 13,000 rpm for 10 min to clear cell debris. The supernatant was removed and stored at -20°C for future use.

2.2.4 Cell lysate protein quantification (Pierce™ BCA protein assay)

To quantify proteins in the cell lysate, the Pierce™ BCA protein assay kit was used (Thermo Fisher Scientific, 23225, UK), applying the standard curve method of pre-determined protein concentrations of bovine serum albumin (BSA). Standard solutions of BSA were prepared in nuclease-free water to obtain the following concentrations: 2,000 µg/ml, 1,500 µg/ml, 1,000 µg/ml, 750 µg/ml, 500 µg/ml, 250 µg/ml, 125 µg/ml, 25 µg/ml, and 0 µg/ml, which served as the blank. A 96-well plate was pre-filled with 8 µl of diluted protein samples (diluted 1:5 in nuclease-free water) and the standards, and each sample was placed in the plate in triplicate. BCA reagent (200 µl) was added to each of the wells containing the diluted protein samples and the standards, and the plate was incubated for 30 min at 37°C before the determination of the protein concentration using a microplate reader (TECAN Infinite F50, UK) at a wavelength of 595 nm. The raw data file was exported to Microsoft Excel for analysis.

2.2.5 Preparation of WB reagents and tools

Prior to processing the samples for WB, all required reagents and experimental tools were prepared for use. The XCell SureLockTM Mini-Cell and XCell IITM Blot Module (Invitrogen, EI0002, UK) were assembled for protein isolation and the transfer of proteins to the nitrocellulose membrane (Sigma, GE10600008, UK). A Bolt^{TM} 4–12% Bis-Tris plus 10-well protein gel (Thermo Fisher Scientific, NW04120BOX, UK) was used for the denatured and reduced protein separations. A working solution of the running buffer was freshly prepared by diluting 50 ml of 20x MOPs Bolt

running buffer (Thermo Fisher Scientific, B000102, UK) with 950 ml of distilled water. The transfer buffer for one gel mixture was prepared by mixing 50 ml of 20x Bolt transfer buffer (Thermo Fisher Scientific, BT00061, UK) with 100 mL of methanol that had previously been diluted in 849 ml of distilled water and 1 ml of Bolt Anti-Oxidant reagent (Thermo Fisher Scientific, BT0005, UK). The wash solution was prepared by dissolving 1 ml of the non-ionic detergent Tween-20 (Sigma, P2287, UK) in 1,000 ml of phosphate-buffered saline (PBS) solution. The blocking solution was prepared by dissolving 2.5 mg of Marvel's dried and fat-free milk at a concentration of 5% in 50 mL of washing solution. All primary antibodies were prepared as needed prior to use and based on previously optimised working conditions. The following fluorescent secondary antibodies prepared in blocking solution at a ratio of 1:15,000 were used for the detection of the primary antibodies: IRDye 800CW donkey anti-rabbit immunoglobulin G (IgG) (H+L) (LI-COR Biotechnology, 926-32213, Cambridge, UK) and IRDye® 680RD donkey anti-mouse IgG (H+L) (LI-COR Biotechnology, 926-68072, Cambridge, UK). β-actin and GAPDH were used as housekeeping and loading controls to ensure equal loading of all samples in the gel. Primary rabbit anti-β-actin antibody (4970S, clone number 13E5, Cell Signalling Technology) and primary anti-mouse β-actin antibody (3700S, clone number 8H10D10, Cell Signalling Technology) were used. Primary rabbit anti-GAPDH monoclonal antibody (ab181602, clone number EPR16891, Abcam) and primary anti-GAPDH monoclonal antibody (ab8245, clone number 6C5, Abcam) were also used as loading controls.

2.2.6 Sample preparation

Each cell line lysate and the PageRuler™ Plus Prestained 10–180 kDa Protein Ladder (Thermo Scientific, 26616, UK) were thawed in a container of crushed ice. Based on the protein concentration of each lysate, the detected amount of each lysate was mixed with its corresponding amount of Bolt LDS Sample Buffer 4x (Invitrogen, Thermo Fisher Scientific, B0007, UK) and Bolt Reducing Agent 10x (Invitrogen, Thermo Fisher Scientific,

B0009, UK). All tubes were spun lightly using a laboratory vortex mixer and denatured on a dry block heater using a protein mixture at 100°C.

2.2.7 Loading of samples and separation of the denatured protein mixture

A 10-well BoltTM 4–12% Bis-Tris plus protein gel (Thermo Fisher Scientific, NW04120BOX, UK) was inserted into the XCell SureLockTM Mini-Cell running tank (Invitrogen, EI0002, UK). The previously prepared running buffer was then slowly poured into the running tank to the maximum fill line and no leakage of the buffer from the running tank was observed. To protect the protein from possible oxidation, 1 mL of the Bolt Antioxidant reagent (Thermo Fisher Scientific, BT0005, UK) was added to the running buffer. The prepared protein samples were loaded into the wells, starting from the first well on either the left- or right-hand side of the gel. The running tank was then connected to an electrical power supply (Invitrogen, Life Technologies, UK) and set to run for 60 min at 160 V.

2.2.8 Protein transfer onto nitrocellulose membranes

After electrophoresis, the proteins were transferred onto a nitrocellulose membrane. To ensure the proper transfer of the protein onto the nitrocellulose membrane, the sponges and filter papers were made to fit at the exact size and specifications of the protein separation gel and were placed in a cassette containing the transfer buffer. The gel was then placed into the cassette contained within an opened blotting module in the following order: 2x sponge layered with 2x filter paper, followed by the gel, 1x nitrocellulose membrane, 2x filter paper, and 2x sponge. To prevent distortion of the protein transfer onto the nitrocellulose membrane, air bubbles were removed using a cylindrical roller passed over the layered gel in filter paper. The blotting module was placed firmly in a cassette containing a previously prepared transfer buffer in the running tank, and electroblotting was performed at 10 V for 90 min.

2.2.9 Blocking and antibody probing

Following the transfer of the proteins onto the nitrocellulose membrane, the membrane was laid in a weighing boat containing a 5% blocking solution. To prevent unspecified binding of the primary antibody, blocking was performed for 60 min at room temperature on a rocker. The membrane was then incubated in 5 mL of primary antibodies diluted to their respective optimised concentrations. A 1:5,000 positive control, either monoclonal anti-β-actin or anti-GAPDH primary antibody, was included in all experiments and incubated overnight with the primary antibody in a Falcon tube at 4°C in a cold room on a roller. Following incubation, the membrane was washed three times for 5 min each on a rocker. The secondary antibodies were diluted at a concentration of 1:15,000 by adding 1 µl of antibody to 15 ml of blocking solution. The nitrocellulose membrane was then immersed in the secondary antibody solution and incubated for 60 min at room temperature with continuous rocking. To avoid bleaching the protein signal, the membrane was shielded from direct light. To eliminate unbound secondary antibodies, the membrane was washed for 5 min three separate times for a total of 15 min, with each wash being 5 min in total.

2.2.10 Visualisation of protein bands

Protein visualisation was achieved using the Odyssey Fc with Image Studio 4.0 (LI-COR Biotechnology, 2800, UK). Visualisation of proteins on the membranes by fluorescent detection was conducted at 600 nm, 700 nm, and 800 nm wavelengths. The IRDye® 680RD secondary antibody, which had an emission wavelength of 700 nm while the IRDye® 800CW secondary antibody had an emission wavelength at 800 nm, were labelled with an infrared dye. Images of both anti-rabbit and anti-mouse secondary antibodies were digitally generated in case the molecular weight of the overlapped with the targeted antigen of a specific biomarker. The positive control, β-actin, had a molecular weight of ~42 kilodaltons (kDa), while GAPDH had a molecular weight of ~36 kDa. The molecular weights of each antibody used in this study are summarised in Table 2.5.

2.3 Tissue handling and processing

2.3.1 FFPE block retrieval from the Nottingham Health Science Biobank (NHSB)

All FFPE donor blocks from the BC patients in this study were retrieved from and approved for use by the NHSB, reference number 15/NW/0685. Details regarding the ethical approval of this study are found in Section 2.7.

2.3.2 TMA construction

TMA is a high-throughput technology that allows for representative tissue cores from over 100 archival tissue blocks to be extracted and embedded in a single recipient block. TMA enables simultaneous high-throughput assessment that saves time and ensures that assay conditions are consistent across all tissue samples, resulting in more reproducible and consistent results. The TMAs used in this study were previously prepared using the steps described in this section.

2.3.2.1 Sectioning in the microtome

Safety procedures and risk assessment guidelines were followed by the researcher prior to the preparation of FFPE blocks for staining. FFPE sections of patient tissues were used for routine histological examination with both H&E and IHC staining. The instruments used to obtain tissue sections included a rotary microtome (Leica Microsystems, RM2125, UK) and a mounting bath for paraffin sections (Thermo Scientific, MH8516, UK), which was filled with deionised water to prevent the distortion of sectioned tissues; in addition, Leica X-tra® or Leica's Apex adhesive slides were used for application of H&E and IHC. Maximum adhesion occurred under hydrophilic conditions due to the presence of positive charges on the surfaces of both slide types. A plastic box with crushed ice, disposable microtome blades (CellPath, S35 Feather, Wales), curved-tip forceps, and a xylene-resistant marker pen or pencil were also made available for the procedure.

Prior to the cutting of the FFPE blocks, the blocks were cooled by placing them face-up on ice. The chilled blocks were then inverted face-down onto the ice for a maximum of 5 min. Before positioning the FFPE blocks in the microtome, water droplets were removed from the block to facilitate the positioning. The face of the block was placed parallel to the blade edge in both axes and, if required, the microtome chuck was unclamped and the block orientation was adjusted appropriately. The thickness of the sections was set at 4 µm. The microtome handle was turned clockwise until the sections were generated. The block face was appropriately trimmed until a 'full-face' section was produced. For optimisation purposes, extra sections were generated for the TMA blocks. Forceps were used to retrieve the ribbon of full-face sections produced by the sectioning. The sections were then carefully placed on the surface of the water bath. Individual sections were separated with forceps following their expansion and single sections were then mounted onto microscope slides by immersing the slides in the water directly beneath the tissue at an angle of about 45°. Sections were realigned over the slide when necessary and slides mounted with tissue sections were then placed vertically in a drying rack and dried overnight in an oven at 37°C. Slides containing cut sections were then placed into a labelled storage box that was subsequently stored in a cold room at 4°C until needed.

2.3.2.2 Preparation of donor blocks and TMA construction

Archival FFPE blocks mounted with invasive BC tissue sections were retrieved from the NHSB. Sections of H&E-stained tissues from these blocks were then evaluated to identify invasive malignant cells for sampling. Trained pathologists were responsible for marking representative tumour areas on each slide alongside their corresponding locations on the FFPE blocks. The marked blocks were then loaded onto a Grand Master® automated TMA machine (3D HISTECH®, Budapest, Hungary) in preparation for mounting onto paraffin blocks. Molten paraffin wax at a temperature of 55–58°C was poured into specially designed moulds

containing the tissue cassettes onto which the tissues had been mounted. The paraffin was solidified on ice and the mould was removed. Moulded paraffin blocks were then stored either at 4°C in the short term or at -20°C in the long term. Prior to the construction of TMA blocks, the blocks were trimmed and marked areas of the tumours from the donor blocks were harvested using a 0.6 mm needle. Each block could hold a maximum of 150 cases and three orientation control tissue cores (kidney, liver, or tonsil) positioned at the start of the first row and two more at the end of the last row to ensure proper alignment and accurate core evaluation. Maps were prepared for each TMA block to display the medical identifiers on each well. The TMA was constructed by Mr. Glynn Donovan (an employee in the Biobank).

2.4 Immunohistochemical staining

2.4.1 Overview

IHC is an invaluable molecular diagnostic technique that is essential in practice for the clinical, diagnostic, and therapeutic management of BC patients. From molecular subtyping to the identification of validated prognostic biomarkers, IHC plays an important role in the diagnostic and pathology reporting of many neoplastic diseases, including invasive BC. The principles of IHC are based on the binding affinity between an antibody and an antigen, which is most often reversible. IHC is an important biomolecular technique that has been used to determine the distribution of a particular protein antigen that is present on the surface of healthy or diseased tissues. It is particularly useful in the diagnosis and staging of cancers due to the *de novo* expression of cancer-specific antigens or overexpression of these antigens in specific neoplasms. IHC is performed on tissue biopsies from patients and the tissue is processed into sections using a microtome before staining with antibodies suitable for IHC. Utilising the biochemical binding between an antibody and an antigen in clinical practice has led to the improvement of patients' clinical profiles and more accurate prognosis by

combining the observed histological features of a tumour with its protein expression patterns.

2.4.2 IHC tools, chemicals, and procedures

In this study, IHC was used for the semi-quantification of protein expression on invasive BC TMA tissue sections that were 4 µm thick. A random number (10–15) of full-face sections of samples from different molecular BC subtypes, histological types, and tumour grades were stained by IHC to assess their pattern of expression and to verify the staining homogeneity before the staining of TMA sections. All tissue sections were acquired from invasive or invasive-mixed FFPE samples. The NovolinkTM Max Polymer Detection System (Leica Biosystems, RE7150-K, UK) was used for the IHC procedure; this detection system is suitable for the detection of protein epitopes on the surface of tissues using both mouse and rabbit IgG antibodies.

2.4.2.1 Labelling of slides

The slides used for the IHC procedure were Leica X-tra® slides. These slides were appropriately labelled to prevent the misidentification of tissue specimens. The label on each slide contained the following information: date of staining, symbol of the target protein, user name, dilution factor of each antibody, and the specimen identification number.

2.4.2.2 Deparaffinisation and rehydration of tissue sections

Dewaxing of the tissues was carried out by placing the TMA sections on a hot plate (Leica Biosystems, 14042321474, UK) to melt the paraffin. The BC tissue was firmly fixed to the slides face-up and incubated at 60°C for 10 min. Following the heating step, the slides were allowed to cool for 5 min. The slides were then placed into a Leica's slide rack (Leica Biosystems, 14047533750, UK) which was then immersed in a reagent vessel filled with fresh xylene (Fisher Scientific, X/0250/PB17, UK) contained inside the chamber of the Leica AXL. The dewaxing programme was activated and carried out under the following conditions: impregnation of the slides for 5

min in xylene followed by transfer of the slides into a fresh solution of xylene and incubation for an additional 5 min. The slides were washed three times with industrial methylated spirits (IMS) (Thermo Fisher Scientific, M/4450/25), with the first two baths in 100% IMS for 2 min and the third bath in 70% IMS for 2 min. Rehydration of the slides with de-ionised water was carried out three times for 5 min each.

2.4.2.3 Heat-induced epitope retrieval (HIER) using citrate buffer at pH 6

Formalin or aldehyde fixation and subsequent paraffin embedding may lead to extensive physical and chemical alterations of the proteins in tissue biopsies. Chemical modification of antigens present in FFPE tissues can take the form of transpositions or alterations to the antigenic site. Poorly exposed epitopes may lead to failure of IHC staining. Therefore, this study used optimal conditions to provide the best reexposure of the epitopes located on the surface of the FFPE tissues. Numerous HIER methods have been tested under various conditions. The citrate buffer antigen retrieval method was chosen as the optimal antigen retrieval method for all antibodies used in this study, as recommended by antibody manufacturers.

2.4.2.4 Preparation of sodium citrate buffer at pH 6

Sodium citrate buffer was prepared by dissolving 2.94 mg of Tris-sodium citrate dihydrate in 1,000 mL of distilled water. The pH was adjusted to 6 by adding 1 N of HCl solution dropwise to the buffer solution, and the pH was determined using a calibrated pH meter. The buffer was then stored at 4°C. A total of 500 mL of pH 6 citrate buffer was then placed into a plastic Leica AXL vessel and the rack from step 2.4.2.2 containing the slides placed in the buffer. To prevent evaporation and to ensure homogeneous distribution of the buffer to all slides, the Leica AXL vessel was covered with a plastic lid with no metal handles. The slides in the citrate (pH 6) vessel were placed inside a laboratory microwave (Whirlpool JT 359 Jet Chef 1000 W, UK) and heated at an optimum temperature for 20 min. To prevent eye

injury, protective glasses were worn by the researcher and the vessel was removed from the microwave using heat-resistant gloves. The lid of the vessel was removed in the sink and flowing water was applied into the vessel with a rubber hose for 5 min to cool the vessel and remove the citrate buffer from the slides.

The following steps were carried out manually, using either Sequenza trays for full-face sections or a flatbed for TMA sections. Slides were inserted into Sequenza plates (Thermo Scientific) under running tap water, inspected for air bubbles, and placed into Sequenza trays for staining. The slides were washed twice with Tris-buffered saline (TBS, pH 7.6). When using flatbed staining, the slides were placed on a flatbed tray and a hydrophobic barrier marker was used across the breadth of the slides to limit the soaking area and to ensure that the chemicals remained within the TMA cores. TBS (pH 7.6) containing 0.1% Tween 20 (TBS-T) was used to wash the slides twice. The remaining steps were identical, except for the use of TBS for washing in the Sequenza staining method and TBS-T for flatbed staining.

2.4.2.5 Blocking of endogenous peroxidases

Endogenous peroxidase activity is a biochemical feature of all tissues, with continuous peroxidase activity occurring even after the fixation of tissues with formalin and subsequent embedding of the tissues in paraffin. As such, the application of horseradish peroxidase (HRP)-conjugated secondary antibodies to the tissues may lead to the cross reaction of tissue peroxidases causing non-specific binding and background staining. It is therefore important to block endogenous peroxidases prior to incubation with HRP antibodies. The importance of the endogenous peroxidase-blocking step was tested by incubating the tissues with a 3,3'-diaminobenzidine (DAB) substrate, the specific details of which are provided in step 2.4.2.10. A change in the colour of the tissues to brown demonstrated the need for the blocking of endogenous peroxidases. Then, 300 µl of Novolink peroxidase blocking reagent was added to each slide. All slides were incubated at room temperature to permit the blocking of

endogenous tissue peroxidases, then the slides were washed with TBS-T three times at 5 min intervals.

2.4.2.6 Blocking non-specific reacting proteins

Non-specific staining of tissue sections may also occur due to the presence of epitopes that display protein sequence similarity with the target protein that might be recognised by the applied antibody. The non-specific binding of the applied antibody on near-identical epitopes present on the tissue surface may occur in different forms, such as charge-based and hydrophilic/hydrophobic attractions. To prevent non-specific binding, 300 µl of Novolink protein block reagent was added to the surface of the tissue sections and the slides were further incubated for 5 min and washed with TBS-T using the same washing procedures as in Section 2.4.2.5.

2.4.2.7 Applying the optimised primary antibody

Primary antibodies were optimised for IHC staining of tissue sections (specific details of antibody optimisation will be discussed in section 2.4.3). BondTM primary antibody diluent (Leica Biosystems, AR9352, UK) was the universal diluent used for all primary antibodies. Using a fixed-volume micropipette, 300 µl of the optimally diluted primary antibody was applied to slides containing the tissues, with a few slides serving as blanks or controls to test the user's accuracy; the positive and negative controls are detailed in Section 2.4.4. The primary antibodies were incubated for 15 min or 60 min at room temperature or overnight in a cold room at 4°C. The tissues were then washed with TBS-T, as in the previous steps.

2.4.2.8 Application of the post-primary antibody

A total of 300 µl of the post-primary antibody was applied to all slides. The slides were incubated for an additional 30 min, then washed three times in TBS-T for 5 min each.

2.4.2.9 Application of NovolinkTM polymer

The Novolink™ polymer functions by binding to the epitope, and detecting the post-primary and monoclonal or polyclonal primary antibodies. Each slide was incubated at room temperature for 30 min with 300 µl of the Novolink™ polymer solution containing the primary antibody solution. Each slide was then washed three times with TBST for 5 min.

2.4.2.10 Detection with DAB

A solution of DAB peroxidase substrate was prepared by diluting DAB chromogen with Novolink™ DAB substrate buffer polymer. The dilution ratio was 1:20 and the solution was freshly prepared for each experiment. The diluted DAB solution was always kept in the dark until use. A total of 300 µl of the solution was applied to each slide and incubated for 5 min, and excess chromogen was removed by washing with TBS-T three times for 5 min.

2.4.2.11 Counterstaining of tissue slides with haematoxylin and tissue dehydration

Counterstaining of slides is essential to provide clear and sharp nuclear stained images to facilitate semi-quantification of the IHC-stained protein. Each of the slides was incubated with 300 µl of haematoxylin reagent for 6 min at room temperature. The cover plates on each slide were then removed in a water tank, and the excess haematoxylin solution was removed from the slides. The slides were reassembled in the Leica AXL rack and placed into a plastic vessel filled with distilled water. The slides were then dehydrated using the Leica AXL system: washing of the slides with IMS three times for 2 min each time, and clearing the slides with xylene twice for 5 min each time. For proper visualisation, dibutylphthalate polystyrene xylene (DPX) medium (BDH, Leica Microsystems, Newcastle, UK) was used to mount the stained tissues, and transparent coverslips were used to seal the slides. The slides were stored at room temperature overnight to dry.

2.4.3 Optimisation of primary antibodies

Antibody optimisation was carried out on the BC TMA sections before staining the whole series; this process was carried out through the modification of different variables in the protocol of the staining until optimal staining with the primary antibodies was observed. The variation in protocol steps included three dilutions of the stock solution obtained from the manufacturer. Optimisation of the conditions used for the staining was based on the dilution recommended by the supplier in addition to three or more dilutions above and below the recommended dilution. Optimisation was also performed by alteration other staining conditions, such as incubation time.

2.4.4 Negative and positive controls

Negative and positive controls were included in all IHC experiments. Negative control slides were used to test the binding specificity of the secondary antibody together with the fragment crystallisable region (Fc region) of the primary antibody. Specific tissues that showed no expression of the target from the human protein atlas (http://www.proteinatlas.org/) were chosen as a negative control. BC tissue with no applied primary antibodies was used when specific negative control tissues were not available. As advised by the antibody manufacturer, or by the selection of a specific tissues from the human protein atlas, specific positive controls were included in the experiments. Positive and negative controls, the dilution factor, and the incubation time for each primary antibody are shown in Table 2.5.

Table 2.5 Details of the primary antibodies used in this study.

Western blot (WB)						Immunohistochemistry (IHC)			
Antibody	**Supplier**	**Species**	**Product code**	**Dilution**	**Molecular weight (kDa)**	**Dilution**	**Incubation time**	**Negative control**	**Positive control**
MX1	Abcam	Rabbit polyclonal	Ab95926	1:1,500	~64	1:100	1 h	BC tissue	Kidney
IDH2	Abcam	Mouse monoclonal	Ab55271	1:500	~47	1:500	1 h	BC tissue	Tonsil
HMGB3	Sigma-Aldrich	Rabbit polyclonal	HPA062583	1:1,000	~30	1:500	1 h	Kidney	Ovary
CCNB1	Abcam	Mouse monoclonal	Ab72	1:1,000	~40	1:5,000	15 min	Liver	Tonsil
CCNB2	Abcam	Rabbit monoclonal	ab185622	1:1,000	~42	1:50	Overnight	Kidney	Colon
RRM2	Abcam	Mouse monoclonal	Ab57653	1:1,000	~39	1:1,000	15 min	Liver	Stomach

2.5 Imaging of tissue sections and the histological scoring (H-score) system

2.5.1 Whole-slide imaging

Storage of IHC-stained tissues for a prolonged time in sub-optimal conditions, such as with direct exposure to light, can lead to a decrease in the intensity and the quality of the IHC staining. Modern clinical practices utilises digitalisation of histological tissue sections, which has aided pathologists in adequately storing pathological information from FFPE tissues of cancer patients for longer periods of time. Digitalisation also allows for sharing of the sections with healthcare colleagues and the editing of the patient's clinical data when necessary. In this study, all IHC-stained tissue sections were imaged at the electron microscopy and imaging unit of the Cellular Pathology Department, Queen's Medical Centre (QMC), A Floor, West Block, NUH. A specific request form for the scanning and imaging of the slides was made and completed by the researcher. A Nanozoomer Digital Pathology scanner (Hamamatsu Photonics, Welwyn Garden City, UK) was used to obtain high-resolution digital images (0.45 μm/pixel) of whole slides. Images were obtained for scoring of the IHC-stained tissues using a magnification power of 20x. Aperio ImageScope software (Aperio Technologies, Milton Keynes, UK) or Xplore 5.1.0 software (Philips, UK) was used to view digital slides.

2.5.2 H-score

The selected scoring method for interpretation of IHC-stained slides and for the results needed to be credible and reproducible. The semi-quantitative scoring method used to quantify the results from IHC-stained tissues also required expert and well-trained observers to score the TMA cores with a high degree of concordance. The H-score test is a common method that was used to score IHC-stained tissue sections in this study. The H-score is generated using four different scoring intensities from randomly scanned TMA sections from different slides: negative intensity (0), weak intensity

(1), moderate intensity (2), and strong intensity (3) [178]. The H-score is defined as the sum of individual IHC-stained intensities multiplied by the total percentage of positively stained invasive BC cells that are observed to express the protein of interest. The H-score can be obtained according to the following formula:

[0 × (% cells expressing intensity 0)] + [1 × (% cells expressing intensity 1)] + [2 × (% expressing intensity 2)] + [3 × (% expressing intensity 3)].

All non-representative cores were removed from scoring, including cores that had <15% tumour cells relative to the total core area, cores that included only stroma or normal breast tissue, and cores that had become folded during processing or staining. Scoring of IHC TMA slides was blind and performed individually by a trained pathologist alongside the main researcher for at least 20% of the cohort under investigation. The scoring results were analysed only upon completion of the H-scoring of each slide by the principal researcher. A threshold agreement of >75% was set for concordance and intra-class correlation between the scores for each slide. In the event of a lower scoring concordance, the slides were rescored, and the differences between the scores were discussed by the main researcher and a clinical pathologist. If a heterogeneous pattern of protein expression was observed in IHC-stained tissues, the results could be presented as a single TMA core; in this case, IHC tissues with a heterogeneous IHC staining pattern could not be scored. As such, a small panel of full-face tissue sections were stained by IHC to assess whether TMA was suitable for this study.

2.5.3 Identification of the cut-off points of IHC staining

All biomarker expression data were dichotomised at the mRNA and protein levels to determine a cut-off point that allowed the cohort to be divided into two categories: expression greater than the defined cut-off point (high-expression group) and expression less than or equal to the identified cut-off point (low-expression group). One method used to detect appropriate

dichotomisation for biomarkers was dependent on the distribution of mRNA or protein expression. The cut-off points for IHC data were determined based on either the mean or median of the data, and this was dependent on the normal or skewed distribution of the data. Frequency histograms were also used to visualise the distribution of the samples and to determine apparent cut-off points. The antibody expression used for IHC staining and the respective cut-off points applied to these antibodies will be discussed in the following chapters.

2.6 Cell culture (*in vitro*)

In vitro assessment was carried out to investigate the potential role of the key targeted proteins (HMGB3 and CCNB2) in LVI and other biological functions using the following cell lines and assays.

2.6.1 Cell lines, general growth conditions, and media

2.6.1.1 BC cell lines

Based on the expression level, three BC cell lines were used in this study for *in vitro* assessments: MCF-7 ($ER^+/PR^+/HER2^-$), ZR-75-1 ($ER^+/PR^{\pm}/HER2^-$), and SK-BR-3 ($HER2^+$). MCF-7 and ZR-75-1 represent luminal molecular classes, while SK-BR-3 represents the HER2-enriched class. These two molecular classes are associated with the highest incidence of LVI [179]; furthermore, they were chosen based on the findings of an IHC investigation of the key targeted proteins (HMGB3 and CCNB2) in various BC subtypes. The cells were initially acquired from the American Type Culture Collection (ATCC, Manassas, VA USA). These cells were cultured at a temperature of 37°C inside a humidified incubator that contained approximately 5% CO_2. RPMI 1640 medium was used to maintain MCF-7 and ZR-75-1, while McCoy's 5A medium was used to maintain SK-BR-3, as described in Section 2.2.3. Table 2.6 summarises the characteristics of the cell lines used in the *in vitro* experiments.

Table 2.6 Characteristics of BC cell lines used in *in vitro* experiments.

Cell lines	Histological type	Receptor status	Subtype	Source	Reference
MCF-7	Human breast adenocarcinoma	ER^+, PR^+, and $HER2^-$	Luminal A-like	Isolated from a pleural effusion of a 69-year-old Caucasian woman in 1970.	[180, 181]
ZR-75-1	Human breast ductal carcinoma	ER^+, PR^-, and $HER2^-$	Luminal B-like	Established in 1978 from ascitic fluid of a 63-year-old Caucasian woman.	[180-182]
SK-BR-3	Human breast adenocarcinoma	ER^-, PR^-, and $HER2^+$	HER2-enriched	Established in 1970 from a pleural effusion of a 43-year-old Caucasian female.	[180, 181, 183]

2.6.1.2 Endothelial cell lines

Endothelial cell monolayers have been used to study tumour–endothelium interactions. Primary human umbilical vein endothelial cells (HUVECs) and dermal lymphatic endothelial cells (DLECs) were used in this study as *in vitro* models of tumour–endothelium interactions. HUVECs and DLECs were purchased from Promocell (C12218, Heidelberg, Germany) and cultured in MV2 endothelial cell growth medium (Promocell, C-22022, Germany). HUVECs and DLECs were used between passages 2 and 10. Being primary cell lines, short passage windows were used due to accelerated senescence and spontaneous apoptosis.

All tissue culture procedures were carried out in a Class II biological safety cabinet, and 70% ethanol was used to clean the cabinet before and after use. T75 75 cm^2 tissue culture flasks (Corning, Fisher Scientific, UK) were used for routine cell culturing. Other sizes of flasks and plates were used based on the specific requirements of various experiments. The reagents that were used for tissue culture had to be pre-warmed in a water bath maintained at the temperature of 37°C. Routine monthly mycoplasma testing for all BC cell lines in this study using the PlasmoTest™ mycoplasma detection test (InvivoGen, rep-ptrk) was conducted by a team of researchers from the Bio-Discovery Institute.

2.6.1.3 Subculturing of adherent cell lines

Pre-warming was carried out in a 37°C water bath for 15–30 min before the reagents could be used. The first steps of sub-culturing were identical to those described in section 2.2.3. Tissue culture flasks were used to culture adherent cells using standard culture media until 80% to 90% confluence was reached.

Following centrifugation and removal of the remaining trypsin, cells were resuspended in fresh culture medium and the desired proportion was plated onto new flasks. The cells in the flasks were labelled with their name, passage number, and date of subculturing and carefully placed in a 37°C humidified incubator with 5% CO_2. The total number of passages for all cell lines was no greater than 25.

2.6.1.4 Cryopreservation and thawing of cell lines

To maintain low passage numbers for the cells, they were stored in liquid nitrogen for future experiments. Adherent tumour cells were trypsinised, as previously described, and resuspended in freezing media (10% dimethyl sulfoxide [DMSO]; Sigma Life Science, SHBK2703, USA) with complete cell growth medium for cryopreservation. The steps for freezing endothelial cells were identical to those used to freeze tumour cells, except the Cryo-

SFM freezing medium (Promocell, C29910, Germany) was used instead of DMSO with cell growth medium.

The final cell concentration in every cryovial was 1×10^6 cells/ml. The cells were placed into pre-labelled cryovials and frozen at -80°C overnight using CoolCell alcohol-free cell-freezing containers. Of note, these containers can control the rate at which the internal temperature decreases to 1°C/min which prevents ice crystal formation in the cells. After approximately 24 h, the cryovials were transferred to liquid nitrogen storage at -180° C for preservation.

To thaw the cells from the liquid nitrogen, the cryovials were exposed to either room temperature or a water bath at 37°C immediately after removal from storage. After thawing the cells, the DMSO residue was extracted by moving them into a centrifuge tube containing 9 ml of media and centrifuging for 5 min at 170 g. After centrifugation, the supernatant was discarded, and the cell pellet was resuspended in the growth medium and cultured under standard conditions.

2.6.2 Small interfering RNA (siRNA) transfection of HMGB3 and CCNB2

2.6.2.1 Overview

To explore the potential functional implications of depletion of the selected gene and to study its involvement in BC development, a siRNA-based approach using BC cell lines was used. siRNA is a form of RNA interference (RNAi) that causes gene silencing through the degradation of mRNA. In this approach, gene silencing is achieved by introducing siRNA into the cytoplasm where it is integrated into a multi-protein complex called the RNA-induced silencing complex (RISC). When the RISC is activated, one of the siRNA strands is cleaved and released from the complex, while the remaining siRNA strand acts as a template for the targeted mRNA. Once this occurs, a ribonuclease cleaves and degrades the mRNA [184]. To

achieve transient knockdown of the targeted gene in this study, siRNA was transfected into cells using Lipofectamine.

In this study, the forward transfection approach was used to knock down the genes, as it is one of the standard transfection practices. This method is effective for most adherent cell types that are seeded at least 1 day before transfection. The aim of this approach is to produce a cell population that is actively dividing at the time of transfection. In contrast, reverse transfection refers to a practice in which seeding of different cells occurs at the time of transfection.

2.6.2.2 Cell line selection

MCF-7, ZR-75-1, and SK-BR-3 cell lines were used in this study, as discussed in Section 2.6.1.1. Before knockdown of the studied gene, evaluation of protein expression in different BC cell lines was completed by WB. Cell lysate of the BC cell lines was freshly prepared, as described in Section 2.2.3, to quantify the proteins of interest in each selected cell line. Protein quantification was performed, as described in Section 2.2.4, before loading the cell lysate on the WB to ensure that the same amount of protein was loaded for each cell line. After the blot image was generated by the Odyssey Fc Imager, as explained previously in Section 2.2.9, protein band intensity analysis was performed by drawing a rectangle around each band using Image Studio software (LI-COR Imaging Systems) to show the intensity of each band, representing the amount of protein in the band. The net values of the proteins for each experimental condition were obtained by dividing the protein intensity by the intensity of the loading control (β-actin or GAPDH) in the same lane. After plotting the net values, a graph was generated using Excel software to evaluate the expression of the selected protein in the selected cell line.

2.6.2.3 Subculture of the selected cell line and cell plating

After confirming the expression of the target protein in each cell line, the cells were sub-cultured, as described in Section 2.6.1.1. Cells were

regularly inspected prior to transfection to ensure that they were growing well and were mycoplasma free. When the cells reached 80% confluency, they were harvested, suspended in trypsin-EDTA, and counted using a haemocytometer, as described in Section 2.2.3. Cells were then seeded on 6-well plates at different cell densities to determine the optimal number of cells needed to achieve knockdown. ZR-75-1 and SK-BR-3 were seeded at 0.25×10^6 cells/well, while MCF-7 was seeded at 0.30×10^6 cells/well. The cells were incubated with 1,250 µl/well of ordinary cell culture media. Three 6-well plates were prepared to assess the effect of the transfection on the cells on different days (day 3, day 5, and day 7) prior the transfection. The cells were incubated overnight at 37°C in a humidified incubator with 5% CO_2 to allow for adhesion.

2.6.2.4 Preparation of transfection reagents

Pre-validated Silencer® Select siRNA constructs at a concentration of 5 nmol were used for *HMGB3* (115646) and *CCNB2* (S17446); the constructs were purchased from ThermoFisher Scientific. The antisense sequences were 5'-GCACCCUGAAACUGUAUCAtt-3' and 5'-GGCGAACUGUUUUAGAAGAtt-3' for *HMGB3* and *CCNB2*, respectively. The siRNAs were tested at different concentrations to determine the knockdown efficiency. A scrambled negative-control siRNA (Silencer® Select siRNA, AM4611, ThermoFisher Scientific) was included in all experiments. Stock concentrations of *HMGB3* and *CCNB2* siRNAs were prepared at a concentration of 100 µM by resuspending 5 nmol lyophilised siRNA in 50 µl nuclease-free water after briefly spinning the siRNA tubes. For immediate use, 10-µM working stock was prepared via 1/10 dilution of the 100 µM stock using nuclease-free water. A transfection master mix of the required reagents was created to avoid pipetting small amounts and to limit pipetting error. Finally, these reagents were stored at -20°C freezer until use. Once reconstituted in nuclease-free water, siRNA was transfected at the optimal concentration. Lipofectamine® RNAiMAX (ThermoFisher Scientific,

13778150, UK) was in its ready-to-use format and required no additional preparation.

2.6.2.5 Adding transfection reagents

siRNAs and negative controls were optimised using different siRNA and Lipofectamine concentrations to achieve the maximum knockdown efficiency determined by WB. Inside the Class II biological safety cabinet, 5 µL Lipofectamine® RNAiMAX reagent was diluted in 250 µL Opti-MEM® medium (ThermoFisher Scientific, 31985062, UK). The siRNA (10 µM) was diluted in 250µL Opti-MEM® medium. Diluted siRNA was added to diluted Lipofectamine® RNAiMAX reagent (1:1 ratio) to produce the siRNA–lipid complex. The complex was then incubated for 5 min at room temperature. A total of 250 µL siRNA–lipid complex per well was added to cells (the final siRNA concentration per well was equivalent to 25 pmol, and the final Lipofectamine® RNAiMAX volume per well was 5 µL). Transfected cells were incubated for 3–7 days at 37°C in a 5% CO_2 incubator. The medium was replaced with a cell culture medium after 3 days of transfection.

2.6.2.6 Cell lysate collection and protein quantification for transfected cells

Cell lysates were collected at day 3, day 5, and day 7 after addition of the transfection reagents to evaluate the effect of the transfection on cells. To collect the cell lysate, the plate of cells was placed on wet ice and the cells were washed with 1 ml PBS per well after the spent media was discarded. Another 1 ml of PBS was pipetted into each well and the cells were gently scraped off using a 1 ml syringe plunger. The cell suspension from each well was then transferred into a corresponding, pre-labelled 1.5 ml Eppendorf tube on ice. The cells were then centrifuged for 10 min at 4,000 rpm. The supernatant was removed from each tube without disturbing the pellet, and 50 µl of RIPA buffer (as prepared in section 2.2.3) was added to the pellet (on ice) for 30 min with repeated agitation. This was followed by centrifugation of the tubes at 13,000 rpm for 10 min in a cold centrifuge.

The supernatant was transferred into new 0.5 ml labelled tubes and stored at -20°C.

2.6.2.7 Analysing the transfection results

The protein concentration of the collected cell lysate was quantified, as described in Section 2.2.4. The proteins extracted from the transfected cells were analysed by WB to verify differences in the protein expression between the transfected and negative-control cells. Protein band intensity analysis of the generated WB images was performed, as described in Section 2.6.2.2. The optimal concentration of siRNA for both *HMGB3* and *CCNB2* that showed high transfection efficiency was 10 μM and was used to carry out the *in vitro* experiments.

2.6.3 CellTitre 96 Aqueous One Solution Cell Proliferation Assay (MTS assay)

2.6.3.1 Overview and principles

The MTS assay is a colorimetric technique for determining cell viability. MTS solution (Promega, G3580, UK) comprises a tetrazolium compound 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) and a reagent for electron coupling (phenazine ethosulfate [PES]). By the action of dehydrogenase enzymes, metabolically active (viable) cells degrade this compound to produce coloured formazan that is soluble in cell culture media. This reaction produces reducing equivalents, such as nicotinamide adenine dinucleotide phosphate hydrogen (NADPH) and nicotinamide adenine dinucleotide hydrogen (NADH). The amount of formazan produced is directly proportional to the number of viable cells present [185].

2.6.3.2 MTS assay procedure

The MTS assay was used to evaluate the impact of silencing *HMGB3* and *CCNB2* on cell proliferation. It is important to note that optimisation of transfected cells and negative control seeding density were performed before carrying out the MTS cell proliferation assay. The assay was

conducted based on the recommendations of the manufacturer. In 100 µl of the growth medium, approximately 500–5,000 cells were seeded into each well of the 96-well plate. The MTS assay was carried out on the third day of transfection, during which time the control and transfected cells were seeded in triplicate at a density of 3,000 cells/well, and were given time to adhere by incubation at a temperature of 37°C in a humidified incubator. A multichannel 20-µl pipet was used to place CellTiter 96® AQueous One Solution reagent in each of the wells at three time points post-transfection (T24 h, T48 h, and T72 h). The plate was then incubated in the dark at 37°C and for 1 h in a humidified incubator containing 5% CO_2.

The MTS reagent was added to the wells, as per the manufacturer's instructions, and the absorbance at 490 nm was recorded at 24-, 48-, and 72-hour time intervals. The optical density (OD) was measured using a microplate reader (TECAN Infinite F50). The OD of this absorbance was evaluated by comparing the results of the transfected cells to the negative control. The assay was performed three times, and each was done in triplicate. Graphs of the results and statistical analyses were produced using GraphPad Prism software version 9.2.0.

2.6.4 Clonogenic survival assay

2.6.4.1 Overview

Clonogenic survival assays evaluate the survival of cells after they have been exposed to transfection agents or other cytotoxic agents. The assay measures viable cells that have undergone several (minimum 5–6) replication rounds to produce a colony, which indicates their ability to survive siRNA exposure [186].

2.6.4.2 Clonogenic assay procedure

Prior to performing the clonogenic assay, the transfected cells and negative control seeding density were optimised at different intensities to achieve ≥50 colonies/cell. On the third day of transfection, the negative-control cells and transfected cell lines were placed in a 6-well plate at low density.

The tissue culture plates that had been seeded with 800–1,000 cells/well were left untouched in culture medium in the incubator for 12–14 days. Following incubation, the media was removed and the wells were fixed for 30 min using 75% (v/v) methanol and 25% (v/v) acetic acid. Crystal violet dye was used to stain the plates for 60 min at room temperature. The stain was discarded, the plate was washed in running water to remove the excess stain, and the cells were allowed to air-dry overnight. Blue-stained colonies were visible, and approximately 50 cells/colony were counted using the cell counter. To avoid double counting, the grid was overlaid before counting. The assay was repeated three times, each in triplicate. The survival fraction was then calculated as follows:

Plating efficiency (PE) = number of colonies formed/number of cells seeded

Surviving fraction (SF) = PE of transfected cells/PE of control cells

Graphs were produced from the results, and the analysis was conducted using GraphPad Prism version 9.2.0.

2.6.5 Wound-healing assay

2.6.5.1 Overview

Cell migration is expected to occur in various physiological processes. For example, controlled cell migration allows for wound healing, tissue injury, and embryonic development. Cell migration is also a hallmark of cancer metastasis and has thus been studied in various fields. The wound-healing assay allows for the study of cell migration *in vitro*. Through the procedure, an artificial gap is created on a single cell layer; cells that appear on the edges of this layer would have migrated in attempt to establish new cell–cell contact. In this study, scratch wound-migration assays were performed to investigate the impact of the knockdown of the studied genes on tumour cells and how the knockdown affects cell migration. Control cells were used for comparison.

2.6.5.2 Migration assay procedure

To determine the migratory ability of the cells, a Culture-Insert 2 Well (Thistle Scientific Ltd, IB-81176) with built-in gap was used. The insert ensured that the gap size was defined and thus improved reproducibility. To optimise this assay, cell seeding and culturing occurred at various densities on the third day of transfection until a confluent monolayer was achieved. Overall, 70 µl of cell suspension containing 7.5×10^5 cells was seeded in each well of the Culture-Insert 2 Well. The cells were then placed in an incubator with 5% CO_2 at 37°C for 24 h until a confluent monolayer was formed. The density of the cells was evaluated after 24 h using a light microscope. If a monolayer was not achieved in the first 24 h, then the cells were further incubated for up to 12–24 h. When confluence was achieved, the Culture-Insert 2 Wells were gently removed using a sterile tweezer. Then, the cell layer was washed in a cell-free medium to remove cellular debris and other unattached cells. To reduce the risk of cell proliferation, which might confound the study of migration, serum starvation (low serum concentration in the cell medium) was used, as it is the most common method for suppressing cell proliferation [187]. Therefore, the µ-Dish contained about 2 ml of 1% serum media. Figure 2.1 illustrates the workflow of the experiment. The wounds were observed by taking multiple images at 10× microscopic magnification via light microscopy (Leica Microsystems, Lecia DMI 3000B, Germany) at the following time points: T0, T24, T48, and T72 h. Image J software (version 1.52) was used to measure the wound area and calculate the percentage of wound closure. The findings were graphed and analysed using GraphPad Prism 9.2.0. Three independent experiments were conducted for this assay.

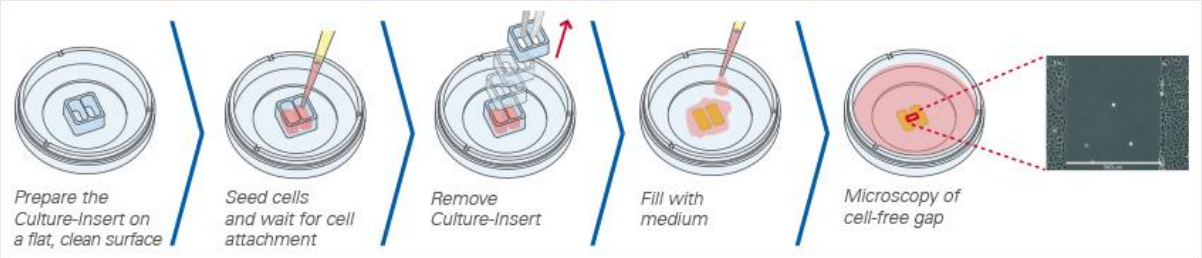


Figure 2.1 Experimental workflow for a wound-healing assay using the ibidi Culture-Insert 2 Well. Figure obtained from [188], courtesy of ibidi GmbH.

*** Permission to reuse this figure was obtained (Appendix B).**

2.6.6 Cell cycle assay

2.6.6.1 Overview and principles

Flow cytometry is a commonly used technique for determining cell cycle distribution. The cell cycle assay used in this study was based on measuring the cell cycle distribution via quantification of DNA content, utilising propidium iodide (PI) as a nucleic acid staining dye followed by flow cytometry analysis. PI dye binds in proportion to the cell's DNA content. This technique identifies the distribution of cells in three distinct phases of the cell cycle: G1, S, and G2/M. Given that PI binds to both RNA and DNA, the kit used in this study includes RNaseA (ribonuclease A) to degrade cellular RNA and therefore reduces background RNA staining. Due to the membrane impermeability of PI, ethanol was employed to fix and permeabilise the cells.

2.6.6.2 Cell cycle assay procedure

The manufacturer's instructions and recommendations were followed to conduct this assay. Briefly, 100,000 cells were harvested on the fifth day after transfection to identify and analyse the cell cycle distribution using a PI flow cytometry kit (Abcam, ab139418, UK). To produce a cell suspension, trypsinisation of the harvested cells was performed, as described in section 2.2.3. However, the culture media and PBS were not discarded form the suspension as they contained loosely adherent cells. The cell suspension containing culture medium, PBS, and trypsinised cells

was then centrifuged. The cell pellet was then resuspended in 1× PBS and centrifuged again. The cells were then fixed with 66% ethanol on ice and stored at 4°C for at least 2 h to allow for fixation. Prior to flow cytometry analysis, the fixed cells were allowed to reach room temperature before gentle resuspension and centrifugation. After centrifugation, the cells were washed with 1x PBS and centrifuged again to pellet them. The cell pellet was resuspended in 1x PI and RNase staining solution and incubated in the dark at 37°C for 30 min. Samples were analysed using a MACSQuant® flow cytometer analyser (Miltenyi Biotec, Bergisch Gldbach, Germany). At least 50,000 cells were analysed from each cell suspension. This assay was performed three times. The data were analysed using FlowJo software (version 14.0.0.0., USA). GraphPad Prism software version 9.2.0 was used to create graphs and perform statistical analysis.

2.6.7 Apoptosis assay

2.6.7.1 Overview and principles

Apoptosis (i.e., programmed cell death) is a common type of cell death that also occurs during embryonic development in eukaryotes. It is characterised by chromatin condensation, cell volume changes, and the translocation of membrane phosphatidylserine (PS) from the inside to the outside of the plasma membrane. Annexin V is a phospholipid-binding protein that is $Ca2^{+}$-dependent and has a high affinity for PS. For the identification of early and late apoptotic cells, the conjunction of Annexin V with a second dye, such as PI, is used for staining. Dead cells with damaged membranes are positive (permeable) to PI, whereas viable cells with intact membranes are negative to PI. Therefore, early apoptotic cells are Annexin V-positive and PI-negative, late apoptotic cells are Annexin V- and PI-positive, viable cells are both Annexin V- and PI-negative, and necrotic cells are Annexin V-negative and PI-positive (Figure 2.2) [189, 190].

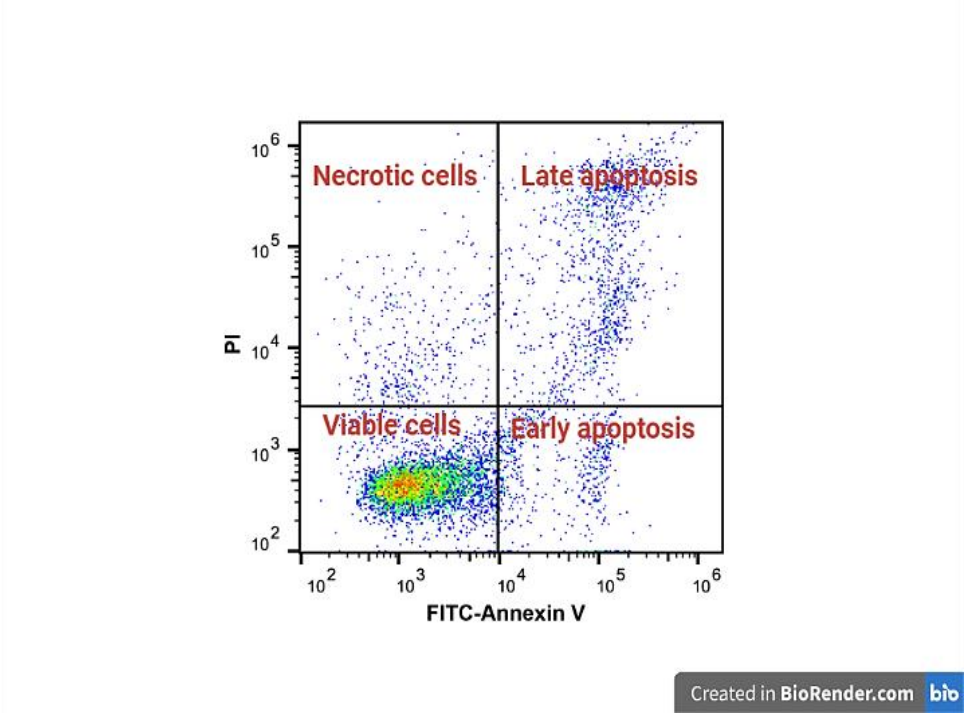


Figure 2.2 Illustration of staining with Annexin V and propidium iodide (PI) by flow cytometry. Illustration created using BioRender.

2.6.7.2 Apoptosis assay procedure

The detection of apoptosis was performed using an Annexin V detection kit (Abcam, ab14082, UK) based on the manufacturer's recommendations. On the fifth day after transfection, 100,000 cells were harvested by trypsinisation and centrifugation. Resuspension of the cell pellet was performed in 1× PBS and centrifuged. The pellet was then resuspended with 1× binding buffer. Annexin V and PI were then added to the cell solution on ice and incubated in the dark for 5 min at room temperature. Three independent assays were performed. The samples were analysed using a MACSQuant® flow cytometer analyser (Miltenyi Biotec, Bergisch Gldbach, Germany), and the data were analysed using FlowJo software (version 14.0.0.0., USA). Graphs were produced and statistical analyses were performed using GraphPad Prism software version 9.2.0.

2.6.8 *In vitro* models of tumour–endothelium interactions

2.6.8.1 Overview

Endothelial cell monolayers have been used *in vitro* as a model system to study tumour–endothelium interactions to mimic *in vivo* interactions [191]. In this study, HUVECs and DLECs were used to investigate BC tumour cell adhesion to the endothelium and the transmigration of tumour cells across vascular and/or lymphatic endothelial cells. This *in vitro* model is useful in understanding the role of *HMGB3* and *CCNB2* in tumour cell adhesion and the transmigration of cancer cells across the endothelium, which ultimately results in the initiation of the LVI process.

2.6.8.2 Adhesion assay

2.6.8.2.1 Adding a coating agent

The adhesion assay was carried out using 24-well plate. Before adding endothelial cells (HUVECs and DLECs) to the 24-well plate, the culture surface was coated with matrix elements to enhance cell attachment and adhesion. The coating agent used in this assay was gelatine (0.2%) (Sigma, G1393, UK). The wells were coated with 0.2% gelatine dissolved in PBS and incubated at 37^{o}C for 30 min. After 30 min, the gelatine solution was discarded, and the well was washed twice with PBS and allowed to dry before adding the endothelial cells.

2.6.8.2.2 Seeding endothelial cells

After coating the 24-well plate with gelatine, the endothelial cells were cultured until confluent. Seeding density was optimised at 1.0×10^5 cells/ml, 1.5×10^5 cells/ml, 2.0×10^5 cells/ml, and 2.5×10^5 cell/ml for HUVECs and DLECs cells to achieve the requisite confluent cell monolayer. For the experiments, both HUVECs and DLECs were seeded at a density of 2.5×10^5 cells/ml and incubated for 24 h.

2.6.8.2.3 Labelling and seeding BC tumour cells

To differentiate tumour cells from endothelial cells, tumour cells were fluorescently labelled prior to adding them to endothelial cells. Cell Tracker Green CMFDA (Invitrogen, C2925, USA) was used for labelling. Optimisation was carried out using different concentrations (0.5–2.0 μM) of Cell Tracker Green CMFDA based on the manufacturer's recommendation. Tumour cells were trypsinised to prepare single-cell suspensions of 1×10^6 cells/ml. The suspensions were then centrifuged, and the supernatant was discarded. To label the tumour cells, Cell Tracker Green CMFDA at a concentration of 1 μM was diluted in serum-free medium at a concentration $1x10^6$ cells/ml at 37°C for 30 min. After 30 min, labelled tumour cells were centrifuged at 170 g for 10 min, and the supernatant was discarded. Cells were then resuspended in 1 ml of media containing serum. Endothelial cell media was then removed and 0.5×10^5 tumour cells per well were cultured on the endothelial cell monolayers and incubated for 35 min at 37°C. The optimal time for tumour cell adhesion (for cell lines MCF-7, ZR-75-1, and SK-BR-3) was also optimised between 10 and 60 min; at the completion of the assay, the number of adherent cells was compared. For all endothelial cells, the number of adherent BC cells lines for MCF-7, ZR-75-1, and SK-BR-3 increased significantly to 35 min before plateauing. Therefore, 35 min was chosen as the optimal time for the cell adhesion experiments. Non-adherent cells were washed with tumour cell media, and adherent tumour cells were counted at a magnification of 10× using a fluorescent microscope (Lecia DMI 3000B, Leica Microsystems, Germany). Five fields of view were counted in the well's core area which was manually determined using the grid lines on the bottom of the plate. Three independent experiments were conducted, all in triplicate. The results were expressed as the absolute number of cells adhering to the endothelial layer and as the percentage of cells adhering to the endothelial layer, compared to the control. GraphPad Prism software version 9.2.0 was used for graph creation and statistical analysis.

2.6.8.3 Transmigration assay

2.6.8.3.1 Adding a coating agent

Millicell® hanging cell culture insert (MERCK, MCEP24H48, UK) with polyethylene terephthalate membrane filters of 8-µm pore size that were used to divide a 24-well tissue culture plate into two compartments were used for the transmigration experiment. Hanging cell culture inserts are made of polyethylene terephthalate membrane, making it difficult to culture endothelial cells to obtain a monolayer of cells. Therefore, to enhance the attachment of cells to the hanging cell culture insert, gelatine (0.2%) was added to the insert, as described in Section 2.8.8.2.1. The only difference in the procedure between the adhesion and transmigration assays was that, in the adhesion assay, the gelatine was added directly to the 24-well plate, while in the transmigration assay, the gelatine was added to the hanging cell culture insert.

2.6.8.3.2 Seeding endothelial cells

Prior to conducting the transmigration assays, the total number of endothelial cells (HUVECs and DLECs) added to each insert was optimised at different seeding densities (7.5×10^4, 1×10^5, 2×10^5, and 2.5×10^5 cells/ml) to create the intact, confluent cell monolayer required for this experiment. HUVEC and DLEC cells (2.5×10^5 cells/ml) were added to the coated insert with 600 µl of complete media in the lower compartment and incubated at 37°C for 24 h; the medium was then removed, and the cells were washed twice with PBS. Before conducting the transmigration assay, the intact cell monolayer was assessed by adding 100 µg/ml of Lucifer yellow (Invitrogen, L453, UK) to the inserts and the empty inserts as a control. PBS (600 µl) was added to the bottom of the control well. The plate was incubated at 37°C for 30 min. The fluorescence leakage of each well was assessed using a plate reader (Fluorostar Optima, BMG LabTech) at an excitation of 458 nm and emission of 520 nm with Fluorostar Optima software version 1.30 R3. For an intact cell monolayer, the Lucifer yellow leakage range is typically 0.3–2.0% of the control; a leakage value >2% of

the control value indicates that the cell monolayer was possibly compromised during the assay. The transmigration experiment proceeded if a monolayer of intact endothelial cells was observed.

2.6.8.3.3 Labelling and seeding BC tumour cells

The labelling process of the tumour cells (MCF-7, ZR-75-1, and SK-BR-3) was similar to the adhesion assay described earlier in Section 2.8.8.2.3, with the only difference being that 1×10^5 tumour cells were seeded in each well on top of the endothelial cell monolayer and the lower compartment was supplemented with 600 µl of tumour cell medium (Figure 2.3). The cells were allowed to transmigrate for 16 h at 37°C (the time for tumour cell transmigration was optimised between 8 h and 24 h). After 16 h, the upper chamber, which contained the endothelial cells and the non-migrated tumour cells, was cleaned with a cotton swab. Tumour cells that had transmigrated to the lower compartment were assessed by counting the cells on the bottom of the chamber using a fluorescence microscope (Lecia DMI 3000B, Leica Microsystems, Germany). Five fields were counted in each well at a magnification of 10×. Experiments were conducted three times, each in duplicate. The results were expressed in terms of both the absolute number of transmigrated cells and the percentage of transmigrated cells relative to the control. GraphPad Prism software version 9.2.0 was used for graph production and statistical analysis.

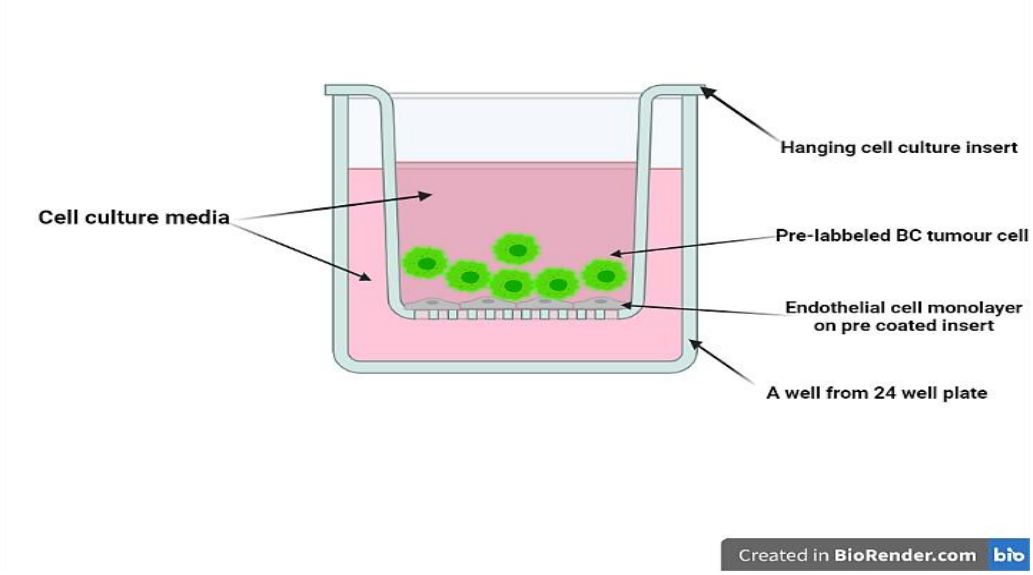


Figure 2.3 Illustration of the transwell migration model. This image was created using BioRender.

2.7 Statistical analysis

2.7.1 Introduction

Several statistical parameters were calculated in this study. The presence of nominal variables, including categorical and ordinal variables, continuous scale data, and independent or dependent variables required the use of multiple statistical tests, which were performed using the Statistical Package for the Social Sciences (SPSS) Version 24.0 (SPSS IBM Corp, Chicago, IL, USA). To plot graphs and analyse data collected in the *in vitro* experiments, GraphPad PRISM software Version 9.2.0 (GraphPad Prism Inc., San Diego, CA, USA) was used. In this study, the threshold of significance (p-value) was set at <0.05.

2.7.2 Descriptive statistics

The analysis of the H-scores of IHC TMA in this study was important for identifying the features of a specific dataset. For example, histograms were used to represent the calculated H-scores for the population, mean, median, and mode, with the acquisition of a normal distribution curve. The central tendency statistics of the datasets of these studies could best be represented with these three statistical tests (mean, median, and mode).

For continuous variables, cut-off points were obtained based on their normal distribution status. The mean was used to determine the cut-off values if the data were normally distributed; in cases where the samples were not normally distributed, the median was designated as a suitable cut-off point. Descriptive analysis involved the use of other variables, such as percentages and total number of cases.

2.7.3 Intra-class correlation coefficient (ICC)

Quality control and H-score assessment were performed using the ICC test, which was based on continuous H-score data for an immune-stained protein obtained by two individual researchers. This test sought to quantitatively measure the similarity between the two datasets and assign a degree of scoring reliability between the two scorers. The ICC agreement test was used to score at least 20% of all valid TMA cores and the stained IHC slides.

2.7.4 Univariate and correlation analyses

For normalised continuous variables, a one-way analysis of variance (ANOVA) with post-hoc Tukey multiple comparison test (for parametric data) or Kruskal–Wallis test (for non-parametric distribution) was used to investigate differences between three or more groups. Analysis of differences between the two groups was conducted using either Student's t-test (parametric data) or the Mann–Whitney U test (non-parametric data distribution). For non-normalised continuous variables, univariate analysis of data that constituted clinicopathological variables from specific biomarkers was conducted using the Mann–Whitney–Wilcoxon (MWW) test.

The strength of the linear correlation between the two normalised transcripts of continuous variables was established using the Pearson correlation coefficient (r). The Spearman rank test was used to evaluate patterns of correlation in non-normalised data. The value of r is in the range of +1 to -1, a value of 0 indicates that there was no correlation between the two variables; an r value closer to +1 indicates a strong correlation for IHC of the target protein between the two slides, whereas a negative r

value suggests an inverse association of co-expression. An *r* value between -1.0 and -0.5 denoted a strong correlation, while an *r* value between -0.5 and -0.3 or 0.3 and 0.5 denoted a moderate correlation, and the correlation was considered weak if the *r* value was between -0.3 and -0.1 or 0.1 and 0.3. The correlation of the IHC results of a target protein between two slides was deemed non-existent if the *r* value was between -0.1 to 0.1.

2.7.5 Survival analysis

The survival distribution of the two populations of BC patients was obtained using the log-rank test. Kaplan–Meier curves were used to represent the survival of BC patient. SPSS was used to obtain the survival curves of the BC patients, and patients whose survival status (dead due to the disease) was validated were denoted as '1' to indicate this event. The Cox regression model was used for multivariate analysis of confounders and the prognostic independency of biomarkers. Tumour grade, LN stage, tumour size, and LVI are the predictive factors used in this study. The endpoints of this study were 10-year BCSS or TTDM.

2.8 Ethical approval

This study was approved by the Nottingham Research Ethics Committee 2 under the title 'Development of a molecular genetic classification of breast cancer' and the North West – Greater Manchester Central Research Ethics Committee under the title 'Nottingham Health Science Biobank (NHSB)', reference number 15/NW/0685. The study was conducted in compliance with the current ethical and legal guidelines of the United Kingdom.

Chapter 3: Bioinformatic analysis for gene selection

3.1 Introduction

Patients with early-stage BC have experienced better outcomes in recent decades as a result of enhanced diagnostic accuracy, targeted drug therapies, and early cancer detection [192]. Despite the major improvement in outcomes, metastasis remains the leading cause of BC-related mortality, affecting approximately 20% of BC patients [193]. Various histopathological characteristics have been investigated as prognostic factors in BC, including tumour size, lymph node status and histology grade, all of which are strongly correlated with mortality [194, 195]. LVI is considered an early event in the development of metastasis and represents a significant prognostic factor [46]. While the molecular profiles involved in tumour differentiation, such as histological type, grade, and the development of lymph node metastasis, have been well studied [196, 197], the molecular mechanisms underlying LVI and its associated genes, which may serve as therapeutic targets or biomarkers, remain unknown. Therefore, targeting LVI is a promising approach for detecting tumour dissemination in early-stage BC.

METABRIC [198] and TCGA [170] are now the largest genomic and transcriptomic datasets of early-stage BC patients with clinical follow-up. In addition to employing these large transcriptome datasets in combination with thorough histological examination of LVI, this study used WAD and ANN bioinformatic analyses to identify the drivers of LVI. Based on the fold-change (FC) method, the target gene list was ranked for the WAD analysis and the gene list was generated for the ANN analysis based on the root-mean-square error (RMS) method. It was attempted to reduce the target gene list below 100 genes; however, validating such a large number of genes and determining which of them were key genes remained a challenge. Therefore, the key molecular determinants identified by the WAD and ANN approaches were further investigated at the protein level using IHC of a well-annotated series of early-stage BCs with long-term follow-up and *in vitro* mechanistic studies using a panel of BC cell lines.

3.2 WAD analysis for ranking the differential gene expression of transcriptomic cohorts

Identifying key genes across a variety of experimental conditions is a critical task for many microarray studies. Based on a previous study, a new FC-based method known as 'WAD' was used to rank differentially expressed genes (DEGs), and it outperformed other approaches in terms of accuracy [199]. Based on a previous study, WAD surpassed other FC approaches, such as average difference and rank products methods in terms of both sensitivity and specificity. Furthermore, WAD yielded consistent findings for various preprocessing algorithms [199]. Therefore, WAD was the first approach used in this study to identify key genes in both LVI-positive and LVI-negative cases [199, 200].

3.3 ANN analysis of the transcriptomic cohorts

Machine learning (ML) is an innovative field with various applications including the bioinformatical analysis of large datasets (e.g., proteomic and transcriptomic data). ML is based on the use of a set of algorithms to analyse such data to reach specific conclusions that address a given question [201, 202]. ANNs serve as the framework for a variety of ML algorithms that process large datasets [201]. ANNs are a type of artificial intelligence that can model complicated systems with high predictive accuracy using blind data [203]. ANNs can learn and model non-linear and complicated associations between input variables. Previously published research has demonstrated that ML can be used to successfully identify novel biomarkers linked to specific clinical outcomes in different cancers [204, 205], including BC [206]. Therefore, an ANN method was the second approach used in this study to overcome the challenges in analysing and profiling the transcriptomic data set for identification of key molecular determinants potentially controlling LVI.

3.4 Hypothesis

Key genes involved in the development of LVI can be discovered using WAD and ANN methodologies.

3.5 Aims

1. To identify the key molecular determinants driving LVI in early-stage BC using WAD- and ANN-based bioinformatic analysis of large transcriptomic data derived from BC cases with histologically and immunohistochemically determined LVI status.
2. To validate the association between the genes identified in the bioinformatic analysis and the presence of LVI at the mRNA level in both METABRIC and TCGA data.

3.6 Methods

3.6.1 Study cohorts

3.6.1.1 METABRIC cohort

In the METABRIC cohort, the Illumina TotalPrep RNA Amplification Kit and Illumina Human HT-12 v3 Expression BeadChips (Ambion, Warrington, UK) were used to extract mRNA from the primary tumours of female patients [198]. Using H&E-stained slides of 1,565 patients from the METABRIC cohort, the LVI status of patients was histologically assessed. LVI status was further examined using IHC with D2-40 for the Nottingham subset (n=288/1,565) [44]. The final LVI status was validated using a combination of several H&E tumour sections and IHC. The METABRIC cohort is detailed in Section 2.1.1.

3.6.1.2 TCGA cohort

The Genomic Data Commons Data Portal and cBioPortal [168, 169] websites were used to extract data from the TCGA cohort of female BC patients [170]. Briefly, the RNASeqV2 mRNA expression datasets were accessed together with de-identified clinical data for several clinicopathological variables and outcomes. The cBioPortal website provided

access to digital H&E-stained slides from the TCGA BC cohort, while an expert breast pathologist (Dr. Leslie Dalton) assessed LVI status. Details of the TCGA cohort are presented in Section 2.1.3.

3.6.1.3 WAD bioinformatic analysis for DEGs

WAD analysis in this study was performed by Dr. Sasagu Kurozumi, a clinical research fellow at the University of Nottingham. The METABRIC cohort was used as a discovery dataset, while the TCGA cohort was used as a validation cohort. Briefly, DEGs between LVI-positive and LVI-negative cases were identified using WAD methodology and subsequently identified based on WAD rankings [199, 200]. Overall, 41% and 35% of cases were LVI-positive in the METABRIC and TCGA cohorts, respectively. Moreover, 59% and 65% of cases were LVI-negative in the METABRIC and TCGA cohorts, respectively. Both the Nottingham cases in the METABRIC cohort and other METABRIC cases were analysed in the WAD test. The resulting genes from the METABRIC cohort were subsequently investigated using hierarchical clustering based on expression similarity. Hierarchical clustering was also performed on the TCGA cohort using the genes identified from the METABRIC cohort to confirm the findings [200]. Figure 3.1 represents WAD bioinformatic analysis for DEGs.

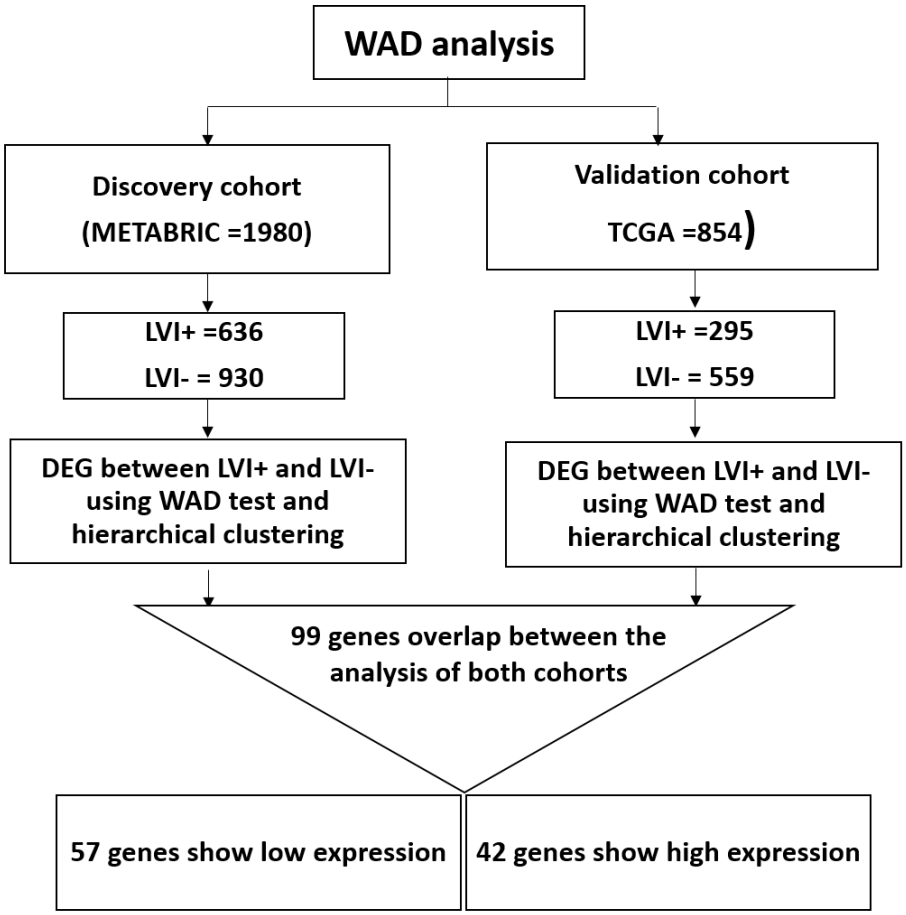


Figure 3.2 Schematic diagram represents WAD bioinformatic analysis for DEGs.

3.6.1.4 ML for biomarker set enrichment

In two large BC datasets (METABRIC and TCGA), ANN was used to determine the DEGs between LVI positivity and LVI negativity. These data were studied using bioinformatic analysis via ANN software in collaboration with Prof. Graham Ball from Nottingham Trent University (NTU). Regarding the genomic expression data obtained from both datasets, ANN-based neural data mining was performed to identify an enriched set of concordant biomarkers associated with LVI. In this case, an ML strategy was implemented based on an ANN and was integrated it with a concordance analysis performed across many splits of Monte Carlo data using the previously described methodology [206]; this method has shown to be effective in reducing over-fitting and erroneous discovery while improving the generalisability of the discovered biomarkers. Specifically, a sparse Multilayer perceptron architecture comprising a single input connecting to

a single hidden layer with two hidden nodes and initial weights, and bias's constrained to a Standard deviation of 0.1 was used. The utilisation of a sparse network was our first strategy used to prevent overfitting and reduce the risk of false discovery. The hidden nodes and the output node used a sigmoid activation function. The output classes used binary annotation and one hot encoding. The weights of the network were updated with a feed-forward – backpropagation algorithm. Early stopping based on a test data set utilising Monte-Carlo cross-validation (60% train: 20% test and 20% validation) as a second strategy to prevent overfitting and eliminate false discovery was implemented. Using this strategy, each transcript was considered singly in the above process, and a model developed for each one in the data set. The rank order of these inputs was then determined based on the RMS error of the predictions for the unseen validation set.

Based on the LVI status, the data were divided into five random groups and the ANN algorithm was run individually for each group, with each run consisting of 20 loops. In this approach, the samples were classified into three subsets: training, test, and validation. The samples were randomly reshuffled for the next training cycle. The training was run for a maximum of 3,000 epochs and stopped when there was no further improvement in the RMS value of the test subset at a threshold of 0.01. The results were sorted (for each loop) based on the lowest predictive error of the test subset. The subset with the lowest test error was placed at the top of the list. Then, integrative analysis was performed using Microsoft Excel to identify the top 100 predictors that were common between the 20 loops. Figure 3.2 represents the workflow of machine learning strategy grounded in an artificial neural network.

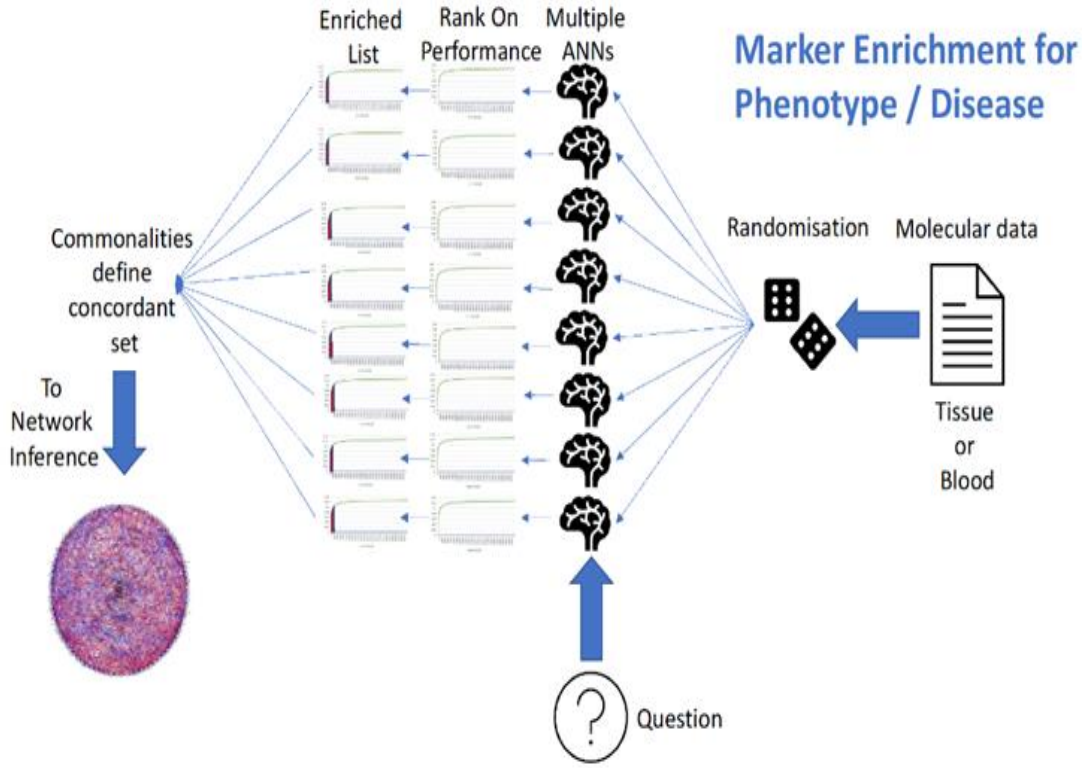


Figure 3.2 Workflow of machine learning strategy grounded in an artificial neural network. Based on the genomic expression data obtained from mRNA datasets, ANN-based neural data mining was used to identify an enriched set of concordant biomarkers associated with LVI.

3.6.1.5 Pathway analysis using an accessible online resource

Functional enrichment analysis is crucial for evaluating the results of high-throughput experiments, which usually yield a list of novel genes or proteins. WebGestalt (Web-based GEne SeT AnaLysis Toolkit) is one of the most extensively used tools for the enrichment analysis of gene sets; it assists users in deriving biological insights from their target genes [207]. WebGestalt (http://www.webgestalt.org/) is an accessible web-based resource that was used in this study to explore the common pathways of the identified genes via bioinformatic analyses. WebGestalt was chosen because it is continuously updated with expanding functionality, a new R package and web APIs to meet the demands of biologists working in a variety of research fields [207, 208]. WebGestalt also includes many

databases, such as KEGG, Panther, Reactome, and WikiPathways [207]. The pathway analysis was based on transcripts that were differentially expressed at the $p<0.05$ level using the KEGG pathway database.

3.7 Results

3.7.1 Genes associated with LVI from WAD analysis

Based on the WAD bioinformatic analysis, a group of 99 genes was discovered to have significantly different expression between LVI-positive and LVI-negative cases. Of those 99 genes, 42 were significantly upregulated and 57 were downregulated in the METABRIC cohort. The 99 genes identified in the METABRIC cohort were then subjected to hierarchical clustering to validate the results of the TCGA cohort; similar results were obtained in both cohorts [200]. Table 3.1 presents the 99 genes identified as being significantly associated with LVI status.

Table 3.1 List of 99 genes correlated with LVI status in the METABRIC cohort.

Genes with high expression		**Genes with low expression**		
APOC1	*KRT7*	*ACTG2*	*FCGBP*	*SERPINA3*
APOE	*KRT8*	*ANG*	*FGD3*	*SERPINE2*
CALML5	*LAPTM4B*	*ANXA1*	*FOS*	*SGCE*
CCNB2	*LRRC26*	*C1S*	*FST*	*SLC40A1*
CDCA5	*LY6E*	*CDC42EP4*	*GAS1*	*SLC44A1*
COX6C	*MMP11*	*CEBPD*	*GSTP1*	*SRPX*
DNAJA4	*MX1*	*CFB*	*HBA2*	*STC2*
EEF1A2	*NME1*	*CFD*	*HBB*	*SUSD3*
ELF3	*NOP56*	*CLIC6*	*HLA-DQA1*	*TNS3*
ERBB2	*PGAP3*	*CXCL12*	*IL17RB*	*TPM2*
GNAS	*PITX1*	*CXCL14*	*MAOA*	*TXNIP*
HMGA1	*PTTG1*	*CYBRD1*	*MFAP4*	*UBD*
HMGB3	*S100P*	*CYP4X1*	*MGP*	*VIM*
HSPB1	*SCD*	*DCN*	*MT1E*	*VTCN1*
IDH2	*SLC52A2*	*DKK3*	*NDP*	*ZBTB20*
IFI27	*SLC9A3R1*	*DPYSL2*	*NINJ1*	
ISG15	*SPDEF*	*DUSP1*	*PDGFRL*	
KRT18	*TM7SF2*	*EEF1B2*	*PLGRKT*	
KRT18P55	*UBE2C*	*FBLN1*	*PYCARD*	
KRT19	*UBE2S*	*FCER1A*	*RPL3*	
UCP2	*YWHAZ*	*S100A4*	*SELENOM*	

3.7.2 ANN analysis of transcriptomic cohorts to discover genes associated with LVI

The ANN findings were ranked according to their predicted error for the unseen cohort in the Monte Carlo cross-validation, with the lowest predictive error ranked higher. As a result, a final model containing the top 100 transcripts that most accurately categorised LVI status (positive or negative) was created. Table 3.2 lists the top 100 concordant genes associated with LVI status, which were identified by ANN analysis of the METABRIC and TCGA cohorts.

Table 3.2 Top 100 concordant genes from the METABRIC and TCGA cohorts associated with LVI, based on ANN analysis.

Top 100 concordant genes		
DSCC1	*AK124197*	*ATAD2*
CCNE2	*TCTN3*	*ATP2A2*
YWHAZ	*GPR34*	*ATP6V1C1*
CCNB2	*TNRC6A*	*AURKA*
LRCH3	*AGAP6*	*AZIN1*
BX116720	*AK129555*	*BBS5*
EIF4H	*LIN52*	*BDH2*
CCNB1	*ACIN1*	*BRI3BP*
AK022140	*CLASP1*	*C1orf91*
DPRXP4	*AL359560*	*C1RL*
AK130741	*PHKA2*	*CASC5*
AK130706	*ANAPC10*	*CCDC130*
AK025793	*BX093900*	*CDC25A*
BX101409	*DA572426*	*CDC6*
BX111162	*EBLN2*	*CDCA5*
AL133627	*PTBP1*	*CENPI*
AX747098	*ZNF493*	*CKAP5*
RRM2	*TYK2*	*CLK1*
MLL4	*BU624523*	*CLK4*
SPTLC1	*PUM1*	*CPNE3*
MLL	*DLC1*	*CSE1L*
ERP29	*CXorf56*	*DCAF13*
NKTR	*DNAH1*	*DERL1*
PML	*DR979451*	*DFNB59*
RBM33	*ARPC3*	*DTL*
ASXL2	*TRPV1*	*ECHDC2*
CD237904	*EPG5*	*ESRP1*
CRYGS	*AP1S1*	*FAM83D*
BROX	*C11orf73*	*FANK1*
SLC25A3	*C1R*	*FBXW4*
ALDOB	*SOLH*	*FES*
MAT2A	*AI655567*	*FLVCR1*
AK129699	*ADC*	
SRRT	*ANKRD10*	

3.8 Validation of identified gene signatures from the bioinformatic analysis

3.8.1 Validation of identified gene signatures from the WAD bioinformatic analysis

For the 99 genes identified from the WAD analysis, further validation of the 42 upregulated genes was conducted to confirm their association with LVI status in both the METABRIC and TCGA cohorts. In this study, downregulated genes were excluded because they cannot be targeted like upregulated genes can in therapy, due to their lack of oncogenic effect. Validation of the WAD analysis for the 42 upregulated genes was performed by analysing the association of each individual gene with LVI status. Of the 42 upregulated genes, 32 genes were significantly overexpressed with the presence of LVI in both the METABRIC and TCGA cohorts ($p<0.05$). Of the remaining upregulated genes, 8 were significantly associated with the presence of LVI in the METABRIC cohort ($p<0.05$), but their association was weak in the TCGA cohort ($p\geq0.05$). In both cohorts, two genes were overexpressed in LVI-negative cases ($p<0.05$). Table 3.3 presents the association between LVI status and the 42 upregulated genes.

Table 3.3 Validation of the association between 42 upregulated genes identified via WAD analysis and the presence of LVI.

Top 42 highly expressed genes in the METABRIC and TCGA cohorts		
Presence of LVI		**Absence of LVI**
APOC1	***KRT8***	*ERBB2*
APOE	***LAPTM4B***	*KRT7*
CALML5	*LRRC26*	
CCNB2	***LY6E***	
CDCA5	***MMP11***	
COX6C	***MX1***	
DNAJA4	***NME1***	
EEF1A2	***NOP56***	
ELF3	*PGAP3*	
GNAS	***PITX1***	
HMGA1	***PTTG1***	
HMGB3	***S100P***	
HSPB1	***SCD***	
IDH2	***SLC52A2***	
IFI27	***SLC9A3R1***	
ISG15	***SPDEF***	
KRT18	***TM7SF2***	
KRT18P55	***UBE2C***	
KRT19	***UBE2S***	
UCP2	***YWHAZ***	

Genes in bold are significantly associated with the presence of LVI in both the METABRIC and TCGA cohorts ($p<0.05$)

3.8.2 Validation of gene signatures identified from ANN bioinformatic analysis

The top 100 genes generated by ANN analysis were further analysed to validate their association with LVI status in both the METABRIC and TCGA cohorts.

In both cohorts, 22 genes were significantly associated with the presence of LVI ($p<0.05$), while 11 genes were associated with the absence of LVI ($p<0.05$). However, 8 of the top 100 genes were associated with the presence of LVI in only the TCGA cohort ($p<0.05$). In addition, 2 genes in the METABRIC cohort and 17 genes in the TCGA cohort exhibited a significant association with the absence of LVI ($p<0.05$).

The remaining genes showed a weak association with the presence of LVI in both cohorts; 22 genes were weakly associated with the presence of LVI and 18 genes were weakly associated with the absence of LVI ($p\geq0.05$).

Table 3.4 summarises the associations of 100 genes identified by ANN analysis with LVI status.

Three genes overlapped between the WAD and the ANN analyses: *CCNB2*, *CDCA5*, and *YWHAZ*. Figure 3.3 presents the overlap of DEGs from the WAD and ANN analyses. The overlapping genes were all significantly associated with the presence of LVI in both transcriptomic cohorts ($p<0.05$).

Table 3.4 Validation of the association between the 100 genes identified via ANN analysis and LVI status.

METABRIC and TCGA cohorts		**METABRIC and TCGA cohorts**		**METABRIC cohort**	**TCGA cohort**	
Genes significantly associated with the presence of LVI ($p<0.05$)	Genes significantly associated with the absence of LVI ($p<0.05$)	Genes weakly associated with the presence of LVI ($p\geq0.05$)	Genes weakly associated with the absence of LVI ($p\geq0.05$)	Genes significantly associated with the absence of LVI ($p<0.05$)	Genes significantly associated with the presence of LVI ($p<0.05$)	Genes significantly associated with the absence of LVI ($p<0.05$)
DSCC1	*TYK2*	*LRCH3*	*BX116720*	*EIF4H*	*SLC25A3*	*TNRC6A*
CCNE2	*CCDC130*	*DPRXP4*	*AK130706*	*SPTLC1*	*ANAPC10*	*AGAP6*
YWHAZ	*ECHDC2*	*AK130741*	*ERP29*		*ARPC3*	*ACIN1*
CCNB2	*FES*	*AK025793*	*CD237904*		*ATP2A2*	*PHKA2*
CCNB1	*MLL4*	*BX101409*	*BROX*		*AZIN1*	*ZNF493*
AK022140	*BDH2*	*BX111162*	*AK129699*		*CKAP5*	*DNAH1*
RRM2	*C1RL*	*AL133627*	*SRRT*		*FAM83D*	*TRPV1*
ATAD2	*NKTR*	*AX747098*	*AK124197*		*FLVCR1*	*SOLH*
ATP6V1C1	*RBM33*	*MLL*	*TCTN3*			*ADC*
AURKA	*CRYGS*	*PML*	*GPR34*			*ANKRD10*
BRI3BP	*C1R*	*ASXL2*	*LIN52*			*BBS5*
CASC5		*ALDOB*	*EBLN2*			*C1orf91*
CDC25A		*MAT2A*	*BU624523*			*CLK1*
CDC6		*AK129555*	*PUM1*			*CLK4*
CDCA5		*CLASP1*	*DLC1*			*DFNB59*
CENPI		*AL359560*	*CXorf56*			*FANK1*

CPNE3		*BX093900*	*DR979451*			*FBXW4*
CSE1L		*DA572426*	*C11orf73*			
DCAF13		*PTBP1*				
DERL1		*EPG5*				
DTL		*AP1S1*				
ESRP1		*AI655567*				

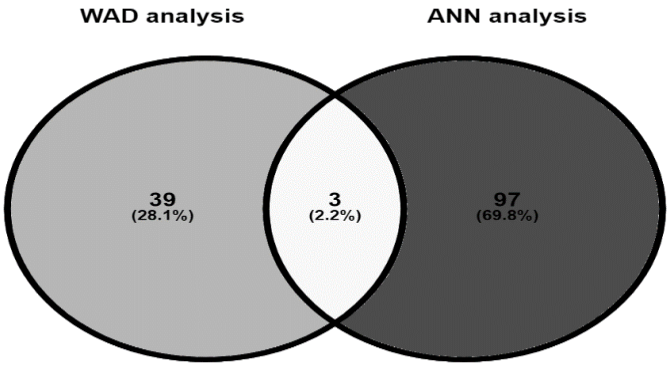


Figure 3.3 Overlapping differentially expressed genes (DEGs) from the weighted average difference (WAD) and artificial neural network (ANN) analysis. The numbers in each circle represent the total number of genes within each analysis, while the overlap represents the genes that were expressed in both analyses (n=3).

3.9 Pathway analysis results

Based on the pathway analysis, the cell cycle, oocyte meiosis, progesterone mediated oocyte maturation and p53 signaling pathway were common pathways of the identified genes in both bioinformatic analyses. Table 3.5 presents the pathway analysis for the 42 upregulated genes and the top 100 genes.

Table 3.5 Pathway analysis results from WebGestalt, used to identify differentially regulated pathways in 42 upregulated genes and the top 100 concordant genes (adapted from http://www.webgestalt.org/).

Gene set	Description	Size	Expect	Ratio	P value	FDR
hsa04110	Cell cycle	124	0.49806	14.055	4.3488e-7	0.00014177
hsa04114	Oocyte meiosis	124	0.49806	12.047	0.0000079466	0.0012953
hsa04914	Progesterone mediated oocyte maturation	99	0.39764	12.574	0.000040500	0.0044010
hsa04115	P53 signaling pathway	72	0.28920	13.831	0.00018001	0.014670

3.10 Gene selection criteria for protein expression studies

The genes in this study were carefully selected. This study aimed to discover a set of novel genes whose expression might be essential in the function of candidate genes that represent potential key drivers of LVI. These distinct candidate genes were chosen based on WAD and/or ANN analysis of the METABRIC and TCGA cohorts using the following criteria: only genes that were highly expressed and significantly associated with LVI positivity in both transcriptomic cohorts were chosen, and genes that exhibited low expression or were associated with LVI negativity were excluded. Genes that exhibited high expression with LVI-positive cases in one cohort were also excluded. Association of these genes with other poor prognostic characteristics of BC, such as advanced stage, high grade, HER2 receptor positivity, and hormone receptor negativity were also considered in the selection. These genes were prioritised based on review of the literature. The rationale for selecting these biomarkers from the myriad of literature while excluding others was based on their role in tumour progression or involvement in the carcinogenic cascade, such as in adhesion, migration, or invasion. In addition, the selection of these genes was based on their novelty, as their role in regulating LVI in invasive BC had not yet been studied or published. The selected genes were validated by performing statistical analyses for each gene at the mRNA level in both transcriptomic cohorts. Based on these criteria, *MX1*, *IDH2*, and *HMGB3* were selected from the WAD analysis, while *CCNB1* and *RRM2* were selected

from the ANN analysis. In addition, *CCNB2* was selected because it overlapped both the WAD and ANN analyses. Although *CDCA5* and *YWHAZ* also overlapped both analyses, they were excluded from this study as they had been studied by other groups conducting research on LVI regulatory genes. Further details on the selected genes are presented in subsequent chapters.

3.11 Discussion

The molecular determinants of LVI remains unknown and key genes and proteins driving LVI have yet to be discovered. The availability of large, robust, and well-characterised cohorts of invasive BC patients, such as the METABRIC, TCGA, and NPS cohorts, encourage the investigation of the unknown molecular mechanisms and genes underlying LVI in invasive BC patients. The primary aim of this study was to identify the key transcripts associated with LVI in invasive BC patients. Various strategies were applied to search for the key molecular determinants of LVI, including a combination of robust cohorts, bioinformatic analysis, and gene selection criteria.

In previous studies, gene sets associated with VI were identified in other cancer types, such as hepatocellular carcinoma [153] and endometrial cancer [151]. Mannelqvist et al. discovered an 18-gene signature associated with vascular involvement in endometrial cancer using microarray and qPCR studies [152]. In the current study, WAD analysis of the METABRIC dataset was used to identify a 99-gene set associated with LVI status in invasive BC; moreover, the findings were confirmed using the TCGA dataset. WAD employs the average difference and relative average signal intensity to ensure that highly expressed genes are ranked highly on average among different conditions [199]. From the 99-gene set generated in this study by WAD analysis, only distinct isoforms of *MMP* and serpin family E member (*SERPINE*) were common to the 18-gene set discovered by Mannelqvist et al. [152]. There is thus a need to develop novel approaches for identifying LVI drivers. Among the ML-based methods,

ANNs have proven to be valuable for this purpose. ANN is a collection of algorithms that attempt to mimic the human brain and identify associations between sets of data. ANN contains several layers, including an input layer, a hidden layer, and an output layer, and each layer performs a specific function [201, 202]. In this study, an ANN methodology was also used to overcome the challenges faced when analysing large genomic and transcriptomic datasets (i.e., METABRIC and TCGA). A target gene list containing the top 100 genes was generated using this approach.

Of those 100 genes generated by the ANN analysis, only three overlapped with the WAD analysis: *CCNB2*, *CDCA5*, and *YWHAZ*. Additionally, some of the 100 genes were found to be significantly associated with the presence of LVI in only one cohort, while others were found in both cohorts. Ultimately, different genes were identified using disparate (statistical) approaches, and multiple factors can contribute to these differences. For example, regarding the diverse methods used to quantify and statistically analyse gene expression and evaluate LVI status, the comparison of METABRIC and TCGA cohorts was a limiting factor in this investigation. A method was previously devised for accurately detecting LVI using D2-40 immunostaining [44]. In the Nottingham cases, rigorous criteria based on morphology and IHC were applied to determine LVI status; in the TCGA cohort, however, H&E-stained slides from the cBioPortal database were used to assess LVI status. Although H&E-stained slides can be used to evaluate LVI, it can be challenging to clearly identify LVI-negative cases [209]. Despite the use of different LVI evaluation methods, the LVI positivity rates in the Nottingham cases, remaining METABRIC cases, and TCGA cases were very similar in this study. Although the findings suggest that LVI evaluation using a single H&E-stained slide is adequate, larger cohorts must be studied to assess LVI status utilising both H&E-stained and IHC slides to accurately report LVI status. Another factor that may have contributed to these variations was the use of microarrays to assess mRNA expression in the METABRIC study, while RNA-seq with next-generation

sequence (NGS) was used in the TCGA analysis. Microarray systems have been utilised and validated for nearly two decades, and this approach is frequently employed for analysing multi-gene expression. In contrast, the unbiased, genome-wide RNA-seq approach allows for the analysis of all identified transcripts in addition to the discovery of novel transcripts, splice junctions, and noncoding RNAs. These technological and methodological variations may contribute to the well-documented difficulties associated with correlating microarray and RNA sequencing results across research [210, 211]. For example, the different techniques may have varying detection thresholds or may cover only specific genetic areas. As a result, it is difficult to assume that the methods are interchangeable as this would necessitate extensive cross-assay comparison [212]. While statistical agreement exists between multiple cohorts in this study, additional analysis using equivalent technologies (e.g., microarray and/or NGS tests) may provide more definitive validation of the LVI gene signature. Increasing the number of target genes to over 100 could also increase the likelihood of overlap between different bioinformatical analyses.

Although different gene sets were generated by the WAD and ANN analyses, the efficacy of these two approaches in identifying genes has been shown in previous studies [199, 203, 206]. Based on the criteria mentioned in Section 3.10, MX1, IDH2, and HMGB3 were selected from the WAD analysis, while CCNB1 and RRM2 were selected from the ANN analysis; CCNB2 was also selected because it overlapped both the WAD and ANN analyses. Although the roles of these genes might be understood in certain types of cancer, their roles and prognostic significance in LVI remains unknown.

3.12 Conclusion

Key genes associated with the development of LVI in invasive BC patients can be identified by utilising WAD and ANN analyses. These key drivers may clarify the complicated biological pathways regulating LVI. The results of

this study indicate the need for further investigation of these drivers. Several of these genes are discussed in subsequent chapters.

Chapter 4: The prognostic significance of MX1 and IDH2 in invasive breast cancer

This chapter has been published in Breast Cancer Research and Treatment (2019, 2020) under the title 'Myxovirus resistance 1 (*MX1*) is an independent predictor of poor outcome in invasive breast cancer' and 'The prognostic significance of wild-type isocitrate dehydrogenase 2 (*IDH2*) in breast cancer' with the corresponding DOI: 10.1007/s10549-020-05646-x and 10.1007/s10549-019-05459-7. The copyright approval for reuse of these materials is provided in (Appendix C/D).

4.1 Introduction

Invasive BC is one of the deadliest diseases threatening the lives of women globally. While early detection of BC may aid in providing more effective clinical interventions, there are still some BC-associated factors that may affect the therapeutic responses of BC patients. Decades ago, invasive BC was assessed morphologically, without the ability to analyse the key molecular drivers of unfavourable impacts on invasive BC patients' survival. However, by utilising modern genomic, transcriptomic, and proteomic technologies, it is possible to unravel key molecular drivers of invasive BC [9, 213].

In invasive BC, the presence of LVI is a clear indicator of whether invasive BC cells will spread to other parts of the body. LVI is associated with aggressive tumour behaviour, shorter patient survival, and the development of distant metastases [46]. Identification of LVI driver genes or proteins with independent and prognostic features could decrease BC deaths by halting cancer spread to other parts of the body through identification of the biological processes involved in early cancer VI and developing strategies to block these processes. As discussed in Chapter 3, the use of bioinformatics analysis methodologies (including WAD and ANN) makes it possible to identify key genes that may dictate LVI mechanisms. This study identified several potential prognostic and therapeutic candidates for LVI. *MX1* and *IDH2* were the candidate genes selected by WAD analysis for further investigation at the protein level using IHC in a well-annotated cohort of early-stage BC with long-term follow-up. *MX1* and *IDH2* were selected (based on the criteria mentioned in Chapter 3) because they were among the top identified genes, their overexpression was

associated with the presence of LVI in both the METABRIC and TCGA cohorts, and there were no published studies assessing their ability to drive LVI in invasive BC.

4.1.1 MX1

The MX1 protein, also known as MXA, is an interferon (IFN)-induced GTPase that belongs to the dynamin superfamily of large GTPases. MX1 has many features, including the tendency to self-assemble, low affinity for guanosine triphosphate (GTP), and a high intrinsic rate of GTP hydrolysis. The *MX1* gene encodes the MX1 GTPase protein, which is mainly localised in the cytoplasm [214] and exhibits a granular staining pattern that may be associated with cytoskeletal structures [215]. MX1 releases GTP, which plays a role in metabolising the protein that contributes to the antiviral mechanism of the cell [216].

MX1 is overexpressed and appears to play a role in a number of cancers, but its effect is still unclear. In most tissues, *MX1* is highly induced by the interferon-stimulated gene (ISG), particularly IFN-α and -β. Many ISGs can inhibit the motility of transformed cells and the invasion of normal tissues. *MX1* plays a role in the inhibition of motility and invasiveness in some cancers, such as in prostate carcinoma and melanoma [217, 218]; however, *MX1* seems to have different effects in other cancers. For example, it was observed that a low level of *MX1* could inhibit apoptosis during cancer development [218]. A study that measured the abundance of over 9,000 proteins using high-resolution mass spectrometry found that *MX1* was overexpressed in lymph node metastasis in colorectal cancer and played a significant role compared to non-metastatic colorectal cancers. The study also demonstrated that, in colorectal carcinoma, knockdown of *MX1* in colon cancer cells inhibited tumour cell migration, invasion, and metastasis [219]. These findings suggest that *MX1* may potentially be involved in LVI.

MX1 is overexpressed in certain IHC subtypes of BC, such as HER2+ tumours, and is correlated with the immune response and tumour-infiltrating lymphocytes (TILs) [220]. *MX1* responds to type I IFN and acts as a mediated signalling pathway [221]. Decreased *MX1* expression leads to lymphocyte imperfection in early and advanced stages of BC, which is a consequence of IFN-α signalling in T and B cells [222]. According to Han, in the advanced stage of BC, the blood is concentrated with highly tumour-associated plasmacytoid dendritic cells that undermine the IFN-α-producing capacity, resulting in a tumour-favouring environment [222]. Type I IFNs influence tumour reduction and development by acting on tumour, immune, and even endothelial cells [223]. By reacting to the tumour cells, type I IFNs may lead to tumour regression through cell surface expression of major histocompatibility complex 1 (MHC-1) and tumour-associated antigen (TAAs) by instigating both apoptosis and superior immunogenicity. In contrast, type I IFNs promote tumour progression by inducing resistance, EMT, tumour cell stemness, apoptosis, and the inability to control immune-inhibitory signals. However, by reacting to the lymphatic and vascular systems, type I IFNs hinder angiogenesis through *VEGF* downregulation [221]. These findings indicate that *MX1* could be a key driver of LVI, potential prognostic factor, and possible therapeutic target in invasive BC.

4.1.2 IDH2

IDH2 is a key metabolic enzyme and a member of the isocitrate dehydrogenase family. It is involved in cellular metabolism and serves as an $NADP^+$-consuming enzyme in the tricarboxylic acid (TCA) cycle, producing NADPH. In the reverse TCA cycle, where IDH2 causes reductive carboxylation of 2-oxoglutarate, it consumes NADPH during hypoxia to survive lower glucose levels [224].

The TCA cycle generates cellular energy and metabolic intermediates that are upregulated in metastasised cancer cells. Additionally, cancer cells upregulate glycolysis to produce the biosynthetic intermediates and energy

required for cell proliferation and survival. Circulating tumour cells are predisposed to anoikis as a result of impaired glucose uptake; therefore, metastasised tumour cells prevent anoikis by upregulating the TCA cycle [225]. Previous studies have shown that wild-type *IDH2* is upregulated in lung cancer, ovarian cancer, endometrioid carcinoma, and advanced colorectal cancer [226]. Gain-of-function mutations in *IDH2* result in an increase in the onco-metabolite 2-hydroxyglutarate (2-HG), which is believed to link aberrant metabolism, aberrant epigenetics, and gene regulation in cancer [227] . A well-known function of mutant IDH2 has been demonstrated in cancers, such as glioma, cholangiocarcinoma, and breast solid papillary carcinoma with reverse polarity (SPCRP) [228]. Wild-type *IDH2* overexpression is an indicator of poor outcome in lung cancer due to its stimulation of the Warburg effect to aid in the maintenance of cancer cells via activation of hypoxia inducible factor 1α (HIF1α), which promotes tumour growth in hypoxic environments [229]. However, the role of wild-type *IDH2* in invasive BC and LVI remains unknown.

4.2 Hypothesis

The identified genes, *MX1* and *IDH2*, are overexpressed in invasive BC and have potential LVI involvement and prognostic implications.

4.3 Aims

1. To evaluate the expression of *MX1* and *IDH2* at the mRNA and protein levels in large, well-characterised, annotated cohorts of invasive BC, and to investigate the association of the genes with clinicopathological parameters and several relevant biological markers.

2. To assess the prognostic significance of *MX1* and *IDH2* at the mRNA and protein levels in different BC molecular subtypes.

4.4 Methods

4.4.1 Study cohort

4.4.1.1 METABRIC cohort

The METABRIC cohort was utilised to assess the expression of *MX1* and *IDH2* at the mRNA level. This cohort included 1,980 patients with primary operable invasive BC. Clinical history, outcome data, and tumour characteristics, including tumour size, grade, LN stage, NPI, and LVI, were available in the METABRIC cohort. In this study, the Nottingham subset (n=288) of the METABRIC cohort was used to evaluate the association between mRNA and protein expression. The characteristics of the METABRIC cohort are detailed in Section 2.1.1, and details of the Nottingham subset are found in Section 2.1.2.

4.4.1.2 TCGA cohort

The TCGA breast carcinoma cohort (n=854) was also used to evaluate *MX1* and *IDH2* mRNA expression. De-identified clinical patient information was available, including patient age, tumour size, histological grade, LVI, and outcome data. The expression of ER, PR, and HER2 was also available. The characteristics of the TCGA are detailed in Section 2.1.3.

4.4.1.3 Nottingham BC cohort

MX1 and IDH2 protein expression were evaluated using the Nottingham BC cohort. This is a well-characterised cohort of patients with early-stage, primary operable invasive BC. The clinicopathological profile of the Nottingham BC cohort included patient age, tumour size, LN stage, tumour grade, NPI, and LVI status. The hormone receptor expression profile and outcome data (BCSS and TTDM) were also available. Section 2.1.4 describes the clinicopathological characteristics of the Nottingham cohort.

The expression of *MX1* and *IDH2*, with clinicopathological characteristics, molecular subtypes, relevant biological markers, and the patient outcomes were evaluated at the mRNA and protein levels.

4.4.2 WB

The specificity of MX1 and IDH2 primary antibodies was validated using WB. Whole-cell lysates from MCF-7 and HEK cell lines were used for MX1, and lysates from MCF-10A, MCF-7, and MB-MDA-231 cell lines were used for IDH2. The methodology, antibody concentrations, and incubation times for MX1 and IDH2 are described in Section 2.2 and Table 2.5.

4.4.3 IHC

MX1 and IDH2 protein expression were assessed using the Nottingham BC cohort. To assess the pattern of MX1 and IDH2 expression, a representative full-face tissue section of invasive BC for each of the different molecular BC subtypes, histological types, LVI statuses, and tumour grades were stained to verify the staining homogeneity prior to staining of TMA. The staining pattern was demonstrated to be homogeneous, verifying the applicability of the immunohistochemical detection method in TMAs. Antigen retrieval was performed according to the manufacturer's recommendations (pH 6 citrate buffer at 1,000 W for 20 min using a microwave). The Novolink polymer detection system was used to stain TMA sections, as described in Section 2.4. Details regarding MX1 and IDH2 concentrations, incubation times, and positive and negative controls are found in Table 2.5.

4.5 Results

4.5.1 Validation of MX1 and IDH2 antibodies by WB

WB analysis detected specific bands at the predicted molecular weights of ~64 kDa and 47 kDa, corresponding to MX1 and IDH2 proteins, respectively (Figure 4.1A/B).

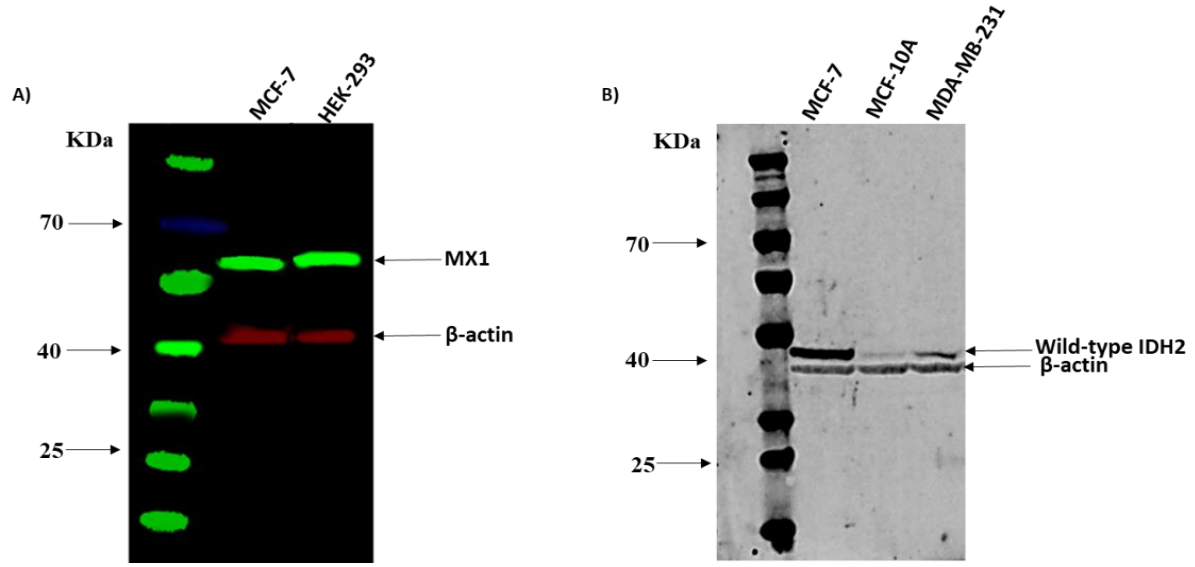


Figure 4.1 Western blot (WB) reflecting antibody specificity of A) MX1 and B) IDH2. β-actin was used as a positive control.

4.5.2 Expression of MX1 and IDH2 in BC

MX1 and *IDH2* gene expression were evaluated using the METABRIC and TCGA cohorts. The *MX1* and *IDH2* mRNA expression data were normally distributed in the METABRIC cohort and were dichotomised using cut-off points generated from the mean, while the *MX1* and *IDH2* mRNA expression data were right-skewed in the TCGA cohort and were dichotomised using cut-off points generated from the median (Figure 4.2A–D). High *MX1* and *IDH2* mRNA expression was observed in 990/1,980 (50%) of the METABRIC BC cases. High *MX1* and *IDH2* expression was observed 427/854 (50%) of the TCGA cases.

MX1 and IDH2 proteins were expressed in the cytoplasm of the invasive tumour with no discernible membranous or nuclear staining, with varying

degrees of intensity, from absent to high (Figure 4.3C/D, Figure 4.3G/H). A total of 901 and 859 cases were considered valid for evaluating MX1 and IDH2 proteins, respectively. MX1 and IDH2 protein expression data were not normally distributed, but were multimodal, and the cut-off points for MX1 and IDH2 positivity were set using their respective medians (Figure 4.2E/F). A strong concordance was found between the scorers (Abrar Aljohani, Mohammed Aleskandarany, and Islam Abdelaziz) regarding MX1 and immunoscoring (ICC=0.959, $p<0.0001$) and IDH2 immunoscoring (ICC=0.855, $p<0.0001$). High MX1 protein expression (H-score >90) was observed in 411/901 (45.6%) of invasive BC cases, while 385/859 (44.8%) of cases were high for IDH2 protein expression (H-score >90).

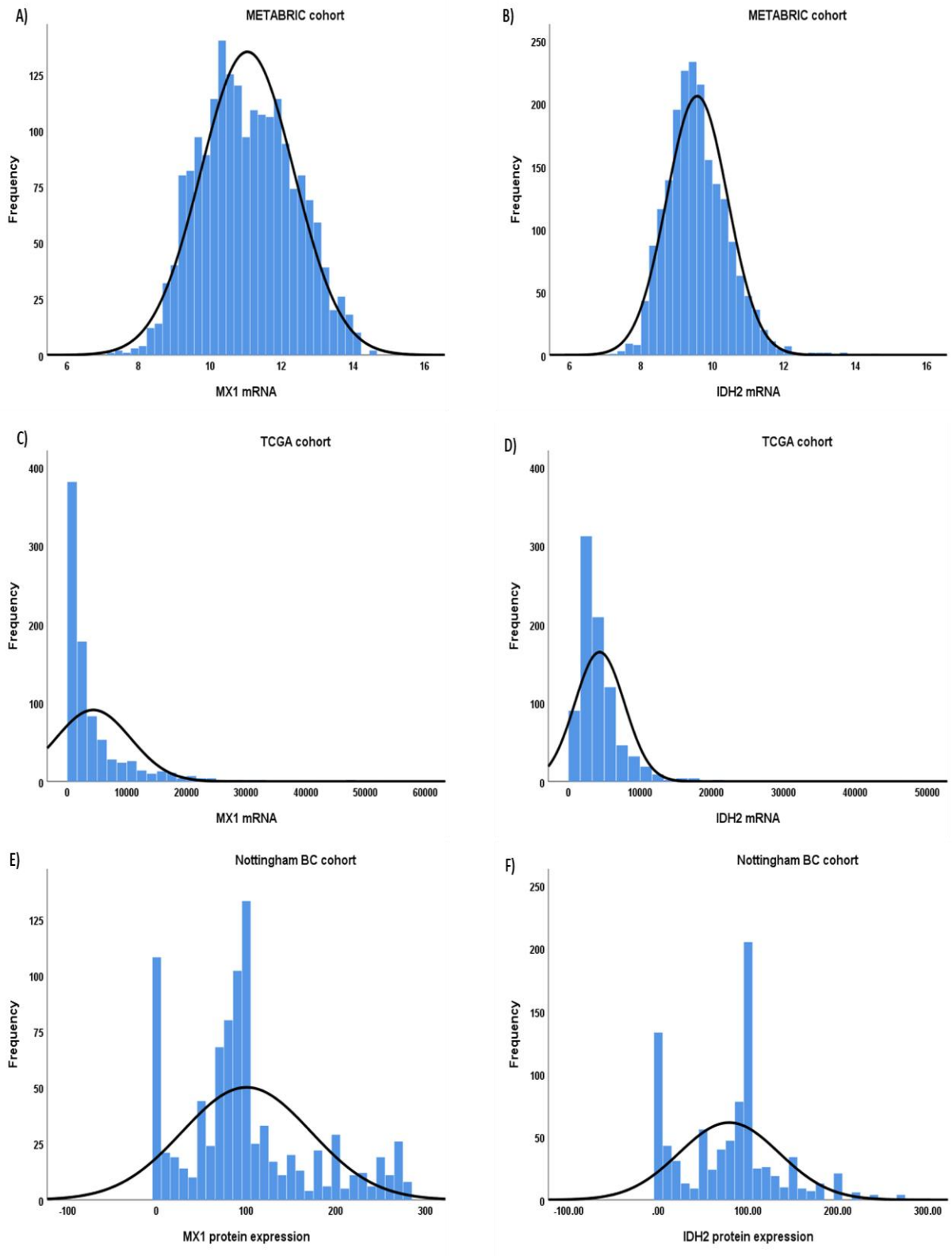


Figure 4.2 The distribution of the expression of A) *MX1* mRNA in the METABRIC cohort, B) *IDH2* mRNA in the METABRIC cohort, C) *MX1* mRNA in the TCGA cohort, D) *IDH2* mRNA in the TCGA cohort, E) MX1 protein in the Nottingham BC cohort, and F) IDH2 protein in the Nottingham BC cohort.

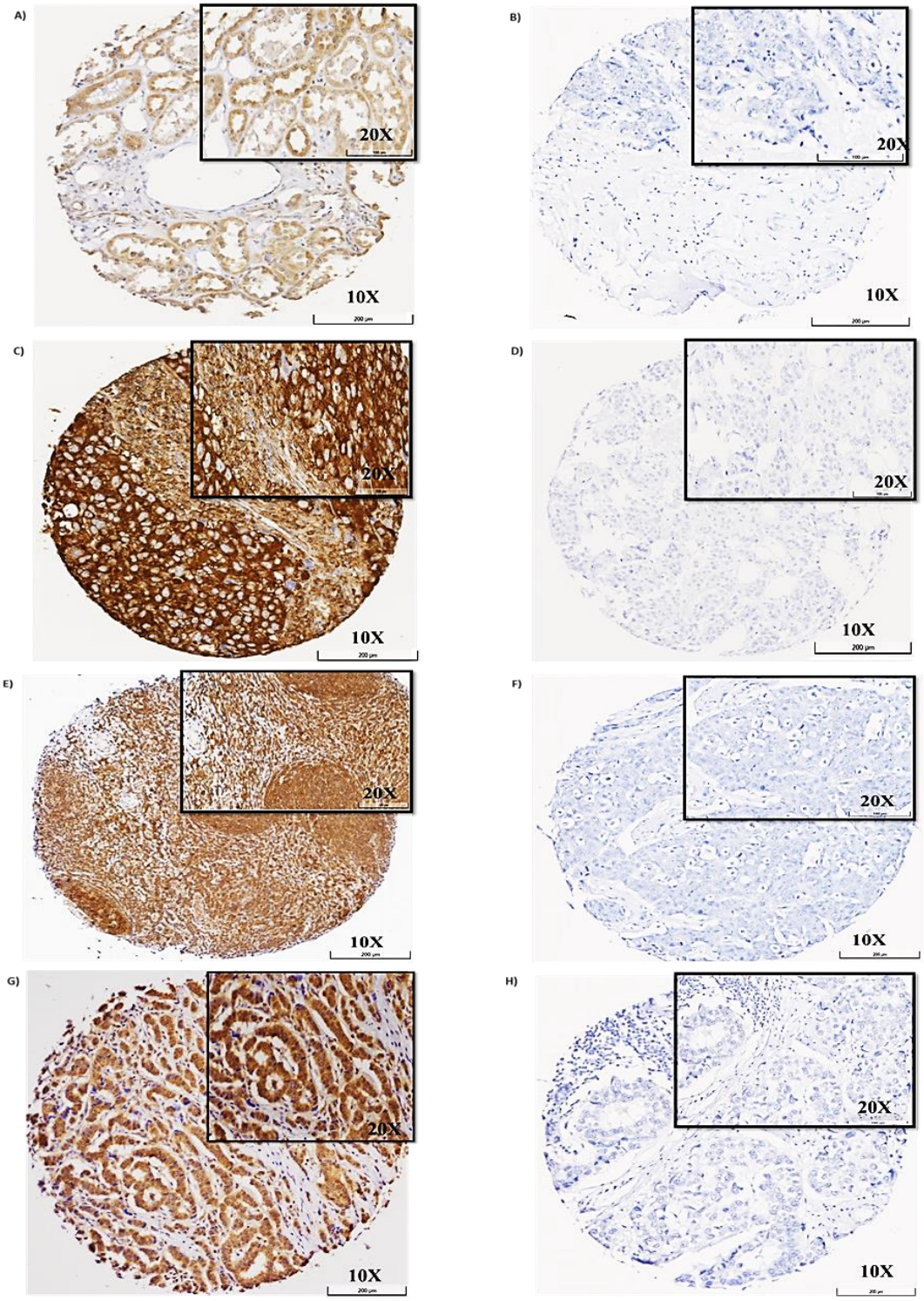


Figure 4.3 Cytoplasmic expression of MX1 and IDH2 protein in invasive breast cancer. A) Positive control normal kidney tissue with MX1 staining, B) negative control breast tissue verified by omission of MX1 antibody, C) MX1-positive IHC expression, D) MXI-negative IHC expression, E) positive control tonsil tissue with IDH2 staining, F) negative control breast tissue verified by omission of IDH2 antibody, G) IDH2-positive IHC expression, and H) IDH2-negative IHC expression. Magnification 10×, scale bars = 200 μm. Inset: magnification 20×, scale bars = 100 μm

4.5.3 Association of *MX1* and *IDH2* with clinicopathological parameters

4.5.3.1 *MX1* and *IDH2* mRNA expression

Association with clinicopathological parameters was evaluated using continuous data on *MX1* and *IDH2* mRNA and protein level. In the METABRIC cohort, high *MX1* mRNA expression was significantly correlated with younger age (p=0.002), high tumour grade (p<0.0001), poor NPI (p<0.0001), LN stage (p=0.010), and LVI positivity (p=0.016). In the TCGA cohort, high *MX1* mRNA expression was significantly associated with high tumour grade (p<0.0001) and LVI positivity (p=0.025) (Table 4.1).

In the METABRIC dataset, overexpression of *IDH2* mRNA was significantly associated with younger age (p=0.016), large tumour size (p=0.001), high histological grade (p<0.0001), poor NPI (p<0.0001), LN stage (p<0.0001), and LVI positivity (p=0.002). In the TCGA dataset, high *IDH2* mRNA levels were associated with high histological grade and LVI positivity (both p=0.002) (Table 4.2).

4.5.3.2 MX1 and IDH2 protein expression

High MX1 protein expression was significantly associated with high histological grade, high mitotic count, high pleomorphism (all p<0.0001), and high tubule formation (p=0.007) (Table 4.3). No significant associations were observed at the protein level between MX1 protein expression and other clinicopathological parameters, including patient age, tumour size, LN stage, NPI, or LVI status (all p>0.05). Based on Spearman's rank correlation coefficient, a weak correlation between *MX1* mRNA and MX1 protein expression was observed in the Nottingham subset (n=288) of the METABRIC cohort (r=0.120, p=0.042). Regarding IDH2 protein expression, in the line with the mRNA, high IDH2 protein expression was associated with younger age (p=0.034), large tumour size (p=0.009), high histological grade (p<0.0001), high mitotic count (p<0.0001), high pleomorphism (p<0.0001), high tubule formation (p=0.011), positive LVI

status (p=0.001), and high NPI (p<0.0001) (Table 4.4). No significant association was observed between IDH2 protein expression and LN stage (p>0.05). There was a weak positive linear correlation between *IDH2* mRNA expression and IDH2 protein expression (r=0.240, p=0.002) in the Nottingham subset of the METABRIC cases (n=288).

4.5.4 *MX1* and *IDH2* expression in molecular BC subtypes

In the METABRIC cohort, high expression of *MX1* or *IDH2* mRNA was significantly correlated with hormone receptor (ER/PR) negativity (all p<0.0001) and HER2 positivity (p=0.002 for *MX1* and p<0.0001 for *IDH2*). In the TCGA cohort, high *MX1* and *IDH2* mRNA levels were significantly associated with hormone receptor negativity: high *MX1* with ER negativity (p<0.0001), PR negativity (p=0.004), and HER2 positivity (p=0.004); and high *IDH2* with hormone receptor negativity and HER2 positivity (all p<0.0001) (Tables 4.1 and 4.2). Investigation of *MX1* mRNA in the METABRIC cohort in the intrinsic (PAM50) subtypes showed that overexpression was observed in normal-like, basal-like, $HER2^+$, luminal B, and luminal A subtypes, in descending order (all p<0.0001, Table 4.1). Evaluation of *IDH2* mRNA in the intrinsic (PAM50) subtypes revealed that high level of *IDH2* was associated with HER2-enriched, basal-like, luminal B, normal-like, and luminal A subtypes, in descending order (all p<0.0001, Table 4.2).

MX1 protein expression in the Nottingham BC cohort showed a significant association with hormone receptor negativity (all p<0.0001). No significant association was observed between high MX1 protein level and HER2 positivity. There was a significant association between overexpression of IDH2 and hormone receptor negativity and HER2 positivity (both p<0.0001). In addition, both MX1 and IDH2 protein levels were highly expressed in TN tumours compared with non-TN tumours (p<0.0001) (Tables 4.3 and 4.4). The results of MX1 protein level association with the IHC subtypes were similar to the above findings, where MX1 protein expression was highest in the $ER^-/HER2^-$ subtype, followed by $HER2^+$

tumours. Within the $ER^+/HER2^-$ group, expression levels were significantly higher in the high-proliferation class than in the low-proliferation class ($p<0.0001$, Table 4.3). Among the IHC subtypes, high IDH2 protein expression showed a significant association with HER2 positivity, followed by the $ER^-/HER2^-$ subtype. IDH2 protein expression levels were significantly higher in the high-proliferation class than in the low-proliferation class ($p<0.0001$, Table 4.4).

Table 4.1 Statistical associations between *MX1* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets.

Parameters	*MX1* mRNA (METABRIC)			*MX1* mRNA (TCGA)		
	Number (%)	**Mean rank**	***p*-value**	**Number (%)**	**Mean rank**	***p*-value**
Patient age (years)						
≤50	424 (21.4)	11.21	**0.002**	231(27)	444.1	0.232
>50	1556 (78.6)	10.99		623 (73)	421.4	
Tumour size						
≤2 cm	622 (31.7)	10.95	0.053	239 (28)	404.27	0.086
>2 cm	1338 (68.3)	11.07		615 (72)	436.53	
Tumour grade						
I	170 (9.0)	10.42		89 (11)	275.7	
II	770 (40.6)	10.83	**<0.0001**	375 (46)	379.1	**<0.0001**
III	952 (50.3)	11.34		352 (43)	473.4	
Nottingham prognostic index (NPI)						
Good	680 (34.3)	10.72	**<0.0001**	Not available		
Moderate	1101 (55.6)	11.19				
Poor	199 (10.1)	11.27				
Lymph node stage						
I	1035 (52.5)	10.95	**0.010**	Not available		
II	622 (31.5)	11.12				
III	316 (16.0)	11.15				
Lymphovascular invasion (LVI)						
Negative	930 (59)	10.99	**0.016**	559 (65)	413.7	**0.025**
Positive	635 (41)	11.15		295 (35)	453.6	
Oestrogen receptor (ER)						
Negative	474 (23.9)	11.43	**<0.0001**	185 (22)	475.7	**<0.0001**
Positive	1506 (76.1)	10.91		639 (78)	394.2	
Progesterone receptor (PR)						
Negative	940 (47.4)	11.24	**<0.0001**	272 (33)	443.5	**0.004**
Positive	1040 (52.6)	10.85		546 (67)	392.6	
Human epidermal growth factor receptor 2 (HER2)			**0.002**			**0.004**
Negative	1733 (87.5)	11.01		567 (81)	339.9	
Positive	247 (12.5)	11.27		133 (19)	395.6	
PAM50 subtype						
Luminal A	718 (36.4)	10.70				
Luminal B	488 (24.7)	10.76				
$HER2^+$-enriched	240 (12.1)	11.11	**<0.0001**			
Basal-like	329 (16.7)	11.18		Not available		
Normal-like	199 (10.1)	11.72				

***p*-values in bold are statistically significant**

Table 4.2 Statistical associations between *IDH2* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets.

Parameters	***IDH2* mRNA (METABRIC)**			***IDH2* mRNA (TCGA)**		
	Number (%)	**Mean rank**	***p*-value**	**Number (%)**	**Mean rank**	***p*-value**
Patient age (years)						
≤50	424 (21.4)	9.66	**0.016**	231 (27)	446.1	0.180
>50	1556 (78.6)	9.55		623 (73)	420.6	
Tumour size						
≤2 cm	622 (31.7)	9.47	**0.001**	239 (28)	419.9	0.575
>2 cm	1338 (68.3)	9.61		615 (72)	430.4	
Tumour grade						
I	170 (9.0)	9.11		89 (11)	288.9	
II	770 (40.6)	9.35	**<0.0001**	375 (46)	352.3	**<0.0001**
III	952 (50.3)	9.85		352 (43)	498.7	
Nottingham prognostic index (NPI)						
Good	680 (34.3)	9.28	**<0.0001**	Not available		
Moderate	1101 (55.6)	9.67				
Poor	199 (10.1)	10.01				
Lymph node stage						
I	1035 (52.5)	9.49	**<0.0001**	Not available		
II	622 (31.5)	9.61				
III	316 (16.0)	9.75				
Lymphovascular invasion (LVI)						
Negative	930 (59)	9.49	**0.002**	559 (65)	408.4	**0.002**
Positive	635 (41)	9.62		295 (35)	463.7	
Oestrogen receptor (ER)						
Negative	474 (23.9)	10.12	**<0.0001**	185 (22)	539.9	**<0.0001**
Positive	1506 (76.1)	9.40		639 (78)	375.6	
Progesterone receptor (PR)						
Negative	940 (47.4)	9.83	**<0.0001**	272 (33)	507.4	**<0.0001**
Positive	1040 (52.6)	9.33		546 (67)	360.7	
Human epidermal growth factor receptor 2 (HER2)						
Negative	1733 (87.5)	9.46	**<0.0001**	567 (81)	332.4	**<0.0001**
Positive	247 (12.5)	10.36		133 (19)	427.7	
PAM50 subtype						
Luminal A	718 (36.4)	9.28				
Luminal B	488 (24.7)	9.42				
HER2^{+}-enriched	240 (12.1)	10.43	**<0.0001**	Not available		
Basal-like	329 (16.7)	9.93				
Normal-like	199 (10.1)	9.33				

***p*-values in bold are statistically significant**

Table 4.3 Statistical associations between MX1 protein expression and clinicopathological factors in the Nottingham BC cohort (n=901).

Nottingham BC cohort			
Parameters	**MX1 protein**		
	Number (%)	Mean rank	*p*-value
Patient age (years)			
≤50	328 (36.2)	468.6	0.190
>50	578 (63.8)	444.9	
Tumour size			
≤2 cm	451 (50.1)	437.8	0.128
>2 cm	449 (49.9)	464.2	
Tumour grade			
I	144 (16)	366.9	**<0.0001**
II	273 (30.2)	413.3	
III	485 (53.8)	498.1	
Mitosis			
I	289 (33)	379.1	**<0.0001**
II	149 (17)	396.2	
III	438 (50)	492.1	
Pleomorphism			
I	18 (2.1)	310.8	**<0.0001**
II	312 (35.7)	376.5	
III	543 (62.2)	475.9	
Tubule formation			
I	46 (5.4)	367.5	**0.007**
II	289 (32.9)	413.1	
III	541 (61.7)	458.1	
Lymphovascular invasion (LVI)			
Negative	592 (66)	447.2	0.660
Positive	307 (34)	455.3	
Lymph node stage			
I	561 (62.1)	455.7	0.764
II	268 (29.6)	449.4	
III	74 (8.3)	432.9	
Nottingham prognostic index (NPI)			
Good	561 (62.1)	455.7	0.764
Moderate	268 (29.7)	449.5	
Poor	74 (8.2)	432.9	
Oestrogen receptor (ER)			
Negative	237 (26.2)	537.1	**<0.0001**
Positive	667 (73.8)	422.4	
Progesterone receptor (PR)			
Negative	378 (42.7)	494.1	**<0.0001**
Positive	507 (57.3)	404.9	
Human epidermal growth factor receptor 2 (HER2)			0.676
Negative	762 (87)	436.6	
Positive	113 (13)	447.2	
Triple-negative			
No	720 (80.9)	416.5	**<0.0001**
Yes	169 (19.1)	566.2	
Immunohistochemistry subtype			
$ER^+/HER2^-$ low proliferation	207 (26.5)	328.9	**<0.0001**
$ER^+/HER2^-$ high proliferation	290 (37.2)	371.3	

Triple-negative HER2^{+}	170 (21.8) 112 (14.5)	491.2 397.7	

***p*-values in bold are statistically significant**

Table 4.4 Statistical associations between IDH2 protein expression and clinicopathological factors in the Nottingham BC cohort (n=859).

Nottingham BC cohort			
Parameters	**IDH2 protein**		
	Number (%)	Mean Rank	*p*-value
Patient age (years)			
≤50	310 (36.2)	451.0	**0.034**
>50	544 (63.8)	414.1	
Tumour size			
≤2 cm	440 (51.6)	405.8	**0.009**
>2 cm	413 (48.4)	449.6	
Tumour grade			
I	130 (15.2)	348.9	**<0.0001**
II	275 (32.1)	316.7	
III	452 (52.7)	492.9	
Mitosis			
I	273 (32.6)	321.1	**<0.0001**
II	161 (19.2)	415.2	
III	403 (48.2)	486.8	
Pleomorphism			
I	14 (1.6)	377.7	**<0.0001**
II	306 (36.7)	345.1	
III	514 (61.7)	461.7	
Tubule formation			
I	42 (5)	334.8	
II	281 (33.6)	401.9	**0.011**
III	514 (61.4)	435.2	
Lymphovascular invasion (LVI)			
Negative	561 (66)	406.3	**0.001**
Positive	289 (34)	462.8	
Lymph node stage			
I	540 (63)	427.4	0.871
II	248 (28.9)	428.2	
III	69 (8.1)	443.8	
Nottingham prognostic index (NPI)			
Good	253 (29.6)	358.8	**<0.0001**
Moderate	462 (54.2)	450.4	
Poor	138 (16.2)	473.5	
Oestrogen receptor (ER)			
Negative	213 (25)	518.6	**<0.0001**
Positive	642 (75)	397.9	
Progesterone receptor (PR)			
Negative	347 (41.4)	468.9	**<0.0001**
Positive	491 (58.6)	384.5	
Human epidermal growth factor receptor 2 (HER2)			**<0.0001**
Negative	727 (87.5)	402.1	
Positive	113 (12.5)	513.0	
Triple-negative			**<0.0001**
No	689 (81.8)	405.5	
Yes	153 (18.2)	493.5	
Immunohistochemistry subtype			
$ER^+/HER2^-$ low proliferation	204 (27.4)	299.5	**<0.0001**
$ER^+/HER2^-$ high proliferation	280 (37.7)	359.6	

Triple-negative HER2$^+$	155 (20.9) 104 (14)	433.8 455.7	

***p*-values in bold are statistically significant**

4.5.5 *MX1* and *IDH2* expression and other associated biomarkers

The correlation between *MX1* and *IDH2* mRNA expression and other genes related to various LVI-associated process was investigated using the METABRIC, TCGA, and Nottingham BC datasets (Tables 4.5 and 4.6). Based on previous publications, EMT or MMPs genes were selected that are involved in invasion, EMT, and adhesion processes [128, 129, 172, 174, 179].

4.5.5.1 Associations with *MX1*

In the METABRIC cohort, there was a significant weak positive linear association between *MX1* mRNA expression and the expression of EMT-related genes, including *P-cadherin*, *ZEB1*, *ZEB2*, *TWIST1*, *TWIST2,* glycogen synthase kinase 3 beta *(GSK3B)*, and catenin beta 1 *(CTNNB1)* (all $p<0.05$) (Table 4.5). A similar correlation was observed in the TCGA cohort between *MX1* mRNA expression and the expression of *P-cadherin*, *ZEB1*, *ZEB2*, *GSK3B*, *CTNNB1*, and *TGFβ* (all $p<0.05$) (Table 4.5). In both cohorts, there was a significant weak positive correlation between *MX1* mRNA expression and other MMP-related genes, including *MMP1*, *MMP7*, *MMP9*, *MMP12*, and *MMP25* (all $p<0.05$). In the Nottingham cohort, there was also a significant weak positive linear correlation between high MX1 protein expression and levels of N-cadherin, P-cadherin, and TWIST2 (all $p<0.05$) (Table 4.5). No significant association was observed between *MX1* expression and the other genes listed in Table 4.5 ($p>0.05$).

4.5.5.2 Associations with *IDH2*

In both the METABRIC and TCGA datasets, a weak to moderate positive linear correlation was seen between *IDH2* mRNA expression and the expression of EMT-related genes, including *P-cadherin*, *ZEB1*, *ZEB2*, *NFKB1*, and *CTNNB1* (all $p<0.05$); however, the correlation with *E-cadherin*

was negative ($p<0.05$). In the METABRIC cohort only, *N-cadherin*, *TWIST2*, and *GSK3B* expression was positively correlated with *IDH2* expression (all $p<0.05$). There was also a weak to moderate significant association between *IDH2* expression and *MMP1*, *MMP7*, *MMP9*, *MMP11*, *MMP12*, *MMP15*, and *MMP25* expression in both cohorts (all $p<0.05$). Finally, there was a significant positive linear correlation between high IDH2 protein expression and N-cadherin, P-cadherin, and TWIST2 (all $p<0.05$) (Table 4.6). No significant association was observed between *IDH2* expression and the other genes listed in Table 4.6 ($p>0.05$).

4.5.5.3 Associations between *MX1* and *IDH2*

The associations between *MX1* and *IDH2* at the mRNA and protein levels were evaluated to explore their possible function in LVI. There was a significant weak positive correlation between expression of these genes at the mRNA and protein levels (all $p<0.0001$) (Table 4.7).

Table 4.5 Correlations between *MX1* expression and mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes.

Gene names	**METABRIC cohort (mRNA)**		**TCGA cohort (mRNA)**		**Nottingham cohort (protein)**	
	Correlation value	*p*-value	Correlation value	*p*-value	Correlation value	*p*-value
EMT-related genes						
E-cadherin	- 0.013	0.550	- 0.006	0.857	- 0.044	0.194
N-cadherin	0.013	0.571	0.036	0.300	0.092	**0.015**
P-cadherin	0.215	**<0.0001**	0.112	**0.001**	0.093	**0.009**
TGFβ1	0.018	0.435	0.132	**<0.0001**	0.005	0.972
TWIST1	0.060	**0.007**	0.006	0.852	**Not available**	
TWIST2	0.089	**<0.0001**	0.064	0.061	0.138	**0.001**
ZEB1	0.168	**<0.0001**	0.100	**0.003**	**Not available**	
ZEB2	0.047	**0.038**	0.074	**0.030**		
NFKB1	0.000	0.989	0.005	0.886		
GSK3B	0.240	**<0.0001**	0.126	**<0.0001**		
CTNNB1	0.168	**<0.0001**	0.117	**0.001**		
MMP-related genes						
MMP1	0.130	**<0.0001**	0.161	**<0.0001**	**Not available**	
MMP7	0.096	**<0.0001**	0.199	**<0.0001**		
MMP9	0.258	**<0.0001**	0.220	**<0.0001**		
MMP11	0.066	**0.003**	0.055	0.105		
MMP12	0.150	**<0.0001**	0.163	**<0.0001**		
MMP15	0.109	**<0.0001**	0.036	0.299		
MMP20	0.040	0.077	0.095	**0.005**		
MMP25	0.146	**<0.0001**	0.087	**0.011**		

***p*-values in bold are statistically significant**

Table 4.6 Correlations between *IDH2* expression and mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes.

Gene names	**METABRIC cohort (mRNA)**		**TCGA cohort (mRNA)**		**Nottingham cohort (protein)**	
	Correlation value	*p*-value	Correlation value	*p*-value	Correlation value	*p*-value
EMT-related genes						
E-cadherin	- 0.117	**<0.0001**	- 0.099	**0.004**	0.029	0.405
N-cadherin	0.124	**<0.0001**	0.018	0.594	0.146	**<0.0001**
P-cadherin	0.340	**<0.0001**	0.252	**<0.0001**	0.253	**<0.0001**
TGFβ1	0.021	0.350	0.028	0.415	0.025	0.889
TWIST1	0.006	0.803	0.026	0.447	**Not available**	
TWIST2	0.127	**<0.0001**	0.042	0.225	0.198	**<0.0001**
ZEB1	0.228	**<0.0001**	0.286	**<0.0001**	**Not available**	
ZEB2	0.139	**<0.0001**	0.207	**<0.0001**		
NFKB1	0.036	**0.005**	0.199	**<0.0001**		
GSK3B	0.506	**<0.0001**	0.014	0.689		
CTNNB1	0.192	**<0.0001**	0.165	**<0.0001**		
MMP-related genes						
MMP1	0.258	**<0.0001**	0.299	**<0.0001**	**Not available**	
MMP7	0.082	**<0.0001**	0.113	**0.001**		
MMP9	0.209	**<0.0001**	0.188	**<0.0001**		
MMP11	0.154	**<0.0001**	0.130	**0.060**		
MMP12	0.229	**<0.0001**	0.193	**<0.0001**		
MMP15	0.385	**<0.0001**	0.320	**0.002**		
MMP20	0.044	**0.052**	0.014	0.677		
MMP25	0.110	**<0.0001**	0.106	**0.002**		

***p*-values in bold are statistically significant**

Table 4.7 Correlations between *MX1* and *IDH2* in mRNA and protein expression levels.

METABRIC cohort (mRNA)		**TCGA cohort (mRNA)**		**Nottingham cohort (protein)**	
Correlation value	*p*-value	Correlation value	*p*-value	Correlation value	*p*-value
0.195	**<0.0001**	0.115	**0.001**	0.246	**<0.0001**

***p*-values in bold are statistically significant**

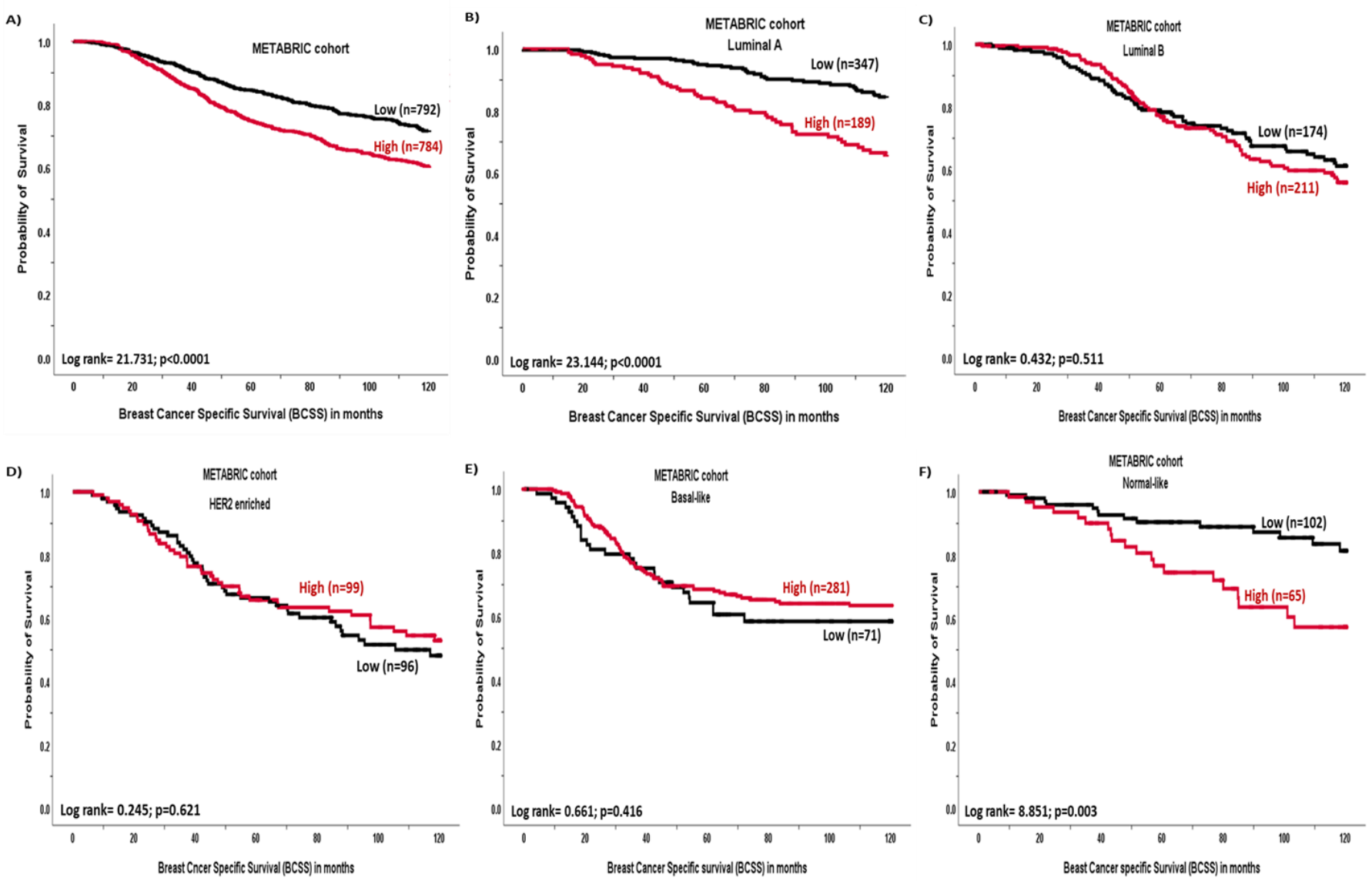
4.5.6 Association of *MX1* and *IDH2* with patient outcome

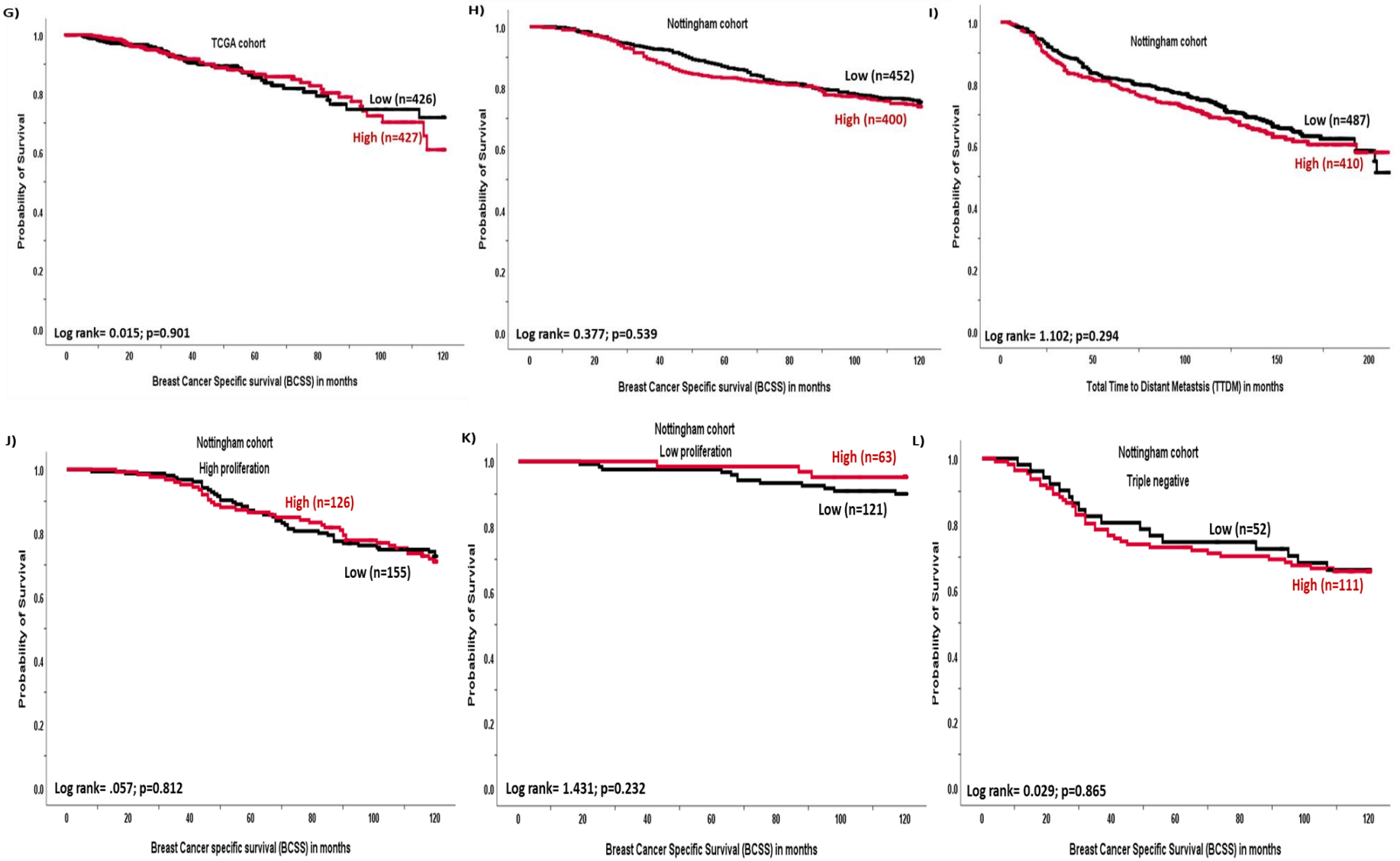
In the METABRIC cohort, survival analyses of both *MX1* and *IDH2* mRNA revealed that higher expression of *MX1* and *IDH2* was associated with shorter BCSS (all p<0.0001) (Figures 4.4A and 4.5A). However, there was no significant association between high *MX1* or *IDH2* mRNA expression and BCSS in the TCGA cohort (all p=0.901) (Figures 4.4G and 4.5G). At the protein level, poor patient outcome was associated with high expression of IDH2 (p=0.038) but not MX1 (p>0.05) (Figures 4.4H and 4.5H). Protein expression of MX1 and IDH2 was not associated with shorter TTDM (all p>0.05) (Figures 4.4I and 4.5I).

When analysis was performed in the intrinsic molecular classes in the METABRIC dataset, high expression of *MX1* and *IDH2* mRNA was only associated with shorter BCSS in the normal-like subtype (p=0.003 and p=0.013, respectively) (Figures 4.4F and 4.5F). High expression of *MX1* mRNA was associated with shorter BCSS in the luminal A subtype (p<0.0001, Figure 4.4B) while no significant association was observed between *MX1* and *IDH2* mRNA expression and shorter BCSS in the other subtypes (all p>0.05) (Figures 4.4C–E and 4.5B–E). Furthermore, none of the subtypes showed a significant association between BCSS and protein expression of either MX1 or IDH2 (all p>0.05) (Figures 4.4J–M and 4.5J–M).

In the METABRIC cohort, multivariate Cox regression analysis revealed that *MX1* and *IDH2* mRNA expression predicted poor BCSS survival independent of the tumour size, LN stage, tumour grade, and LVI (HR 1.3; 95%CI=1.01-1.6*;* p=0.038 and HR 1.3; 95%CI=1.1-1.7*;* p=0.007 respectively) (Tables 4.8 and 4.9); when the same analysis was performed in the TCGA cohort, no significant association was observed between *MX1* and *IDH2* mRNA expression and poor BCSS survival (all p>0.05, Tables 4.8 and 4.9). At the protein level, Cox regression analysis showed that high expression of IDH2,

but not MX1, was a significant predictor of shorter BCSS, regardless of the same variables (HR 1.3; 95%CI=1.0-1.8*;* p=0.049 and p>0.05, respectively) (Tables 4.8 and 4.9).


A)
METABRIC cohort
Low (n=792)
High (n=784)
Probability of Survival
Log rank= 21.731; p<0.0001
Breast Cancer Specific Survival (BCSS) in months
B)
METABRIC cohort
Luminal A
Low (n=347)
High (n=189)
Probability of Survival
Log rank= 23.144; p<0.0001
Breast Cancer Specific Survival (BCSS) in months
C)
METABRIC cohort
Luminal B
Low (n=174)
High (n=211)
Probability of Survival
Log rank= 0.432; p=0.511
Breast Cancer Specific Survival (BCSS) in months
D)
METABRIC cohort
HER2 enriched
High (n=99)
Low (n=96)
Probability of Survival
Log rank= 0.245; p=0.621
Breast Cncer Specific Survival (BCSS) in months
E)
METABRIC cohort
Basal-like
High (n=281)
Low (n=71)
Probability of Survival
Log rank= 0.661; p=0.416
Breast Cancer Specific survival (BCSS) in months
F)
METABRIC cohort
Normal-like
Low (n=102)
High (n=65)
Probability of Survival
Log rank= 8.851; p=0.003
Beast Cancer specific Survival (BCSS) in months


G)
TCGA cohort
Low (n=426)
High (n=427)
Probability of Survival
Log rank= 0.015; p=0.901
Breast Cancer Specific survival (BCSS) in months
H)
Nottingham cohort
Low (n=452)
High (n=400)
Probability of Survival
Log rank= 0.377; p=0.539
Breast Cancer Specific survival (BCSS) in months
I)
Nottingham cohort
Low (n=487)
High (n=410)
Probability of Survival
Log rank= 1.102; p=0.294
Total Time to Distant Metastsis (TTDM) in months
J)
Nottingham cohort
High proliferation
High (n=126)
Low (n=155)
Probability of Survival
Log rank= .057; p=0.812
Breast Cancer specific survival (BCSS) in months
K)
Nottingham cohort
Low proliferation
High (n=63)
Low (n=121)
Probability of Survival
Log rank= 1.431; p=0.232
Breast Cancer Specific Survival (BCSS) in months
L)
Nottingham cohort
Triple negative
Low (n=52)
High (n=111)
Probability of Survival
Log rank= 0.029; p=0.865
Breast Cancer Specific Survival (BCSS) in months

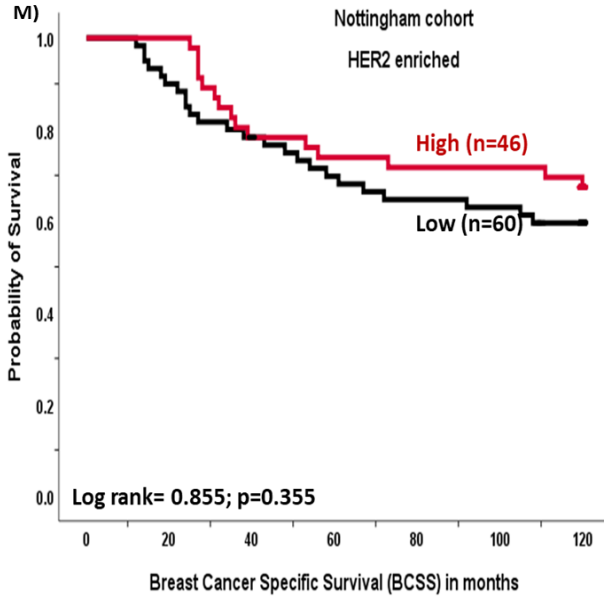
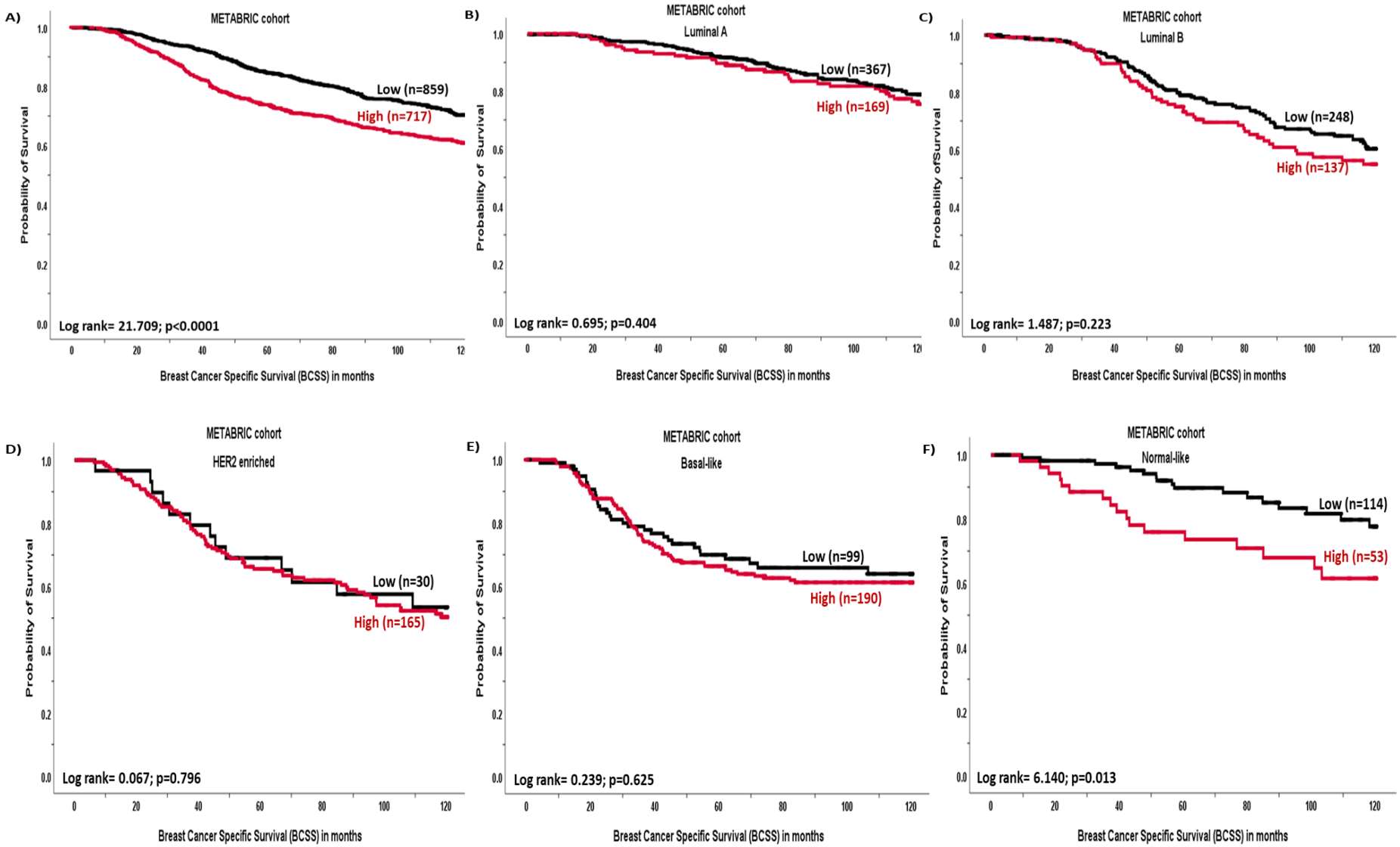
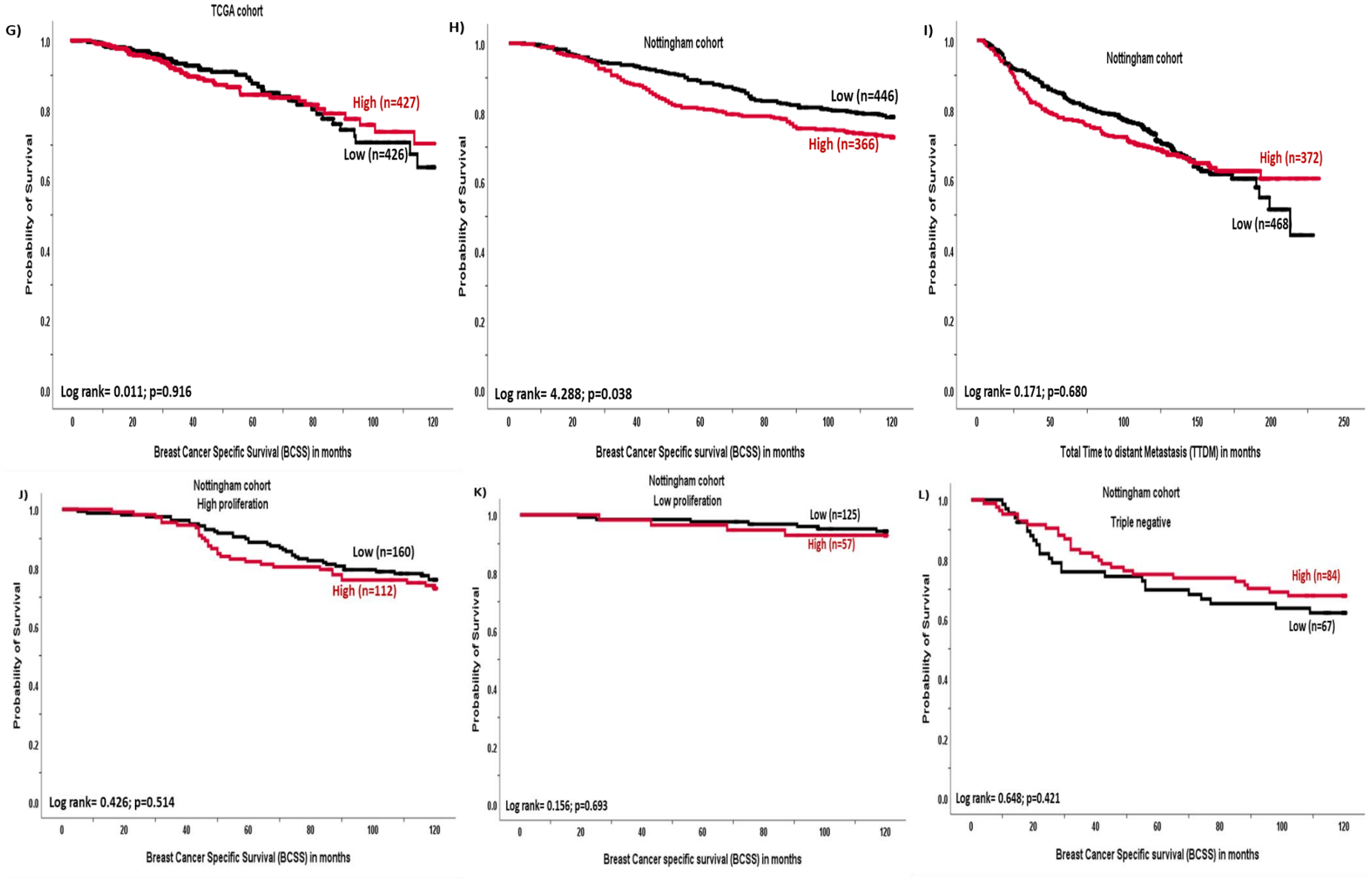


Figure 4.4 Kaplan–Meier survival plots showing the association between *MX1* mRNA expression and breast cancer-specific survival (BCSS) in A) the whole METABRIC cohort, B) luminal A (METABRIC), C) luminal B (METABRIC), D) HER2-enriched (METABRIC), E) basal-like (METABRIC), F) normal-like (METABRIC), and G) the whole TCGA cohort. For Nottingham cohort data, associations are shown between MX1 protein expression and H) BCSS, I) total time to distant metastasis (TTDM), J) BCSS in high-proliferation tumours, K) BCSS in low-proliferation tumours, L) BCSS in triple-negative tumours, and M) BCSS in HER2-enriched tumours.


A)
METABRIC cohort
Low (n=859)
High (n=717)
Log rank= 21.709; p<0.0001
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
B)
METABRIC cohort
Luminal A
Low (n=367)
High (n=169)
Log rank= 0.695; p=0.404
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
C)
METABRIC cohort
Luminal B
Low (n=248)
High (n=137)
Log rank= 1.487; p=0.223
Probability ofSurvival
Breast Cancer Specific Survival (BCSS) in months
D)
METABRIC cohort
HER2 enriched
Low (n=30)
High (n=165)
Log rank= 0.067; p=0.796
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
E)
METABRIC cohort
Basal-like
Low (n=99)
High (n=190)
Log rank= 0.239; p=0.625
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
F)
METABRIC cohort
Normal-like
Low (n=114)
High (n=53)
Log rank= 6.140; p=0.013
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months


G)
TCGA cohort
High (n=427)
Low (n=426)
Log rank= 0.011; p=0.916
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
H)
Nottingham cohort
Low (n=446)
High (n=366)
Log rank= 4.288; p=0.038
Probability of Survival
Breast Cancer Specific survival (BCSS) in months
I)
Nottingham cohort
High (n=372)
Low (n=468)
Log rank= 0.171; p=0.680
Probability of Survival
Total Time to distant Metastasis (TTDM) in months
J)
Nottingham cohort
High proliferation
Low (n=160)
High (n=112)
Log rank= 0.426; p=0.514
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
K)
Nottingham cohort
Low proliferation
Low (n=125)
High (n=57)
Log rank= 0.156; p=0.693
Probability of Survival
Breast Cancer specific survival (BCSS) in months
L)
Nottingham cohort
Triple negative
High (n=84)
Low (n=67)
Log rank= 0.648; p=0.421
Probability of Survival
Breast Cancer Specific survival (BCSS) in months

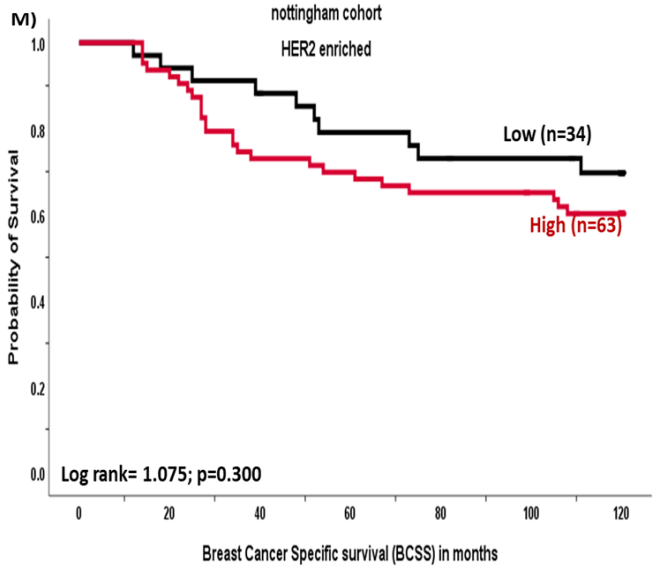


Figure 4.5 Kaplan–Meier survival plots showing the association between *IDH2* mRNA expression and breast cancer-specific survival (BCSS) in A) the whole METABRIC cohort, B) luminal A (METABRIC), C) luminal B (METABRIC), D) HER2-enriched (METABRIC), E) basal-like (METABRIC), F) normal-like (METABRIC), and G) the whole TCGA cohort. For Nottingham cohort data, associations are shown between IDH2 protein expression and H) BCSS, I) total time to distant metastasis (TTDM), J) BCSS in high-proliferation tumours, K) BCSS in low-proliferation tumours, L) BCSS in triple-negative tumours, and M) BCSS in HER2-enriched tumours.

Table 4.8 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *MX1* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort.

METABRIC cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		***p*-value**
		Lower	**Upper**	
***MX1* mRNA expression**	1.268	1.014	1.585	**0.038**
Tumour size	1.443	1.100	1.893	**0.008**
Lymph node stage	1.781	1.532	2.071	**<0.0001**
Tumour grade	1.355	1.109	1.655	**0.003**
Lymphovascular invasion (LVI)	1.377	1.088	1.741	**0.008**
TCGA cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		***p*-value**
		Lower	**Upper**	
***MX1* mRNA expression**	0.891	0.572	1.390	0.891
Tumour size	1.465	0.840	2.555	0.179
Lymph node	1.205	0.712	2.041	0.487
Tumour grade	1.301	0.921	1.838	0.136
Lymphovascular invasion (LVI)	1.934	1.176	3.179	**0.009**
Nottingham BC cohort (protein)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		***p*-value**
		Lower	**Upper**	
MX1 protein expression	1.048	0.795	1.382	0.738
Tumour size	1.345	0.999	1.810	**0.051**
Lymph node stage	1.869	1.525	2.289	**<0.0001**
Tumour grade	2.164	1.666	2.810	**<0.0001**
Lymphovascular invasion (LVI)	1.290	0.963	1.727	0.088

***p*-values in bold are statistically significant**

Table 4.9 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *IDH2* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort.

METABRIC cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		***p*-value**
		Lower	**Upper**	
***IDH2* mRNA expression**	1.364	1.089	1.709	**0.007**
Tumour size	1.446	1.103	1.896	**0.008**
Lymph node stage	1.792	1.540	2.084	**<0.0001**
Tumour grade	1.302	1.061	1.597	**0.012**
Lymphovascular invasion (LVI)	1.374	1.085	1.738	**0.008**
TCGA cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		***p*-value**
		Lower	**Upper**	
***IDH2* mRNA expression**	0.927	0.591	1.453	0.740
Tumour size	1.458	0.837	2.542	0.183
Lymph node	1.218	0.722	2.056	0.460
Tumour grade	1.297	0.914	1.840	0.145
Lymphovascular invasion (LVI)	1.929	1.174	3.169	**0.010**
Nottingham BC cohort (protein)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		***p*-value**
		Lower	**Upper**	
IDH2 protein expression	1.339	0.990	1.828	**0.049**
Tumour size	1.283	0.939	1.753	0.118
Lymph node stage	2.210	1.707	2.634	**<0.0001**
Tumour grade	1.948	1.495	2.538	**<0.0001**
Lymphovascular invasion (LVI)	1.287	0.949	1.744	0.104

***p*-values in bold are statistically significant**

4.6 Discussion

Despite breakthroughs in genetic and molecular analysis, the mechanisms underlying the progression of BC to invasive disease, distant metastasis, and LVI remain unclear. Identification of the set of genes or proteins that control LVI has thus far remained elusive. However, through stringent bioinformatic analysis in this study, *MX1* and *IDH2* were identified using the METABRIC and TCGA transcriptomic datasets as candidate genes associated with LVI positivity. Studies that address the prognostic significance of *MX1* and *IDH2* in invasive BC and their association with aggressive tumour behaviour, including LVI, remain limited; therefore, this study evaluated the clinicopathological and the prognostic significance of *MX1* and *IDH2* at the transcriptomic and proteomic levels using large BC cohorts with long-term follow-up.

4.6.1 MX1

Although MX1 protein activity is essential for the antiviral activity of the immune system [230, 231], it is also part of a signalling pathway that contributes significantly to cancer progression. MX1 and its association with carcinogenesis is poorly understood. It has been shown that type I, type II, and type III IFNs can induce *MX1* transcription [232, 233]. Both type I and type II IFNs have been used to treat a variety of cancers, including melanoma, chronic myeloid leukaemia, and renal cell carcinoma, but the antitumoural mechanism of these IFNs remains unknown [234].

To evaluate mRNA and protein expression of MX1 using IHC, multiple well-annotated BC cohorts were used to assess their association with clinicopathological parameters and patient outcomes. The results revealed an association between *MX1* mRNA and protein expression and the clinicopathological features characteristic of aggressive tumour behaviour, such as high histological grade and hormonal receptor negativity; these findings strengthen the putative role of *MX1* in tumour progression. At the mRNA level, high *MX1* levels were significantly associated with LVI

positivity, LN stage, poor NPI, and HER2 positivity, but these findings were not reflected at the protein level. In addition, the correlation between the mRNA and protein expression of *MX1* was weak. The weak correlation between the mRNA and the protein level could support the contradicting results in LVI between protein and mRNA.This may have resulted from different biological and technical factors. One of these biological factors is the rate of mRNA translation into protein, often termed 'translational efficiency', which has been shown to significantly impact mRNA and protein levels [235]. Another reason might be attributed to the subjectivity of the H-score method in the interpretation of expression in IHC staining sections [236] and the fact that the METABRIC cases used whole tissue comprising many different cell types.

ISGs are largely expressed in different types of cancers, and IFNs play a significant role in various pathways associated with malignancies. A study revealed that high histological grade and intense infiltration of TILs were correlated with the expression of MX1 protein. Type I IFNs cause T-cell exhaustion by inducing the overexpression of tumour-programmed cell death ligand 1 (PD-L1) and the increase of PD-L1 levels, which can interact with programmed death 1 in T-cells [222].

In this study, high *MX1* mRNA and protein levels were significantly associated with EMT markers, such as P-cadherin and TWIST-2, that regulate migration, EMT, and invasion [237]. High *MX1* level was also associated with MMP markers, such as *MMP1*, *MMP7*, and *MMP9*. MMPs are involved in numerous essential carcinogenic processes of carcinogenesis, including migration, invasion, metastasis, and angiogenesis, which are all tightly connected to the extracellular environment. For example, MMP-9 can cleave a large number of ECM proteins to regulate ECM remodelling; in addition, it is capable of cleaving a large number of plasma surface proteins, releasing them from the cell surface [238]. Therefore, through association with EMT and MMP markers that are important to the mechanism of

metastasis in breast carcinoma, *MX1* may play an essential role in the regulation of tumour progression.

The results of the METABRIC cohort indicated that *MX1* mRNA is a potential prognostic biomarker of invasive BC, particularly for the luminal A and normal-like tumour subtypes. In addition, in the same cohort, *MX1* predicted poor BCSS survival independent of tumour size, LN stage, tumour grade, and LVI. However, no significant association between BCSS and *MX1* was found in the TCGA cohort at the mRNA level or in the Nottingham cohort at the protein level. This inconsistency could be explained by the fact that the number of events and follow-ups in each dataset may affect the results, in addition to tumour-specific differences in *MX1* mRNA/protein stability or post-transcriptional regulation. There are also cohort-based differences that could contribute to the inconsistency observed in some results. For example, RNA sequencing was employed to acquire gene expression data for TCGA, while microarray technology was used for METABRIC. The method used to normalise the mRNA expression in each cohort represented another difference between the cohorts.

In conclusion, *MX1* may play a role in invasive BC. However, further functional *in vitro* and/or *in vivo* studies of the biological role of *MX1* in BC cases and LVI are necessary to assess its potential as a therapeutic target in BC.

4.6.2 IDH2

Wild-type *IDH2* has been previously reported to be highly expressed in some cancer types, including lung cancer and ovarian cancer [229]; however, the prognostic significance of *IDH2* in invasive BC has not yet been described. This is the first study addressing the role of wild-type IDH2 protein in BC using IHC and well-annotated patient cohorts. This study included large cohorts of invasive BC in its assessment of the transcriptomic and proteomic expression of IDH2 and the correlation with clinicopathological parameters and patient outcomes. The analysis of IDH2

expression in invasive BC showed that high IDH2 expression was associated with features of aggressive tumour behaviour, including LVI, large tumour size, high grade, and poor NPI. Furthermore, *IDH2* mRNA and protein was also highly expressed in TN and $HER2^+$ BC. TN BC is highly malignant, prone to metastasis and recurrence, and is associated with a worse prognosis and a higher risk of mortality when compared with other BC subtypes [239]. $HER2^+$ tumours were more likely to have IDH2 protein, which is perhaps unsurprising as high-grade progressive LVI is significantly associated with HER2 positivity [240]. High *IDH2* expression in BC might enhance HER2 signalling pathway activity, which can induce the TME to support tumour cell growth, stimulating invasion, LVI, and metastasis [241]. These findings suggest that *IDH2* has an oncogenic effect, plays a role in tumourigenesis, and could be a key gene in LVI.

This study revealed that IDH2 was associated with poor prognostic characteristics and short-term survival outcomes in invasive BC. High *IDH2* expression was also an independent predictor of shorter BCSS in both proteomic and transcriptomic datasets. The METABRIC and Nottingham BC datasets supported the significance of high *IDH2* expression as predictor of poor prognosis.

Cellular energy produced by the TCA cycle is upregulated in highly proliferative and metastasised cancer cells. Energy metabolism and biosynthetic intermediates, such as alpha-ketoglutarate (αKG) produced in the TCA cycle from isocitrate by IDH2, are essential in tumour progression and metastasis. Wild-type *IDH2* can reduce αKG levels and increase 2-HG production, disrupting normal epigenetic regulation of transcription [229, 242]. For example, in hypoxic conditions in BC, IDH2 carboxylates αKG from glutamine to citrate and elevates 2-HG levels, acting as an oncometabolite [243]. Moreover, IDH2-mediated reduction of αKG can have an oncogenic impact on cellular differentiation [229]. αKG may have a role in EMT, which is a key step of metastasis. It has been shown that blocking αKG inhibits cellular transformation and cancerous cell invasion

through transamination or the reverse TCA cycle [227]. Research shows that IDH2 is associated with EMT and MMP proteins, which support its role in disease progression and metastasis. This study demonstrated that high levels of N-cadherin were associated with high levels of IDH2 protein. Previous studies elucidated that N-cadherin has a role in motility, invasion, and metastasis by increasing MMP-9 production, allowing tumour cells to penetrate matrix barriers and to adhere to the endothelium; this may also increase the likelihood of tumour cells entering and exiting the vasculature [237]. Therefore, the observed association between high levels of IDH2 and N-cadherin elucidates the role of IDH2 in LVI, as N-cadherin can promote two properties of metastasised cells: loss of cellular adhesion and invasion. IDH2 was also associated with the TWIST2, a marker of EMT. EMT has an important role in cancer invasion and metastasis. TWIST2 is overexpressed in BC and lymph node metastases as it results in morphological transformation, downregulation of epithelial markers, and upregulation of mesenchymal markers [244]. Therefore, the interaction of IDH2 with EMT markers suggests its role in BC cell invasion through lymphovascular channels. EMT is a known mechanism utilised by breast carcinoma cells to achieve basement membrane invasion and LVI [170]. Further mechanistic studies are needed to understand the underlying molecular mechanisms.

One of the major hallmarks of cancer is the reprogramming of energy metabolism [245]. Glucose and glutamine are taken up by cancer cells at a high rate to produce the NADPH needed to survive and grow, and to decrease the levels of intracellular αKG [246]. Many studies have provided evidence that oncogenes can alter cancer cell metabolism, including alteration of glutamine transporters, which were associated with high IDH2 in this study [246, 247]. This finding suggests that IDH2 could potentially reprogram metabolic pathways in BC and support tumour cell survival, highlighting its prognostic significance in the disease.

It has been reported that IDH2 plays a critical role in cell proliferation through the alteration of NADP levels. In addition to the role of NADP as an

antioxidant, it has been reported to have an important role in cell death. NADP links cell survival with biological properties, such as energy metabolism and oxidative stress, which are both factors that determine cell death [248]. Primary tumour cells must proliferate and invade the adjacent tissue to establish an invasion and metastasis cascade that comprises many steps, including basement membrane degradation and LVI. Proliferation lasts until the invasion of blood vessels or lymphatic channels occur. Tumour cells at this stage evade apoptosis and immune responses [249]. Therefore, IDH2 may play an important role in cell proliferation, which is a prerequisite step of the invasion and metastatic processes, and can lead to the development of invasive cancer, subsequent LVI, and distant metastasis.

In conclusion, high IDH2 expression is an independent poor prognostic factor for BC. High IDH2 expression may have a key role in BC progression and in the development of LVI and metastasis. Further functional assessment of IDH2 in BC is warranted to detect its underlying mechanistic roles in LVI and its therapeutic potential.

4.7 Conclusion

Altogether, this study revealed that high expression of *IDH2*, but not *MX1*, was significantly associated with LVI positivity and poor patient outcomes, suggesting that *IDH2* might be an LVI-associated gene. In addition, high expression of *MX1* and *IDH2* at both the mRNA and protein levels was associated with poor clinicopathological parameters, such as higher tumour grade and hormone receptor negativity. A significant association of *MX1* and *IDH2* with EMT and MMP markers was also observed. Therefore, these two genes may be involved in one or more pathways of LVI development and may play a role in BC progression, highlighting their potential as prognostic markers. Functional assessment is essential for elucidating the specific roles of these genes in LVI and to discover their potential as therapeutic targets in BC.

Chapter 5: The prognostic significance and potential pathogenetic role of HMGB3 in invasive BC

5.1 Introduction

The second most abundant protein in the cell is the high mobility group (HMG) proteins involved in the global assembly of chromatin domains. HMG enhances transcriptional fine-tuning in response to abrupt environmental changes by interacting with nucleosomes, transcription factors, nucleosome remodelling complexes, and histone H1 [250]. The high mobility group box (HMGB) family belongs to the HMG superfamily and comprises chromosomal proteins that participate in DNA replication, transcription, recombination, and repair [251, 252]. Abnormal expression of HMGB members is correlated with various cancer hallmarks, including uncontrolled replicative capacity, resistance to apoptosis, tissue invasion, and metastasis, indicating the importance of the HMGB family in human malignancies [250, 253]. The HMGB family includes *HMGB1*, *HMGB2*, *HMGB3*, and *HMGB4* [254]. These members have 80% amino acid homology; however, their biological roles in cells are distinct [250, 253]. Numerous studies have revealed that *HMGB1*, *HMGB2*, and *HMGB3* play a significant role in a variety of malignancies, including hepatocellular carcinoma [255], pancreatic cancer [256], and colon cancer [257]. Following WAD bioinformatic analysis, HMGB3 was determined to be one of the highly expressed genes associated with LVI positivity (as discussed in Chapter 3). *HMGB3* is mainly distributed in the nucleus, cytoplasm and chromosomes [255]. Normally, *HMGB3* expression is high during embryogenesis, with low or absent expression in normal adult tissue [258, 259]. It is widely expressed in mammalian cells and has roles in many physiological and pathological pathways, including DNA repair and cell death. Although *HMGB3* has the potential to maintain homeostasis, it can also affect disease progression [260]. High expression of *HMGB3* is strongly associated with the occurrence of numerous tumours and poor prognosis of advanced tumours in lung [261], bladder [262], and prostate cancer [263]. To enhance tumour development, *HMGB3* can control the cell cycle and stimulate the proliferation and invasion of cancer cells via the Wnt/-

catenin, MAPK, and other signalling pathways. *HMGB3* can also increase reactive oxygen species (ROS) formation and tumour cell growth by stimulating the expression of *HIF-1α* [260]. However, the exact role of *HMGB3* in LVI in BC is unclear as it has not been previously investigated in BC studies. This study aimed to investigate the clinicopathological significance of *HMGB3* at the transcriptomic and proteomic levels using large BC cohorts with long-term follow-up and to assess *in vitro* mechanistic role of *HMGB3* in BC cell lines.

5.2 Hypothesis

HMGB3 is a key driver for LVI in invasive BC, has prognostic importance, is a possible therapeutic target, and is involved in a variety of LVI-related pathways.

5.3 Aims

1. Assess the clinicopathological and prognostic significance of *HMGB3* at the transcriptomic and protein levels using well-characterised BC cohorts.

2. Assess the effect of *HMGB3* suppression on tumour cell characteristics, including cell proliferation, migration, cell cycle and apoptosis, employing panels of relevant BC cell lines in *in vitro* experimental studies.

3. Assess the functional importance of *HMGB3* in LVI by co-culturing BC cell lines with lymphatic and vascular endothelial cell models, and by performing adhesion and transmigration assays to investigate the interaction of BC cells with endothelial cells.

5.4 Methods

5.4.1 Study cohort

5.4.1.1 METABRIC cohort

The METABRIC cohort was used to evaluate the expression of *HMGB3* at the transcriptomic level. This study enrolled a total of 1,980 patients with primary operable invasive BC. The clinical history, outcome data, and clinicopathological features of the METABRIC cohort are described in

Section 2.1.1, and the Nottingham subgroup of the METABRIC cohort is described in Section 2.1.2.

5.4.1.2 TCGA cohort

mRNA expression levels of *HMGB3* were also evaluated using the TCGA BC cohort (n=854). The features of the TCGA cohort are detailed in Section 2.1.3.

5.4.1.3 Nottingham BC cohort

The expression of HMGB3 at the protein level was assessed using tissue samples from the well-characterised Nottingham invasive BC cohort. For each patient, a robust clinicopathological profile and outcome data were readily available. Section 2.1.4 discusses the clinicopathological features of the cohort. The associations between HMGB3 and EMT-related markers, including E-cadherin, N-cadherin, P-cadherin, TGFβ1, and TWIST2, were investigated to further understand their molecular interactions with LVI-related process (these biomarkers are available in our database).

5.4.2 WB

WB was performed to validate the specificity of the HMGB3 primary antibody. MCF-7 and SK-BR-3 BC cell line lysates were used for this antibody. The methods, HMGB3 antibody concentration, and incubation period are detailed in Section 2.2 and Table 2.5.

5.4.3 IHC

HMGB3 protein expression levels were determined using the Nottingham BC cohort. Before staining the TMA, representative full-face tissue sections for invasive BC were stained to analyse the pattern of HMGB3 expression. The staining pattern was homogeneous, verifying the applicability of immunohistochemical detection methods in TMAs. TMAs of the study cohort were prepared using a TMA Grand Master®. Antigen retrieval was performed according to the manufacturer's guidelines (pH 6 citrate buffer, at 1,000 W for 20 min using a microwave). The expression of the HMGB3

antibody was assessed by staining the TMAs with a Novolink polymer detection systems kit, as described in Section 2.4. The HMGB3 concentration, incubation time, and positive and negative controls are detailed in Table 2.5.

5.4.4 *In vitro* studies of *HMGB3*

Preclinical investigations using the following cells and assays were conducted to determine the possible involvement of *HMGB3* in LVI and other biological functions.

5.4.4.1 BC tumour cells

The protein expression levels from an IHC analysis of HMGB3 were used to select BC cell lines for *in vitro* investigations in this study. The BC cell lines MCF-7 ($ER^+/PR^+/HER2^-$), ZR-75-1 ($ER^+/PR^{\pm}/HER2^-$), and SK-BR-3 ($HER2^+$) were used. MCF-7 and ZR-75-1 are luminal molecular classes, while SK-BR-3 is a $HER2^+$ molecular class. Luminal and HER2 molecular classes have been reported to be associated with the highest incidence of LVI [179]. WB was performed to detect relative protein expression in all BC cell lines, and the LI-COR Odyssey machine was used for quantification. Cells with the highest levels of *HMGB3* expression were selected for the knockdown and other *in vitro* assays. Chapter 2 provides further information on maintaining cell growth and culture conditions.

5.4.4.2 Endothelial cells

To investigate the impact of *HMGB3* on LVI, primary HUVECs and DLECs were used to study the interaction between endothelial cells and BC cells. Details on the cell growth maintenance and culture conditions of HUVECs and DLECs are reported in Chapter 2.

5.4.4.3 Silencing *HMGB3* using siRNA

An siRNA-based approach using BC cell lines was used to study the potential functional consequences of *HMGB3* knockdown and its role in BC progression and LVI. Pre-validated Silencer Select siRNA constructs, mainly

for *HMGB3,* or scrambled negative-control siRNA that did not target *HMGB3*, MCF-7, ZR-75-1, and SK-BR-3 were transfected using the forward transfection method, as described in Chapter 2. The sequence of *HMGB3*-siRNA was as follows: 5'-GCACCCUGAAACUGUAUCAtt-3'.

5.4.4.4 MTS assay

The effect of *HMGB3* knockdown on the proliferation of tumour cells was evaluated via the AQueous Non-Radioactive Cell Proliferation Assay, as described in Section 2.6.3.2. Cell viability was determined at 24 h, 48 h, and 72 h after transfection using siRNA targeted against *HMGB3*, and cell vitality was assessed by measuring the OD.

5.4.4.5 Colony formation assay

BC cell lines were seeded and grown in culture medium in an incubator for 14 days. Following incubation, colonies were washed with PBS, fixed for 30 min, stained with crystal violet, and counted using a microscope. The assay procedure and conditions are described in Section 2.6.4.2.

5.4.4.6 Wound-healing assay

In this assay, the ability of tumour cells to migrate was assessed by measuring the wound repair rate of an *HMGB3* knockdown and control at the following time points after transfection: T0h, T24h, T48h, and T72h. The assay procedure and conditions are described in Section 2.6.5.2

5.4.4.7 Flow cytometric analysis of the cell cycle

On the fifth day after transfection, the cell cycle distribution was analysed by quantifying DNA content using a propidium iodide flow cytometry kit. The assay procedure and conditions are described in Section 2.6.6.2.

5.4.4.8 Annexin V-based flow cytometric analysis of apoptosis

The detection of apoptosis was performed using an Annexin V detection kit on the fifth day after transfection. Section 2.6.7.2 details the assay procedure and conditions.

5.4.4.9 *In vitro* assessment of lymphatic and blood vessel invasion

5.4.4.9.1 Static adhesion assay

In a 24-well plate, endothelial cells (HUVECs and DLECs) were seeded to confluence. Tumour cell adhesion was determined after cells were labelled with 1 µM Cell Tracker Green CMFDA and incubated for 30 min at 37 °C. Following labelling, the tumour cells were resuspended in tumour cell medium supplemented with serum and incubated for 35 min at 37 °C with endothelial cell monolayers. Non-adherent cells were washed away with tumour cell medium, and adherent tumour cells were counted using a fluorescent microscope at a 10× magnification. The assay procedure and conditions are described in Section 2.6.8.2.

5.4.4.9.2 Transmigration assay

A confluent endothelial cell monolayer was grown on hanging transwell inserts. Tumour cell transmigration was determined following labelling with 1 µM Cell Tracker Green CMFDA. After 16 h, transmigration was observed using a fluorescent microscope by counting cells in the bottom of the chamber. Section 2.6.8.3 contains more information regarding the assay procedure.

5.5 Results

5.5.1 Validation of HMGB3 antibody by WB

WB showed specific bands with anticipated molecular weights of ~30 kDa that corresponded to the HMGB3 protein (Figure 5.1).

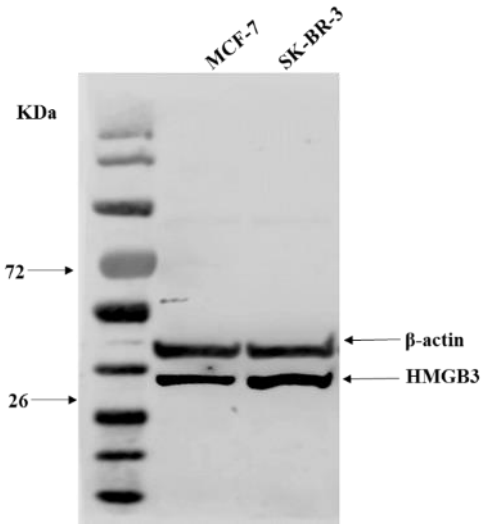


Figure 5.1 Western blot (WB) reflecting the antibody specificity of HMGB3. β-actin was used as a positive control.

5.5.2 Expression of *HMGB3* in BC

The METABRIC and TCGA cohorts were used to assess *HMGB3* mRNA expression. The METABRIC cohort data on *HMGB3* mRNA expression were normally distributed and dichotomised using mean-based cut-off points. The data on *HMGB3* mRNA expression in the TCGA cohort were right-skewed and were categorised using cut-off values derived from the median (Figure 5.2A/B). High *HMGB3* mRNA expression was observed in 901/1,980 (46%) of the METABRIC BC cases and 427/854 (50%) of the TCGA cases.

HMGB3 protein expression ranged from negative to strong in the nucleus of invasive BC cells (Figure 5.3C/D). A total of 1,647 cases were valid for evaluating HMGB3 protein levels. The data on HMGB3 protein expression showed a multimodal distribution, and therefore the median was used to define the cut-off value for HMGB3 positivity (Figure 5.2C). High HMGB3 protein levels (H-score >50) were observed in 787/1,647 (47.8%) invasive BC cases. There was a high degree of concordance between the TMAs scored by both scorers (Abrar Aljohani and Sara Abuelmaaty) in HMGB3 immunoscoring (ICC=0.844, $p<0.0001$).

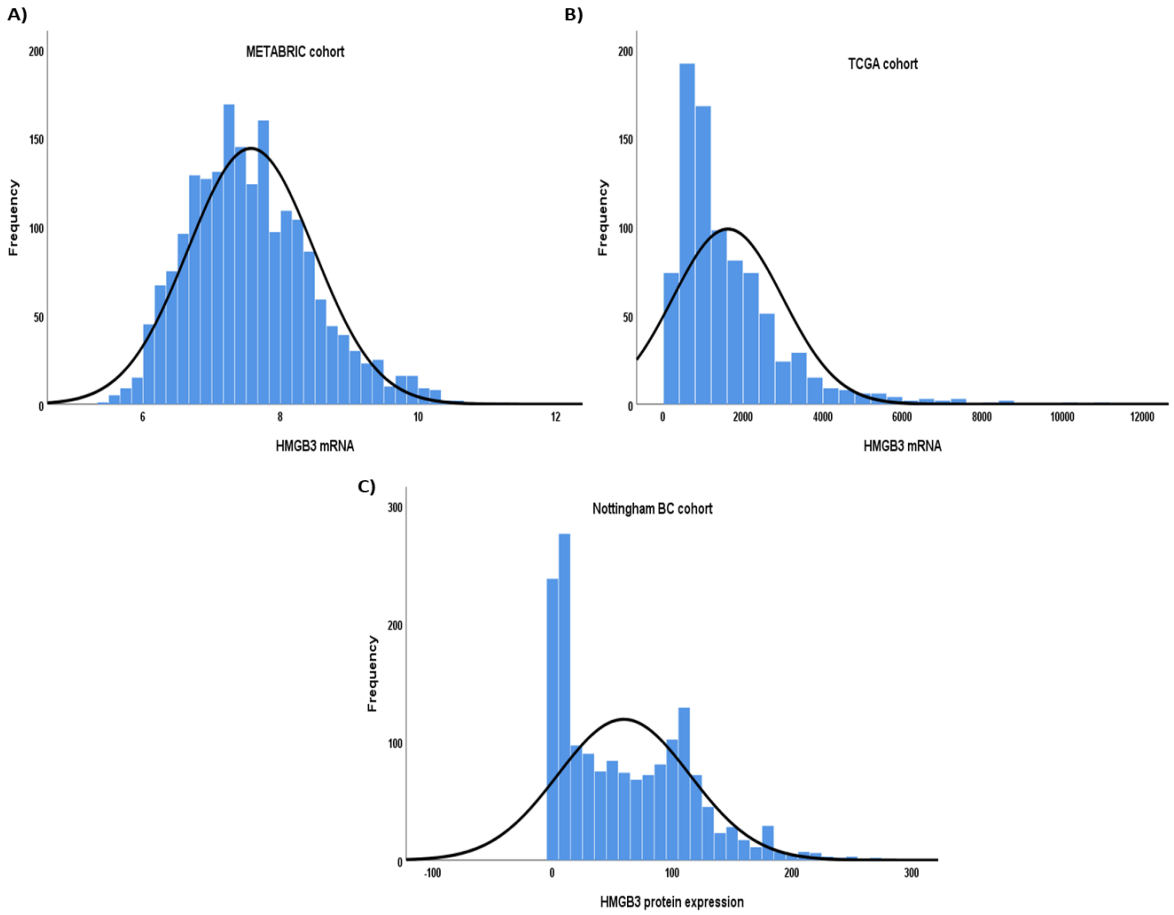


Figure 5.2 The distribution of the expression of A) *HMGB3* mRNA in the METABRIC cohort, B) *HMGB3* mRNA in the TCGA cohort, and C) HMGB3 protein in the Nottingham BC cohort.

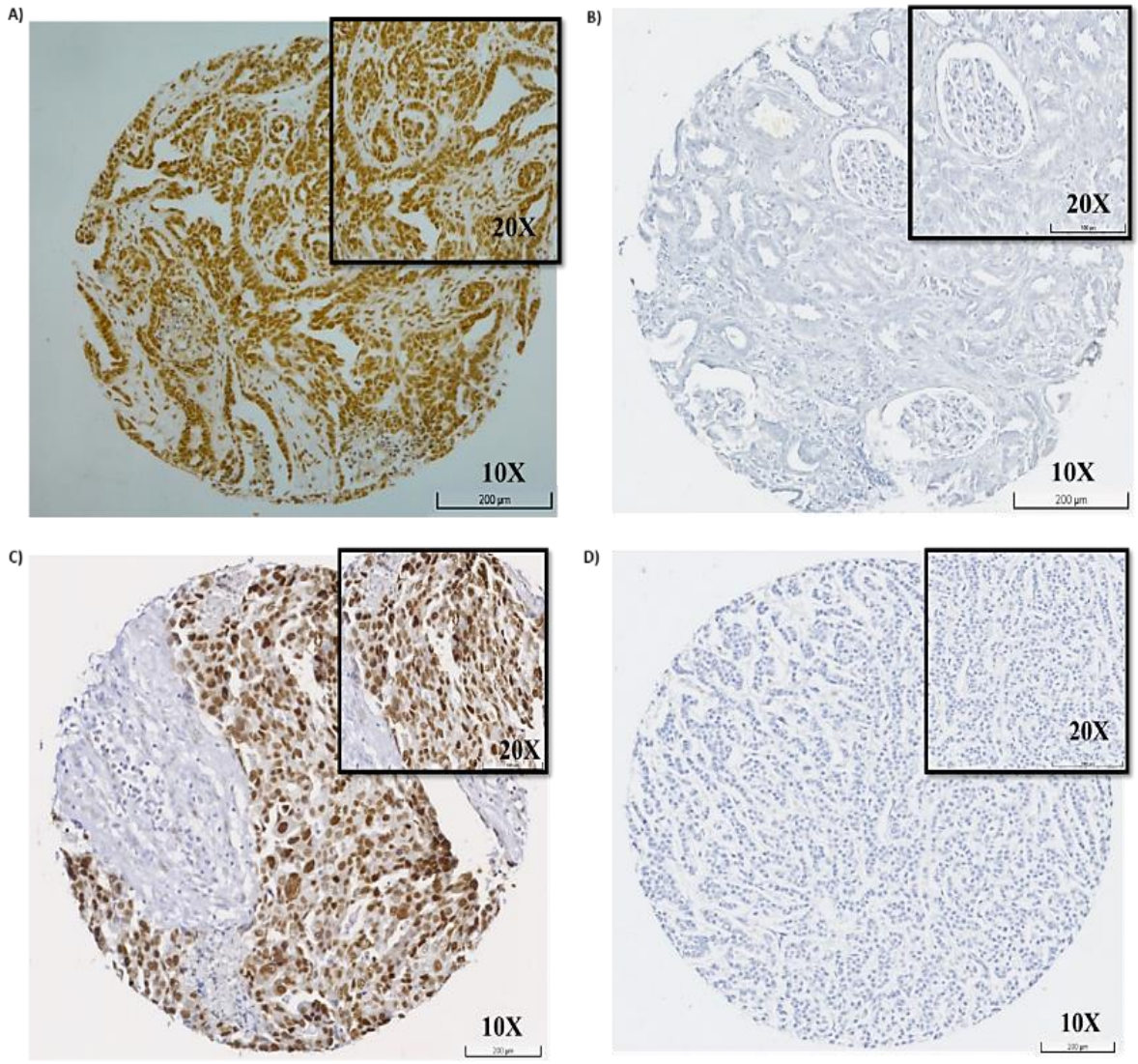


Figure 5.3 Nuclear expression of HMGB3 protein in invasive breast cancer. A) Positive control ovarian tissue stained for HMGB3, B) negative control normal kidney stained for HMGB3, C) HMGB3-positive IHC expression, and D) HMGB3-negative IHC expression. Magnification 10×, scale bars = 200 µm. Inset: magnification 20×, scale bars = 100 µm.

5.5.3 Association of *HMGB3* with clinicopathological parameters

5.5.3.1 *HMGB3* mRNA expression

In both transcriptomic datasets, high expression of *HMGB3* mRNA was significantly associated with LVI positivity and high tumour grade (all $p<0.0001$). In the METABRIC cohort, high expression of *HMGB3* mRNA was significantly correlated with larger tumour size ($p=0.001$), poor NPI ($p<0.0001$), and LN stage ($p=0.048$) (Table 5.1).

5.5.3.2 HMGB3 protein expression

Similar to the transcriptomics results, high HMGB3 protein expression was significantly associated with aggressive tumour features, including LVI positivity, younger age, higher histological grade, poor NPI, high mitotic count (all $p<0.0001$), high pleomorphism ($p=0.005$), high tubule formation ($p=0.001$), and LN stage ($p=0.049$) (Table 5.2). A weak correlation between HMGB3 protein and mRNA expression was observed in the Nottingham subset ($n=288$) of the METABRIC cohort based on Spearman's rank correlation coefficient ($r=0.200$, $p=0.016$).

5.5.4 *HMGB3* expression in molecular BC subtypes

In both METABRIC and TCGA datasets, ER^-, PR^-, and $HER2^+$ tumours all demonstrated significantly high expression of *HMGB3* mRNA (all $p<0.0001$) (Table 5.1). Analysis of the METABRIC cohort regarding the intrinsic (PAM50) subtypes showed that high *HMGB3* mRNA expression was associated with $HER2^+$, luminal B, basal-like, luminal A, and normal-like subtypes, in descending order (all $p<0.0001$) (Table 5.1). The correlation of HMGB3 protein level with IHC subtypes was comparable to previous findings, with HMGB3 protein expression being highest in $HER2^+$ tumours, followed by TN tumours. Within the $ER^+/HER2^-$ group, the high-proliferation class had considerably higher HMGB3 expression levels than the low-proliferation class ($p<0.0001$) (Table 5.2).

Table 5.1 Statistical associations between *HMGB3* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets**.**

Parameters	***HMGB3* mRNA (METABRIC)**			***HMGB3* mRNA (TCGA)**		
	Number (%)	**Mean rank**	***p*-value**	**Number (%)**	**Mean rank**	***p*-value**
Patient age (years)						
≤50	424 (21.4)	7.58	0.695	231 (27)	435.0	0.587
>50	1556 (78.6)	7.56		623 (73)	424.1	
Tumour size						
≤2 cm	622 (31.7)	7.42	**0.001**	239 (28)	411.8	0.245
>2 cm	1338 (68.3)	7.62		615 (72)	433.6	
Tumour grade						
I	170 (9.0)	7.09		89 (11)	267.2	
II	770 (40.6)	7.43	**<0.0001**	375 (46)	338.9	**<0.0001**
III	952 (50.3)	7.77		352 (43)	518.4	
Nottingham prognostic index (NPI)						
Good	680 (34.3)	7.35	**<0.0001**	Not available		
Moderate	1101(55.6)	7.67				
Poor	199 (10.1)	7.78				
Lymph node stage						
I	1035 (52.5)	7.51	**0.048**	Not available		
II	622 (31.5)	7.63				
III	316 (16.0)	7.62				
Lymphovascular invasion (LVI)						
Negative	930 (59)	7.46	**<0.0001**	559 (65)	403.1	**<0.0001**
Positive	635 (41)	7.65		295 (35)	437.7	
Oestrogen receptor (ER)						
Negative	474 (23.9)	7.82	**<0.0001**	185 (22)	580.9	**<0.0001**
Positive	1506 (76.1)	7.49		639 (78)	363.7	
Progesterone receptor (PR)						
Negative	940 (47.4)	7.65	**<0.0001**	272 (33)	510.4	**<0.0001**
Positive	1040 (52.6)	7.49		546 (67)	359.2	
Human epidermal growth factor receptor 2 (HER2)						
Negative	1733 (87.5)	7.48	**<0.0001**	567 (81)	339.9	**0.004**
Positive	247 (12.5)	8.18		133 (19)	395.6	
PAM50 subtype						
Luminal A	718 (36.4)	7.31				
Luminal B	488 (24.7)	7.81	**<0.0001**	Not available		
HER2^{+}-enriched	240 (12.1)	8.15				
Basal-like	329 (16.7)	7.65				
Normal-like	199 (10.1)	7.04				

***p*-values in bold are statistically significant**

Table 5.2 Statistical associations between HMGB3 protein expression and clinicopathological factors in the Nottingham BC cohort (n=1,647).

Nottingham BC cohort			
Parameters	**HMGB3 protein**		
	Number (%)	Mean rank	*p*-value
Patient age (years) ≤50 >50	 574 (35) 1070 (65)	 882.3 790.4	**<0.0001**
Tumour size ≤2 cm >2 cm	 914 (56) 725 (44)	 815.5 825.6	0.668
Tumour grade I II III	 206 (13) 584 (52) 854 (35)	 697.9 779.4 882.0	**<0.0001**
Mitosis I II III	 622 (38.3) 312 (19.2) 686 (42.5)	 715.66 838.60 883.71	**<0.0001**
Pleomorphism I II III	 21 (1.3) 465 (28.7) 1135 (70)	 674.48 759.01 834.83	**0.005**
Tubule formation I II III	 83 (5.1) 472 (29.1) 1066 (65.8)	 619.11 814.69 824.31	**0.001**
Lymphovascular invasion (LVI) Negative Positive	 1096 (67) 541 (33)	 790.1 877.6	**<0.0001**
Lymph node stage I II III	 975 (60) 500 (31) 167 (9)	 799.5 843.9 882.7	**0.049**
Nottingham prognostic index (NPI) Good Moderate Poor	 462 (28) 880 (54) 295 (18)	 733.2 853.1 851.8	**<0.0001**
Oestrogen receptor (ER) Negative Positive	 386 (23) 1262 (77)	 895.1 802.9	**0.001**
Progesterone receptor (PR) Negative Positive	 673 (41) 951 (59)	 834.7 796.8	0.108
Human epidermal growth factor receptor 2 (HER2) Negative Positive	 1402 (86) 225 (14)	 784.5 997.9	**<0.0001**
Triple-negative No Yes	 1359 (83) 276 (17)	 811.1 852.2	0.185
Immunohistochemistry subtype $ER^+/HER2^-$ low proliferation $ER^+/HER2^-$ high proliferation	 528 (34) 522 (33.7)	 655.6 802.2	**<0.0001**

Triple-negative HER2$^+$	275 (17.7) 225 (14.6)	809.2 953.5	

***p*-values in bold are statistically significant**

5.5.5 *HMGB3* expression and other associated biomarkers

To further assess the role of *HMGB3* in BC and its relationships with other genes involved in LVI processes, the METABRIC and TCGA datasets were analysed for correlations between *HMGB3* mRNA and other genes related to invasion, EMT, and adhesion. A significant weak to moderate positive linear correlation was observed between *HMGB3* and *N-cadherin*, *P-cadherin*, *GSK3B, TWIST1, TWIST2, ZEB1, ZEB2, NFKB1*, and *CTNNB1* while the correlation was negative with *E-cadherin* (all $p<0.05$). Moreover, there was a significant weak to moderate positive linear correlation between *HMGB3* and MMPs, including *MMP1*, *MMP7*, *MMP9*, *MMP12, MMP15,* and *MMP25* (all $p< 0.05$) (Table 5.3).

The correlation between HMGB3 and EMT-related markers, such as E-cadherin, N-cadherin, and P-cadherin was further studied at the protein level using the Nottingham BC cohort. There was a significant weak negative correlation between HMGB3 and E-cadherin ($p=0.001$), while the correlation was positive with N-cadherin ($p=0.023$), and P-cadherin ($p=0.002$) (Table 5.3).

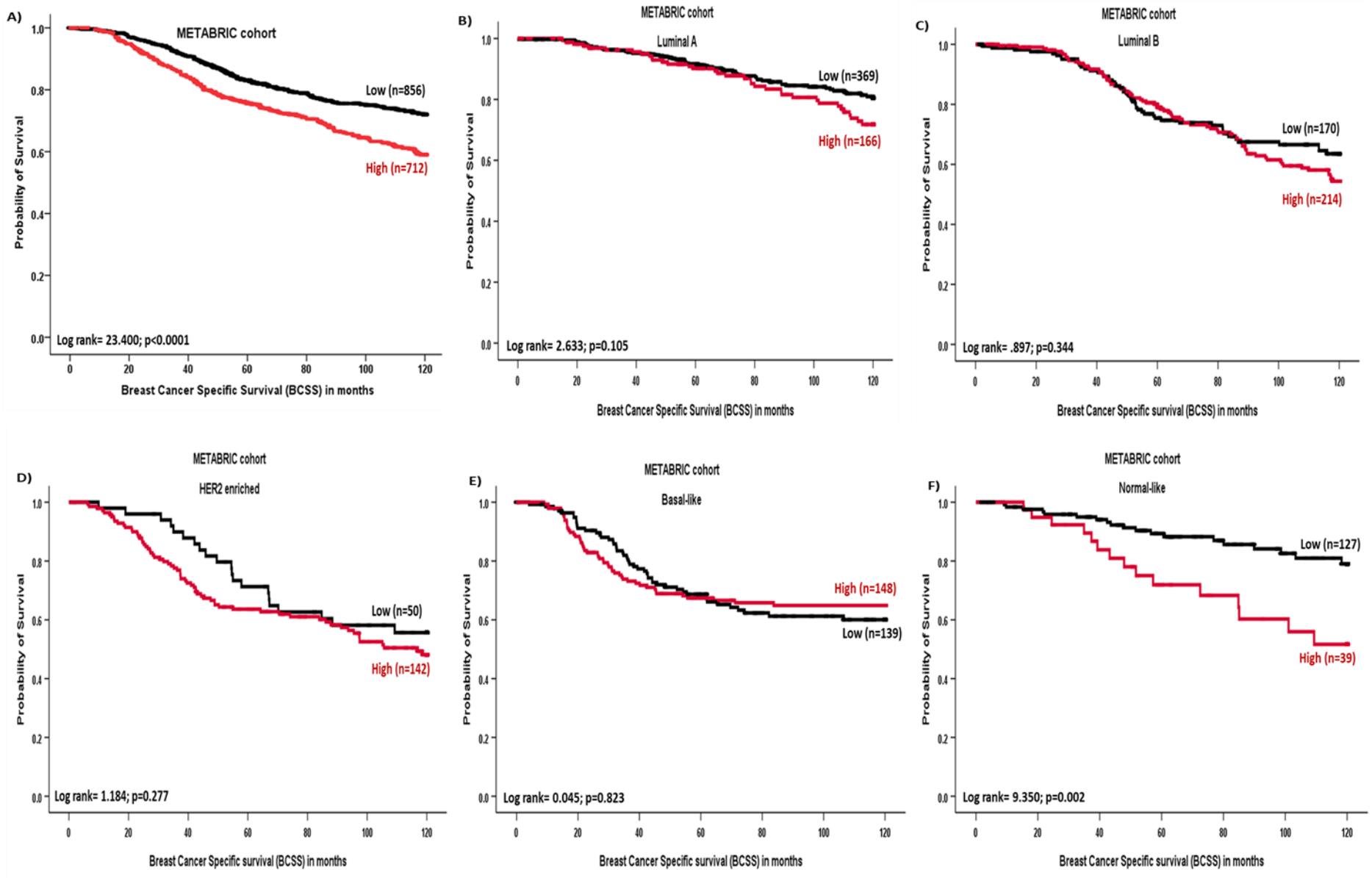
Table 5.3 Correlations between HMGB3 expression and mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes.

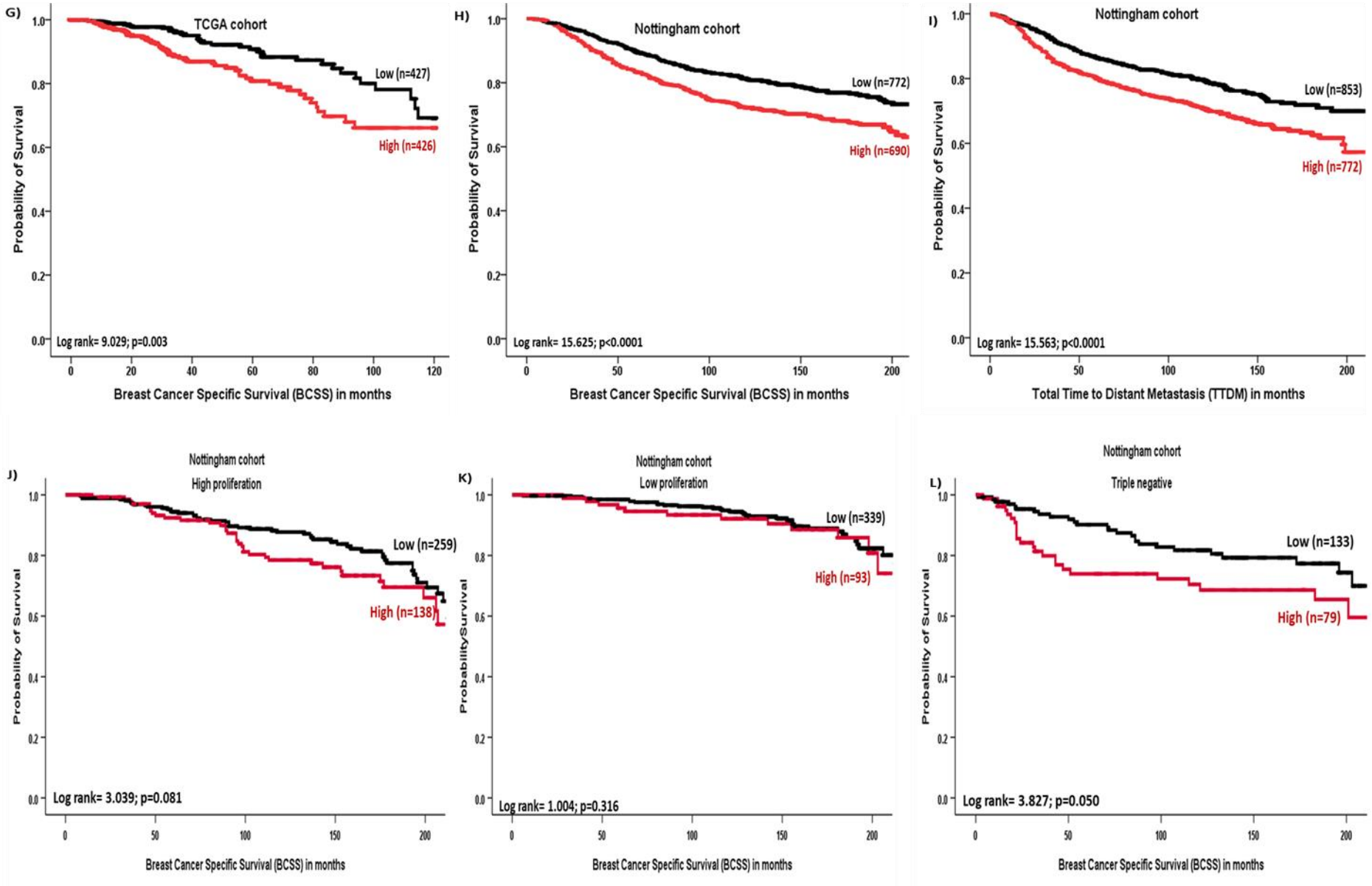
Gene names	**METABRIC cohort (mRNA)**		**TCGA cohort (mRNA)**		**Nottingham cohort (protein)**	
	Correlation value	*p*-value	Correlation value	*p*-value	Correlation value	*p*-value
EMT-related genes						
E-cadherin	- 0.252	**<0.0001**	- 0.197	**<0.0001**	- 0.128	**0.001**
N-cadherin	0.118	**<0.0001**	0.169	**<0.0001**	0.099	**0.023**
P-cadherin	0.064	**0.005**	0.196	**<0.0001**	0.125	**0.002**
TGFβ1	0.001	0.947	0.216	**<0.0001**	0.052	0.827
TWIST1	0.149	**<0.0001**	0.149	**<0.0001**	**Not available**	
TWIST2	0.282	**<0.0001**	0.217	**<0.0001**	0.062	0.186
ZEB1	0.202	**<0.0001**	0.347	**<0.0001**	**Not available**	
ZEB2	0.231	**<0.0001**	0.197	**<0.0001**		
NFKB1	0.114	**<0.0001**	0.150	**<0.0001**		
GSK3B	0.286	**<0.0001**	0.121	**<0.0001**		
CTNNB1	0.138	**<0.0001**	0.074	**0.032**		
MMP-related genes						
MMP1	0.188	**<0.0001**	0.382	**<0.0001**	**Not available**	
MMP7	0.056	**0.013**	0.104	**0.002**		
MMP9	0.133	**<0.0001**	0.184	**<0.0001**		
MMP11	0.138	**<0.0001**	0.016	0.640		
MMP12	0.137	**<0.0001**	0.273	**<0.0001**		
MMP15	0.315	**<0.0001**	0.214	**<0.0001**		
MMP20	0.023	0.299	0.031	0.367		
MMP25	0.051	**0.023**	0.069	**0.043**		

***p*-values in bold are statistically significant**

5.5.6 Association of *HMGB3* with patient outcomes

Survival analysis of *HMGB3* mRNA showed that high expression was associated with shorter BCSS in both the METABRIC and TCGA cohorts ($p<0.0001$ and $p=0.003$, respectively) (Figure 5.4A/G). At the protein level, there was a significant association between the high expression of HMGB3 and shorter BCSS and TTDM (both $p<0.0001$) (Figure 5.4H/I). When the analysis was performed on intrinsic molecular classes in the METABRIC dataset, high expression of *HMGB3* was associated with shorter BCSS in the normal-like subtype ($p=0.002$) (Figure 5.4F). No significant association of *HMGB3* mRNA and protein levels with the other subtypes was observed (all $p>0.05$) (Figure 5.4B–E/J–M). In all cohorts, multivariate Cox regression analysis showed that *HMGB3* mRNA and protein levels predicted poor BCSS survival independent of the tumour size, LN stage, tumour grade, and LVI (HR 1.4; 95%CI=1.1-1.7*;* $p=0.004$ for METABRIC cohort, HR 1.7; 95%CI=1.1-2.8*;* $p=0.028$ for TCGA cohort, and HR 1.4; 95%CI=1.1-1.7*;* $p=0.003$ for Nottingham cohort) (Table 5.4).


A)
METABRIC cohort
Low (n=856)
High (n=712)
Probability of Survival
Log rank= 23.400; p<0.0001
Breast Cancer Specific Survival (BCSS) in months
B)
METABRIC cohort
Luminal A
Low (n=369)
High (n=166)
Probability of Survival
Log rank= 2.633; p=0.105
Breast Cancer Specific Survival (BCSS) in months
C)
METABRIC cohort
Luminal B
Low (n=170)
High (n=214)
Probability of Survival
Log rank= .897; p=0.344
Breast Cancer Specific survival (BCSS) in months
D)
METABRIC cohort
HER2 enriched
Low (n=50)
High (n=142)
Probability of Survival
Log rank= 1.184; p=0.277
Breast Cancer Specific survival (BCSS) in months
E)
METABRIC cohort
Basal-like
High (n=148)
Low (n=139)
Probability of Survival
Log rank= 0.045; p=0.823
Breast Cancer Specific survival (BCSS) in months
F)
METABRIC cohort
Normal-like
Low (n=127)
High (n=39)
Probability of Survival
Log rank= 9.350; p=0.002
Breast Cancer Specific survival (BCSS) in months


G)
TCGA cohort
Low (n=427)
High (n=426)
Probability of Survival
Log rank= 9.029; p=0.003
Breast Cancer Specific Survival (BCSS) in months
H)
Nottingham cohort
Low (n=772)
High (n=690)
Probability of Survival
Log rank= 15.625; p<0.0001
Breast Cancer Specific Survival (BCSS) in months
I)
Nottingham cohort
Low (n=853)
High (n=772)
Probability of Survival
Log rank= 15.563; p<0.0001
Total Time to Distant Metastasis (TTDM) in months
J)
Nottingham cohort
High proliferation
Low (n=259)
High (n=138)
Probability of Survival
Log rank= 3.039; p=0.081
Breast Cancer Specific Survival (BCSS) in months
K)
Nottingham cohort
Low proliferation
Low (n=339)
High (n=93)
ProbabilitySurvival
Log rank= 1.004; p=0.316
Breast Cancer Specific Survival (BCSS) in months
L)
Nottingham cohort
Triple negative
Low (n=133)
High (n=79)
Probability of Survival
Log rank= 3.827; p=0.050
Breast Cancer Specific Survival (BCSS) in months

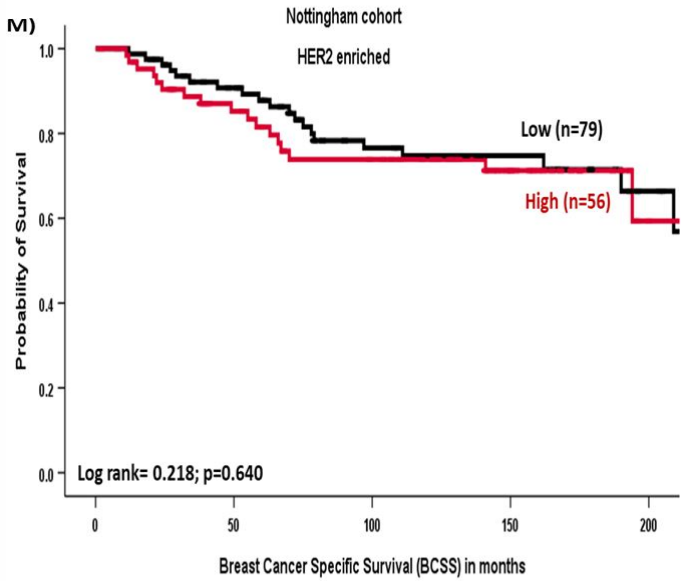


Figure 5.4 Kaplan–Meier survival plots showing the association between *HMGB3* mRNA expression and breast cancer-specific survival (BCSS) in A) the whole METABRIC cohort, B) luminal A (METABRIC), C) luminal B (METABRIC), D) HER2-enriched (METABRIC), E) basal-like (METABRIC), F) normal-like (METABRIC), and G) the whole TCGA cohort. For Nottingham cohort data, associations are shown between HMGB3 protein expression and H) BCSS, I) total time to distant metastasis (TTDM), J) BCSS in high-proliferation tumours, K) BCSS in low-proliferation tumours, L) BCSS in triple-negative tumours, and M) BCSS in HER2-enriched tumours.

Table 5.4 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *HMGB3* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort.

METABRIC cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
***HMGB3* mRNA expression**	1.391	1.113	1.739	**0.004**
Tumour size	1.474	1.126	1.928	**0.005**
Lymph node stage	1.969	1.538	2.520	**<0.0001**
Tumour grade	1.456	1.192	1.779	**<0.0001**
Lymphovascular invasion (LVI)	1.469	1.159	1.862	**0.001**
TCGA cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
***HMGB3* mRNA expression**	1.728	1.061	2.816	**0.028**
Tumour size	1.523	0.873	2.656	0.138
Lymph node	1.196	0.712	2.010	0.498
Tumour grade	1.092	0.761	1.567	0.634
Lymphovascular invasion (LVI)	1.878	1.149	3.070	**0.012**
Nottingham BC cohort (protein)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
HMGB3 protein expression	1.350	1.104	1.650	**0.003**
Tumour size	1.383	1.116	1.714	**0.003**
Lymph node stage	1.708	1.478	1.973	**<0.0001**
Tumour grade	1.649	1.375	1.977	**<0.0001**
Lymphovascular invasion (LVI)	1.620	1.307	2.009	**<0.0001**

***p*-values in bold are statistically significant**

5.5.7 Preclinical studies (*in vitro* investigation of the mechanistic role of *HMGB3* in LVI-related biological processes)

5.5.7.1 Silencing of *HMGB3* expression using siRNA

The expression of the HMGB3 protein was determined in a variety of BC cell lines, including luminal A-like (MCF-7), luminal B-like (ZR-75-1), HER2-enriched (SK-BR-3), and triple-negative (MDA-MB-231) (Figure 5.5A). HMGB3 expression varied among these cell lines, with the highest levels seen in SK-BR-3, ZR-75-1, and MCF-7, and the lowest levels in MDA-MB-231 (Figure 5.5B). Depending on the level of protein expression and the IHC results, the following BC cell lines were employed in this study: MCF-7 ($ER^+/PR^+/HER2^-$), ZR-75-1 ($ER^+/PR^{\pm}/HER2^-$), and SK-BR-3 ($HER2^+$). When *HMGB3* knockdown expression was compared to β-actin expression in MCF-7, ZR-75-1, and SK-BR-3 cells at post-transfection days 3, 5, and 7, complete reduction of HMGB3 protein expression was found using WB analysis (Figure 5.5C–H).

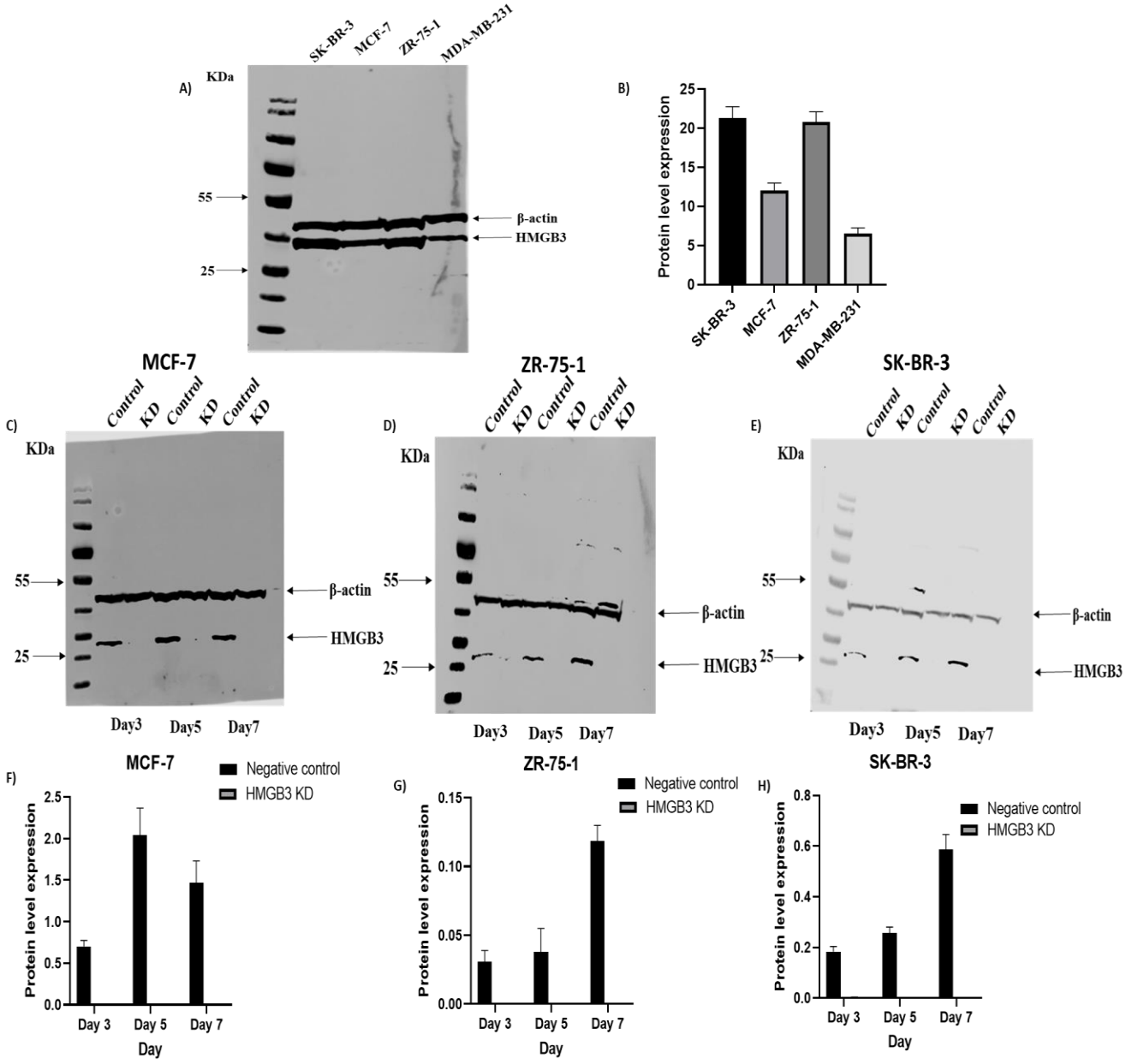


Figure 5.5 HMGB3 protein expression in breast cancer (BC) cell lines. A) Evaluation of the differential expression of HMGB3 in BC cell lines by Western blotting and B) quantification of HMGB3 protein expression level in SK-BR-3, MCF-7, ZR-75-1, and MDA-MB-231 cell lines. Using pre-validated Silencer Select siRNA constructs (mainly for *HMGB3*) or scrambled negative-control siRNA, silencing of *HMGB3* using the forward transfection method showed complete knockdown relative to non-targeting, scrambled control siRNA in C) MCF-7, D) ZR-75-1, and E) SK-BR-3. Quantification of HMGB3 protein expression in the transfected cells and negative control by densitometry and normalisation to β-actin levels revealed complete loss of HMGB3 protein expression in F) MCF-7, G) ZR-75-1, and H) SK-BR-3 cell lines. Each bar represents the mean ± standard error of the mean (SEM) of three or more independent experiments. β-actin was included as a loading control.

5.5.7.2 *HMGB3* silencing suppresses cell proliferation and cell growth among the selected BC cell lines

Deletion of *HMGB3* resulted in a considerable decrease in proliferation rate in MCF-7 ($p=0.0120$ at T24h, $p=0.0247$ at T48h, and $p=0.0160$ at T72h), ZR-75-1 ($p=0.8382$ at T24h, $p=0.0044$ at T48h, and $p=0.0117$ at T72h), and SK-BR-3 ($p=0.0027$ at T24h, $p<0.0001$ at T48h, and $p<0.0001$ at T72h), as compared with the control (Figure 5.6A–C). In addition, a clonogenic experiment was performed to examine how *HMGB3* knockdown affected cell growth and survival. The capacity of a single cell colony to survive after being transfected was much lower than the control in MCF-7 ($p<0.0001$), ZR-75-1 ($p=0.0022$), and SK-BR-3 ($p<0.0001$) (Figure 5.6D–F).

5.5.7.3 *HMGB3* silencing suppresses cell migration among the selected BC cell lines

The knockdown of *HMGB3* resulted in a significantly larger unhealed wound in comparison to the negative controls in MCF-7, ZR-75-1, and SK-BR-3 ($p<0.0001$, $p=0.0389$, and $p=0.0040$, respectively) (Figure 5.7A–C).

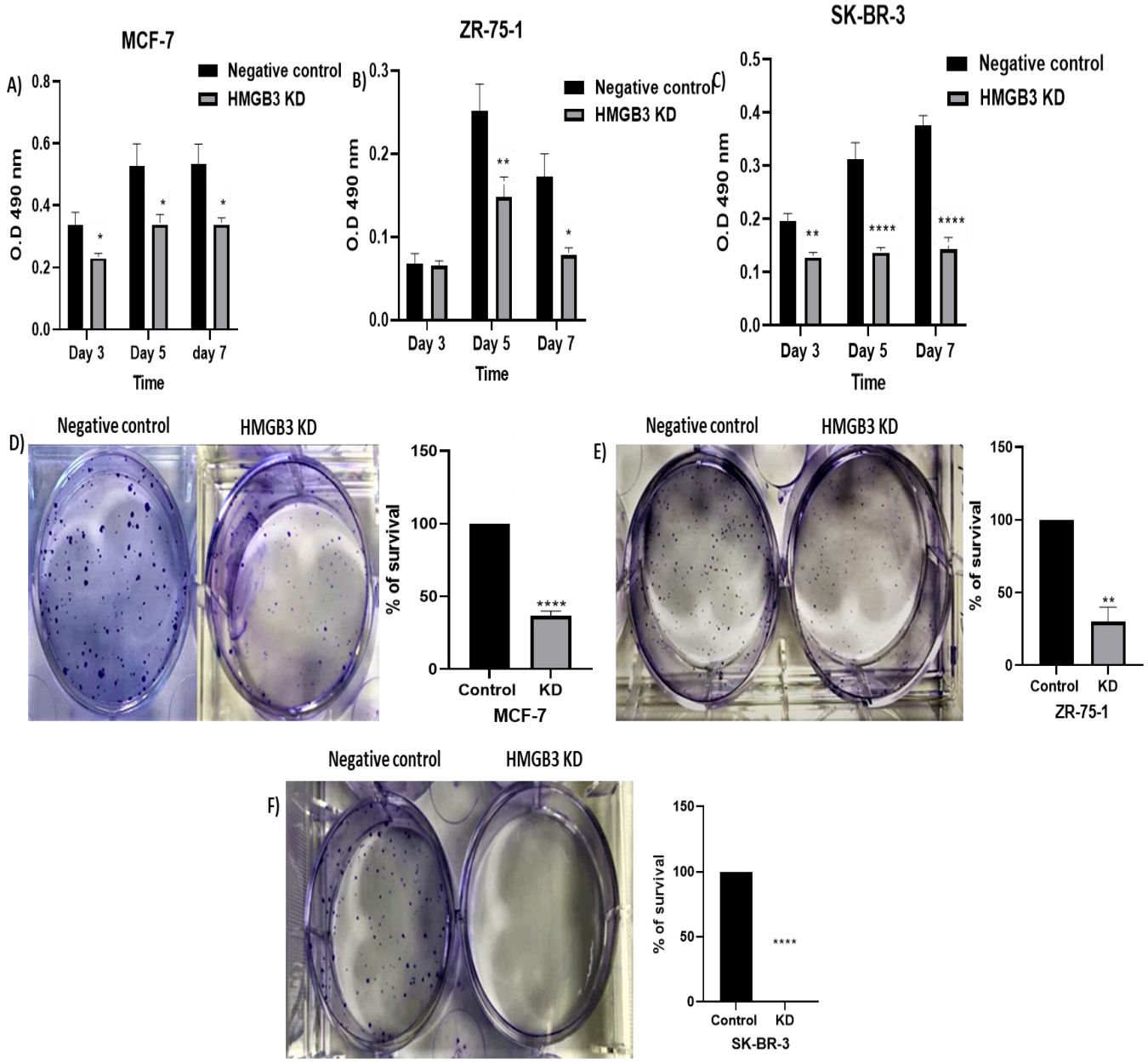


Figure 5.6 The effect of *HMGB3* knockdown (KD) by siRNA on cell proliferation in A) MCF-7, B) ZR-75-1, and C) SK-BR-3 cells. (A–C) Cell proliferation was significantly reduced after KD in BC cell lines, as detected by the MTS assay. *HMGB3* siRNA transfection reduced the colonisation capacity of BC cell lines in D) MCF-7, E) ZR-75-1, and F) SK-BR-3, as detected by the colony formation assay. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

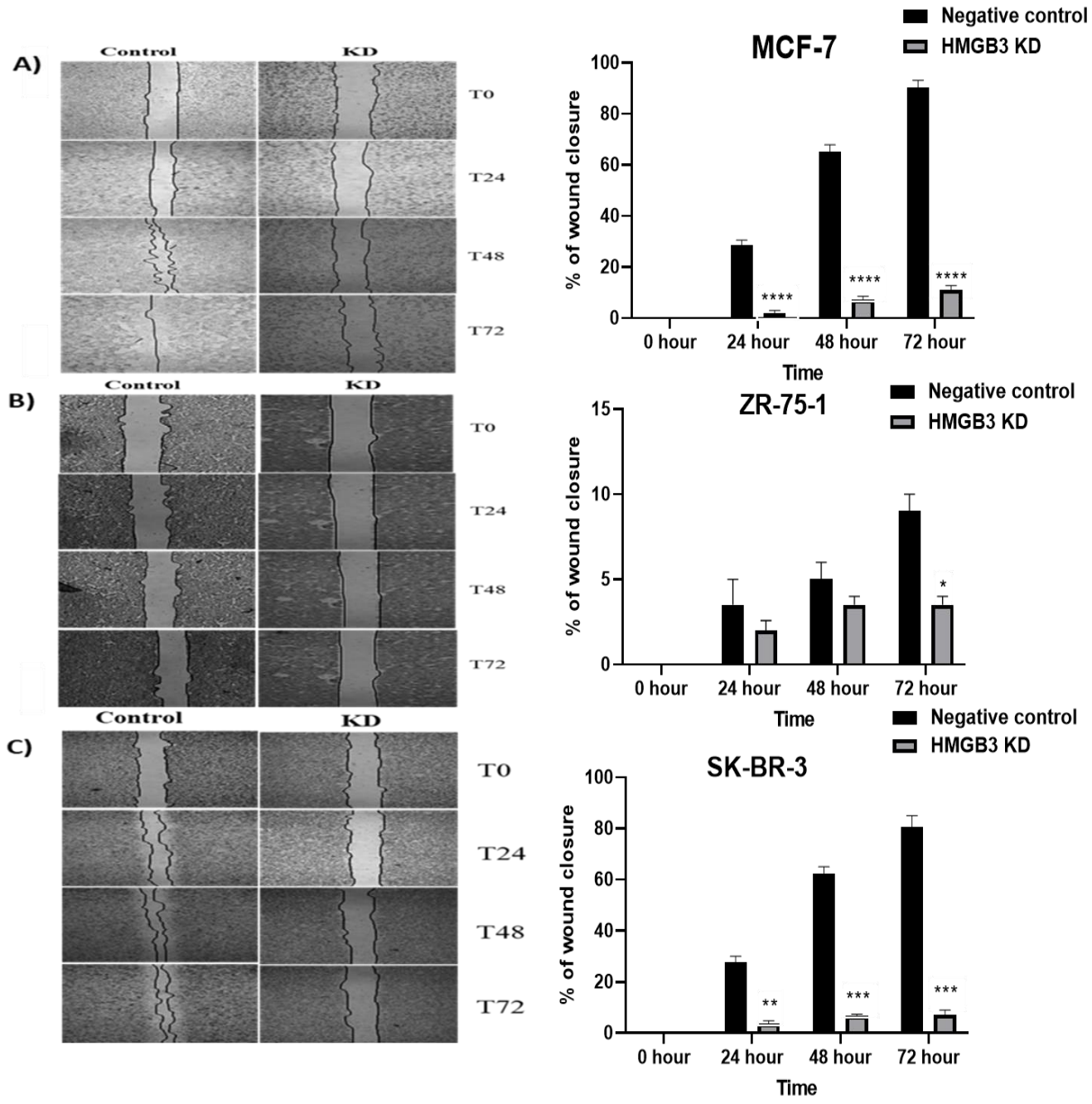


Figure 5.7 The effect of *HMGB3* knockdown (KD) by siRNA on cell migration in A) MCF-7, B) ZR-75-1, and C) SK-BR-3 cells. The wound repair rate of *HMGB3* KD and control cells was observed by measuring the width of the unhealed gap at T0h, T24h, T48h, and T72h. (A–C) Silencing *HMGB3* significantly reduced the migration rate in BC cell lines, as detected by the wound-healing assay. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

5.5.7.4 *HMGB3* silencing arrests cells in the G1 phase

HMGB3 knockdown triggered G1-phase arrest in MCF-7, ZR-75-1, and SK-BR-3 cells, preventing the cell cycle from progressing to the S phase. The percentage of MCF-7 cells with *HMGB3* knockdown in the G1 phase increased (63.6% G1, 14.35% S, and 27% G2/M) compared with the negative control (53.6% G1, 29.86% S, and 22.6% G2/M) ($p=0.0333$ in G1, $p=0.0112$ in S, and $p=0.0106$ in G2/M). Similar observations were made in ZR-75-1 (73.8% G1, 12.5% S, and 29.65% G2/M) compared with the control (45.95% G1, 22.5% S, and 14.8% G2/M) ($p=0.0278$ in G1, $p=0.0430$ in S, and $p=0.0359$ in G2/M). Similarly, the percentage of SK-BR-3 cells in the G1 phase significantly increased in the *HMGB3* knockdown cells (65.45% G1, 18.3% S, and 27% G2/M) compared with the negative control (40% G1, 27.46% S, and 35.3% G2/M) ($p=0.0118$ in G1, $p=0.0062$ in S, and $p=0.0372$ in G2/M) (Figure 5.8A–C).

5.5.7.5 *HMGB3* silencing enhances cell apoptosis

In both early and late apoptosis, the number of apoptotic cells in the *HMGB3* knockdown population rose dramatically when compared with the control. In early apoptosis, the knockdown vs control values were 3.40% vs 2.36% in MCF-7 ($p = 0.0223$), 0.62% vs 0.32% in ZR-75-1 ($p = 0.2457$), and 1.10% vs 0.58% in SK-BR-3 ($p = 0.0424$), respectively. In late apoptosis, the knockdown vs control values were 2.10% vs 1.76% in MCF-7 ($p = 0.0457$), 5.92% vs 3.90% in ZR-75-1 ($p = 0.0202$), and 6.92% vs 1.08% in SK-BR-3 ($p=0.0163$), respectively (Figure 5.9A–C).

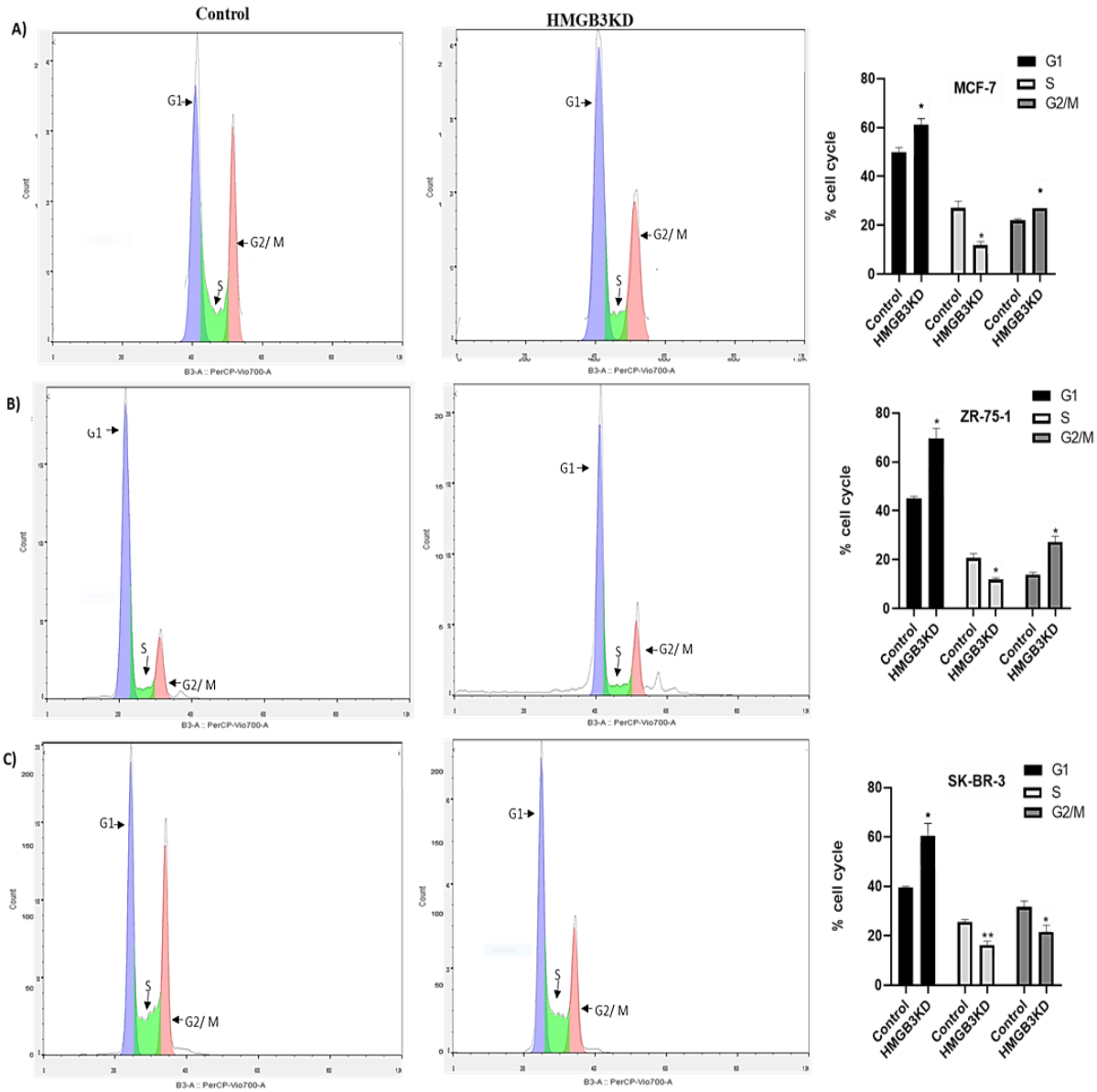


Figure 5.8 The effect of siRNA knockdown (KD) of *HMGB3* on the cell cycle. BC cells showed arrest in the G1 phase of the cell cycle in A) MCF-7, B) ZR-75-1, and C) SK-BR-3 cells. (A–C) Silencing *HMGB3* prevented all BC cell lines from entering the S phase, as detected by flow cytometry. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

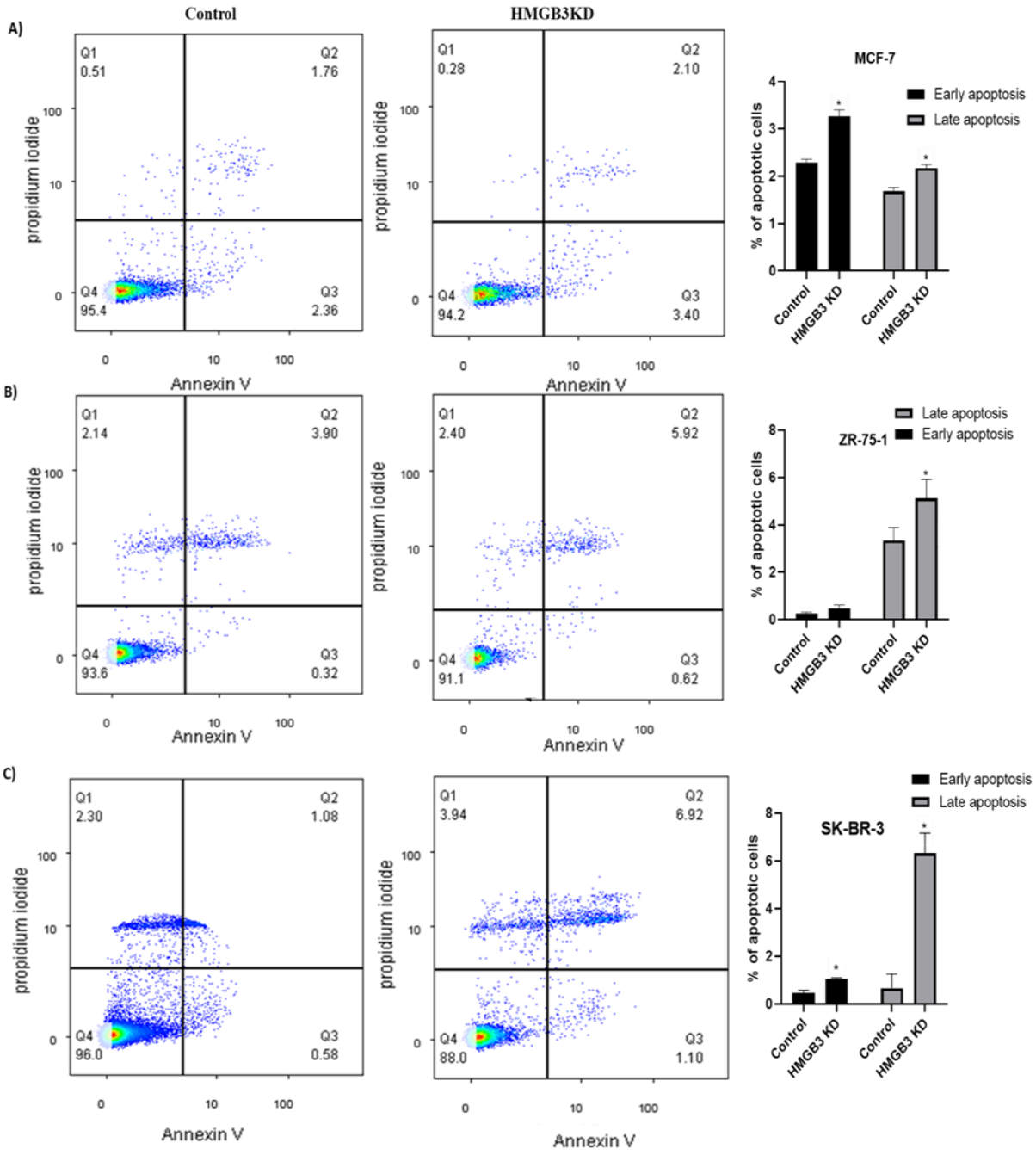


Figure 5.9 The effect of siRNA knockdown (KD) of *HMGB3* on apoptosis. BC cell data showed that KD of *HMGB3* enhanced apoptosis in A) MCF-7, B) ZR-75-1, and C) SK-BR-3 cells. (A–C) More apoptotic cells were observed in the transfected cells than in the control ones, as detected by flow cytometry. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

5.5.7.6 The role of *HMGB3* in the adhesion of BC cells to endothelial cells

Higher adherence to HUVECs in the *HMGB3*-untransfected cells than HMGB3-transfected cells was observed in cell lines MCF-7 (p=0.0002) and SK-BR-3 (p=0.0068), but not ZR-75-1 (p=0.2722); higher adherence to DLECs than HMGB3-transfected cells was observed in cell lines MCF-7 (p=0.0205) and SK-BR-3 (p=0.0006), but not ZR-75-1 (p=0.2899) (Figure 5.10A–C).

5.5.8.7 The role of *HMGB3* in the transmigration of BC cells through endothelial cells

Tumour cell transmigration through HUVECs was higher in the control than in the knockdown group in cell lines MCF-7 (p=0.0208) and SK-BR-3 (p=0.0260), but not ZR-75-1 (p=0.0955); transmigration through DLECs was higher in the control than in the knockdown group in cell lines MCF-7 (p=0.0307) and SK-BR-3 (p=0.0006), but not ZR-75-1 (p=0.4226) (Figure 5.11A/B).

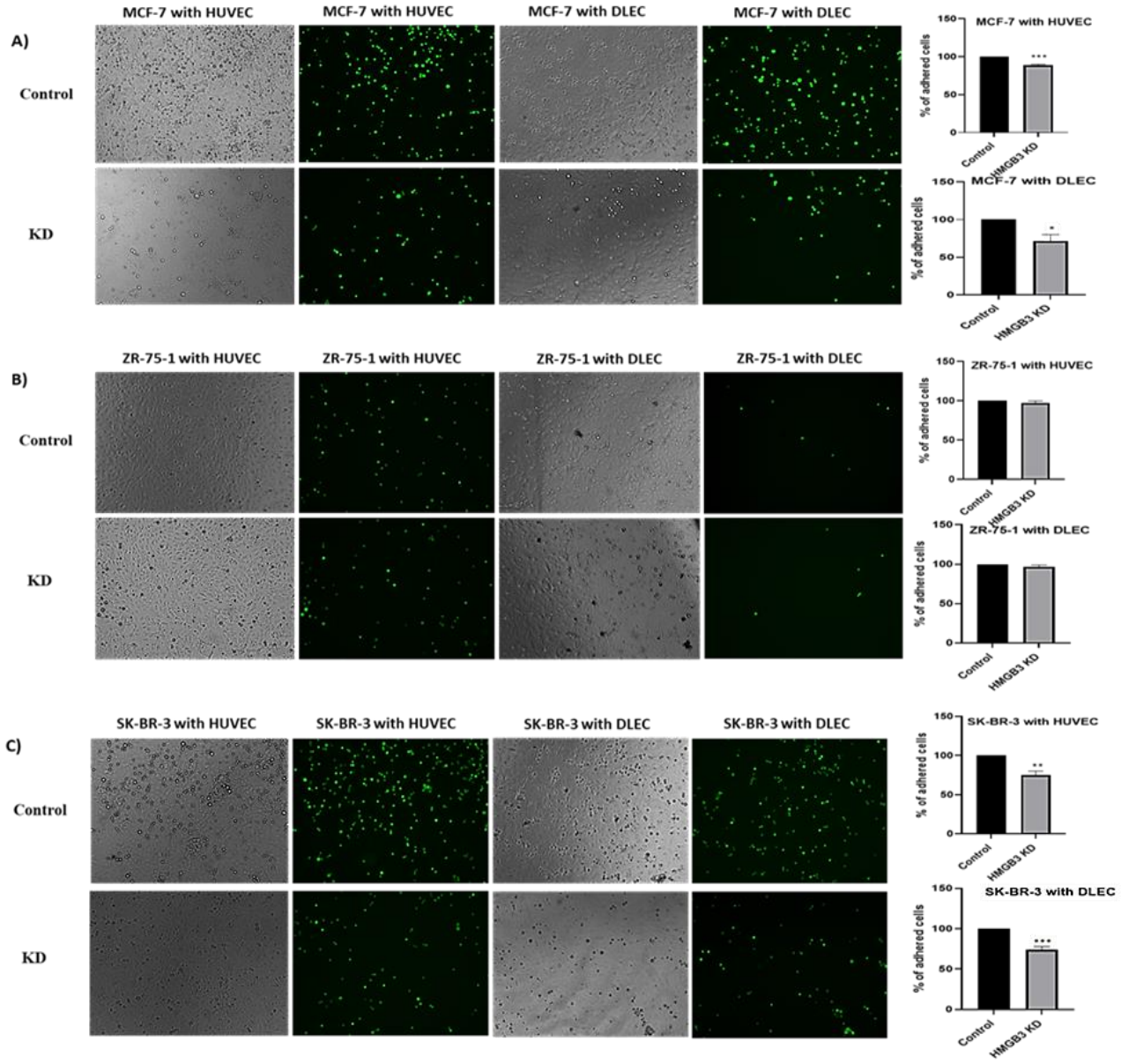


Figure 5.10 Representative photomicrographs of tumour cell adhesion across vascular and lymphatic endothelial cells (HUVECs and DLECs). A) MCF-7, B) ZR-75-1 and C) SK-BR-3 cells. (A-C) Silencing *HMGB3* decreased the number of BC cell lines that adhered to HUVECs and DLECs. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =<0.0001.

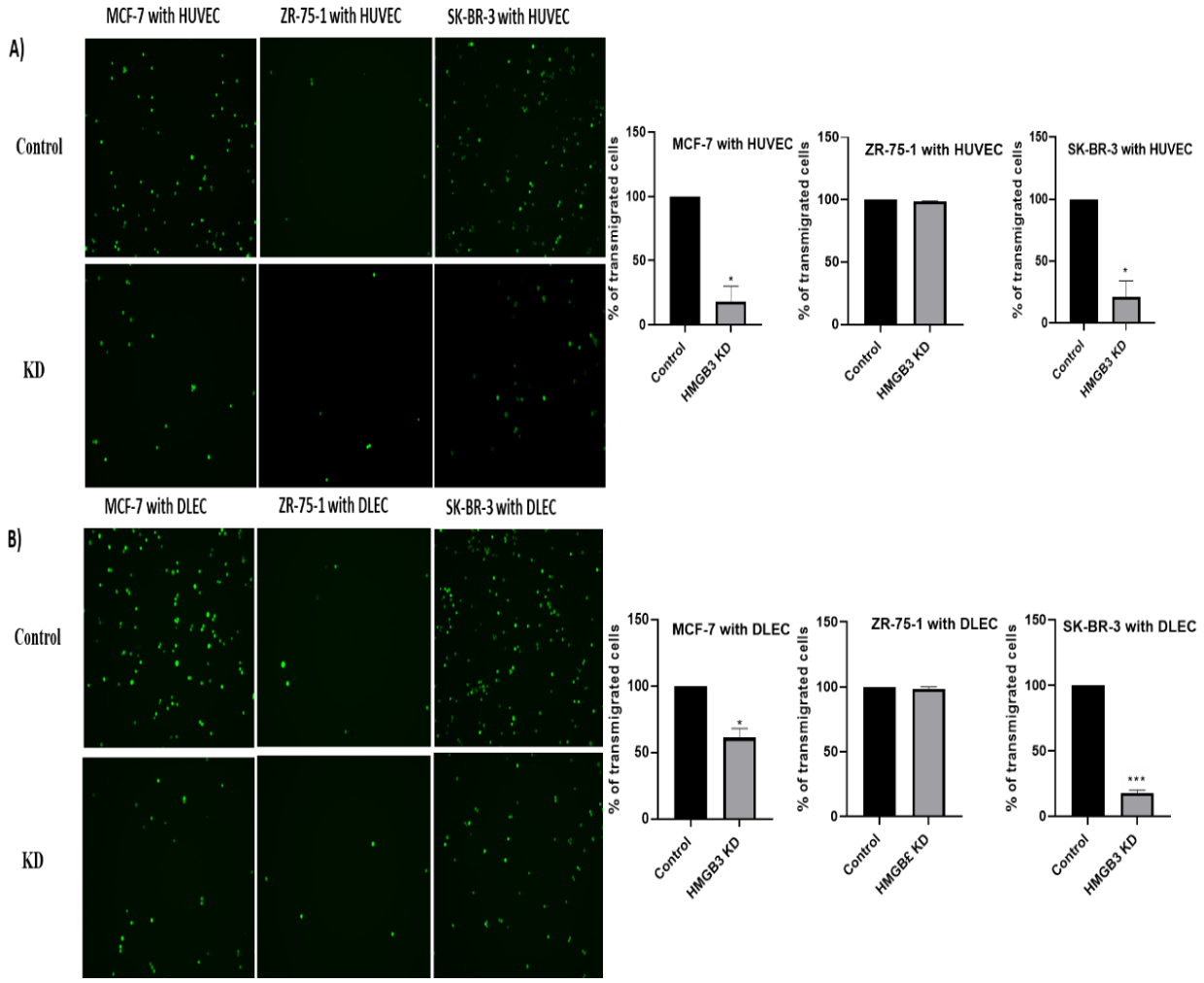


Figure 5.11 Representative photomicrographs of tumour cell (MCF-7, ZR-75-1 and SK-BR-3) transmigration across A) HUVECs and B) DLECs. (A/B) The number of tumour cells that transmigrated across HUVECs and DLECs was higher in the control group than in the KD group. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

5.6 Discussion

HMGB3 is a multifunctional protein that performs a variety of functions in many cellular compartments and has been identified as a critical regulator of tumour development [264]. Recently, aberrant *HMGB3* was identified as a pro-carcinogenes, promoting tumour development, proliferation, invasion, and metastasis in a variety of tumour types, including gastric [265], lung [266], colorectal [254], and urinary bladder [262]. Although *HMGB3* has previously been associated with cancer cell proliferation and migration [254], no studies have been performed to date to investigate the potential role of *HMGB3* in the development of LVI.

In Chapter 3, bioinformatic analysis determined that *HMGB3* was a gene associated with LVI positivity. Subsequent research employing bioinformatic and co-expression analyses revealed a link between *HMGB3* and the development of human cancers, including gastric cancer [250]. This study aimed to investigate the clinicopathological significance of *HMGB3* at the transcriptomic and proteomic levels using large BC cohorts with long-term follow-up, and to study the *in vitro* mechanistic role of *HMGB3* in BC cell lines.

This study evaluated the clinicopathological and prognostic significance of *HMGB3* expression using large, well-annotated BC cohorts. High expression of *HMGB3* at the mRNA and protein levels was associated with aggressive BC features, including LVI positivity, high histological grade, poor NPI, ER/PR negativity, and HER2 positivity. Kaplan–Meier survival analysis revealed that high *HMGB3* expression at the mRNA and protein levels was significantly associated with poor BCSS survival and TTDM. Further analysis using the Cox regression model indicated that *HMGB3* expression was an independent predictor of BCSS in patients with invasive BC, which is consistent with previous studies [257, 266]. A weak correlation between the IHC data and mRNA expression data was observed in the Nottingham subset (n=288) of the METABRIC cohort. This weak association could be explained by the rate at which mRNA is translated into protein, which is

often referred to as 'translational efficiency'. Translational efficiency has been shown to have a significant influence on both mRNA and protein levels [235]. The subjectivity of the H-score method in the interpretation of expression in IHC staining sections was another reason [236], in addition to the use of whole tissue, including different cell types, in the METABRIC cases.

The preclinical experiments conducted in this study showed that silencing *HMGB3* suppressed cell proliferation and growth; as expected, suppressing *HMGB3* also induced cell cycle arrest in the G1 phase of BC cells, and silencing *HMGB3* increased cell apoptosis in BC cell lines. These findings were consistent with those of previous studies [258, 262, 267]. A study reported that *HMGB3* overexpression might be related to cell cycle and DNA replication pathways [264]. After silencing *HMGB3*, both p53 and p21 protein expression levels increased significantly. It is widely known that p53 regulates cell cycle arrest, senescence, apoptosis, and autophagy [268]. p21 is a p53 transcriptional target and is required for cell cycle arrest [269]. Silencing *HMGB3* resulted in G0/G1 phase cell cycle arrest, likely via modification of the p53 and p21 signalling pathways [258]. In addition, the levels of anti-apoptotic factor Bcl-2 and pro-apoptotic factor Bax were assessed following *HMGB3* knockdown, and revealed that Bax levels increased while Bcl-2 levels decreased when compared with the control group [258]. The balance of anti-apoptotic and pro-apoptotic factors is an important determinant of cell survival [270]. *HMGB3* knockdown decreased proliferation, triggered G1 phase arrest, and enhanced apoptosis by altering the p53 and p21 pathways and decreasing the Bcl-2/Bax ratio. Tumour cells must proliferate and evade apoptosis to penetrate surrounding tissues and develop metastatic cascades; proliferation continues until tumour cells invade the vascular or lymphatic channels.

In vitro functional assays showed that silencing *HMGB3* in BC cell lines reduced migration. This finding supports a study that evaluated the expression of β-catenin, a major Wnt pathway protein, to determine that

HMGB3 stimulated colorectal cancer migration via the Wnt/-catenin pathway. *HMGB3* can increase the expression of β-catenin and the downstream genes *c-Myc* and *MMP7* [254].

The malignant features and mechanism of action of *HMGB3* in LVI remain unknown. Importantly, in this study, *HMGB3* knockdown reduced adhesion and transmigration across endothelial cell lines in SK-BR-3 and MCF-7; the positive associations between EMT- and MMP-related markers support these results. Although the association between HMGB3 and these biomarkers ranged from weak to moderate correlation, it was statistically significant which indicates that these markers are contributing to the same oncogenic pathway in the context of LVI process. N-cadherin is associated with EMT, which is required for cell invasion and intravasation into the bloodstream and for ECM destruction induced by protease synthesis. Increased N-cadherin expression thus results in loss of the connection between the BC cell epithelium and other epithelial cells, leading to invasion into the stroma [271]. Additionally, P-cadherin, a critical protein that may activate integrin molecules, enables cancer cells adherence to the ECM and initiates invasion [272]. The microenvironment of tumours exhibiting LVI is closely associated with the expression of MMPs, notably *MMP9* and *MMP1*. The expressions of these MMPs is involved in the intravasation and metastasis of BC cells [179]. This finding also corroborates a previous study reporting that, by limiting the activity of MMP9, downregulation of *HMGB3* expression decreased the invasion of gastric cancer cells [260]. Although the *in vitro* studies revealed that overexpression of *HMGB3* aided migration, adhesion, and transmigration via endothelial cell lines, triggering the LVI process in the MCF-7 and SK-BR-3 cell lines, further studies are needed to determine why the ZR-75-1 cell line is unable to adhere to and transmigrate across endothelial cells. Therefore, more mechanistic investigations into how HMGB3 triggers LVI and *in vivo* animal experiments are warranted.

The findings suggest that *HMGB3* has an oncogenic role in BC and that it may be involved in the pathogenesis of LVI. By influencing the cell cycle

pathway, *HMGB3* may enhance the proliferation of BC cells and could represent one of the factors contributing to LVI in BC. This study showed the impact of *HMGB3* silencing on several processes of tumour development, including cell cycle, migration, adhesion, and transmigration, but did not investigate the molecular mechanisms involved in the development of LVI. Although the findings of this study suggest that *HMGB3* is a critical gene in LVI, especially in light of the lymphatic invasion findings, and a precursor for the metastatic cascade, more research is required to further substantiate these findings.

Chapter 6: RRM2, CCNB1, and CCNB2 as key drivers of LVI in invasive breast cancer: prognostic significance and *in vitro* investigation

This chapter has been published in the American Journal of Cancer Research (2022) under the title 'Upregulation of Cyclin B2 (*CCNB2*) in breast cancer contributes to the development of lymphovascular invasion'. Am J Cancer Res 2022; 12(2):469-489. The copyright approval for reuse of this material is provided in Appendix E.

6.1 Introduction

LVI is a key initial event of the metastatic cascade and serves as a prognostic factor for BC [273]. The lack of targeted therapy to inhibit LVI and prevent metastasis typically stems from the complex underlying molecular mechanisms of LVI that interact with other carcinogenesis pathways, and the subjectivity of assessing LVI in clinical BC cohorts. Therefore, deciphering the underlying molecular mechanisms of LVI and identifying the key genes controlling LVI in BC is necessary.

Through large transcriptomic cohorts of BC and ANN algorithms, *RRM2*, *CCNB1*, and *CCNB2* were identified as genes associated with LVI positivity, as mentioned in Chapter 3. *CCNB2* also overlapped between the WAD and the ANN analyses. This chapter will evaluate the clinicopathological significance of *RRM2*, *CCNB1*, and *CCNB2* at the transcriptomic and proteomic levels using BC cohorts with long-term follow-up to evaluate the *in vitro* mechanistic role of *CCNB2* using a panel of BC cell lines to understand their potential role in LVI development.

6.1.1 RRM2

Human ribonucleotide reductase (RNR) is the enzyme that converts ribonucleotide 5'-diphosphates to the 2'-deoxyribonucleotides necessary for DNA synthesis. RNR consists of two subunits: RRM1 and RRM2 [274]. RRM1 protein expression is largely consistent throughout the life of a cell, but RRM2 protein expression fluctuates dynamically in response to stimuli [275, 276]; therefore, RRM2 levels influence the cell cycle-dependent action of RNR [277]. Many types of cancer have been associated with an unregulated cell cycle [278]. RRM2 is the catalytic subunit of RNR that regulates the enzymatic activity required for DNA replication and repair [279]. Abnormal degradation of RRM2 results in an imbalanced dNTP pool

and subsequent genomic instability [280]. Recent data indicate that RRM2 acts as an oncogene and plays a role in the progression of several types of cancer [281], including glioma, bladder cancer, and colorectal cancer [282, 283]. According to bioinformatic analysis, the *RRM2* gene is overexpressed in BC patients [281]. RRM2 is also involved in various important biological processes, including cell proliferation, invasion, migration, angiogenesis, and ageing [279]. Suppression of *RRM2* in human neuroblastoma cells hinders cell growth, causes cell cycle arrest, and enhances apoptosis [284]. In addition, RRM2 has been shown to promote atypical EMT by increasing the expression of E-cadherin and P-cadherin [285]. Although RRM2 levels have been studied in numerous tumour types, nothing is known about the RRM2 levels in invasive BC and its potential role in LVI development.

6.1.2 CCNB1

Cyclins are proteins that activate certain cyclin-dependent kinases (CDKs) necessary for cell cycle progression [286]. CCNB1 is a member of the cyclin family and a critical initiator and with quality control function in in cellular division [287]. CCNB1 plays a critical role in regulating and complexing with CDK1 to promote transition from the G2 to the mitotic phase of the cell cycle [288]. From G0/G1 through the mid-S phase, CCNB1 is relatively undetectable in cells; it becomes apparent in the cytoplasm in the late S phase. CCNB1 levels rapidly increased in the perinuclear region of the cytoplasm as cells progressed through the G2 phase, and it then appears in the nucleus during the mitotic phase [289, 290]. CCNB1 is essential in checkpoint regulation, as its dysregulation is an early event in carcinogenesis [291]. High expression of CCNB1 results in the continuous cell cycling and division of cancer cells, promoting their migration and metastasis to distant sites [292]. Increasing evidence indicates that CCNB1 is overexpressed in a variety of human malignancies, including colorectal cancer, BC, and prostate cancer [288, 293, 294]. Inhibition of CCNB1 causes cell cycle arrest in various colorectal cancer cell lines by altering the expression of G2/M cell cycle regulators [288]. CCNB1 is also involved in

the proliferation, migration, apoptosis, chemoresistance, and metastasis of tumours [295-297]. However, the role of CCNB1 in LVI is unknown.

6.1.3 CCNB2

Cyclins are crucial elements of the cell-division cycle, and various studies, including a study by Stamatakos et al. [298], have discovered that defects in the functionality of cyclins can lead to carcinogenesis. Although evidence is still emerging, recent studies have revealed more information regarding the mechanistic pathways through which cyclins can influence the oncogenic potential of cells [299, 300]. Cyclins are active at different parts of the cell cycle and contribute to various functions. One of these cyclins is CCNB2. The *CCNB2* gene, a member of the cyclin B family, has a role in the G2/M transition through activation of cell division control (CDC2), and its inhibition leads to cell cycle arrest [301, 302]. *CCNB2* is critical to cell cycle regulation; it is one of the central protein kinases, and when activated, it becomes a master regulator in M-phase transition, phosphorylation, and the activation of other protein kinases [303, 304]. The expression of *CCNB2* is normally strictly regulated [299]; however, high expression of *CCNB2* has been observed in human tumours, including lung [305], liver [299], bladder [306], and BC [300]. *CCNB2* may act as an oncogene and could represent a potential biomarker for unfavourable outcomes [300, 307]. For example, reduced CCNB2 expression *in vivo* and *in vitro* dramatically inhibited bladder cancer, migration, invasion, and metastasis [306]. The role of *CCNB2* in LVI in BC is unclear; to my knowledge, this is the first study to report on the role of *CCNB2* in LVI.

6.2 Hypothesis

RRM2, *CCNB1*, and *CCNB2* are key drivers of LVI in BC, have a prognostic significance, are potential therapeutic targets, and are involved in various mechanisms controlling LVI.

6.3 Aims

1. Evaluate the clinicopathological and prognostic significance of *RRM2*, *CCNB1*, and *CCNB2* at transcriptomic and protein levels using well-characterised BC cohorts.
2. Evaluate the impact of silencing *CCNB2* on tumour cells using panels of relevant BC cell lines in *in vitro* experimental studies, including cell proliferation, migration, cell cycle, and apoptosis.
3. Assess the functional role of *CCNB2* on LVI by conducting an *in vitro* study of BC cell lines with lymphatic and vascular endothelial cell models to elucidate the interaction between BC cells and endothelial cells by performing adhesion and transmigration assays.

6.4 Methods

6.4.1 Study cohort

6.4.1.1 METABRIC cohort

The METABRIC cohort was used to evaluate the expression of *RRM2*, *CCNB1*, and *CCNB2* at the mRNA level. A total of 1,980 patients with mainly operable invasive BC were included in this study. Clinical history, outcome data, and clinicopathological parameters, including tumour size, grade, LN stage, NPI, and LVI, were all available for this cohort. PAM50 was used to establish the intrinsic subtypes of BC, including luminal A, luminal B, HER2-enriched, normal-like, and basal-like. A Nottingham subset (n=288) of the METABRIC cohort was utilised to assess the correlation between mRNA and protein expression. The features of the METABRIC cohort are described in Section 2.1.1, and the Nottingham subgroup is described in Section 2.1.2.

6.4.1.2 TCGA cohort

RRM2, *CCNB1*, and *CCNB2* mRNA expression were also assessed using the TCGA breast carcinoma cohort (n=854). Patient age, tumour size, histological grade, LVI, and outcome data were all available as de-identified clinical information, in addition to information about the expression of ER,

PR, and HER2. The characteristics of the TCGA cohort are described in Section 2.1.3.

6.4.1.3 Nottingham BC cohort

The expression of RRM2, CCNB1, and CCNB2 was evaluated at the protein level using the Nottingham BC cohort. Tissue samples in this cohort were derived from the well-characterised Nottingham invasive BC cohort. A robust clinicopathological profile was readily available for each patient, including age at diagnosis, primary tumour size, LN stage, histological grade, NPI, and LVI status. Data on the hormone receptor profile, BCSS, and TTDM were also available. Section 2.1.4 describes the clinicopathological characteristics of this cohort. To further understand the molecular interactions of RRM2, CCNB1, and CCNB2, their associations with EMT-related markers, including E-cadherin, N-cadherin, P-cadherin, TGFβ1, and TWIST2, were investigated (these biomarkers are available in our database).

6.4.2 WB

WB was used to validate the specificity of the RRM2, CCNB1, and CCNB2 primary antibodies. An array of BC cell line lysates was used for these antibodies, including MCF-7, SK-BR-3, and MB-MDA-231. Section 2.2 and Table 2.5 describe the methods, antibody concentrations, and incubation periods used for RRM2, CCNB1, and CCNB2.

6.4.3 IHC

To assess the expression patterns of RRM2, CCNB1, and CCNB2 before staining of the TMA, a representative full-face tissue section for invasive BC was stained to verify the staining homogeneity. It was confirmed that the staining pattern was homogeneous, verifying the applicability of immunohistochemical detection methods in TMAs. TMAs of the study cohort were prepared using a TMA Grand Master®. Antigen retrieval was performed (pH 6 citrate buffer at 1,000 W for 20 min using a microwave) following manufacturer recommendations for these antibodies. Expression

of these proteins was evaluated by staining the TMAs with a Novolink polymer detection systems kit, as described in Section 2.4. RRM2, CCNB1, and CCNB2 concentrations, incubation times, and positive and negative controls are detailed in Table 2.5.

6.4.4 *In vitro* assessment of *CCNB2*

Preclinical studies were carried out to investigate the potential role of *CCNB2* in LVI and other biological processes using the following cells and assays.

6.4.4.1 BC tumour cells

BC cell lines were chosen for *in vitro* experiments based on the protein expression levels from IHC analysis of CCNB2. MCF-7 ($ER^+/PR^+/HER2^-$), ZR-75-1 ($ER^+/PR^{\pm}/HER2^-$), and SK-BR-3 ($HER2^+$) BC cell lines were used. MCF-7 and ZR-75-1 represent luminal molecular classes, whereas SK-BR-3 represents a $HER2^+$ tumour; these two molecular classes are associated with the highest incidence of LVI [179]. WB was used to determine relative protein expression in all BC cell lines, and quantification was performed using the LI-COR Odyssey machine. The cells with the highest CCNB2 expression levels were chosen for the knockdown and other *in vitro* experiments. MCF-7 and ZR-75-1 were cultured in RPMI 1640 medium with L-glutamine supplemented with 10% FBS. McCoy's 5A medium modified with L-glutamine, and sodium bicarbonate liquid supplemented with 10% FBS were used to culture SK-BR-3. More details regarding the maintenance of cell growth and conditions are in Chapter 2.

6.4.4.2 Endothelial cells

Primary HUVECs and DLECs were used to investigate the role of *CCNB2* on LVI by mimicking the interactions that occur *in vivo* between endothelial and BC cells. HUVECs and DLECs were cultured in endothelial cell growth medium, as described in Chapter 2.

6.4.4.3 Transfection of *CCNB2* using siRNA

An siRNA-based approach using BC cell lines was used to investigate the possible functional consequences of *CCNB2* depletion and to study its role in BC progression, with emphases on its role in LVI. Using pre-validated Silencer Select siRNA constructs (designed for *CCNB2*) and scrambled negative control siRNA, MCF-7, ZR-75-1, and SK-BR-3 cells were transfected with a mixture of Opti-MEM medium, 25 pmol siRNA, and Lipofectamine™ RNAiMAX using the forward transfection method and according to the manufacturer's protocol (Chapter 2). The sequence of the *CCNB2*-siRNA was 5'-GGCGAACUGUUUUAGAAGAtt-3'.

6.4.4.4 Proliferation assay

The effect of *CCNB2* knockdown on the proliferation of tumour cells was assessed via 3-(4,5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium (MTS) assay, as described in Section 2.6.3.2.

6.4.4.5 Clonogenic assay

The cells were seeded in six-well plates for MCF-7, ZR-75-1, and SK-BR-3. Cells were left in the incubator in the cultured medium for 14 days. Following incubation, colonies were washed with PBS, fixed for 30 min, and stained with crystal violet. Finally, the cloned cells were counted using a microscope. The assay procedure and conditions are described in further detail in Section 2.6.4.2.

6.4.4.6 Wound-healing assay

In the scratching assay, a Culture Inserts 2 well, which contains a built-in gap, was used to assess the cell migration by measuring the wound repair rate of the *CCNB2* knockdown group and a control at the following time points: T0h, T24h, T48h, and T72h. The assay procedure and conditions are described in further detail in Section 2.6.5.2.

6.4.4.7 Cell cycle assay

The cells were harvested on the fifth day after transfection to detect and analyse the cell cycle distribution by quantifying DNA content using a propidium iodide flow cytometry kit. More details on the assay procedure and condition are presented in Section 2.6.6.2.

6.4.4.8 Apoptosis assay

On the fifth day after transfection, cells were stained to detect apoptosis using an Annexin V detection kit. Section 2.6.7.2 contains more information on the assay procedure and conditions.

6.4.4.9 Static adhesion assay

Endothelial cells (HUVECs and DLECs) were seeded to be confluent in a 24-well plate. Tumour cell adhesion was assessed following cell labelling with 1 μM Cell Tracker Green CMFDA and incubation at 37 °C for 30 min. After labelling, tumour cell medium supplemented with serum was used to resuspend the tumour cells, which were incubated with endothelial cell monolayers for 35 min at 37 °C. Non-adherent cells were washed with tumour cell media, and the adherent tumour cells were counted using a fluorescence microscope at 10× magnification. More details on the assay procedure and conditions are presented in Section 2.6.8.2.

6.4.4.10 Transmigration assay

A confluent endothelial cell monolayer was grown on hanging transwell inserts. Tumour cell transmigration was assessed after labelling with 1 μM of Cell Tracker Green CMFDA. After 16 h, transmigration was monitored by counting cells on the bottom of the chamber using a fluorescence microscope. The assay procedure and conditions are described in further detail in Section 2.6.8.3.

6.5 Results

6.5.1 Validation of RRM2, CCNB1, and CCNB2 antibodies by WB

WB was used to detect specific bands at the predicted molecular weights of ~39 kDa, 40 kDa, and 42 kDa, which corresponded to the RRM2, CCNB1, and CCNB2 proteins, respectively (Figure 6.1A–C).

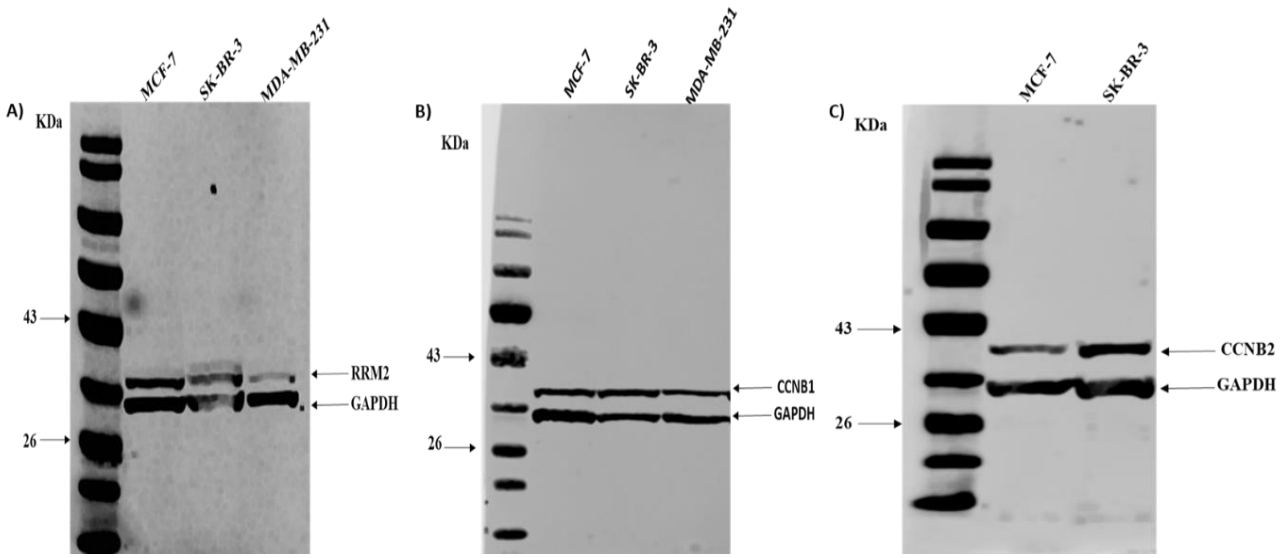
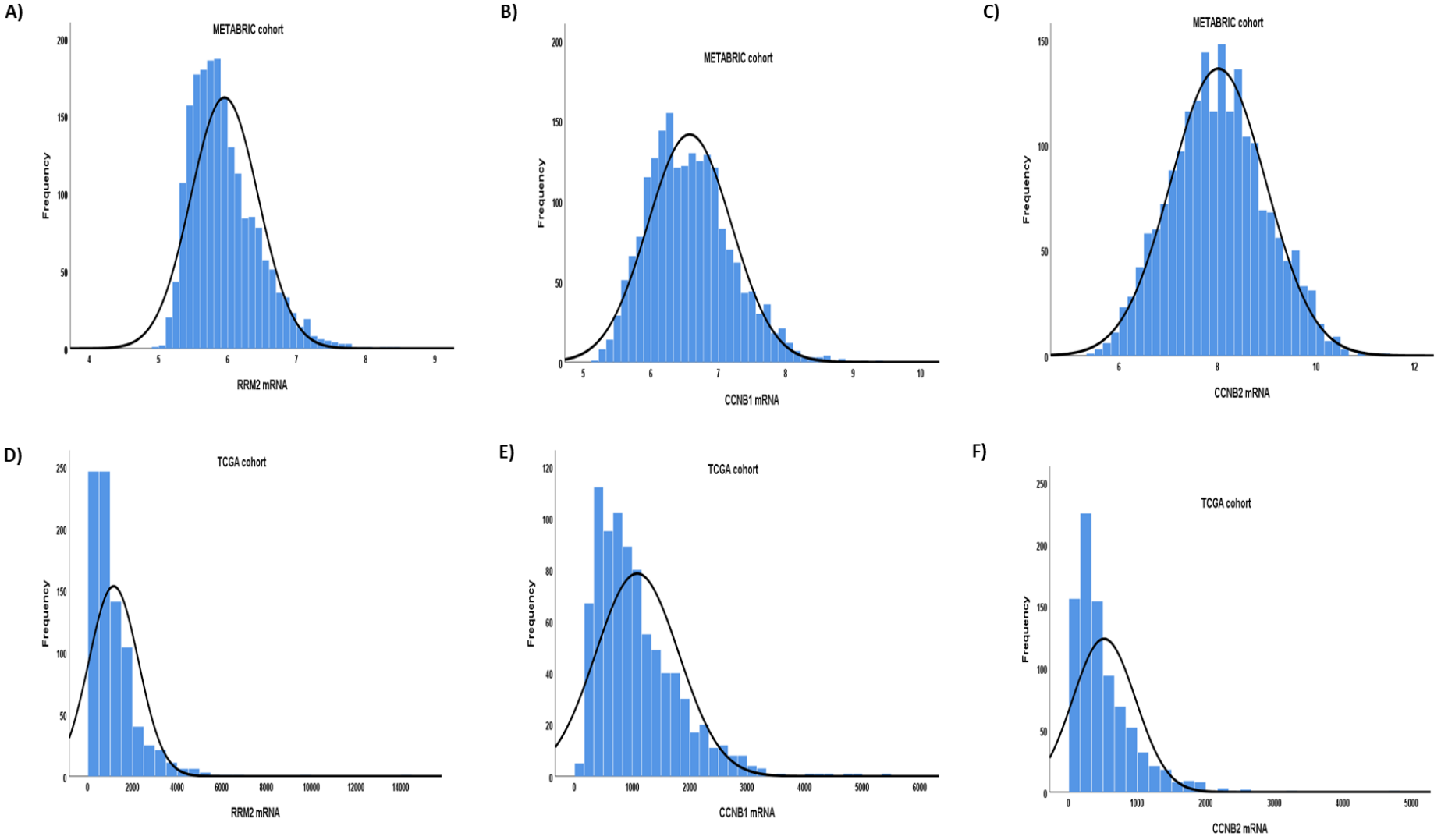


Figure 6.1 Western blot (WB) reflecting the antibody specificities of A) RRM2, B) CCNB1, and C) CCNB2. GAPDH was used as a positive control.

6.5.2 Expression of *RRM2*, *CCNB1*, and *CCNB2* in BC

In the METABRIC cohort, the *RRM2*, *CCNB1*, and *CCNB2* mRNA expression data were normally distributed and were categorised as either 'low' or 'high' expression using cut-off points generated from the mean. In the TCGA cohort, the *RRM2*, *CCNB1*, and *CCNB2* mRNA expression data were right-skewed and were categorised using cut-off points generated from the median (Figure 6.2A–F). High *RRM2*, *CCNB1*, and *CCNB2* mRNA expression was observed in 811/1,980 (41%), 922/1,980 (47%), and 904/1,980 (46%) of the METABRIC BC cases, respectively. High *RRM2*, *CCNB1*, and *CCNB2* expression was observed in 427/854 (50%), 427/854 (50%), and 319/854 (37%) of the TCGA cases, respectively.

RRM2, and CCNB2 protein expression was observed in the cytoplasm of invasive BC cells, with expression levels varying from absent to strong (Figures 6.3C/D and 6.3K/L). CCNB1 protein expression was observed mainly in the cytoplasm of invasive BC cells (Figure 6.3G/H), with occasional cases showing minimal to weak nuclear expression, which were not counted enough to run statistical analysis for. A total of 1,940, 2,480, and 3,178 cases were valid for evaluating RRM2, CCNB1, and CCNB2 proteins, respectively. The RRM2, CCNB1, and CCNB2 protein expression data had a multimodal distribution. The cut-off point for RRM2, CCNB1, and CCNB2 positivity was set using the median (Figure 6.2G–I). A high RRM2 protein level (H-score >110) was observed in 894/1,940 (46%) of invasive BC cases and 1,141/2,480 (46%) of cases showing high CCNB1 protein (H-score >100). High CCNB2 level (H-score >20) was observed in 1,564/3,178 (49%) of invasive BC cases. A strong concordance was observed between both scorers (Abrar Aljohani and Michael Toss) of the TMAs in immunoscoring of RRM2 (ICC=0.893, $p<0.0001$), CCNB1 (ICC=0.880, $p<0.0001$), and CCNB2 (ICC=0.861, $p<0.0001$).


A)
METABRIC cohort
Frequency
RRM2 mRNA
B)
METABRIC cohort
Frequency
CCNB1 mRNA
C)
METABRIC cohort
Frequency
CCNB2 mRNA
D)
TCGA cohort
Frequency
RRM2 mRNA
E)
TCGA cohort
Frequency
CCNB1 mRNA
F)
TCGA cohort
Frequency
CCNB2 mRNA

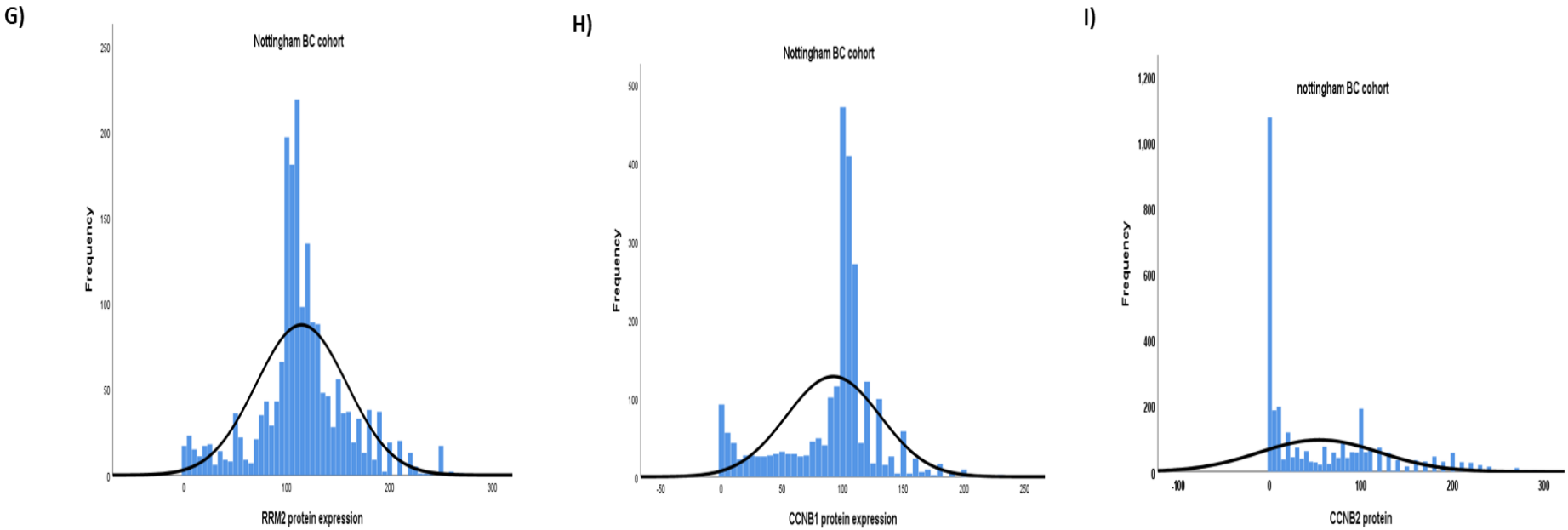
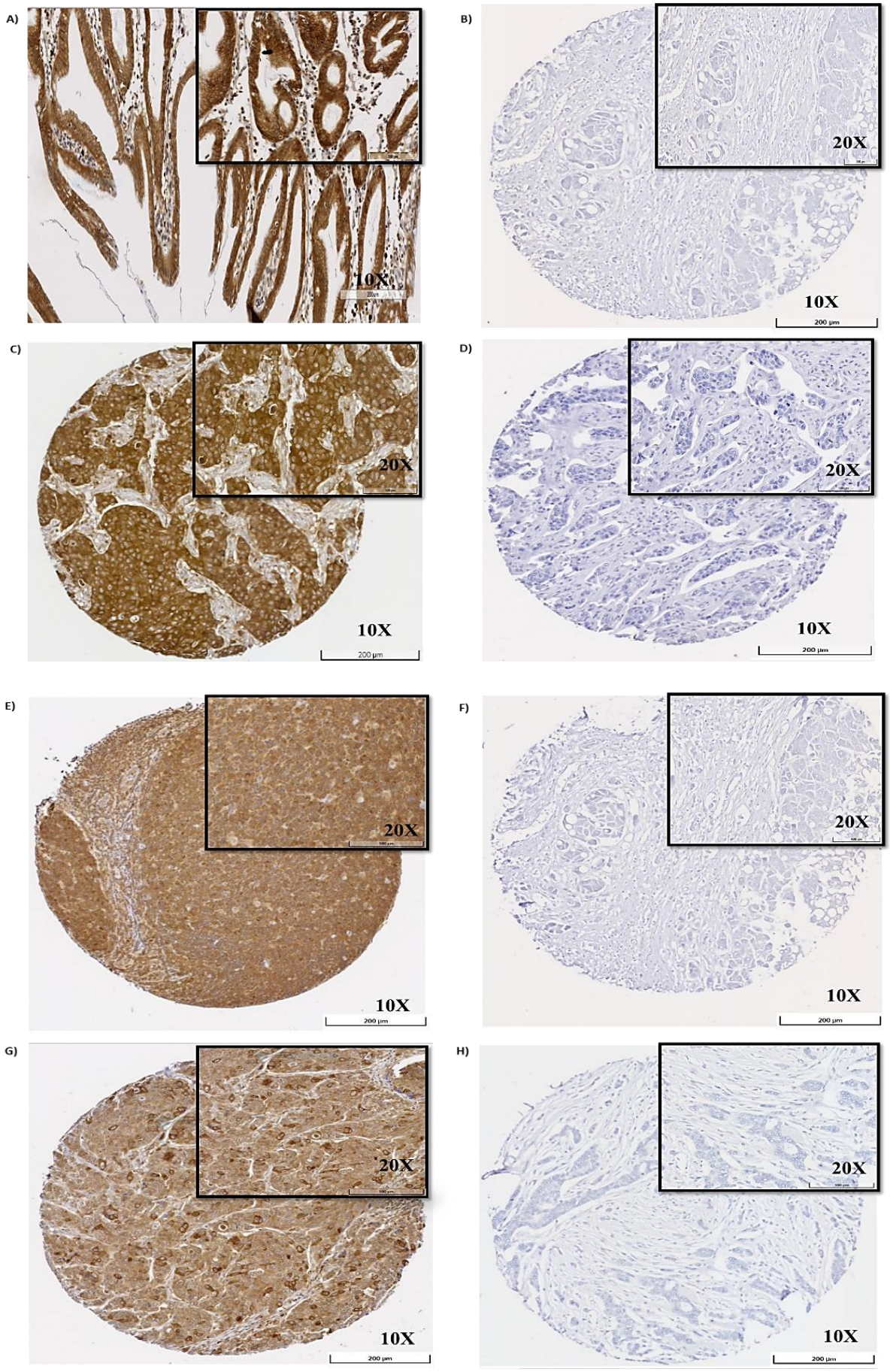


Figure 6.2 The distribution of the expression of A) *RRM2* mRNA in the METABRIC cohort, B) *CCNB1* mRNA in the METABRIC cohort, C) *CCNB2* mRNA in the METABRIC cohort, D) *RRM2* mRNA in the TCGA cohort, E) *CCNB1* mRNA in the TCGA cohort, F) *CCNB2* mRNA in the TCGA cohort, G) RRM2 protein in the Nottingham BC cohort, H) CCNB1 protein in the Nottingham BC cohort, and I) CCNB2 protein in the Nottingham BC cohort.


A)
10X
B)
20X
10X
200 µm
C)
20X
10X
200 µm
D)
20X
10X
200 µm
E)
20X
10X
200 µm
F)
20X
10X
200 µm
G)
20X
10X
200 µm
H)
20X
10X
200 µm

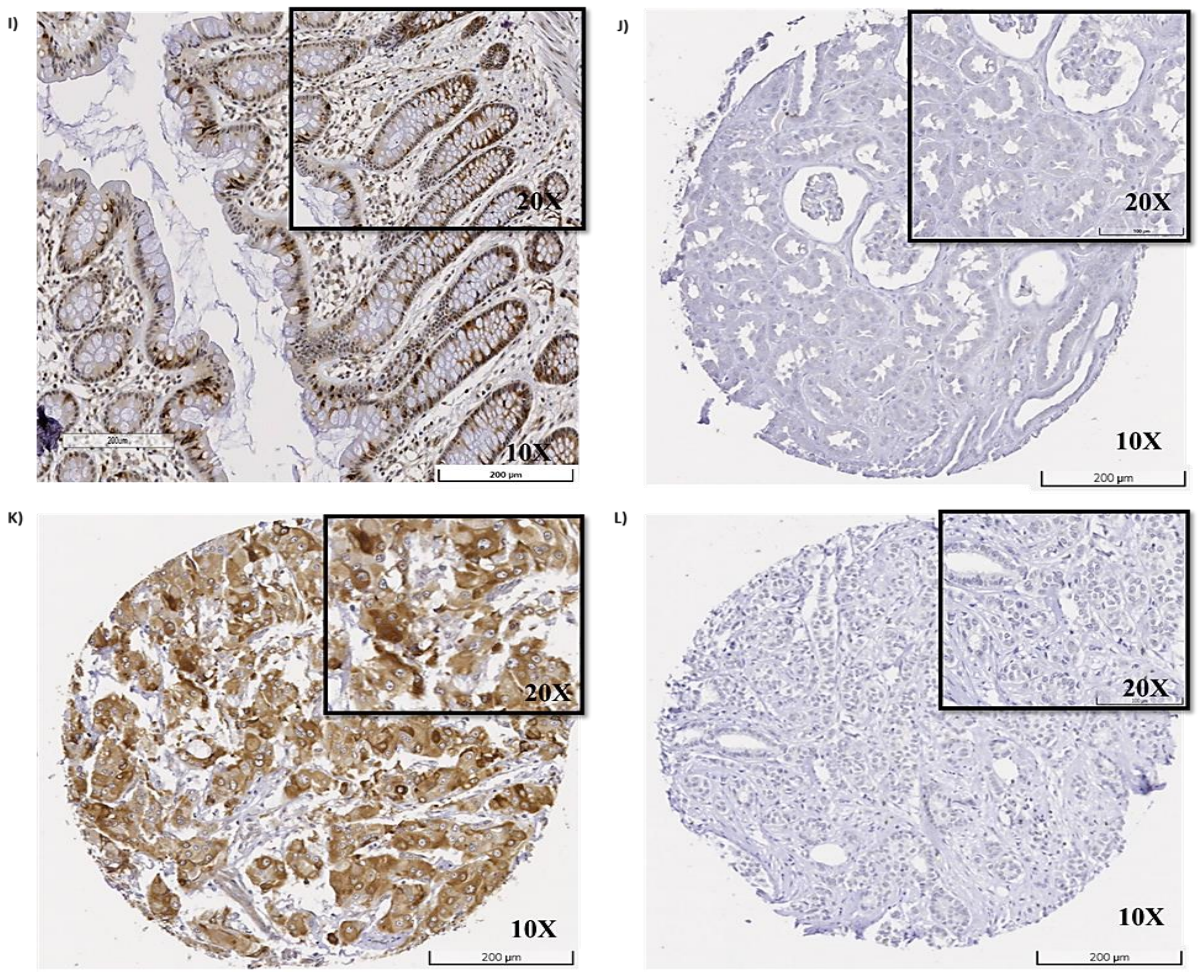


Figure 6.3 Cytoplasmic expression of RRM2, CCNB1, and CCNB2 proteins in invasive breast cancer. A) Positive control normal stomach tissue stained by RRM2, B) negative control normal liver stained by RRM2, C) RRM2-positive IHC expression, D) RRM2-negative IHC expression, E) positive control tonsil tissue stained by CCNB1, F) negative control liver stained by CCNB1, G) CCNB1-positive IHC expression, H) CCNB1-negative IHC expression, I) positive control colon tissue stained by CCNB2, J) negative control normal kidney stained by CCNB2, K) CCNB2-positive IHC expression, and L) CCNB2-negative IHC expression. Magnification 10×, scale bars = 200 µm. Inset: magnification 20×, scale bars = 100 µm.

6.5.3 Association of *RRM2*, *CCNB1*, and *CCNB2* with clinicopathological parameters

6.5.3.1 *RRM2*, *CCNB1*, and *CCNB2* mRNA expression

Association with clinicopathological parameters was evaluated using continuous data on *RRM2*, *CCNB1*, and *CCNB2* mRNA and protein levels. In the METABRIC and TCGA datasets, a significant association was observed between high *RRM2*, *CCNB1*, and *CCNB2* mRNA expression and LVI positivity (all $p<0.0001$). High *RRM2*, *CCNB1*, and *CCNB2* mRNA expression was also correlated with features of aggressive tumour behaviour; these features included large tumour size and high histological grade (all $p<0.0001$). In the METABRIC cohort, high expression of *RRM2*, *CCNB1*, and *CCNB2* mRNA was significantly associated with poor NPI and LN stage (all $p<0.0001$) (Tables 6.1, 6.2, and 6.3).

6.5.3.2 RRM2, CCNB1, and CCNB2 protein expression

At the protein level, high RRM2 expression was associated with large tumour size ($p=0.021$), high tumour grade ($p=0.010$), LN stage ($p=0.035$), high pleomorphism ($p<0.0001$), and moderate mitotic count ($p=0.008$). No significant association was observed at the protein level between the expression of RRM2 and patient age, tubule formation, or LVI status (all $p>0.05$) (Table 6.4). The Spearman's rank correlation coefficient showed no significant correlation between *RRM2* mRNA expression in the Nottingham subset ($n=288$) of the METABRIC cohort and protein level ($r=0.110$, $p=0.225$).

High CCNB1 protein levels were significantly correlated with high tumour grade, poor NPI, high pleomorphism, high mitotic count (all $p<0.0001$), and LVI negativity ($p=0.014$). No significant correlation was observed between the protein-level expression of CCNB1 and patient age, tubule formation, and tumour size (all $p>0.05$) (Table 6.5). No significant correlation was noted between CCNB1 protein expression and *CCNB1*

mRNA expression in the Nottingham subset (n=288) of the METABRIC cohort (r=0.136, p=0.080).

In a similar pattern to that of the transcriptomic results, high CCNB2 protein expression was significantly associated with clinicopathological parameters related to aggressive tumour behaviour, including LVI positivity (p=0.004), younger age (p<0.0001), high histological grade (p<0.0001), high mitotic count (p<0.0001), high pleomorphism (p<0.0001), high tubule formation (p<0.0001) , poor NPI (p<0.0001), and large tumour size (p=0.032) (Table 6.6). The Spearman's rank correlation coefficient showed a weak correlation between CCNB2 protein and *CCNB2* mRNA expression in the Nottingham subset (n=288) of the METABRIC cohort (r=0.127, p=0.031).

Table 6.1 Statistical associations between *RRM2* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets**.**

Parameters	***RRM2* mRNA (METABRIC)**			***RRM2* mRNA (TCGA)**		
	Number (%)	**Mean rank**	***p*-value**	**Number (%)**	**Mean rank**	***p*-value**
Patient age (years)						
≤50	424 (21.4)	5.99	0.114	231(27)	456.90	**0.034**
>50	1556 (78.6)	5.94		623 (73)	416.60	
Tumour size						
≤2 cm	622 (31.7)	5.89	**<0.0001**	239 (28)	367.48	**<0.0001**
>2 cm	1338 (68.3)	5.98		615 (72)	450.82	
Tumour grade						
I	170 (9.0)	5.68		89 (11)	225.97	
II	770 (40.6)	5.83	**<0.0001**	375 (46)	329.30	**<0.0001**
III	952 (50.3)	6.12		352 (43)	539.7	
Nottingham prognostic index (NPI)				Not available		
Good	680 (34.3)	5.78	**<0.0001**			
Moderate	1101 (55.6)	6.02				
Poor	199 (10.1)	6.18				
Lymph node stage				Not available		
I	1035 (52.5)	5.91				
II	622 (31.5)	5.97	**<0.0001**			
III	316 (16.0)	6.06				
Lymphovascular invasion (LVI)						
Negative	930 (59)	5.92	**0.013**	559 (65)	396.01	**<0.0001**
Positive	635 (41)	5.98		295 (35)	487.17	
Oestrogen receptor (ER)						
Negative	474 (23.9)	6.21	**<0.0001**	185 (22)	577.75	**<0.0001**
Positive	1506 (76.1)	5.87		639 (78)	364.66	
Progesterone receptor (PR)						
Negative	940 (47.4)	6.08	**<0.0001**	272 (33)	536.64	**<0.0001**
Positive	1040 (52.6)	5.84		546 (67)	346.16	
Human epidermal growth factor receptor 2 (HER2)			**<0.0001**			**<0.0001**
Negative	1733 (87.5)	5.92		567 (81)	330.16	
Positive	247 (12.5)	6.20		133 (19)	437.22	
PAM50 subtype						
Luminal A	718 (36.4)	5.71				
Luminal B	488 (24.7)	6.11	**<0.0001**			
HER2^{+}-enriched	240 (12.1)	6.15				
Basal-like	329 (16.7)	6.25		Not available		
Normal-like	199 (10.1)	5.70				

***p*-values in bold are statistically significant**

Table 6.2 Statistical associations between *CCNB1* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets**.**

Parameters	*CCNB1* mRNA (METABRIC)			*CCNB1* mRNA (TCGA)		
	Number (%)	**Mean rank**	***p*-value**	**Number (%)**	**Mean rank**	***p*-value**
Patient age (years)						
≤50	424 (21.4)	6.62	**0.041**	231(27)	444.66	0.216
>50	1556 (78.6)	6.55		623 (73)	421.14	
Tumour size						
≤2 cm	622 (31.7)	6.47	**<0.0001**	239 (28)	362.98	**<0.0001**
>2 cm	1338 (68.3)	6.61		615 (72)	452.57	
Tumour grade						
I	170 (9.0)	6.20		89 (11)	204.56	**<0.0001**
II	770 (40.6)	6.43	**<0.0001**	375 (46)	327.36	
III	952 (50.3)	6.77		352 (43)	546.50	
Nottingham prognostic index (NPI)			**<0.0001**	Not available		
Good	680 (34.3)	6.36				
Moderate	1101 (55.6)	6.67				
Poor	199 (10.1)	6.77				
Lymph node stage						
I	1035 (52.5)	6.51		Not available		
II	622 (31.5)	6.63	**0.001**			
III	316 (16.0)	6.66				
Lymphovascular invasion (LVI)						
Negative	930 (59)	6.53	**<0.0001**	559 (65)	395.47	**<0.0001**
Positive	635 (41)	6.66		295 (35)	488.20	
Oestrogen receptor (ER)						
Negative	474 (23.9)	6.73	**<0.0001**	185 (22)	562.99	**<0.0001**
Positive	1506 (76.1)	6.52		639 (78)	368.93	
Progesterone receptor (PR)						
Negative	940 (47.4)	6.65	**<0.0001**	272 (33)	511.06	**<0.0001**
Positive	1040 (52.6)	6.51		546 (67)	358.91	
Human epidermal growth factor receptor 2 (HER2)						
Negative	1733 (87.5)	6.54	**<0.0001**	567 (81)	340.36	**0.006**
Positive	247 (12.5)	6.81		133 (19)	393.73	
PAM50 subtype						
Luminal A	718 (36.4)	6.28				
Luminal B	488 (24.7)	6.94	**<0.0001**			
$HER2^+$-enriched	240 (12.1)	6.79				
Basal-like	329 (16.7)	6.83		Not available		
Normal-like	199 (10.1)	6.03				

***p*-values in bold are statistically significant**

Table 6.3 Statistical associations between *CCNB2* mRNA expression and clinicopathological parameters in the METABRIC (n=1,980) and TCGA (n=854) breast carcinoma datasets**.**

Parameters	***CCNB2* mRNA (METABRIC)**			***CCNB2* mRNA (TCGA)**		
	Number (%)	**Mean rank**	***p*-value**	**Number (%)**	**Mean rank**	***p*-value**
Patient age (years)						
≤50	424 (21.4)	8.25	**<0.0001**	231(27)	466.3	**0.005**
>50	1556 (78.6)	7.95		623 (73)	413.1	
Tumour size						
≤2 cm	622 (31.7)	7.58	**<0.0001**	239 (28)	362.5	**<0.0001**
>2 cm	1338 (68.3)	8.09		615 (72)	452.8	
Tumour grade						
I	170 (9.0)	7.16		89 (11)	198.5	**<0.0001**
II	770 (40.6)	7.66	**<0.0001**	375 (46)	308.3	
III	952 (50.3)	8.50		352 (43)	568.4	
Nottingham prognostic index (NPI)						
Good	680 (34.3)	7.51	**<0.0001**	Not available		
Moderate	1101 (55.6)	8.26				
Poor	199 (10.1)	8.41				
Lymph node stage						
I	1035 (52.5)	7.90				
II	622 (31.5)	8.12	**<0.0001**	Not available		
III	316 (16.0)	8.15				
Lymphovascular invasion (LVI)						
Negative	930 (59)	7.93	**<0.0001**	559 (65)	392.5	**<0.0001**
Positive	635 (41)	8.10		295 (35)	493.8	
Oestrogen receptor (ER)						
Negative	474 (23.9)	8.77	**<0.0001**	185 (22)	625.7	**<0.0001**
Positive	1506 (76.1)	7.78		639 (78)	350.8	
Progesterone receptor (PR)						
Negative	940 (47.4)	8.53	**<0.0001**	272 (33)	552.1	**<0.0001**
Positive	1040 (52.6)	7.71		546 (67)	338.4	
Human epidermal growth factor receptor 2 (HER2)			**<0.0001**			**0.002**
Negative	1733 (87.5)	7.94		567 (81)	339.1	
Positive	247 (12.5)	8.52		133 (19)	399.1	
PAM50 subtype						
Luminal A	718 (36.4)	7.34				
Luminal B	488 (24.7)	8.39	**<0.0001**			
$HER2^+$-enriched	240 (12.1)	8.52				
Basal-like	329 (16.7)	9.03		Not available		
Normal-like	199 (10.1)	7.20				

***p*-values in bold are statistically significant**

Table 6.4 Statistical associations between RRM2 protein expression and clinicopathological factors in the Nottingham BC cohort (n=1,940).

Nottingham BC cohort			
Parameters	**RRM2 protein**		
	Number (%)	Mean Rank	*p*-value
Patient age (years) ≤50 >50	 655 (33.7) 1285 (66.30	 937.8 987.2	0.066
Tumour size ≤2 cm >2 cm	 1141 (59) 799 (41)	 935.4 995.1	**0.021**
Tumour grade I II III	 285 (14.7) 740 (38.1) 915 (47.2)	 951.1 928.5 1010.5	**0.010**
Mitosis I II III	 852 (43.9) 378 (19.5) 710 (36.6)	 940.1 1047 966.2	**0.008**
Pleomorphism I II III	 31 (1.6) 583 (30.1) 1326 (68.3)	 811.9 847.5 1028.3	**<0.0001**
Tubule formation I II III	 125 (6.4) 585 (30.1) 1230 (63.5)	 1042.7 987.9 954.9	0.164
Lymphovascular invasion (LVI) Negative Positive	 1353 (69.7) 587 (30.3)	 969.3 973.2	0.890
Lymph node stage I II III	 1197 (61.7) 555 (28.6) 187 (9.7)	 955.1 969 1068.6	**0.035**
Nottingham prognostic index (NPI) Good Moderate Poor	 628 (32.4) 1005 (51.8) 306 (15.8)	 947.8 967.4 1024.2	0.143
Oestrogen receptor (ER) Negative Positive	 426 (22) 1512 (78)	 898.6 989.5	**0.003**
Progesterone receptor (PR) Negative Positive	 782 (40.7) 1137 (59.3)	 951.1 966.1	0.558
Human epidermal growth factor receptor 2 (HER2) Negative Positive	 1662 (86.6) 256 (13.4)	 951.2 1013.6	0.092
Triple-negative No Yes	 1614 (84.5) 296 (15.5)	 970.9 871	**0.004**
Immunohistochemistry subtype $ER^+/HER2^-$ low proliferation $ER^+/HER2^-$ high proliferation	 893 (46.7) 619 (32.4)	 943 1015.2	**0.002**

Triple-negative HER2^{+}	296 (15.5) 102 (5.8)	871 948.2	

***p*-values in bold are statistically significant**

Table 6.5 Statistical associations between CCNB1 protein expression and clinicopathological factors in the Nottingham BC cohort (n=2,480).

Nottingham BC cohort			
Parameters	**CCNB1 protein**		
	Number (%)	Mean rank	*p*-value
Patient age (years)			
≤50	843 (34)	1273.9	0.093
>50	1637 (66)	1223.3	
Tumour size			
≤2 cm	1371 (55.3)	1261.6	0.101
>2 cm	1109 (44.7)	1214.5	
Tumour grade			
I	371 (20)	371	**<0.0001**
II	875 (35.3)	875	
III	1234 (44.7)	1234	
Mitosis			
I	971 (39)	1114.6	**<0.0001**
II	498 (20.1)	1210.4	
III	1011 (40.9)	1376.2	
Pleomorphism			
I	51 (2.1)	1112.3	**<0.0001**
II	766 (30.9)	1038.9	
III	1663 (67)	1337.3	
Tubule formation			
I	167 (8.6)	1256.3	0.070
II	775 (31.3)	1191.7	
III	1538 (60.1)	1263.4	
Lymphovascular invasion (LVI)			
Negative	1703 (69)	1264.1	**0.014**
Positive	777 (31)	1188.8	
Lymph node stage			
I	1535 (61.9)	1240.2	0.994
II	706 (28.5)	1242.4	
III	239 (9.6)	1237	
Nottingham prognostic index (NPI)			
Good	769 (31)	1140.9	**<0.0001**
Moderate	1299 (52.4)	1278.8	
Poor	412 (16.6)	1305.8	
Oestrogen receptor (ER)			
Negative	620 (25.1)	1404.5	**<0.0001**
Positive	1848 (74.9)	1177.5	
Progesterone receptor (PR)			
Negative	1046 (43)	1281.2	**<0.0001**
Positive	1382 (57)	1163.9	
Human epidermal growth factor receptor 2 (HER2)			**0.011**
Negative	2084 (86)	1197.9	
Positive	340 (14)	1301.8	
Triple-negative			
No	1995 (82.4)	1164.3	**<0.0001**
Yes	426 (17.6)	1429.9	
Immunohistochemistry subtype			
$ER^+/HER2^-$ low proliferation	1063 (46.4)	1078.8	**<0.0001**
$ER^+/HER2^-$ high proliferation	785 (34.3)	1251.4	

Triple-negative HER2^{+}	426 (18.6) 147 (0.7)	1429.9 1317.2	

***p*-values in bold are statistically significant**

Table 6.6 Statistical associations between CCNB2 protein expression and clinicopathological factors in the Nottingham BC cohort (n=3,178).

Nottingham BC cohort			
Parameters	**CCNB2 protein**		
	Number (%)	Mean rank	*p*-value
Patient age (years) ≤50 >50	 1042 (33) 2131 (67)	 1672.3 1545.3	**<0.0001**
Tumour size ≤2 cm >2 cm	 1768 (56) 1403 (44)	 1555.6 1624.3	**0.032**
Tumour grade I II III	543 (17) 1131 (36) 1496 (47)	1314.0 1503.9 1745.7	**<0.0001**
Mitosis I II III	1284 (42) 598 (19) 1211 (39)	1361.82 1597.05 1718.63	**<0.0001**
Pleomorphism I II III	58 (1.8) 1084 (35) 1949 (36.2)	1213.87 1368.07 1654.48	**<0.0001**
Tubule formation I II III	205 (6.6) 984 (31.7) 1906 (61.7)	1282.29 1510.29 1596.05	**<0.0001**
Lymphovascular invasion (LVI) Negative Positive	2205 (70) 954 (30)	1549.6 1650.1	**0.004**
Lymph node stage I II III	1989 (63) 903 (28) 277 (9)	1563.5 1607.3 1666.6	0.136
Nottingham prognostic index (NPI) Good Moderate Poor	1041 (33) 1632 (51) 496 (16)	1416.0 1659.3 1695.0	**<0.0001**
Oestrogen receptor (ER) Negative Positive	704 (22) 2453 (78)	1677.0 1550.9	**0.001**
Progesterone receptor (PR) Negative Positive	1256 (41) 1813 (59)	1585.5 1500.0	**0.007**
Human epidermal growth factor receptor 2 (HER2) Negative Positive	2677 (87) 402 (13)	1513.8 1714.3	**<0.0001**
Triple-negative No Yes	2630 (84) 483 (16)	1538.7 1656.5	**0.007**
Immunohistochemistry subtype $ER^+/HER2^-$ low proliferation $ER^+/HER2^-$ high proliferation	1051 (36.9) 906 (31.9)	1264.1 1497.8	**<0.0001**

Triple-negative HER2$^+$	483 (17) 402 (14.2)	1499.5 1567.4	

***p*-values in bold are statistically significant**

6.5.4 *RRM2*, *CCNB1*, and *CCNB2* expression in molecular BC subtypes

In both the METABRIC and TCGA cohorts, high expression of *RRM2*, *CCNB1*, and *CCNB2* mRNA was significantly correlated with hormone receptor (ER/PR) negativity (all p<0.0001) and HER2 positivity (p<0.0001 [METABRIC, all]; p<0.0001 [TCGA, *RRM2*]; p=0.006 [TCGA, *CCNB1*]; p=0.002 [TCGA, *CCNB2*]) (Tables 6.1, 6.2, and 6.3).

Investigation of *RRM2* and *CCNB2* mRNA in the intrinsic (PAM50) subtypes of the METABRIC cohort revealed that high expression was observed in basal-like, HER2$^+$, luminal B, luminal A, and normal-like subtypes, in descending order (all p<0.0001) (Tables 6.1 and 6.3). Assessment of *CCNB1* mRNA in the intrinsic (PAM50) subtypes showed that overexpression of *CCNB1* was correlated with luminal B, basal-like, HER2$^+$, luminal A, and normal-like subtypes, in descending order (p<0.0001) (Table 6.2).

In the Nottingham BC cohort, RRM2 protein expression showed a significant association with ER positivity (p=0.003). No significant association was observed between high RRM2 protein level and PR or HER2 status (p>0.05) (Table 6.4). However, overexpression of CCNB1 was associated with hormone receptor (ER/PR) negativity (p<0.0001) and HER2 positivity (p=0.011) (Table 6.5). Regarding the CCNB2 protein level, there was a significant association between overexpression of CCNB2 and ER negativity (p=0.001), PR negativity (p=0.007), and HER2 positivity (p<0.0001) (Table 6.6). Furthermore, protein levels of both CCNB1 and CCNB2 were more highly expressed in TN tumours than in non-TN tumours (p<0.0001, CCNB1; p=0.007, CCNB2) (Tables 6.5 and 6.6). RRM2 was significantly associated with non-TN tumours (p=0.004) (Table 6.4). In the IHC subtypes, high RRM2 protein expression showed a significant association

with high proliferation class and the $HER2^+$ subtype, followed by low proliferation class and $ER^-/HER2^-$ subtype (p=0.002) (Table 6.4). High expression of CCNB1 was associated with $ER^-/HER2^-$, $HER2^+$, and a high proliferation subtype, followed by low proliferation subtype (p<0.0001) (Table 6.5).

The association of CCNB2 protein level with the IHC subtypes was similar to the above findings, where the $HER2^+$ subtype had the highest levels of CCNB2 protein expression, followed by TN tumours. Within the $ER^+/HER2^-$ group, expression levels were significantly greater in the high proliferation class than in the low proliferation class (p<0.0001) (Table 6.6).

6.5.5 *RRM2*, *CCNB1*, and *CCNB2* expression and other associated biomarkers

To further evaluate the role of *RRM2*, *CCNB1*, and *CCNB2* in BC and their interactions with other genes related LVI process, the METABRIC and TCGA datasets were interrogated for the correlation between *RRM2*, *CCNB1*, and *CCNB2* mRNA and other genes involved in invasion, EMT, and adhesion. The genes were selected based on previous publications [128, 129, 172, 174, 179]. At the protein level, the association of RRM2, CCNB1, and CCNB2 with EMT-related markers, including E-cadherin, N-cadherin, P-cadherin, TGFβ1, and TWIST2, that were available in the Nottingham BC cohort were also investigated.

6.5.5.1 Associations with *RRM2*

In both the METABRIC and TCGA cohorts, there was a significant weak to positive linear association between *RRM2* mRNA expression and the expression of EMT-related genes, including *P-cadherin, ZEB1, ZEB2, GSK3B, NFKB1,* and *CTNNB1* (all p<0.05). *E-cadherin* was negatively correlated with *RRM2* in both cohorts. In the TCGA cohort, there was a significant weak positive correlation between *RRM2* and *TGFB1, TWIST1,* and *TWIST2* (all p<0.05). In both cohorts, *RRM2* mRNA level showed a significant weak to moderate positive linear correlation with other MMP-

related genes, including *MMP1, MMP7, MMP9,* and *MMP12* (all p<0.05); in the METABRIC cohort, a positive correlation was also observed with *MMP20* and *MMP25,* whereas in the TCGA cohort, the same correlation was seen with *MMP15* (all p<0.05) (Table 6.7). In the Nottingham cohort, increased RRM2 protein levels were also correlated with the absence of E-cadherin, and the presence of N-cadherin, TGFB1, and TWIST2 (all p<0.05) (Table 6.7). No significant association was observed between RRM2 and the other genes listed in Table 6.7 (p>0.05).

Table 6.7 Correlations of RRM2 expression with mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes.

Gene names	**METABRIC cohort (mRNA)**		**TCGA cohort (mRNA)**		**Nottingham cohort (protein)**	
	Correlation value	*p*-value	Correlation value	*p*-value	Correlation value	*p*-value
EMT-related genes						
E-cadherin	- 0.232	**<0.0001**	- 0.223	**<0.0001**	- 0.140	**0.002**
N-cadherin	0.044	0.051	0.158	**<0.0001**	0.198	**<0.0001**
P-cadherin	0.115	**<0.0001**	0.164	**<0.0001**	0.081	0.080
TGFβ1	0.037	0.100	0.217	**<0.0001**	0.117	**0.030**
TWIST1	0.021	0.361	0.140	**<0.0001**	**Not available**	
TWIST2	0.005	0.815	0.223	**<0.0001**	0.239	**<0.0001**
ZEB1	0.054	**0.016**	0.263	**<0.0001**	**Not available**	
ZEB2	0.063	**0.005**	0.083	**0.015**		
NFKB1	0.174	**<0.0001**	0.150	**<0.0001**		
GSK3B	0.091	**<0.0001**	0.234	**<0.0001**		
CTNNB1	0.290	**<0.0001**	0.116	**0.001**		
MMP-related genes						
MMP1	0.163	**<0.0001**	0.443	**<0.0001**	**Not available**	
MMP7	0.166	**<0.0001**	0.067	**0.050**		
MMP9	0.194	**<0.0001**	0.199	**<0.0001**		
MMP11	0.003	0.911	0.035	0.305		
MMP12	0.224	**<0.0001**	0.301	**<0.0001**		
MMP15	0.036	0.106	0.099	**0.004**		
MMP20	0.057	**0.011**	0.031	0.368		
MMP25	0.062	**0.006**	0.004	0.912		

***p*-values in bold are statistically significant**

6.5.5.2 Associations with *CCNB1*

Both transcriptomic cohorts showed a significant weak to moderate positive linear correlation between *CCNB1* mRNA expression and the expression of EMT-related genes, including *N-cadherin, GSK3B, TWIST1, TWIST2, ZEB1, ZEB2, NFKB1 and CTNNB1, while a negative linear correlation was observed with E-cadherin* (all $p<0.05$). In the TCGA cohort, a similar correlation was observed between *CCNB1* mRNA expression and the expression of *P-cadherin* and *TGFB1* (both $p<0.05$). In addition, in both cohorts, *CCNB1* mRNA expression was positively correlated with the expression of MMP-related genes, including *MMP1, MMP7, MMP9, MMP12, MMP15,* and *MMP20* (all $p<0.05$) (Table 6.8). In the Nottingham cohort, high CCNB1 protein level showed a negative correlation with E-cadherin and a positive correlation with N-cadherin, P-cadherin and TWIST2 (all $p<0.05$) (Table 6.8). No significant association was observed between RRM2 and the other genes listed in Table 6.8 ($p>0.05$).

6.5.5.3 Associations with *CCNB2*

Both METABRIC and TCGA cohorts showed a significant weak to moderate positive linear correlation between *CCNB2* and numerous genes involved in EMT and stromal degradation biomarkers, including *N-cadherin*, *P-cadherin*, *GSK3B, TWIST1, TWIST2, ZEB1, ZEB2, NFKB1, and CTNNB1* (all $p<0.05$). There was also a significant weak negative linear correlation with *E-cadherin*. Moreover, there was a significant weak to moderate positive linear correlation between *CCNB2* and MMPs, including *MMP1*, *MMP7*, *MMP9*, *MMP11 MMP12, MMP15, MMP20,* and *MMP25* (all $p< 0.05$) (Table 6.9). At the protein level, there was a significant correlation between levels of CCNB2 and E-cadherin ($p=0.0001$), N-cadherin ($p=0.048$), P-cadherin, ($p=0.030$), TGFβ1 ($p=0.008$), and TWIST2 ($p=0.004$) (Table 6.9).

6.5.6 Associations between *RRM2*, *CCNB1*, and *CCNB2*

To further explore the potential role of *RRM2*, *CCNB1*, and *CCNB2* in LVI, the correlation between these genes at the mRNA and protein levels was investigated. There was a significant weak to strong positive linear correlation between these genes at the mRNA and protein levels (all p<0.0001) (Table 6.10).

Table 6.8 Correlations of CCNB1 expression with mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes.

Gene names	**METABRIC cohort (mRNA)**		**TCGA cohort (mRNA)**		**Nottingham cohort (protein)**	
	Correlation value	*p*-value	Correlation value	*p*-value	Correlation value	*p*-value
EMT-related genes						
E-cadherin	- 0.090	**<0.0001**	- 0.128	**<0.0001**	- 0.091	**0.003**
N-cadherin	0.079	**<0.0001**	0.096	**0.005**	0.083	**0.021**
P-cadherin	0.023	0.311	0.158	**<0.0001**	0.145	**<0.0001**
TGFβ1	0.014	0.522	0.300	**<0.0001**	0.053	0.163
TWIST1	0.153	**<0.0001**	0.157	**<0.0001**	**Not available**	
TWIST2	0.289	**<0.0001**	0.252	**<0.0001**	0.079	**0.036**
ZEB1	0.203	**<0.0001**	0.396	**<0.0001**	**Not available**	
ZEB2	0.162	**<0.0001**	0.259	**<0.0001**		
NFKB1	0.264	**<0.0001**	0.296	**<0.0001**		
GSK3B	0.134	**<0.0001**	0.122	**<0.0001**		
CTNNB1	0.188	**<0.0001**	0.178	**<0.0001**		
MMP-related genes						
MMP1	0.300	**<0.0001**	0.357	**<0.0001**	**Not available**	
MMP7	0.068	**0.002**	0.113	**0.001**		
MMP9	0.116	**<0.0001**	0.124	**<0.0001**		
MMP11	0.035	0.115	0.036	0.290		
MMP12	0.195	**<0.0001**	0.291	**<0.0001**		
MMP15	0.103	**<0.0001**	0.070	**0.042**		
MMP20	0.064	**0.004**	0.065	**0.057**		
MMP25	0.053	**0.018**	0.022	0.519		

***p*-values in bold are statistically significant**

Table 6.9 Correlations of CCNB2 expression with mRNA and protein expression of epithelial-mesenchymal transition (EMT)- and matrix metalloproteinase (MMP)-related genes.

Gene names	**METABRIC cohort (mRNA)**		**TCGA cohort (mRNA)**		**Nottingham cohort (protein)**	
	Correlation value	*p*-value	Correlation value	*p*-value	Correlation value	*p*-value
EMT-related genes						
E-cadherin	- 0.125	**<0.0001**	- 0.140	**<0.0001**	- 0.108	**<0.0001**
N-cadherin	0.151	**<0.0001**	0.133	**<0.0001**	0.095	**0.048**
P-cadherin	0.322	**<0.0001**	0.280	**<0.0001**	0.261	**0.030**
TGFβ1	0.025	0.260	0.252	**<0.0001**	0.080	**0.008**
TWIST1	0.045	**0.045**	0.131	**<0.0001**	**Not available**	
TWIST2	0.270	**<0.0001**	0.214	**<0.0001**	0.087	**0.004**
ZEB1	0.281	**<0.0001**	0.420	**<0.0001**	**Not available**	
ZEB2	0.143	**<0.0001**	0.214	**<0.0001**		
NFKB1	0.176	**<0.0001**	0.254	**<0.0001**		
GSK3B	0.330	**<0.0001**	0.129	**<0.0001**		
CTNNB1	0.073	**0.001**	0.172	**<0.0001**		
MMP-related genes						
MMP1	0.328	**<0.0001**	0.428	**<0.0001**	**Not available**	
MMP7	0.134	**<0.0001**	0.190	**<0.0001**		
MMP9	0.308	**<0.0001**	0.227	**<0.0001**		
MMP11	0.051	**0.022**	0.064	**0.060**		
MMP12	0.382	**<0.0001**	0.350	**<0.0001**		
MMP15	0.199	**<0.0001**	0.104	**0.002**		
MMP20	0.092	**<0.0001**	0.134	**<0.0001**		
MMP25	0.165	**<0.0001**	0.073	**0.034**		

***p*-values in bold are statistically significant**

Table 6.10 Correlations between *RRM2*, *CCNB1*, and *CCNB2* mRNA and protein expression levels.

Gene names	**METABRIC cohort (mRNA)**		**TCGA cohort (mRNA)**		**Nottingham cohort (protein)**	
	Correlation value	*p*-value	Correlation value	*p*-value	Correlation value	*p*-value
RRM2						
CCNB1	0.330	**<0.0001**	0.763	**<0.0001**	0.279	**<0.0001**
CCNB2	0.585	**<0.0001**	0.814	**<0.0001**	0.157	**0.048**
CCNB1						
RRM2	0.330	**<0.0001**	0.763	**<0.0001**	0.279	**<0.0001**
CCNB2	0.650	**<0.0001**	0.870	**<0.0001**	0.263	**<0.0001**
CCNB2						
RRM2	0.585	**<0.0001**	0.814	**<0.0001**	0.157	**<0.0001**
CCNB1	0.650	**<0.0001**	0.870	**<0.0001**	0.263	**<0.0001**

***p*-values in bold are statistically significant**

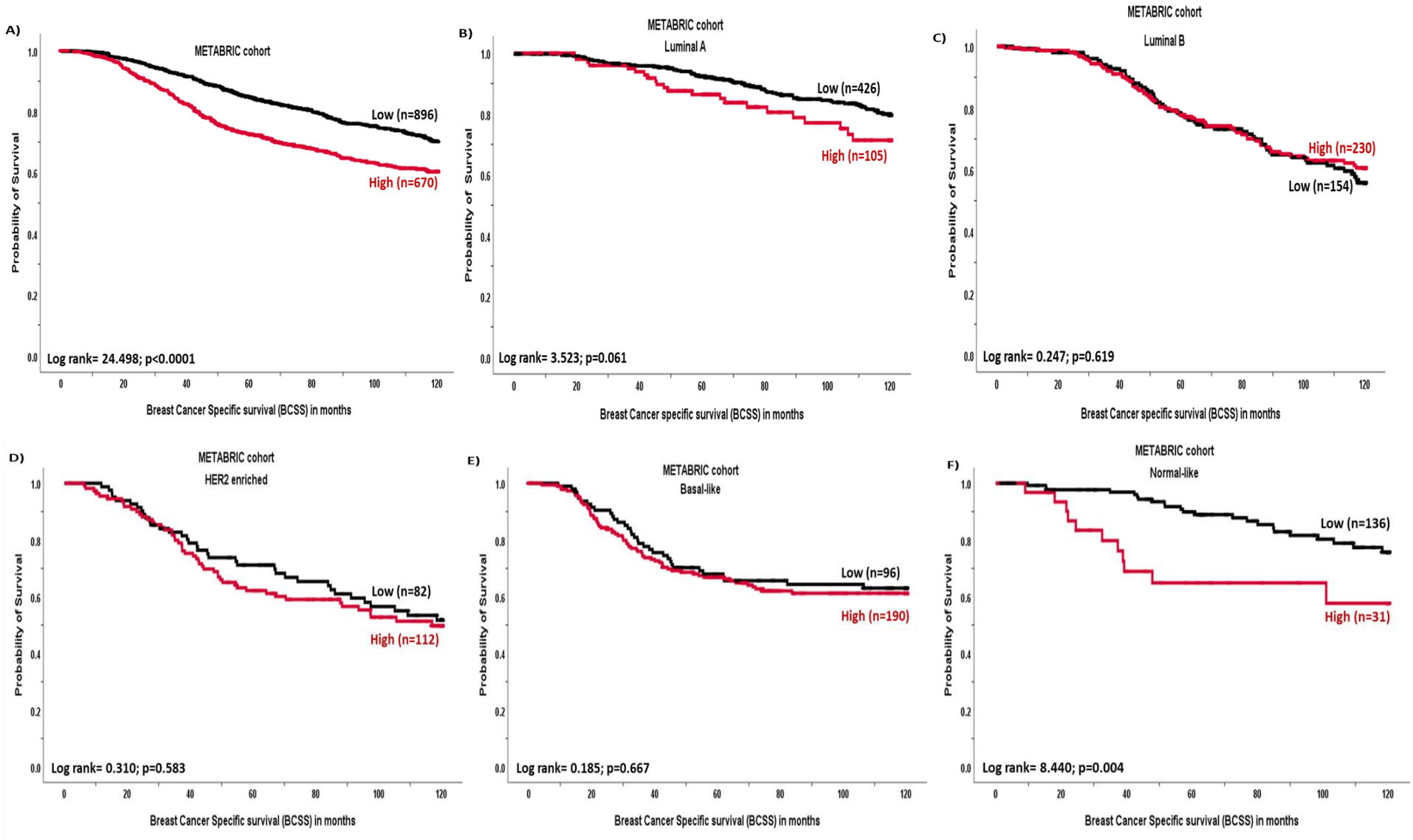
6.5.7 Association between *RRM2*, *CCNB1*, and *CCNB2* and patient outcome

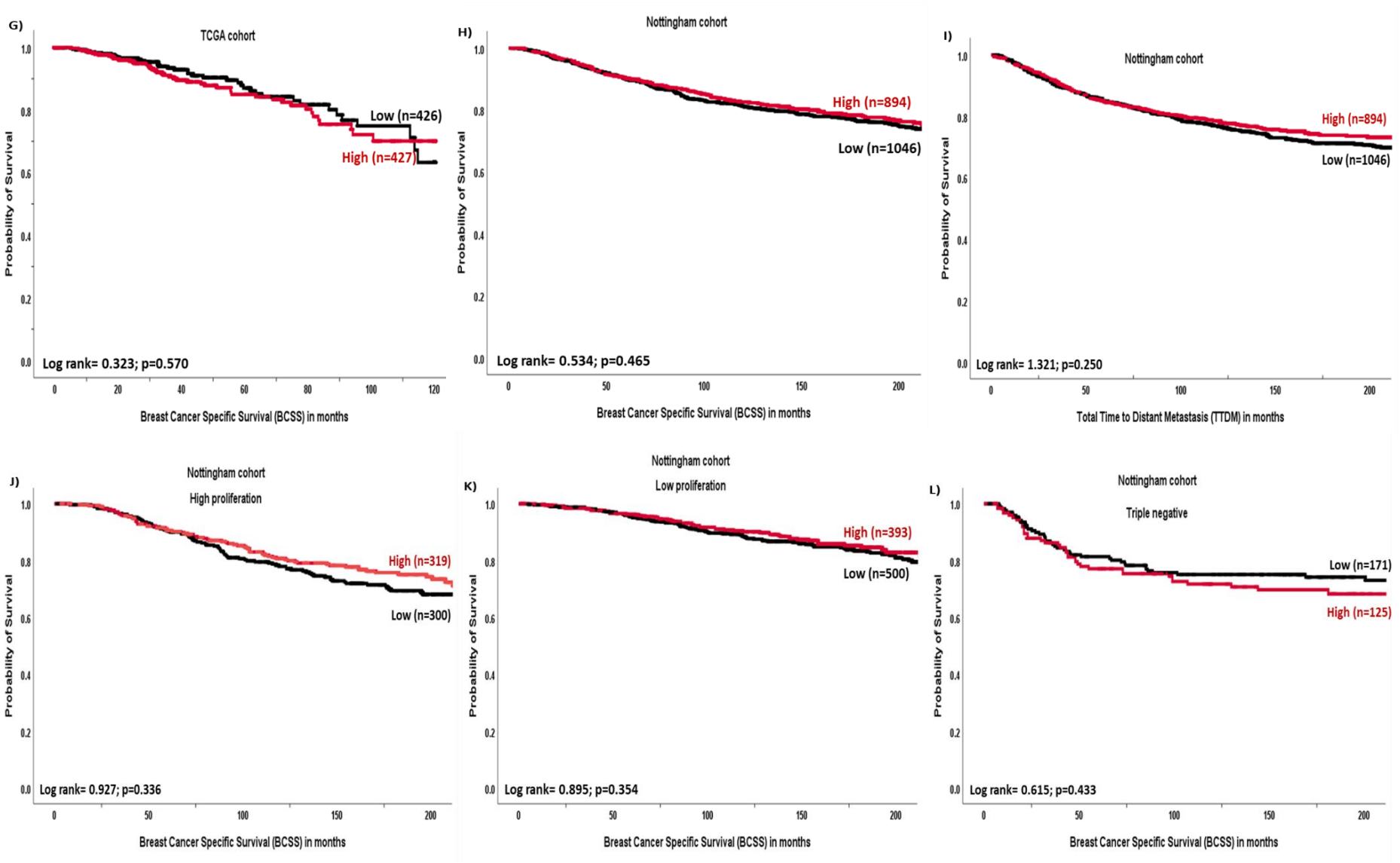
In the METABRIC cohort, survival analyses of *RRM2, CCNB1,* and *CCNB2* mRNA showed that overexpression of *RRM2*, *CCNB1,* and *CCNB2* was associated with shorter BCSS (all p<0.0001) (Figures 6.4A, 6.5A, and 6.6A). In the TCGA cohort, high *CCNB1* and *CCNB2* mRNA expression was significantly associated with poor outcome (p=0.010, *CCNB1*; p<0.0001, *CCNB2*) (Figures 6.5G and 6.6G). However, there was no significant association between the expression of *RRM2* and BCSS in the TCGA cohort (p=0.570) (Figure 6.4G). At the protein level, there was no significant association between the increased expression of RRM2 and CCNB1 and patient outcome (all p>0.05) (Figures 6.4H and 6.5H). The BCSS of BC patients with high CCNB2 protein levels was significantly shorter than patients with a low CCNB2 protein levels in the Nottingham cohort (p=0.042) (Figure 6.6H). RRM2 and CCNB1 protein expression was not associated with shorter TTDM (all p>0.05) (Figures 6.4I and 6.5I). However, increased CCNB2 protein expression was associated with shorter TTDM (p=0.039) (Figure 6.6I).

The analysis was also performed on the intrinsic molecular classes in the METABRIC dataset: high expression of *RRM2*, *CCNB1,* and *CCNB2* mRNA was associated with shorter BCSS in the normal-like subtype (p=0.004, p=0.024, and p<0.0001, respectively) (Figures 6.4F, 6.5F, and 6.6F). In addition, overexpression of *CCNB2* mRNA was associated with shorter BCSS in the HER2-enriched subtype (p<0.0001) (Figure 6.6D); no significant association between overexpression of *RRM2*, *CCNB1,* and *CCNB2* mRNA and shorter BCSS was observed in the other subtypes (all p>0.05) (Figures 6.4B–D, 6.5B–D, and 6.6B–C/E). When the Nottingham cohort was stratified into molecular subtypes, no significant association between the expression of RRM2 and CCNB1 and the molecular subtype was observed (p>0.05) (Figures 6.4J–M and 6.5J–M). However, high expression of CCNB2 was significantly associated with poor outcome in luminal A tumours (p = 0.007) and HER2-positive tumours (p = 0.028), but not in the TN tumours (p = 0.219) (Figure 6.6K–M).

In the METABRIC cohort, multivariate Cox regression analysis showed that *RRM2, CCNB1,* and *CCNB2* mRNA predicted poor BCSS survival independent of tumour size, LN stage, tumour grade, and LVI (HR 1.3; 95%CI=1.02-1.6*;* p=0.030, *RRM2*; HR 1.5; 95%CI=1.2-1.9*;* p<0.0001, *CCNB1,* and HR 1.7; 95%CI=1.3-2.2*;* p<0.0001, *CCNB2*) (Tables 6.11, 6.12, and 6.13). Similar results were observed when the same analysis was performed for *CCNB2* in the TCGA cohort (HR 2.2; 95%CI=1.3-3.7*;* p=0.004), but no significant association was observed for *RRM2* and *CCNB1* (HR 0.8; 95%CI=0.5-1.4*;* p=0.627, *RRM2;* HR 1.6; 95%CI=0.9-2.8*;* p=0.061, *CCNB1*) (Tables 6.11, 6.12, and 6.13). In the Nottingham cohort, Cox regression analysis showed that high expression of CCNB1, but not RRM2, was a significant predictor of shorter BCSS regardless of LVI status, tumour size, LN stage, and tumour grade (HR= 1.3; 95% CI=1.1-1.5*;* p=0.010, CCNB1;HR= 0.8; 95% CI=0.7-1.0*;* p=0.097, RRM2) (Tables 6.11 and 6.12). Moreover, multivariate survival analysis revealed that high expression of CCNB2 was an independent marker of shorter survival,

regardless of the same variables (HR= 1.2; 95% CI=1.1-1.3*;* p=0.045) (Table 6.13).


A)
METABRIC cohort
Low (n=896)
High (n=670)
Log rank= 24.498; p<0.0001
Probability of Survival
Breast Cancer Specific survival (BCSS) in months
B)
METABRIC cohort
Luminal A
Low (n=426)
High (n=105)
Log rank= 3.523; p=0.061
Probability of Survival
Breast Cancer specific survival (BCSS) in months
C)
METABRIC cohort
Luminal B
High (n=230)
Low (n=154)
Log rank= 0.247; p=0.619
Probability of Survival
Breast Cancer specific survival (BCSS) in months
D)
METABRIC cohort
HER2 enriched
Low (n=82)
High (n=112)
Log rank= 0.310; p=0.583
Probability of Survival
Breast Cancer Specific survival (BCSS) in months
E)
METABRIC cohort
Basal-like
Low (n=96)
High (n=190)
Log rank= 0.185; p=0.667
Probability of Survival
Breast Cancer Specific survival (BCSS) in months
F)
METABRIC cohort
Normal-like
Low (n=136)
High (n=31)
Log rank= 8.440; p=0.004
Probability of Survival
Breast Cancer Specific survival (BCSS) in months


G)
TCGA cohort
Low (n=426)
High (n=427)
Log rank= 0.323; p=0.570
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
H)
Nottingham cohort
High (n=894)
Low (n=1046)
Log rank= 0.534; p=0.465
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
I)
Nottingham cohort
High (n=894)
Low (n=1046)
Log rank= 1.321; p=0.250
Probability of Survival
Total Time to Distant Metastasis (TTDM) in months
J)
Nottingham cohort
High proliferation
High (n=319)
Low (n=300)
Log rank= 0.927; p=0.336
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
K)
Nottingham cohort
Low proliferation
High (n=393)
Low (n=500)
Log rank= 0.895; p=0.354
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
L)
Nottingham cohort
Triple negative
Low (n=171)
High (n=125)
Log rank= 0.615; p=0.433
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months

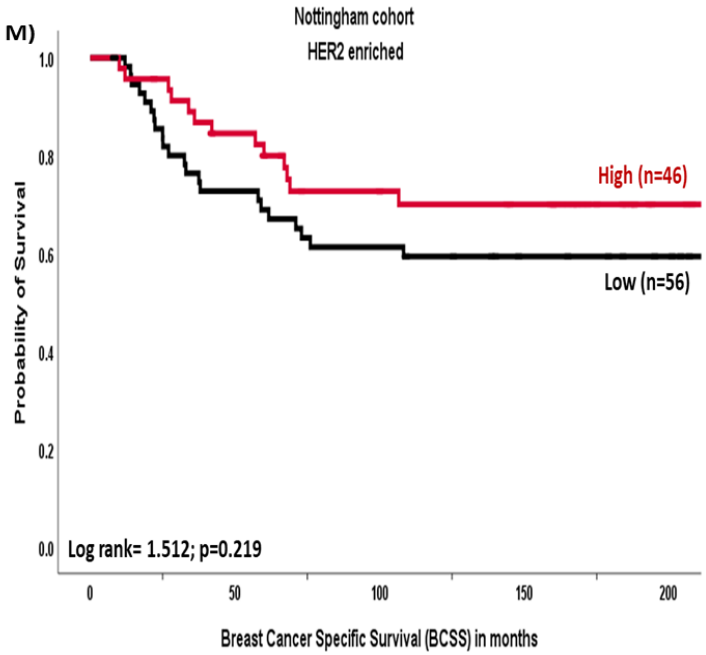
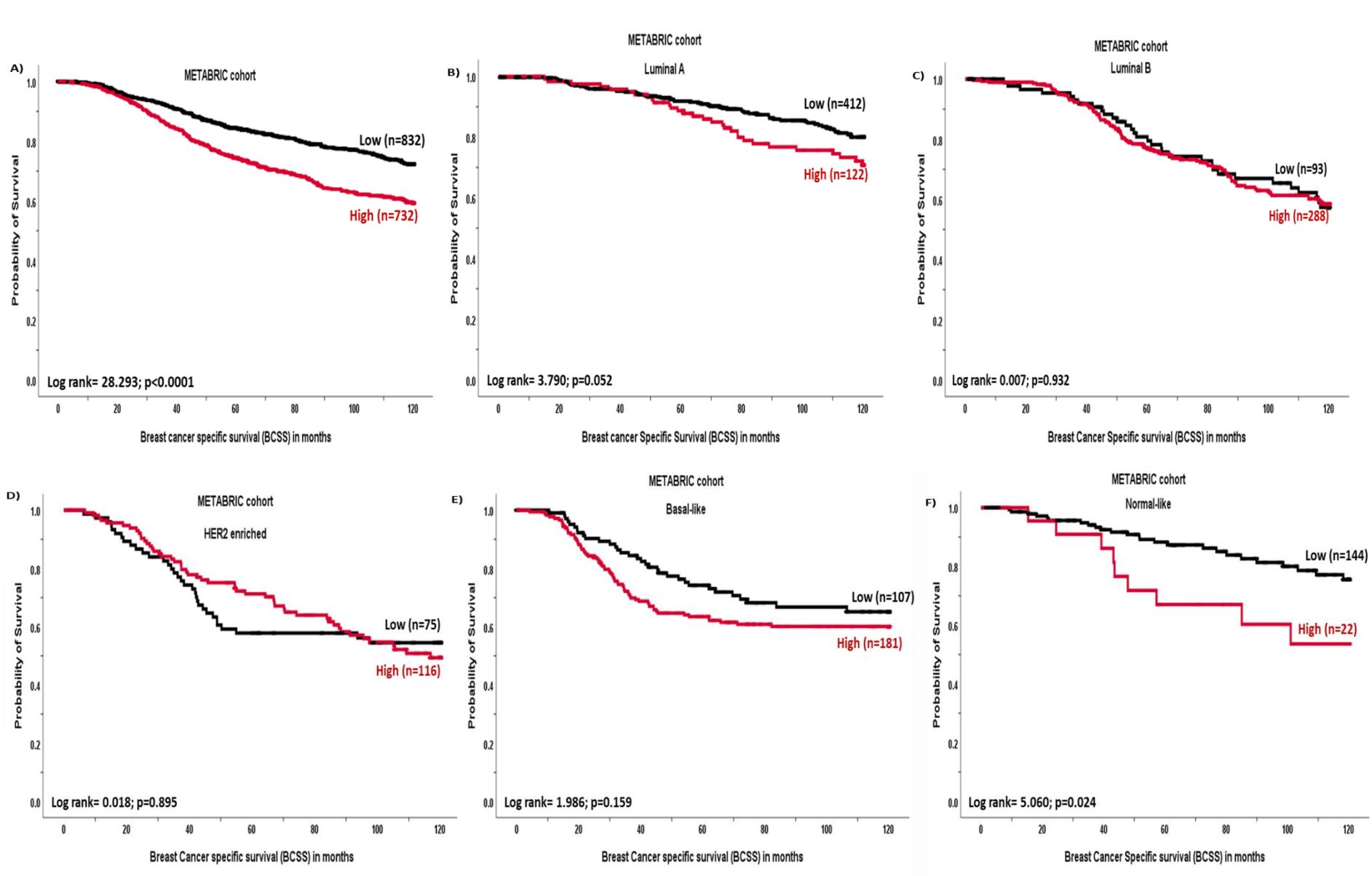
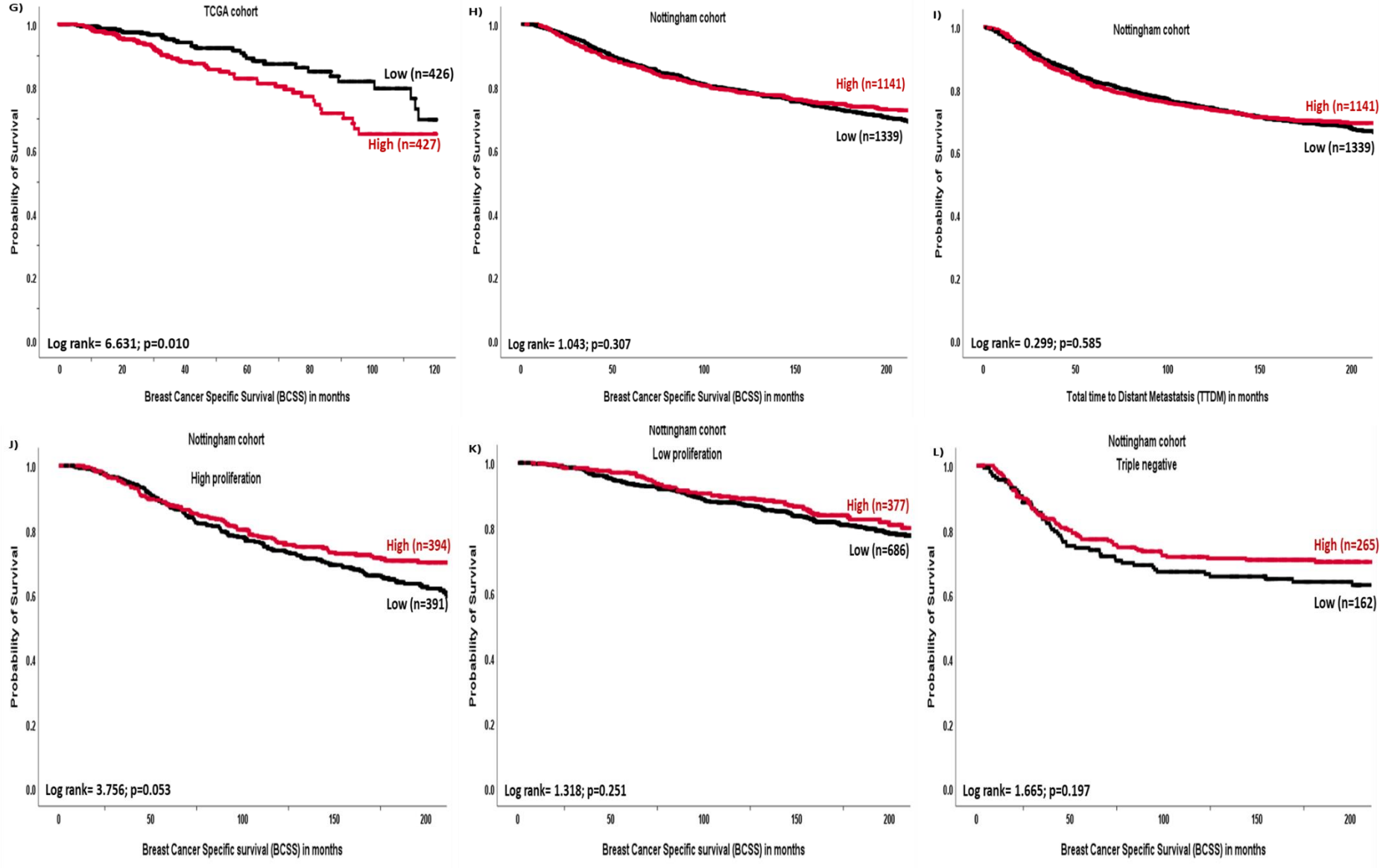


Figure 6.4 Kaplan–Meier survival plots showing the association between *RRM2* mRNA expression and breast cancer-specific survival (BCSS) in A) the whole METABRIC cohort, B) luminal A (METABRIC), C) luminal B (METABRIC), D) HER2-enriched (METABRIC), E) basal-like (METABRIC), F) normal-like (METABRIC), and G) the whole TCGA cohort. For Nottingham cohort data, associations are shown between RRM2 protein expression and H) BCSS, I) total time to distant metastasis (TTDM), J) BCSS in high-proliferation tumours, K) BCSS in low-proliferation tumours, L) BCSS in triple-negative tumours, and M) BCSS in HER2-enriched tumours.


A)
METABRIC cohort
Low (n=832)
High (n=732)
Log rank= 28.293; p<0.0001
Probability of Survival
Breast cancer specific survival (BCSS) in months
B)
METABRIC cohort
Luminal A
Low (n=412)
High (n=122)
Log rank= 3.790; p=0.052
Probability of Survival
Breast cancer Specific Survival (BCSS) in months
C)
METABRIC cohort
Luminal B
Low (n=93)
High (n=288)
Log rank= 0.007; p=0.932
Probability of Survival
Breast Cancer Specific survival (BCSS) in months
D)
METABRIC cohort
HER2 enriched
Low (n=75)
High (n=116)
Log rank= 0.018; p=0.895
Probability of Survival
Breast Cancer specific survival (BCSS) in months
E)
METABRIC cohort
Basal-like
Low (n=107)
High (n=181)
Log rank= 1.986; p=0.159
Probability of Survival
Breast Cancer specific survival (BCSS) in months
F)
METABRIC cohort
Normal-like
Low (n=144)
High (n=22)
Log rank= 5.060; p=0.024
Probability of Survival
Breast Cancer Specific survival (BCSS) in months


G)
TCGA cohort
Low (n=426)
High (n=427)
Probability of Survival
Log rank= 6.631; p=0.010
Breast Cancer Specific Survival (BCSS) in months
H)
Nottingham cohort
High (n=1141)
Low (n=1339)
Probability of Survival
Log rank= 1.043; p=0.307
Breast Cancer Specific Survival (BCSS) in months
I)
Nottingham cohort
High (n=1141)
Low (n=1339)
Probability of Survival
Log rank= 0.299; p=0.585
Total time to Distant Metastatsis (TTDM) in months
J)
Nottingham cohort
High proliferation
High (n=394)
Low (n=391)
Probability of Survival
Log rank= 3.756; p=0.053
Breast Cancer Specific survival (BCSS) in months
K)
Nottingham cohort
Low proliferation
High (n=377)
Low (n=686)
Probability of Survival
Log rank= 1.318; p=0.251
Breast Cancer Specific survival (BCSS) in months
L)
Nottingham cohort
Triple negative
High (n=265)
Low (n=162)
Probability of Survival
Log rank= 1.665; p=0.197
Breast Cancer Specific Survival (BCSS) in months

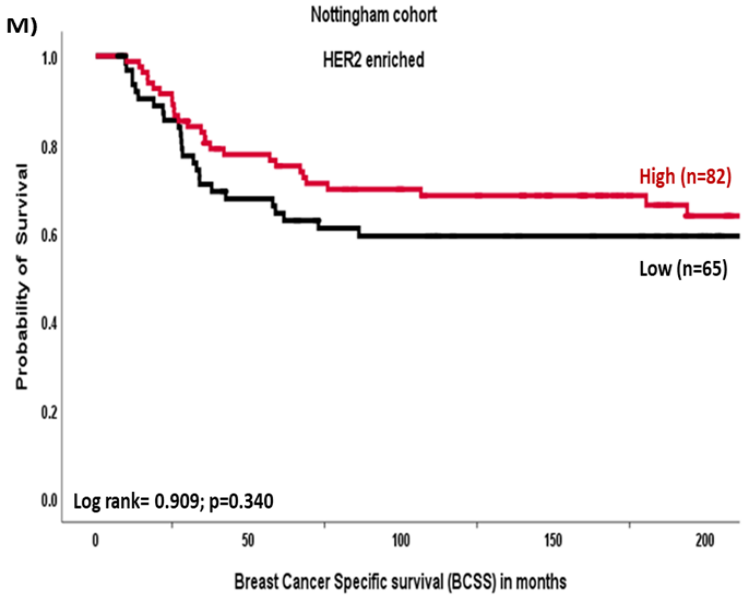
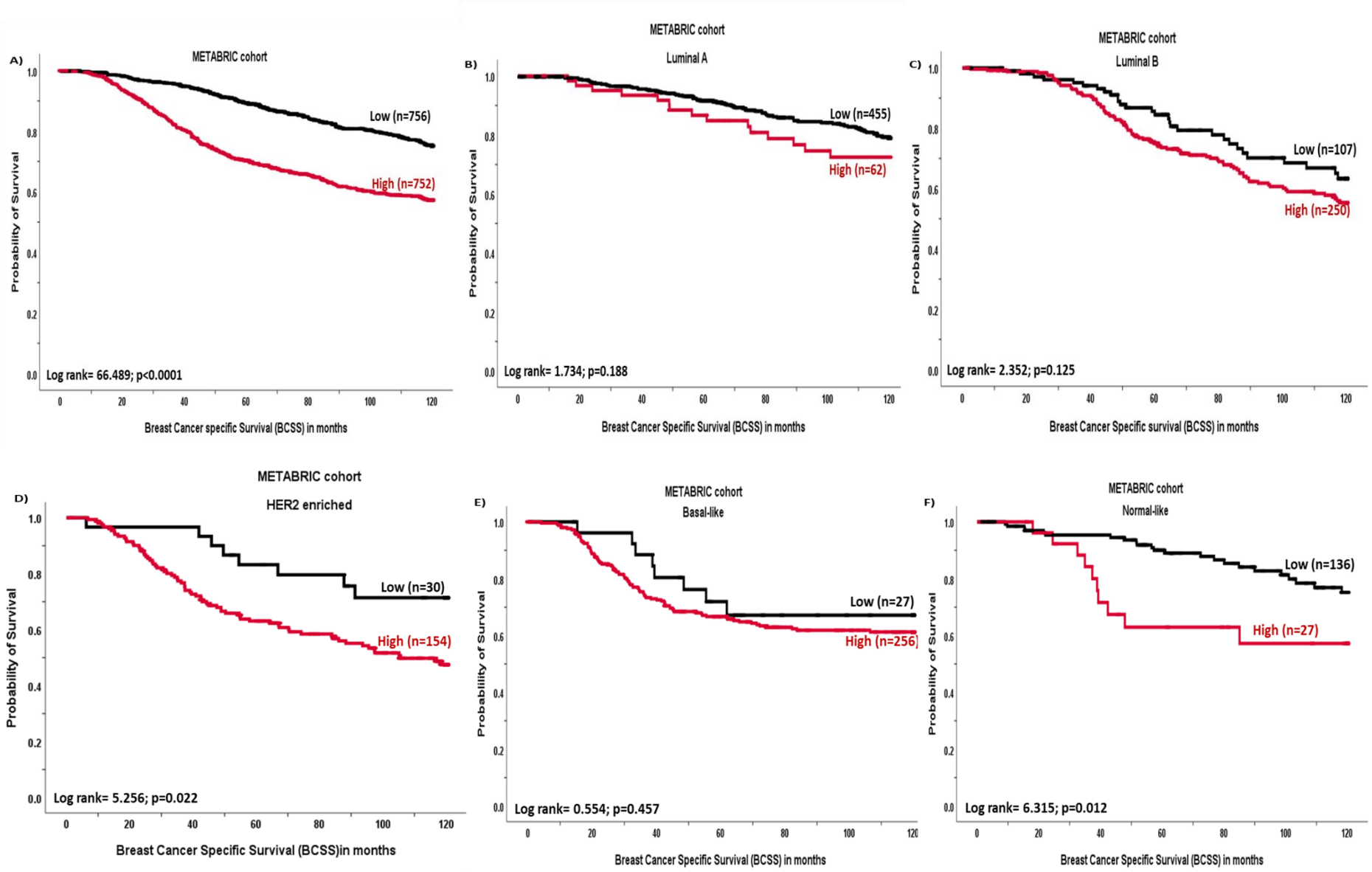
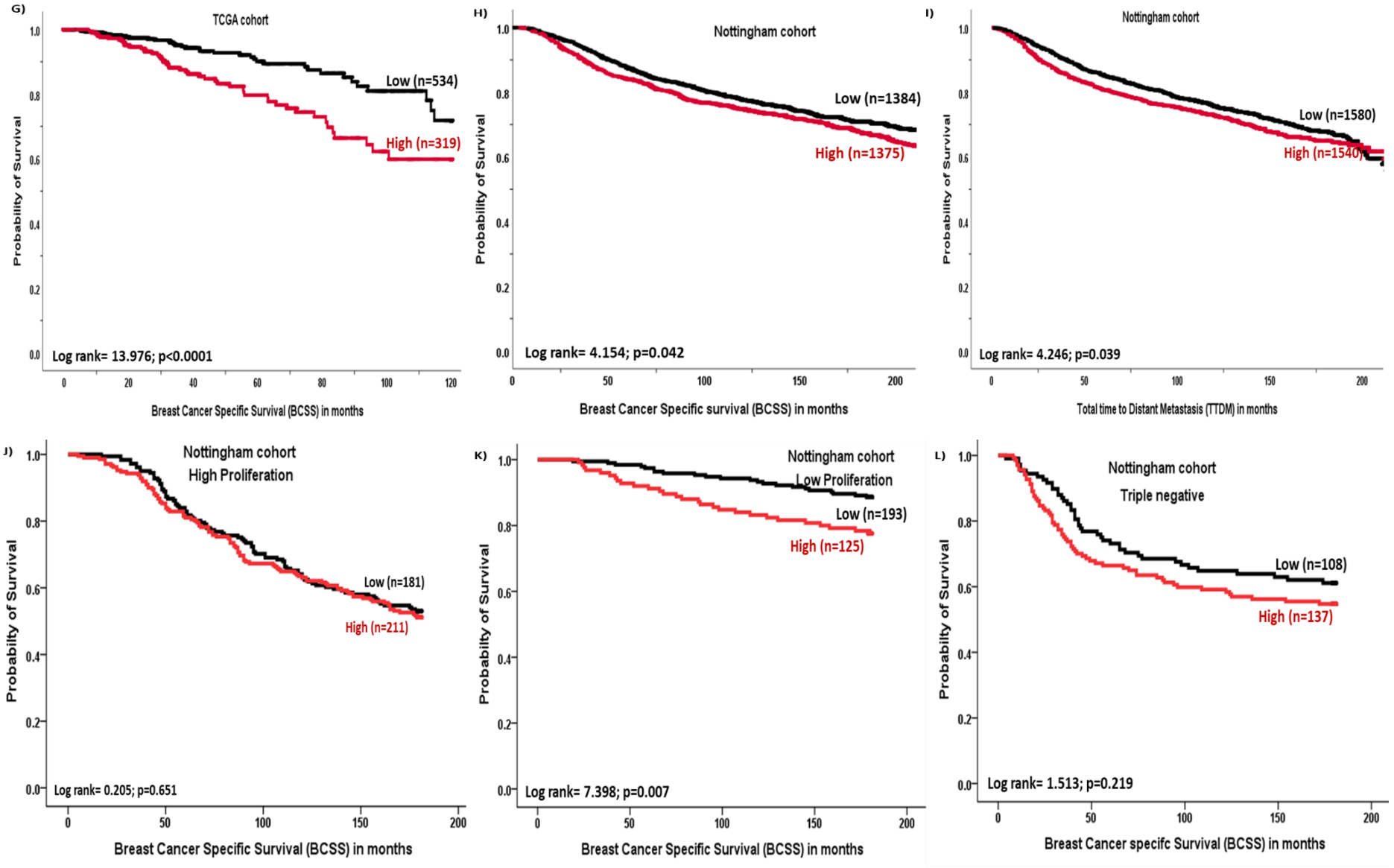


Figure 6.5 Kaplan–Meier survival plots showing the association between *CCNB1* mRNA expression and breast cancer-specific survival (BCSS) in A) the whole METABRIC cohort, B) luminal A (METABRIC), C) luminal B (METABRIC), D) HER2-enriched (METABRIC), E) basal-like (METABRIC), F) normal-like (METABRIC), and G) the whole TCGA cohort. For Nottingham cohort data, associations are shown between CCNB1 protein expression and H) BCSS, I) total time to distant metastasis (TTDM), J) BCSS in high-proliferation tumours, K) BCSS in low-proliferation tumours, L) BCSS in triple-negative tumours, and M) BCSS in HER2-enriched tumours.


A)
METABRIC cohort
Low (n=756)
High (n=752)
Log rank= 66.489; p<0.0001
Probability of Survival
Breast Cancer specific Survival (BCSS) in months
B)
METABRIC cohort
Luminal A
Low (n=455)
High (n=62)
Log rank= 1.734; p=0.188
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
C)
METABRIC cohort
Luminal B
Low (n=107)
High (n=250)
Log rank= 2.352; p=0.125
Probability of Survival
Breast Cancer Specific survival (BCSS) in months
D)
METABRIC cohort
HER2 enriched
Low (n=30)
High (n=154)
Log rank= 5.256; p=0.022
Probability of Survival
Breast Cancer Specific Survival (BCSS)in months
E)
METABRIC cohort
Basal-like
Low (n=27)
High (n=256)
Log rank= 0.554; p=0.457
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months
F)
METABRIC cohort
Normal-like
Low (n=136)
High (n=27)
Log rank= 6.315; p=0.012
Probability of Survival
Breast Cancer Specific Survival (BCSS) in months


G)
TCGA cohort
Low (n=534)
High (n=319)
Probability of Survival
Log rank= 13.976; p<0.0001
Breast Cancer Specific Survival (BCSS) in months
H)
Nottingham cohort
Low (n=1384)
High (n=1375)
Probability of Survival
Log rank= 4.154; p=0.042
Breast Cancer Specific survival (BCSS) in months
I)
Nottingham cohort
Low (n=1580)
High (n=1540)
Probability of Survival
Log rank= 4.246; p=0.039
Total time to Distant Metastasis (TTDM) in months
J)
Nottingham cohort
High Proliferation
Low (n=181)
High (n=211)
Probabilty of Survival
Log rank= 0.205; p=0.651
Breast Cancer Specific Survival (BCSS) in months
K)
Nottingham cohort
Low Proliferation
Low (n=193)
High (n=125)
Probabilty of Survival
Log rank= 7.398; p=0.007
Breast Cancer Specific Survival (BCSS) in months
L)
Nottingham cohort
Triple negative
Low (n=108)
High (n=137)
Probabilty of Survival
Log rank= 1.513; p=0.219
Breast Cancer specifc Survival (BCSS) in months

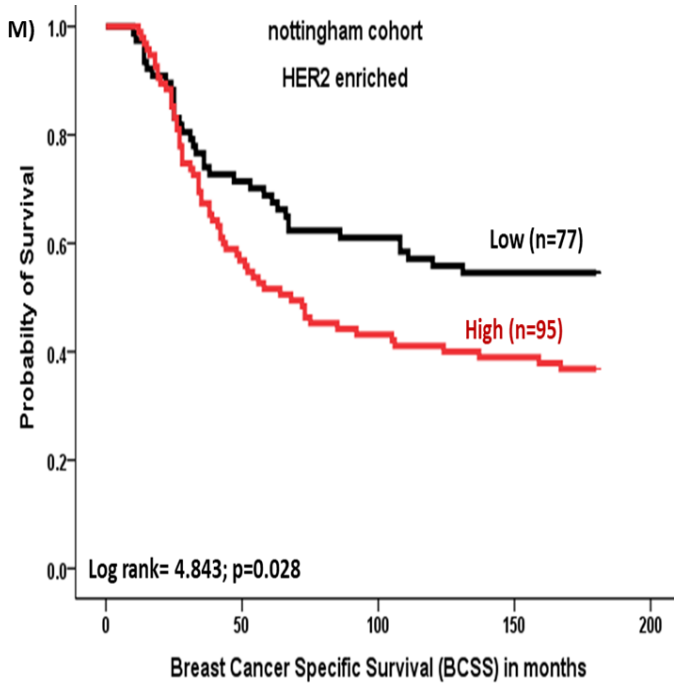


Figure 6.6 Kaplan–Meier survival plots showing the association between *CCNB2* mRNA expression and breast cancer-specific survival (BCSS) in A) the whole METABRIC cohort, B) luminal A (METABRIC), C) luminal B (METABRIC), D) HER2-enriched (METABRIC), E) basal-like (METABRIC), F) normal-like (METABRIC), and G) the whole TCGA cohort. For Nottingham cohort data, associations are shown between CCNB2 protein expression and H) BCSS, I) total time to distant metastasis (TTDM), J) BCSS in high-proliferation tumours, K) BCSS in low-proliferation tumours, L) BCSS in triple-negative tumours, and M) BCSS in HER2-enriched tumours.

Table 6.11 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *RRM2* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort.

METABRIC cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
***RRM2* mRNA expression**	1.280	1.024	1.601	**0.030**
Tumour size	1.410	1.071	1.856	**0.014**
Lymph node stage	1.757	1.510	2.045	**<0.0001**
Tumour grade	1.337	1.093	1.637	**0.005**
Lymphovascular invasion (LVI)	1.384	1.093	1.753	**0.007**
TCGA cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
***RRM2* mRNA expression**	0.886	0.543	1.444	0.627
Tumour size	1.473	0.884	2.570	0.173
Lymph node	1.210	0.715	2.046	0.478
Tumour grade	1.329	0.914	1.934	0.137
Lymphovascular invasion (LVI)	1.944	1.180	3.201	**0.009**
Nottingham BC cohort (protein)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
RRM2 protein expression	0.849	0.699	1.030	0.097
Tumour size	1.522	1.237	1.871	**<0.0001**
Lymph node stage	1.757	1.530	2.018	**<0.0001**
Tumour grade	1.714	1.447	2.031	**<0.0001**
Lymphovascular invasion (LVI)	1.602	1.302	1.973	**<0.0001**

***p*-values in bold are statistically significant**

Table 6.12 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *CCNB1* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort.

METABRIC cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
***CCNB1* mRNA expression**	1.514	1.209	1.895	**<0.0001**
Tumour size	1.448	1.104	1.899	**0.007**
Lymph node stage	1.788	1.537	2.080	**<0.0001**
Tumour grade	1.304	1.066	1.595	**0.010**
Lymphovascular invasion (LVI)	1.393	1.100	1.764	**0.006**
TCGA cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
***CCNB1* mRNA expression**	1.641	0.977	2.758	0.061
Tumour size	1.416	0.812	2.467	0.220
Lymph node	1.279	0.757	2.158	0.357
Tumour grade	1.079	0.740	1.573	0.692
Lymphovascular invasion (LVI)	1.836	1.118	3.014	**0.016**
Nottingham BC cohort (protein)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
CCNB1 protein expression	1.300	1.101	1.500	**0.010**
Tumour size	1.439	1.217	1.701	**<0.0001**
Lymph node stage	1.754	1.565	1.965	**<0.0001**
Tumour grade	1.750	1.524	2.009	**<0.0001**
Lymphovascular invasion (LVI)	1.423	1.200	1.688	**<0.0001**

***p*-values in bold are statistically significant**

Table 6.13 Multivariate Cox regression analysis for predictors of breast cancer-specific survival (BCSS) and *CCNB2* mRNA expression in the METABRIC and TCGA cohorts and protein expression in the Nottingham BC cohort.

METABRIC cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
***CCNB2* mRNA expression**	1.719	1.342	2.202	**<0.0001**
Tumour size	1.367	1.035	1.805	**0.028**
Lymph node stage	1.810	1.551	2.113	**<0.0001**
Tumour grade	1.237	0.996	1.538	**0.050**
Lymphovascular invasion (LVI)	1.373	1.080	1.745	**0.010**
TCGA cohort (mRNA)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
***CCNB2* mRNA expression**	2.164	1.278	3.664	**0.004**
Tumour size	1.340	0.765	2.347	0.307
Lymph node	1.258	0.747	2.119	0.388
Tumour grade	0.947	0.645	1.389	0.780
Lymphovascular invasion (LVI)	1.821	1.110	2.986	**0.018**
Nottingham BC cohort (protein)				
Parameters	**Hazard ratio (HR)**	**95% confidence interval (CI)**		**Significance *p*-value**
		Lower	**Upper**	
CCNB2 protein expression	1.200	1.100	1.300	**0.045**
Tumour size	1.511	1.311	1.743	**<0.0001**
Lymph node stage	1.700	1.540	1.876	**<0.0001**
Tumour grade	1.533	1.377	1.708	**<0.0001**
Lymphovascular invasion (LVI)	1.491	1.291	1.721	**<0.0001**

***p*-values in bold are statistically significant**

6.5.8 Preclinical studies (*in vitro* investigation of the role of *CCNB2* in LVI-related biological processes)

6.5.8.1 Suppression of *CCNB2* expression using siRNA

The expression of CCNB2 was variable among luminal A-like (MCF-7), luminal B-like (ZR-75-1), HER2-enriched (SK-BR-3), and TN (MDA-MB-231) cell lines, with the highest levels observed in MCF-7, followed by SK-BR-3, ZR-75-1 lines, and the lowest levels in the MDA-MB-231 line (Figure 6.7A/B). Using WB analysis, complete loss of CCNB2 protein expression in MCF-7, ZR-75-1, and SK-BR-3 was observed when comparing *CCNB2* knockdown expression relative to GAPDH expression at post-transfection days 3, 5, and 7 (Figure 6.7C–H).

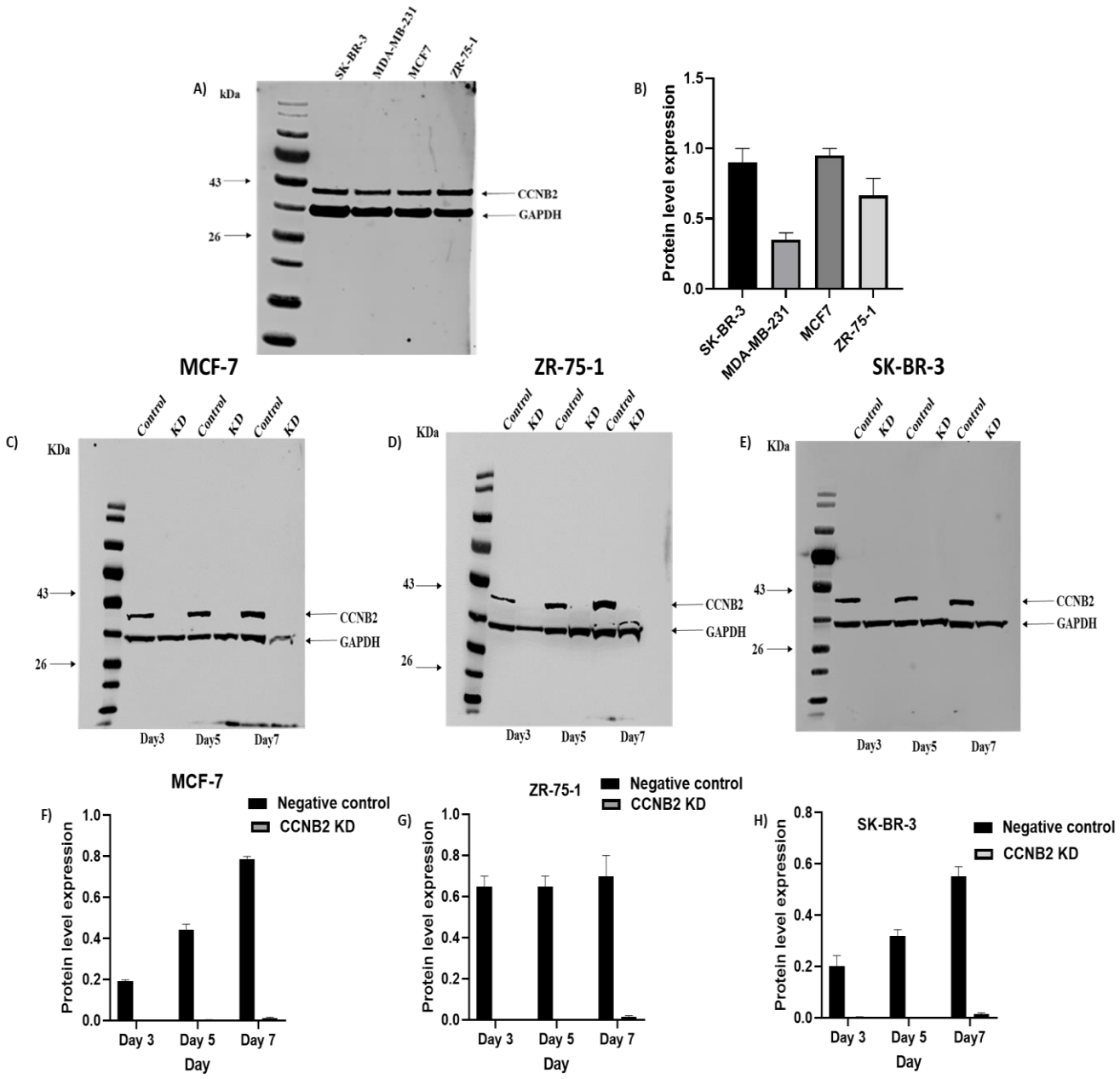


Figure 6.7 CCNB2 protein expression in breast cancer (BC) cell lines. A) Evaluation of differential expression of CCNB2 in BC cell lines by Western blotting, B) quantification of CCNB2 protein level expression in SK-BR-3, MDA-MB-231, MCF-7, and ZR-75-1 cell lines. Using pre-validated Silencer Select siRNA constructs (mainly for *CCNB2*) or scrambled negative-control siRNA, silencing of *CCNB2* using the forward transfection method showed complete knockdown relative to non-targeting, scrambled control siRNA in C) MCF-7, D) ZR-75-1, and E) SK-BR-3. Quantification of CCNB2 protein expression in the transfected cell and negative control by densitometry and normalisation to GAPDH levels revealed complete loss of CCNB2 protein expression in F) MCF-7, G) ZR-75-1, and H) SK-BR-3 cell lines. Each bar represents the mean ± standard error of the mean (SEM) of three or more independent experiments. GAPDH was included as a loading control.

6.5.8.2 Downregulation of *CCNB2* suppresses cell proliferation and cell growth in the selected BC cell lines

CCNB2 knockdown significantly reduced the proliferation rates of the following cell lines, compared to the control: MCF-7 cells at T24h ($p=0.7631$), T48h ($p=0.0003$), and T72h ($p=0.0343$) (Figure 6.8A); ZR-75-1 cells at T24h ($p=0.0016$), T48h ($p<0.0001$), and T72h ($p=0.0015$) (Figure 6.8B); and SK-BR-3 cells at T24h ($p=0.0296$), T48h ($p<0.0001$), and T72h ($p=0.0202$) (Figure 6.8C). The ability of a single cell colony to survive after exposure to the transfection agent was significantly weaker than the control in MCF-7 ($p=0.0075$), ZR-75-1 ($p=0.0034$), and SK-BR-3 cells ($p=0.0198$) (Figure 6.8D–F).

6.5.8.3 Downregulation of *CCNB2* suppresses cell migration in the selected BC cell lines

Knockdown of *CCNB2* resulted in a larger, unhealed wound than in the negative controls in MCF-7, ZR-75-1, and SK-BR-3 cell lines ($p=0.0377$, $p=0.0028$, and $p=0.0344$, respectively) (Figure 6.9A–C).

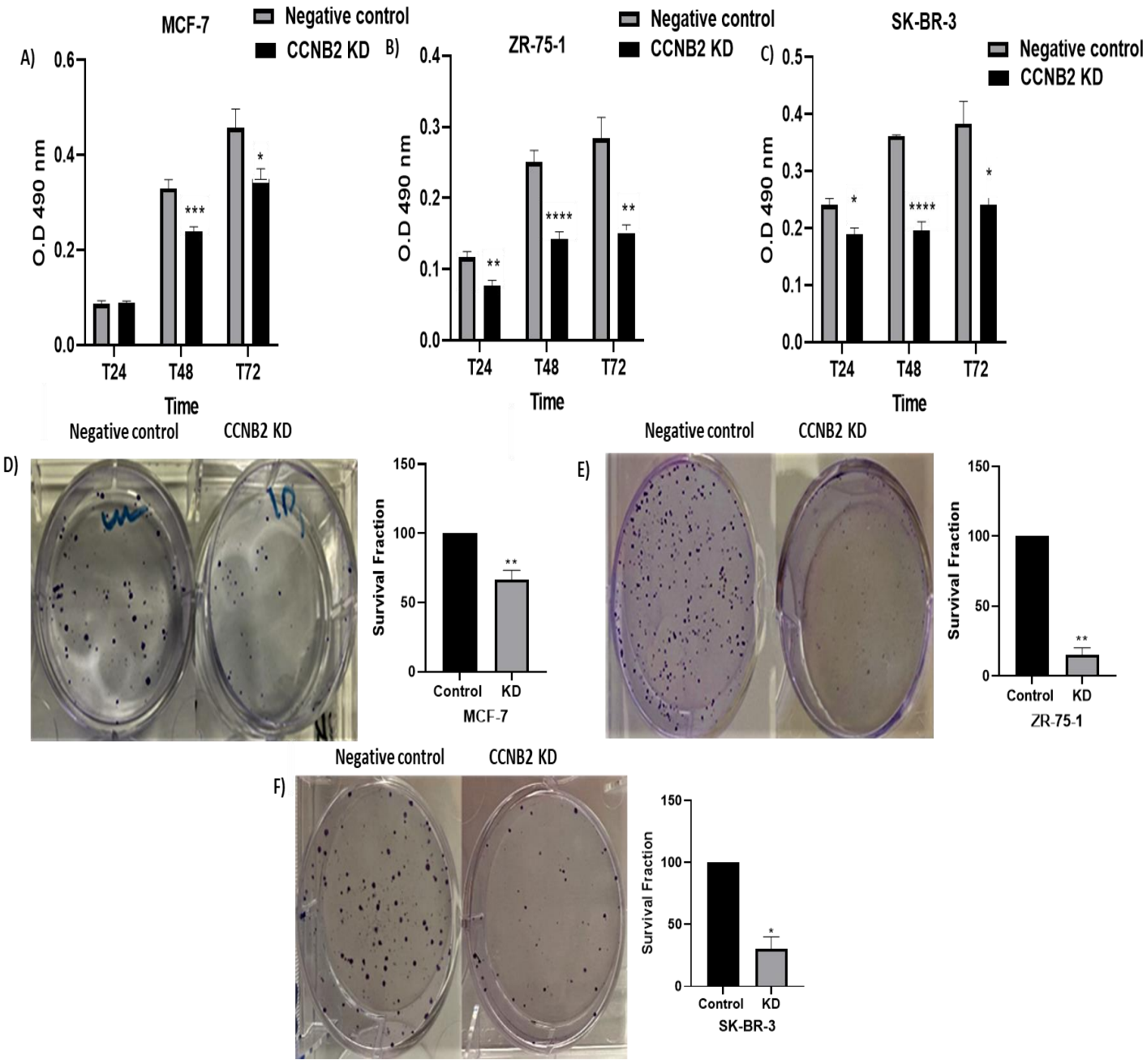


Figure 6.8 The effect of *CCNB2* knockdown (KD) by siRNA on cell proliferation in A) MCF-7, B) ZR-75-1, and C) SK-BR-3 cells. (A–C) Cell proliferation was significantly reduced after KD in BC cell lines, as detected by MTS assay. *CCNB2* siRNA transfection reduced the ability of BC cell lines to colonise in D) MCF-7, E) ZR-75-1, and F) SK-BR-3, as detected by the colony formation assay. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

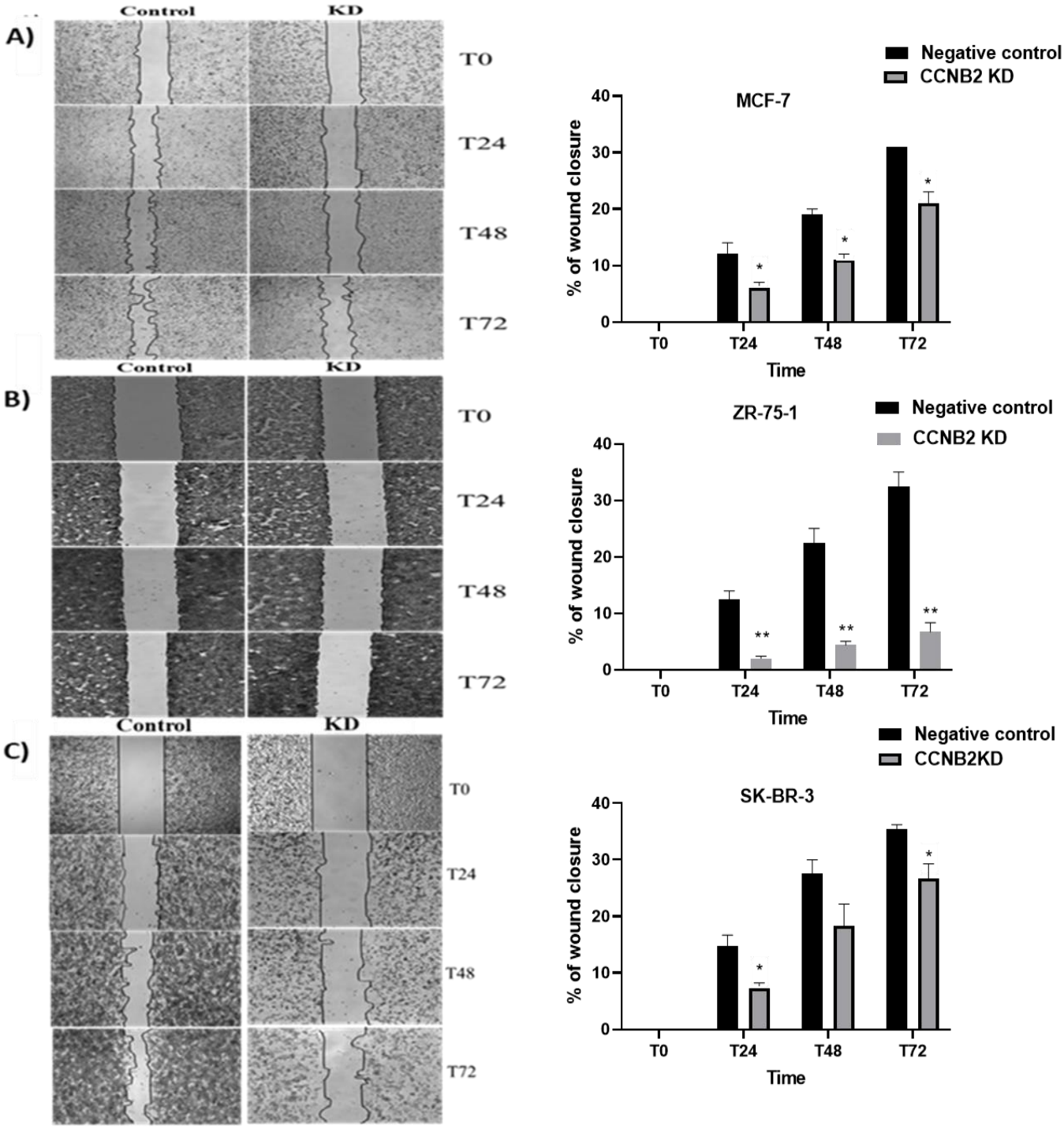


Figure 6. 9 The effect of *CCNB2* knockdown (KD) by siRNA on cell migration in A) MCF-7, B) ZR-75-1, and C) SK-BR-3 cells. The wound repair rate of *CCNB2* KD and control cells was observed by measuring the width of the unhealed gap at T0h, T24h, T48h, and T72h. (A–C) Silencing *CCNB2* significantly reduced the cell migration rate of BC cell lines, as detected by the wound-healing assay. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

6.5.8.4 Downregulation of *CCNB2* arrests cells in S phase

Knockdown of *CCNB2* caused S-phase arrest in MCF-7, ZR-75-1, and SK-BR-3 lines, inhibiting the transition to G2/M phase. The percentage of MCF-7 cells in the S phase significantly increased in the *CCNB2* knockdown cells (50.5% G1, 22% S, and 25.4% G2/M) as compared with the negative control (69.2% G1, 10.5% S, and 13.4% G2/M) ($p=0.032$). Similar observations were noted with ZR-75-1, where the percentage of cells in the S phase significantly increased in the *CCNB2* knockdown cells (70.3% G1, 15% S, and 14% G2/M) as compared with the control (68.3% G1, 8.7% S, and 12.3% G2/M) ($p=0.9916$ in G1, $p=0.0248$ in S, $p=0.0309$ in G2/M). Similarly, the percentage of SK-BR-3 cells in the S phase significantly increased in the *CCNB2* knockdown cells (46% G1, 35.2% S, and 18.5% G2/M) as compared with the negative control (51% G1, 26.5% S, and 15.7% G2/M) ($p=0.6413$ in G1, $p=0.0102$ in S, $p=0.0299$ in G2/M) (Figure 6.10A–C).

6.5.8.5 Downregulation of *CCNB2* enhances cell apoptosis

The number of apoptotic cells in *CCNB2* knockdown cells increased significantly when compared with the scramble control in early and late apoptosis. In early apoptosis, knockdown vs negative control values were 5.23% vs 2.80% in MCF-7 ($p=0.0307$), 23.2% vs 5% in ZR-75-1 ($p<0.0001$), and 4.55% vs 0.88% in SK-BR-3 ($p=0.0035$). In late apoptosis, the knockdown vs control values were 2.87% vs 0.91% in MCF-7 ($p=0.0241$), 12.1% vs 4.93% in ZR-75-1 ($p=0.0132$), and 9.77% vs 5.86% in SK-BR-3 ($p=0.0100$) (Figure 6.11A–C).

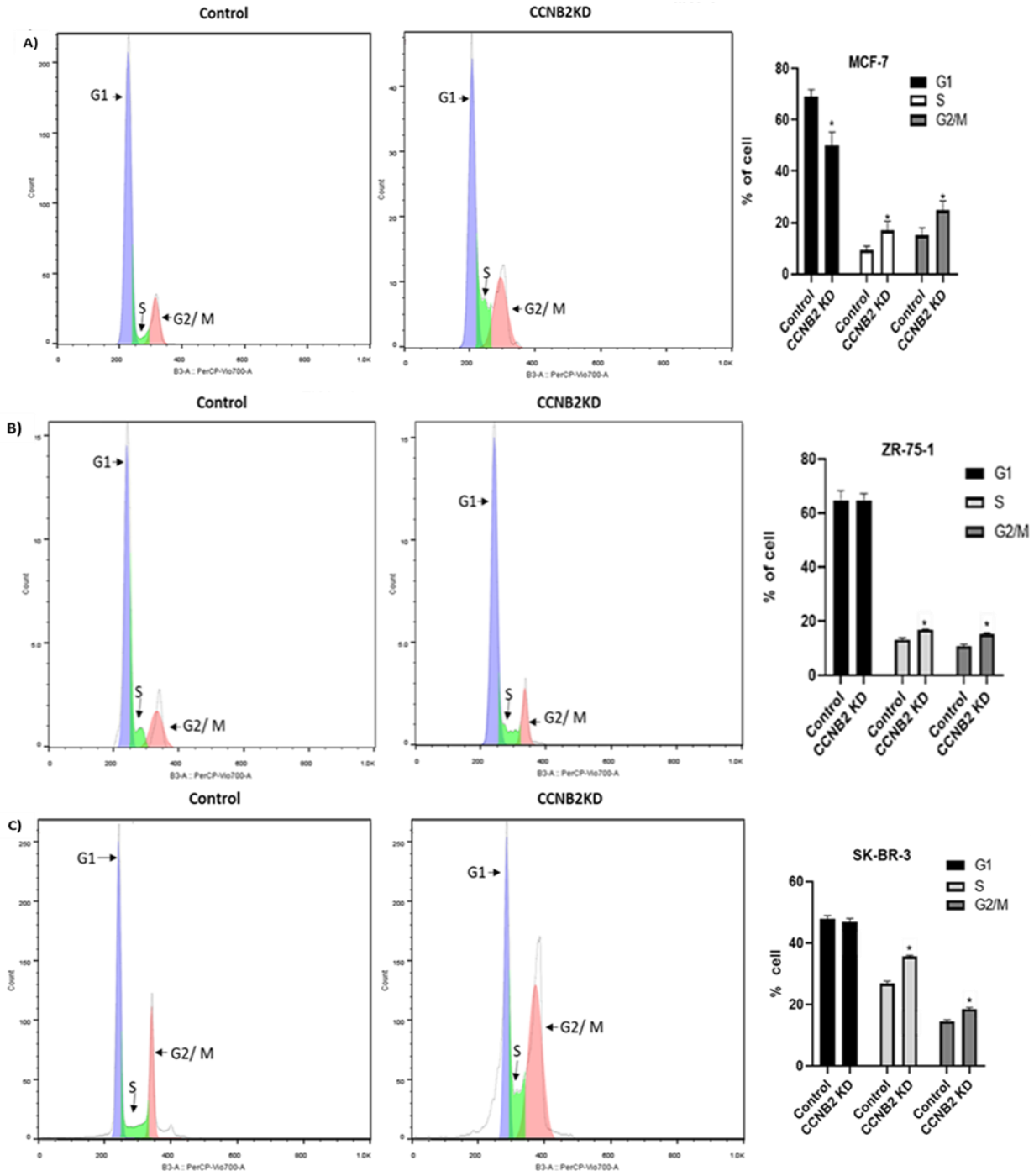


Figure 6.10 The effect of siRNA knockdown (KD) of *CCNB2* on the cell cycle. BC cells showed arrest in the S phase of the cell cycle in A) MCF-7, B) ZR-75-1, and C) SK-BR-3 cells. (A–C) Silencing *CCNB2* inhibited all BC cell lines from entering the G2/M phase, as detected by flow cytometry. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

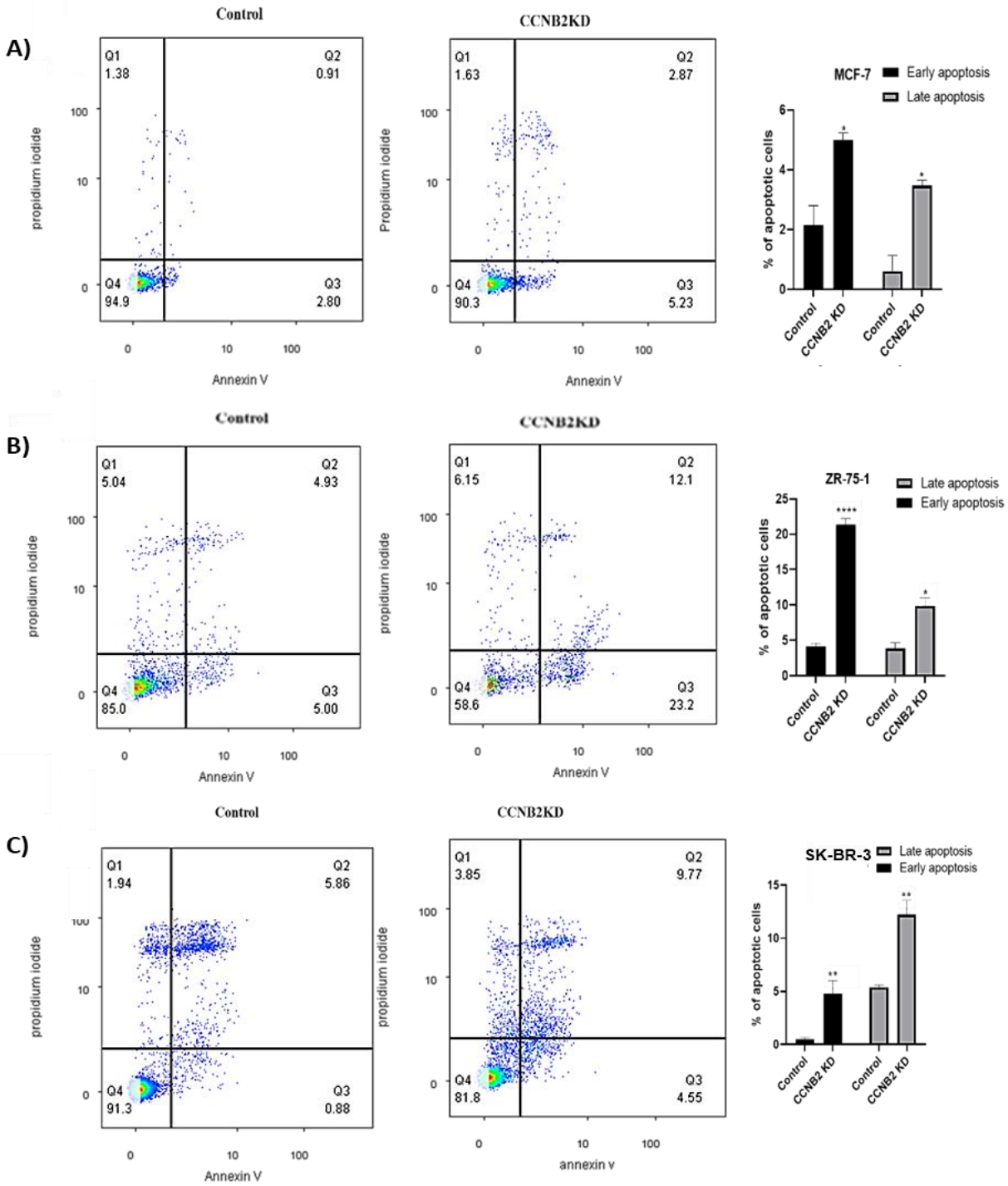


Figure 6.11 The effect of siRNA knockdown (KD) of *CCNB2* on apoptosis. BC cell data showed that KD of *CCNB2* enhanced apoptosis in A) MCF-7, B) ZR-75-1, and C) SK-BR-3 cells. (A–C) More apoptotic cells were observed in the transfected cells than in the control ones, as detected by flow cytometry. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

6.5.8.6 *CCNB2* enhances BC cell adhesion to endothelial cells

All MCF-7, ZR-75-1, and SK-BR-3 cells lines showed higher percentage adherence to both HUVECs (p=0.0003, p=0.0298, and p=0.0377, respectively) and DLECs (p=0.0003, p=0.0026, and p=0.0171, respectively) than to the *CCNB2* transfected cells (Figure 6.12A–C).

6.5.8.7 *CCNB2* enhances BC cell transmigration to endothelial cells

The number of the tumour cells that transmigrated across endothelial cells in the control group was higher than in the knockdown group for MCF-7 cells transmigrating across HUVECs (p=0.0240) and DLECs (p=0.0170), ZR-75-1 cells across HUVECs (p=0.0471 and DLECs (p=0.0002), and SK-BR-3 cells across HUVECs (p=0.0077) and DLECs (p=0.0091) (Figure 6.13A/B).

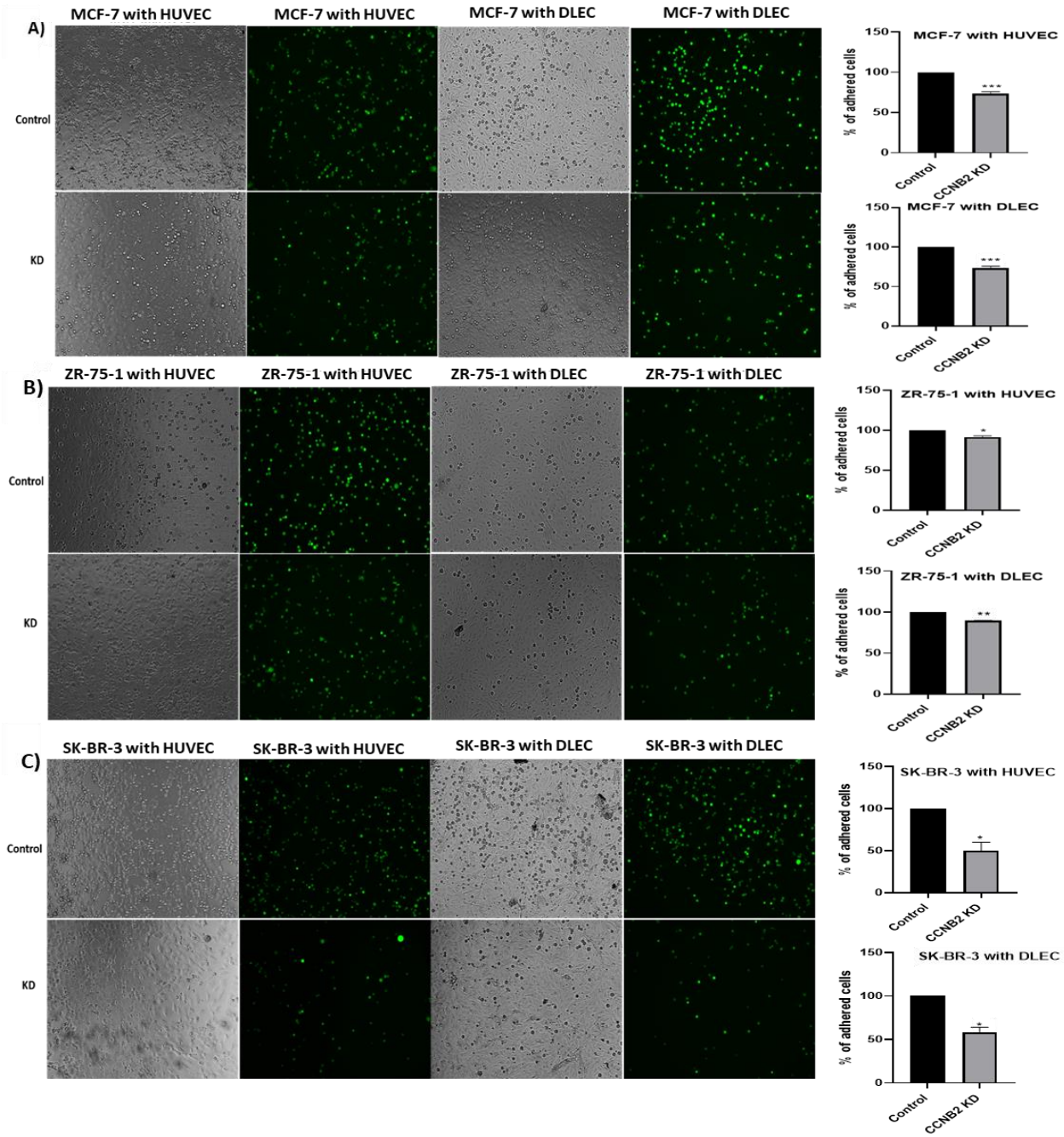


Figure 6.12 Representative photomicrographs of tumour cell adhesion across vascular and lymphatic endothelial cells (HUVECs and DLECs). A) MCF-7, B) ZR-75-1, and C) SK-BR-3 cells. (A–C) Silencing *CCNB2* decreased the number of BC cell lines that adhered to HUVECs and DLECs. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =<0.0001.

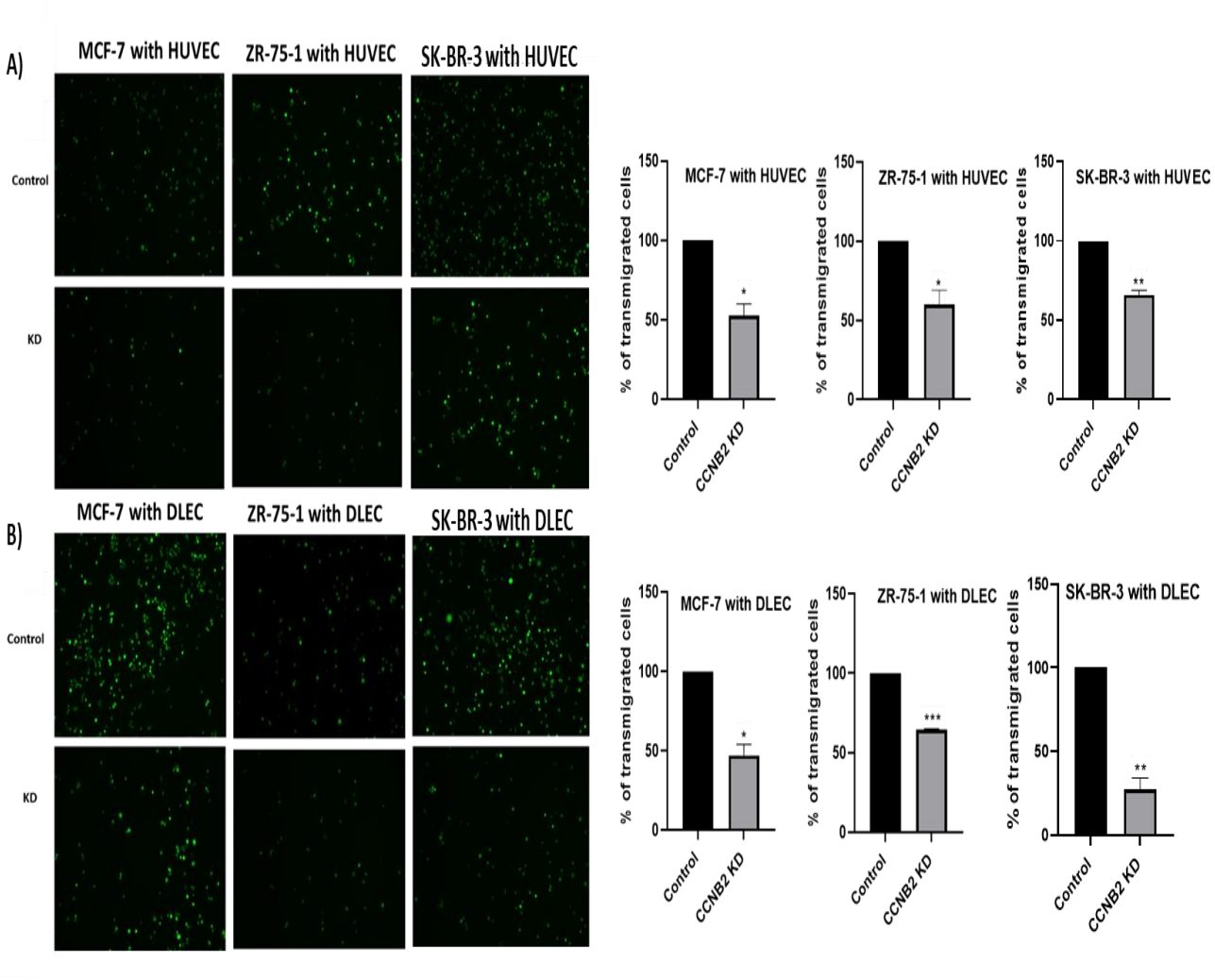


Figure 6.13 Representative photomicrographs of tumour cell (MCF-7, ZR-75-1, and SK-BR-3) transmigration across A) HUVECs and B) DLECs. (A/B) The number of tumour cells that transmigrated across HUVECs and DLECs was higher in the control group than in the KD group. Results are presented as mean ± standard error of the mean (SEM) of three independent experiments. The p-values are * =< 0.05, ** =< 0.01, *** =< 0.001, and **** =< 0.0001.

6.6 Discussion

6.6.1 RRM2

In clinical practice, prognostic or predictive criteria, such as LVI, tumour size, LN stage, histological grade, and hormone receptor status, are frequently utilised to predict outcomes and guide the selection of systemic therapy. LVI develops in BC through complicated molecular processes that include both the surrounding milieu and endothelial cells lining the lymphovascular spaces, and alterations in the malignant epithelial cells with the acquisition of invasive and migratory capabilities [308]. Evidence in the literature suggested that, despite the existence of stromal invasion, LVI could be targeted in BC [308]. Therefore, key drivers that may control the mechanism underlying LVI could be identified as targets of LVI, which was the main aim of this study. Through ANN analysis, *RRM2, CCNB1,* and *CNNB2* were associated with the presence of LVI in both the METABRIC and TCGA cohorts, as described in Chapter 3. Several preclinical studies have demonstrated the potential role of *RRM2* in cancer progression in many type of cancers, including multiple myeloma [309], glioblastoma [310], and BC [311]; however, this is the first study to date conducting a comprehensive analysis of the correlation between *RRM2* and clinicopathological parameters and patient outcomes at the transcriptomic and protein levels, with emphases on whether *RRM2* is correlated with LVI.

The results of *RRM2* evaluation in this study revealed that *RRM2* mRNA and protein expression was associated with poor prognostic variables, including large tumour size, high tumour grade, and LN stage, demonstrating its critical role in the progression of BC. At the mRNA level, there was a significant association between high expression of *RRM2* and LVI-positive tumours, poor NPI, hormone receptor negativity, and HER2 positivity. However, investigation at the protein level using a well-characterised invasive BC cohort showed no significant association between RRM2 and the LVI status, NPI, hormone receptor negativity, and HER2 positivity.

Many studies describe factors that might explain the differences between mRNA and protein levels. One of these factors is the cellular steady state, which describes the average protein and/or mRNA level in a cell that remains largely constant through time [312]. Because mRNA levels principally determine protein levels, if cells are going through long-term dynamic processes, such as continual proliferation and differentiation, there will be variation between cellular mRNA and protein levels [312]. Previous evidence has stated that *RRM2* induces increased proliferation in cancer cells and has the potential to function as a proliferation-related oncogene [313]. Therefore, the lack of significance between LVI status and RRM2 at the protein level could be explained by the long-term proliferation processes induced by *RRM2*. Another factor related to the variation between mRNA and protein levels is the possible delay during state transition in the synthesis of mRNA and protein; stimulated transcription does not instantly result in higher protein levels due to the time required for mRNA maturation, export, and translation [312]. In addition, the abundance of transcript isoforms that may be produced from a single gene through alternative splicing significantly complicates comparisons of protein and mRNA levels [312, 314]. Other factors that could contribute to the difference in the significance between LVI and *RRM2* at the mRNA and protein level is the methodologies used for quantifying and statistically analysing gene expression in the METABRIC and TCGA cohorts, as well as the methods for determining LVI in the transcriptomic and proteomic cohorts, as described in Chapter 3. For example, in the Nottingham cohort, LVI status was determined using morphology and IHC; in contrast, cBioPortal H&E-stained slides were used to determine LVI status in the TCGA cohort. Although H&E slides can be used to evaluate LVI, it might be difficult to distinguish LVI-negative cases [209]. High RRM2 protein levels were not significantly associated with BCSS or TTDM; however, there was a significant association with BCSS in the METABRIC cohort, which could be explained by the aforementioned factors. In addition to these factors, the lack of prognostic significance of *RRM2* could be rationalised by the

subjectivity of the H-score approach in interpreting expression in IHC staining sections [236] in addition to the fact that the METABRIC cohort employed whole tissue with many distinct cell types, including stromal cells, which are in the tumoural compartment. This was not the case in the IHC scoring method, as tumour cells only and not stromal cells were scored.

EMT is governed by a complex network of genes, although the exact mechanism by which it is regulated remains unknown [315]. In this study, the correlation between *RRM2* mRNA and protein levels and MMP-related biomarkers, such as MMP9 and EMT-related markers including E-cadherin, N-cadherin, and TWIST2, was explored to elucidate the potential role of RRM2 in BC development and metastasis. Although the results showed a weak correlation between RRM2 and MMP- and EMT-related markers, the correlation was statistically significant which indicates that these markers are contributing to the same oncogenic pathway, which corroborates previous studies; these multiple associations could collectively play a biological role in LVI process. It has been reported that overexpression of RRM2 is associated with tumour invasiveness, malignant transformation, and metastatic potential in many human cancer cells [316]. Studies of pancreatic cancer cells revealed that overexpression of RRM2 may promote invasion via NF-kB, whereas RNA interference-mediated knockdown of *RRM2* reduced cell invasion and MMP9 activity [317]. E-cadherin dysregulation is regarded as the initiating event of EMT. Han et al. discovered that overexpression of RRM2 induced EMT in nasopharyngeal carcinoma (NPC) cells, which lead to E-cadherin downregulation and promoted NPC cell motility and invasion [318, 319].

This study also revealed a correlation between the EMT biomarkers, such as ZEB1 and ZEB2. Increased ZEB1 and ZEB2 expression may facilitate EMT-dependent invasion and cancer stem cell activity, which are critical for metastatic cancer dissemination [320, 321]. It has been shown that RRM2-induced EMT in NPC cells results in the overexpression of many downstream genes, including ZEB2, and the downregulation of E-cadherin [318]. In

addition, *RRM2 in vivo* experiments induced lung and lymph node metastases in naïve mice [318]. Decreased expression of E-cadherin in prostate cancer *RRM2* cells may contribute to the more aggressive phenotype [285].

All these findings suggest that *RRM2* may play a critical role in the metastatic process. The findings, which partly corroborated those of the preceding investigations, suggest that *RRM2* associated with aggressive clinicopathological parameters in invasive BC could drive the underlying mechanism of LVI development, leading to metastatic processes, such as invasion and migration. Such mechanisms, however, warrant more comprehensive mechanistic research to thoroughly investigate the potential role of *RRM2* in LVI. Although the findings of the current study did not show a significant association between RRM2 protein levels and LVI, they suggest that RRM2 promotes EMT and invasion. More studies are needed to investigate how *RRM2* regulates critical downstream biological processes and signalling to govern tumour initiation and LVI development.

6.6.2 CCNB1

One of the hallmarks of human cancer is abnormal cell cycle regulation [322]. Uncontrolled cell division is a necessary step in the progression of cancer. Several studies have found that cyclins, which orchestrate normal cell cycle, have abnormally increased expression in a variety of human malignancies [323]. Cyclins are required to activate distinct CDKs at different stages of the cell cycle. Among the several cyclin/CDK complexes involved in cell cycle regulation, CCNB1/cdc2 is a well-studied complex that regulates G2/M phase checkpoint surveillance and is required for mitotic initiation [324, 325]. *CCNB1* has been extensively studied in many solid tumours, such as lung cancer [326], hepatocellular carcinoma [327], and pancreatic cancer [294]. However, this is the first study to date investigating the association between *CCNB1*, LVI, and LVI-related biomarkers in invasive BC.

In this study, CCNB1 expression was observed in the cytoplasm in a large number of cases, while only a small number of cases were in the nucleus. CCNB1 cytoplasmic localisation has been previously identified in some types of cancer, including BC [328]. During the late S/G2 phase of normal human cells, CCNB1/cdc2 complexes accumulate in the cytoplasm and must be translocated into the nucleus to initiate mitosis [329]. However, when DNA is damaged, CCNB1/cdc2 complexes are preserved in the cytoplasm, most likely to avoid premature mitosis [330]. On the other hand, cytoplasmic CCNB1 accumulation has been demonstrated to initiate mitosis by passing a p53-mediated G2/M checkpoint [331]. Cytoplasmic CCNB1 expression causes abnormal cell cycle progression at the G2/M checkpoint, enhancing genomic instability and malignant transformation [332]. This support the potential role of the cytoplasmic expression of CCNB1 in BC.

Tumour metastasis is a multistep process that begins with the separation of cancer cells from the initial tumour mass and proceeds with intravasation, extravasation, and the formation of new foci in a distant organ [333, 334]. The siRNA knockdown approach showed significantly reduced cell proliferation, colony formation, and invasion when an endogenous CCNB1 was disrupted in oesophageal squamous cell carcinoma (ESCC) cells. Furthermore, the findings from animal models suggest that high expression of CCNB1 enhances invasive tumour growth *in vivo* and most likely leads to lung metastasis [295]. A study found that *CCNB1* overexpression provided cells with a greater capacity for transmigration through oesophageal carcinoma endothelium cells and human lung endothelium cells, which may have altered the cytoskeletal structure and promoted extravasation [295]. High expression of *CCNB1* resulted in decreased E-cadherin expression and increased N-cadherin expression, which induced EMT, an important mechanism in the metastatic cascade [295, 335]. High CCNB1 levels are associated with TWIST2 in ESCC, suggesting that TWIST2 might play a role in CCNB1-induced EMT [295].

This study showed that LVI-related biomarkers, such as E-cadherin, N-cadherin, P-cadherin, and TWIST2, were significantly associated with CCNB1 at both mRNA and protein levels, which is consistent with the abovementioned findings. Furthermore, CCNB1 was correlated with MMPs biomarkers, such as MMP1, MMP7, and MMP9. The production of MMP ECM-degrading enzymes increases cell escape from the main tumour tissue and subsequent invasion into tumour-adjacent tissues, such as epithelial cell strata and eventually lymphatic vessels [336]. Although the association between CCNB1 and other migration and invasion biomarkers ranged from weak to moderate correlation, it was statistically significant which indicates that these markers are contributing to the same oncogenic pathway in the context of the LVI process. High expression of *CCNB1* results in the continuous cell cycle and division of cancer cells, promoting their migration and metastasis to distant sites [292]. Uncontrolled cell division promoted by *CCNB1* could lead to gaining extracellular genetic instability and mutations that could affect other key genes for cellular migration and invasion that ultimately lead to LVI. Therefore, blocking this cascade from the early prolifartive phase can stop these processes. These findings demonstrate that overexpression of CCNB1 may control one of the mechanisms driving LVI.

High expression of *CCNB1* mRNA in both the METABRIC and TCGA cohorts showed an association with large tumour size, high tumour grade, poor NPI, LN stage, LVI positivity, ER negativity, PR negativity, and HER2 positivity. Similar findings were observed at the protein level; however, high expression of CCNB1 was not associated with presence of LVI at the protein level in contrast to the mRNA level. Such a disparity between mRNA and protein levels can be explained by various mechanisms, as described in Section 6.6.1. Because mRNA levels primarily determine protein levels, there will be variation between cellular mRNA and protein levels if the cell is undergoing long-term dynamic activities, such as continuous proliferation, which refers to the steady state of the cell [312]. *CCNB1* has

been identified as a critical target gene for promoting tumour proliferation [337]. As a result of the proliferation induced by tumour cells, when *CCNB1* is highly expressed in a malignant cell, the cell may not be stable in the long-term, leading to variation between mRNA and protein levels. Another factor that could contribute to the difference in the significance between LVI and CCNB1 at the mRNA and protein level is the methodologies that are used for quantifying and statistically analysing gene expression in the METABRIC and TCGA cohorts, as well as the ways of determining LVI in the transcriptomic and proteomic cohorts, as described in Chapter 3. The weak correlation between mRNA and protein levels could also explain the contradictory results in LVI between protein and mRNA.

In both the METABRIC and TCGA cohorts, high expression of *CCNB1* was associated with worse outcomes. Although no significant association was observed between CCNB1 protein level, BCSS, and TTDM following univariate analysis, it is interesting that when multivariate Cox regression analysis was performed, CCNB1 mRNA and protein levels represented an independent prognostic factor of shorter BCSS, regardless of LVI status, tumour size, LN stage, or tumour grade, which supports its prognostic significance. These findings were consistent with numerous reports that have shown inconsistent findings of the association between CCNB1 expression and prognosis in solid cancers [325, 326, 338]. Several studies using RT-PCR to analyse gene expression at the RNA level discovered that CCNB1 was highly associated with poor outcomes in solid tumours; however, the link between increased CCNB1 expression at the protein level and clinical prognosis in solid tumours remains unclear [325]. These disparities indicate that more research is needed to elucidate the underlying mechanism and function of CCNB1 in tumour development and prognosis in various tumour types.

6.6.3 CCNB2

Metastasis is the main cause of BC-related mortality [193]. Therefore, early detection of metastatic potential in BC patients can be critical in reducing

cancer-related mortality. LVI is an early event in the development of metastasis and a potent prognostic factor [339]. LVI in BC, like other cancers, has attracted attention not only for its role in prognosis, but also for being a potential therapeutic target. However, the complex molecular mechanism of LVI and its overlap with other carcinogenesis pathways contribute to the difficulty in identifying a key driver gene of LVI that can be targeted. This study identified a strong association between LVI and *CCNB2* utilising large BC cohorts and ANN algorithms, as described in Chapter 3, a powerful technique with low testing error. Likewise, recent studies using different bioinformatic and co-expression analyses confirmed the association of *CCNB2* with human cancer progression [299, 340, 341].

This study has explored the prognostic significance of *CCNB2* expression using multiple well-annotated BC cohorts. Overexpression is associated with aggressive clinicopathological BC features, including LVI positivity, high histological grade, large tumour size, hormone receptor negativity and HER2 positivity and worse clinical outcomes.

The preclinical studies revealed that knocking down *CCNB2* significantly inhibited cell proliferation and reduced cell growth. As expected, silencing *CCNB2* arrested the S phase of the cell cycle, which inhibited the BC cells from entering the G2/M transition. Similarly, knocking down *CCNB2* promoted cell apoptosis. Failure to remove the dead cells due to inhibited cell apoptosis caused the malignant proliferation of cancer cells [342]. A study showed that this occurs through the interference of proteolysis and the obstruction of normal intracellular trafficking [306]. In Lei et al.'s investigation of the role of *CCNB2* in bladder cancer, the gene's tumourigenic function was linked to the dysregulation and abhorrent expression of cell cycle-associated proteins [306]. Li et al. investigated the effect of *CCNB2* in liver cancer and arrived at a similar conclusion, stating that *CCNB2* may promote cell apoptosis by causing S phase arrest and instigating the functional failure of the G2/M DNA and genetic damage checkpoint during cell division [299]. Checkpoints are important in

repairing damaged DNA and maintaining genomic integrity; defects in this domain can lead to mutations and oncogenesis [343]. It is essential that the primary tumour cells proliferate so that they can invade the surrounding tissue and establish metastatic cascades, leading to the degradation of the basement membrane and LVI. Proliferation continues until the invasion of vascular or lymphatic channels occur; at this stage, tumour cells can successfully evade apoptosis [249]. Therefore, *CCNB2* is one of the essential regulators of cell proliferation and a prerequisite step in the development of LVI and distant metastasis.

Although the role of *CCNB2* in the cell cycle and proliferation is well known, its role in driving LVI in BC has not yet been characterised. The results of this study revealed that the silencing of *CCNB2* suppressed migration and reduced the number of tumour cells that adhered to and transmigrated across vascular and lymphatic endothelial cell lines. This finding was supported by the positive correlation between EMT- and MMP-related markers. N-cadherin is associated with EMT, which plays a vital role in the invasion and intravasation of cells into the bloodstream and in the degradation of the ECM via production of proteases. Therefore, stromal invasion is a result of loss of the linkage between the epithelial layer of the BC cell and other epithelial cells due to the upregulation of N-cadherin [344]. Moreover, P-cadherin is an important molecule that can activate integrin molecules, allow cancer cells to attach to the ECM, and activate the invasion of cancer cells [345], which could explain the role of *CCNB2* in LVI as *CCNB2* has a positive correlation with P-cadherin. To invade the tissue, tumour cells must penetrate the tissue boundaries by breaking down the ECM with MMPs and the urokinase plasminogen activator (uPA) system. Inhibition of uPA halts the invasion and expression of MMP9 [42]. The LVI TME is strongly correlated with the expression of MMPs, specifically MMP9 and MMP1; this expression is responsible for cancer cell intravasation and metastasis in BC [179]. The *in vitro* assessment indicated the contribution of upregulated *CCNB2* expression to migration, adhesion, and

transmigration through endothelial cell lines, which induced the LVI process. This finding is also consistent with a previous bladder cancer study that reported that silencing *CCNB2* could restrain migration and invasion, leading to inhibition of tumour metastasis and prolonged survival time [306].

Because *CCNB2* is a cell cycle regulator, the cell cycle may be one of the mechanisms contributing to LVI in BC. However, the molecular mechanisms that cause BC tumour cells to invade and disseminate in vascular spaces are still largely unknown and require more research. Cancer treatment strategies that induce cell cycle arrest in cancer cells are effective [346]. This observation suggests that *CCNB2* is a key gene in BC, particularly given the findings regarding lymphatic invasion, which is a prerequisite for the metastatic cascade. Therefore, *CCNB2* may represent a potential target for LVI in BC patients.

This study provides evidence that *CCNB2* is a key modulator of LVI in BC and could be a potential therapeutic target to suppress LVI and improve patient outcomes. However, further *in vitro* and *in vivo* experiments investigating the differential gene expression and pathways observed in *CCNB2* knockdown cells are needed.

6.7 Conclusion

BC progression is a complex process. Deciphering the transcriptomics and proteomics that govern the invasive and metastatic cascades of BC is essential in understanding the mechanisms of cancer progression and can lead to the discovery of novel therapeutic targets. LVI is a prerequisite step of BC metastasis is controlled by specialised molecular pathways; however, the drivers underlying the molecular mechanisms of LVI in BC are yet to be defined. Understanding the molecular mechanisms underlying BC progression (LVI in particular) and unveiling their driving molecular pathways would ultimately improve patient outcomes.

Evidence from this chapter demonstrated that *RRM2*, *CCNB1*, and *CCNB2* are important in invasive BC progression. Because *RRM2*, *CCNB1,* and

CCNB2 are significantly correlated with each other at the mRNA and protein levels, share similar associations with several aggressive tumour features, and are important in regulating the cell cycle, the cell cycle may be one of the molecular mechanisms driving LVI in BC. However, the exact functional and mechanistic role of these biomarkers in LVI might not have been sufficiently evaluated in this study. Collectively, the significance and value of *RRM2*, *CCNB1*, and *CCNB2* as prognostic biomarkers in invasive BC is apparent because they are significantly associated with migration, invasion, and metastasis. Targeting the underlying molecular mechanisms of these genes may present promising novel therapeutic targets.

Chapter 7: General discussion and conclusion

7.1 General discussion

As previously stated, invasive BC is one of the most common cancers diagnosed in women and a leading cause of death among women globally. Therefore, it is essential to improve upon the existing options for invasive BC prevention, diagnosis, prognosis, and treatment to improve patient outcomes and quality of life. This endeavour is challenging because BC is a heterogeneous condition involving a wide variety of morphological and molecular entities, and it has been established that each BC case has a unique molecular architecture, with inter-tumour or even intra-tumour heterogeneity [347, 348].

Cancer metastasis is a multistep process that begins with the separation of tumour cells from the primary tumour mass and progresses through intravasation, extravasation, and the development of new foci in a distant organ [333, 334]. The progression of BC to metastatic disease is strongly correlated with poor outcomes in invasive BC patients and requires certain molecular conditions involving several proteins and converging pathways [85, 86]. The main molecular drivers of BC can influence cell growth, migration, adhesion, and invasion. LVI is one of the initial events in metastasis and has been identified as a prognostic factor that has been shown to be highly associated with BCSS in clinical studies [349]. BC cells can use specific molecular-level biological processes to facilitate an invasive and metastatic cascade of events via LVI [47, 88]. Exploration of the key molecular determinants of LVI development is crucial to elucidating the mechanisms underlying LVI and potentially identifying novel prognostic factors and therapeutic targets for invasive BC.

The key genes involved in the development of LVI in invasive BC are poorly characterised. The development of high-throughput molecular and bioinformatic tools has enabled large-scale analysis of genes related to LVI [150], as discussed in detail in Chapter 1, Section 1.7.6. Although previous studies have explored genes correlated with LVI in certain types of cancer, such as endometrial carcinoma and hepatocellular carcinoma [151, 153],

these studies had limitations. An example of these limitations was the nature of the cohorts used in these studies and the lack of accurate or incomplete LVI status characterisation. To overcome these limitations, the present study used two large transcriptomic cohorts with a well-characterised LVI status. This study hypothesised that by performing bioinformatic analyses, key genetic drivers of LVI in BC could be identified, some of which may have prognostic value, serve as potential therapeutic targets, and play a role in the mechanisms controlling the development of LVI. To test this hypothesis, bioinformatic analyses, including ANN and WAD methodologies, were performed to identify key genes associated with LVI by employing two large transcriptomic datasets (METABRIC and TCGA) with a well-characterised LVI status. Then, the association between clinicopathological and prognostic significance of the identified genes at the mRNA and protein levels was evaluated. To further explore the role of the identified genes in BC progression and the development of LVI, *in vitro* experimental studies were conducted using a panel of relevant BC cell lines, along with *in vitro* investigations of BC cell lines with lymphatic and vascular endothelial cell models to mimic *in vivo* interactions.

7.2 Major findings of the study

To date, this is the first study to use two large transcriptome cohorts with a well-defined LVI status to generate a list of genes associated with LVI, evaluate their association with LVI and other prognostic factors at the protein level, and investigate their potential role in LVI development using *in vitro* models. In this study, a 99-gene set associated with LVI status in invasive BC was identified using WAD analysis; moreover, a target gene list containing 100 genes was generated using ANN analysis. Some of the genes identified in the WAD and ANN analyses were associated with LVI in only one cohort, whereas others were observed in both. Only 3 of the 100 genes identified by ANN analysis overlapped with those identified by WAD analysis: *CCNB2*, *CDCA5*, and *YWHAZ*. Many factors could explain this unexpected finding. The comparability of the METABRIC and TCGA cohorts

was a limiting factor in this study due to the varied methodologies employed by these cohorts to measure and analyse mRNA expression and evaluate LVI status, as discussed in Section 3.10. Another factor that may have contributed to these variations was the methods used to measure mRNA expression. For example, the METABRIC study employed microarrays to measure mRNA expression, while the TCGA study used RNA-seq with next-NGS; therefore, it is difficult to assume that that microarrays and RNA-seq approaches are interchangeable. Although the WAD and ANN analyses produced lists of distinct genes, their applicability for gene identification has been previously demonstrated [199, 203, 206].

In this study, the genes *MX1*, *IDH2*, *HMGB3*, *RRM2*, *CCNB1,* and *CCNB2* were carefully selected based on the criteria described in Section 3.10. Although all the selected genes were significantly associated with LVI positivity at the mRNA level in both the METABRIC and TCGA cohorts, some candidates, such as *MX1*, *RRM2,* and *CCNB1*, were not associated with LVI at the protein level. Different biological and technical factors could contribute to these variations, as discussed in Sections 4.6.1 and 6.6.1.

In this study, *MX1*, *IDH2,* and *HMGB3* were identified as candidates of interest in the WAD analysis*. MX1* was significantly associated with aggressive tumour features at the mRNA and protein levels, including high tumour grade and hormone receptor negativity. In addition, an association was observed between *IDH2* and high tumour grade, large tumour size, poor NPI, hormone receptor negativity, and HER2 positivity. An association between *MX1* and poor outcome was observed at the mRNA level but not at the protein level; however, *IDH2* showed a significant association with shorter BCSS at both the mRNA and protein levels. The number of events and follow-up in each dataset may have affected the results. These discrepancies may also be explained by tumour-specific variations in *MX1* mRNA and protein stability or post-transcriptional regulation. Regarding *HMGB3*, at the mRNA and protein levels, increased expression of this gene was associated with aggressive tumour behaviour, such as high histological

grade, poor NPI, ER and PR negativity, HER2 positivity, poor BCSS survival, and TTDM. Additionally, *MX1*, *IDH2*, and *HMGB3* were significantly associated with EMT and MMP markers at the transcriptomic and proteomic levels. *In vitro* assessment of *HMGB3* was carried out in concert with the aforementioned findings to elucidate the possible functional role of *HMGB*3 in LVI development in BC, as described in Chapter 5. These preclinical investigations showed that knockdown of *HMGB3* in BC cell lines reduced cell proliferation, migration, and, importantly, adherence and transmigration across endothelial cells. Knockdown of *HMGB3* also enhanced cell apoptosis and arrested the cell cycle in the G1 phase. Taken together, these findings suggest a role for these genes in BC metastasis and involvement in LVI development. Therefore, *IDH2*, *HMGB3,* and potentially *MX1* are driver genes of LVI.

As described in Chapter 3, ANN analysis showed that *RRM2* and *CCNB1* were associated with the presence of LVI in both the METABRIC and TCGA cohorts. *RRM2* mRNA and protein expression was associated with large tumour size, high tumour grade, and LN stage. High *RRM2* mRNA levels but not protein levels were associated with poor outcomes. An association was also noted between *CCNB1* mRNA and protein overexpression and high tumour grade, poor NPI, ER and PR negativity, and HER2 positivity. Increased expression of *CCNB1* mRNA, but not protein, was correlated with poor outcome.

In this study, *CCNB2* was one of the genes in the overlap between the WAD and ANN analyses. High *CCNB2* mRNA and protein levels were associated with clinicopathological parameters characteristic of aggressive tumour behaviour, including high histologic grade, large tumour size, ER negativity, HER2 positivity, and shorter BCSS. The expression of *RRM2*, *CCNB1*, and *CCNB2* was significantly correlated with the expression of EMT- and MMP-related genes, which are thought to play a role in cancer cell invasion and intravasation into the bloodstream. To further explore the possible functional role of *CCNB2, in vitro* functional assays were performed.

Knockdown of *CCNB2* in BC cell lines reduced cell proliferation, migration, and, importantly, reduced adherence and transmigration across endothelial cell lines. Knockdown of *CCNB2* in BC cell lines enhanced apoptosis and arrested the cell cycle in the S phase. These findings suggest that *CCNB2* has an oncogenic role and prognostic significance in BC and may be involved in the development of LVI.

Although there were differences between the gene lists produced by the WAD and ANN methodologies and differences in the results at the mRNA and protein levels, it is important to note that several factors likely contributed to these variations, as discussed in Chapters 3, 4, and 6. Nevertheless, this study revealed intriguing prognostic biomarkers with clinicopathological significance and further highlighted the potential functional role of *CCNB2* and *HMGB3* in LVI development in BC.

Altogether, this thesis identified potential drivers of LVI that showed statistically significant association with LVI and other prognostic factors at the mRNA or protein levels. As the selected genes were similarly associated with certain aggressive tumour features, were known to be important for cell cycle regulation, and correlated with each other in some cases, the cell cycle may be mechanistically involved in driving LVI in BC. However, the exact functional and molecular pathways of LVI in which these biomarkers are involved were not sufficiently investigated in this study.

7.3 Conclusion

The present study confirmed the hypothesis that WAD and ANN methodologies could identify key drivers of LVI that showed prognostic significance in invasive BC. The potential roles of these genes in BC progression and LVI development were also identified. Evidently, targeting *HMGB3* and *CCNB2* significantly reduced the adherence and transmigration of BC cells across lymphatic and vascular endothelial cells, suggesting that these genes play a role in controlling LVI development. This thesis also provided insight into the proteins involved in the development of LVI in

invasive BC patients by examining major regulators of the cell cycle and other pathways. These findings also pave the way for a potential novel treatment strategy using these genes as therapeutic targets for BC that may extend the lives of patients with aggressive BC. More *in vitro* and *in vivo* functional investigations are recommended for further validation of this study.

7.4 Strengths of the study

1- The use of samples from patients with invasive BC rather than cell lines or xenografts.
2- The use of large, well-characterised BC cohorts with a well-defined LVI status at the mRNA and protein levels in addition to clinicopathological and pertinent biomarker data.
3- The use of two large transcriptomic LVI datasets derived from the METABRIC and TCGA cohorts for the bioinformatic analysis.
4- The discovery of reliable prognostic biomarkers that have potential roles in invasive BC.
5- The use of WB to confirm the specificity of the markers under investigation.
6- The inclusion of positive and negative controls with each IHC experiment to confirm effective staining.
7- The inclusion of *in vitro* experiments with several functional assays and models, and the selection of optimal BC cell lines based on the IHC findings.
8- The publication of several articles based on this study.

7.5 Limitations of the study

1- LVI has not previously been extensively investigated in invasive BC and other malignancies; therefore, an exploratory research approach was used rather than focusing on only a single pathway or group of LVI-related molecules.

2- H-scoring for IHC-stained slides is a semi-quantitative method that relies on the subjectivity of scorers even though standardisation was performed to minimise this constraint.
3- Although the expression patterns of the IHC-stained proteins were demonstrated to be homogeneous in full-face sections, it would have been optimal to include more than one core per case to aid in their evaluation.
4- The TMA blocks utilised for IHC staining were historical and, in some cases, TMA cores were either missing or the tumour burden was at the margin of the slide.
5- The number of BC and endothelial cell lines used for *in vitro* experiments was limited in their ability to accurately represent various BC subtypes.
6- Although the functional assays were conducted on days 3–7 after knockdown and the WB showed that the inhibition was successful, siRNA is transient and lentivirally transduced shRNA should have been used to silence the RNA sequence and construct stable cell lines. Alternatively, CRISPR-Cas9-based knockout cell lines should have been constructed to better validate key functional assays.
7- *In vivo* experiments in mouse models should have been performed to substantiate the findings.

7.6 Future work

1- Investigating the other lists of genes identified by the WAD and ANN analyses to explore their association with LVI and prognostic significance in invasive BC patients.
2- Utilising selective inhibitors in *in vitro* experiments and performing additional functional assays.
3- Investigating the differential gene expression and pathways observed in the candidate gene knockdown cells via *in vitro* and/or *in vivo* experiments.
4- Evaluating the current LVI genes in terms of their subcellular localisation and co-expression *in vitro*.
5- Studying the predictive significance of the selected biomarkers to evaluate their utility and predictive importance.

6- Collaborating and publishing multiple articles on the molecular regulators of LVI.

7.7 Impact of COVID-19

The COVID-19 pandemic, which resulted in the closure of the University of Nottingham campus in addition to national and global lockdowns, impacted my thesis work both directly and indirectly. Access to the labs was restricted during the lockdown. Additionally, even when the lab was reopened, it did not operate at full capacity and the return to campus took place in phases. Throughout this period, I faced several moments when I thought I would be unable to complete my thesis due to the COVID-19 pandemic. However, with the assistance of my family and supervisors, I was able to overcome these difficult moments. The work presented in this thesis has been influenced by the COVID-19 pandemic, and as such, its impact should be considered when evaluating the quality and quantity of the experiments detailed herein.

References

1. Going JJ. Ductal-lobar organisation of human breast tissue, its relevance in disease and a research objective: vector mapping of parenchyma in complete breasts (the Astley Cooper project). Breast Cancer Res. 2006;8(4):107.
2. Lumley JSP, Cravens JL, Abrahams PH, Tunstall RG. Bailey & love's essential clinical anatomy. 1st ed. Boca Raton: CRC Press; 2018.
3. Ferguson, DJ. Intraepithelial lymphocytes and macrophages in the normal breast. Virchows Arch A Pathol Anat Histopathol. 1985;407(4):369-78.
4. Fleming JM, Long EL, Ginsburg E, Gerscovich D, Meltzer PS, Vonderhaar BK. Interlobular and intralobular mammary stroma: genotype may not reflect phenotype. BMC Cell Biol. 2008 Aug 18;9:46.
5. Hassiotou F, Geddes D. Anatomy of the human mammary gland: current status of knowledge. Clin Anat. 2013 Jan;26(1)29-48.
6. Kumar V, Abbas AK, Aster JC. Robbins basic pathology [Internet]. Amsterdam: Elsevier Health Sciences; 2017.
7. Pinder SE, Ellis IO. The diagnosis and management of pre-invasive breast disease: ductal carcinoma in situ (DCIS) and atypical ductal hyperplasia (ADH)--current definitions and classification. Breast Cancer Res. 2003;5(5)254-7.
8. Simpson PT, Gale T, Fulford LG, Reis-Filho JS, Lakhani SR. The diagnosis and management of pre-invasive breast disease: pathology of atypical lobular hyperplasia and lobular carcinoma in situ. Breast Cancer Res. 2003;5(5):258-62.
9. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar 4;144(5):646-74.
10. Braunstein LZ, Taghian AG. Molecular phenotype, multigene assays, and the locoregional management of breast cancer. Semin Radiat Oncol. 2016 Jan;26(1):9-16.
11. Sgroi DC. Preinvasive breast cancer. Annu Rev Pathol. 2010;5:193-221.
12. Stingl J, Caldas C. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. Nat Rev Cancer. 2007 Oct;7(10):791-9.
13. Sabiani L, Houvenaeghel G, Heinemann M, Reyal F, Classe JM, Cohen M, Garbay JR, Giard S, Charitansky H, Chopin N, Rouzier R, Daraï E, Coutant C, Azuar P, Gimbergues P, Villet R, Tunon de Lara C, Lambaudie E. Breast cancer in young women: Pathologic features and molecular phenotype. Breast. 2016 Oct;29:109-16.
14. Goksu SS, Tastekin D, Arslan D, Gunduz S, Tatli AM, Unal D, Salim D, Guler T, Coskun HS. Clinicopathologic features and molecular subtypes of breast cancer in young women (age ≤ 35). Asian Pac J Cancer Prev. 2014;15(16):6665-8.
15. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, Jemal A, Siegel RL. Breast cancer statistics, 2019. CA Cancer J Clin. 2019 Nov;69(6):438-51.
16. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021 May;71(3):209-49.
17. Breast Cancer Now. Facts and statistics 2021 [Internet]. c2021 [cited 2021 Oct 21. Available from: https://breastcancernow.org/about-us/media/facts-statistics.
18. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015 Mar 1;136(5):E359-86.
19. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. Cancer Epidemiol. 2012 Jun;36(3):237-48.

20. Assi HA, Khoury KE, Dbouk H, Khalil LE, Mouhieddine TH, El Saghir NS. Epidemiology and prognosis of breast cancer in young women. J Thorac Dis. 2013 Jun;5 Suppl 1(Suppl 1):S2-8.
21. Gierisch JM, Coeytaux RR, Urrutia RP, Havrilesky LJ, Moorman PG, Lowery WJ, Dinan M, McBroom AJ, Hasselblad V, Sanders GD, Myers ER. Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: a systematic review. Cancer Epidemiol Biomarkers Prev. 2013 Nov;22(11):1931-43.
22. Engmann NJ, Golmakani MK, Miglioretti DL, Sprague BL, Kerlikowske K; Breast Cancer Surveillance Consortium. Population-attributable risk proportion of clinical risk factors for breast cancer. JAMA Oncol. 2017 Sep 1;3(9):1228-36.
23. Crandall CJ, Aragaki AK, Cauley JA, McTiernan A, Manson JE, Anderson G, Chlebowski RT. Breast tenderness and breast cancer risk in the estrogen plus progestin and estrogen-alone women's health initiative clinical trials. Breast Cancer Res Treat. 2012 Feb;132(1):275-85.
24. Neuhouser ML, Aragaki AK, Prentice RL, Manson JE, Chlebowski R, Carty CL, Ochs-Balcom HM, Thomson CA, Caan BJ, Tinker LF, Urrutia RP, Knudtson J, Anderson GL. Overweight, obesity, and postmenopausal invasive breast cancer risk: A secondary analysis of the women's health initiative randomized clinical trials. JAMA Oncol. 2015;1(5):611-21.
25. McDonnell DP, Park S, Goulet MT, Jasper J, Wardell SE, Chang CY, Norris JD, Guyton JR, Nelson ER. Obesity, cholesterol metabolism, and breast cancer pathogenesis. Cancer Res. 2014 Sep 15;74(18):4976-82.
26. Liu Y, Nguyen N, Colditz GA. Links between alcohol consumption and breast cancer: a look at the evidence. Women's Health (Lond). 2015 Jan;11(1):65-77.
27. Dossus L, Boutron-Ruault MC, Kaaks R, Gram IT, Vilier A, Fervers B, Manjer J, Tjonneland A, Olsen A, Overvad K, Chang-Claude J, Boeing H, Steffen A, Trichopoulou A, Lagiou P, Sarantopoulou M, Palli D, Berrino F, Tumino R, Vineis P, Mattiello A, Bueno-de-Mesquita HB, van Duijnhoven FJ, Bakker MF, Peeters PH, Weiderpass E, Bjerkaas E, Braaten T, Menéndez V, Agudo A, Sanchez MJ, Amiano P, Tormo MJ, Barricarte A, Butt S, Khaw KT, Wareham N, Key TJ, Travis RC, Rinaldi S, McCormack V, Romieu I, Cox DG, Norat T, Riboli E, Clavel-Chapelon F. Active and passive cigarette smoking and breast cancer risk: results from the EPIC cohort. International J Cancer. 2014 Apr 15;134(8):1871-88.
28. Symmans WF. Breast cancer prognostic and predictive factors. Seminars in Breast Disease. 2004 Sep;7(3):91-100.
29. Makki J. Diversity of breast carcinoma: histological subtypes and clinical relevance. Clin Med Insights Pathol. 2015 Dec;8:23-31.
30. Yang F, Li J. WHO classification of tumors of the breast [in Chinese]. Zhonghua Wai Ke Za Zhi. 2014;52(1):1-3.
31. Ellis IO, Galea M, Broughton N, Locker A, Blamey RW, Elston CW. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. Histopathology. 1992 Jun;20(6):479-89.
32. Rakha EA, El-Sayed ME, Lee AH, Elston CW, Grainge MJ, Hodi Z, Blamey RW, Ellis IO. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. J Clin Oncol. 2008 Jul 1;26(19):3153-8.
33. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology. 1991 Nov;19(5):403-10.
34. Lee AHS, Ellis IO. The Nottingham prognostic index for invasive carcinoma of the breast. Pathol Oncol Res. 2008 Jun;14(2)113-5.
35. Joseph C, Macnamara O, Craze M, Russell R, Provenzano E, Nolan CC, Diez-Rodriguez M, Sonbul SN, Aleskandarany MA, Green AR, Rakha EA, Ellis IO, Mukherjee A. Mediator complex (MED) 7: a biomarker associated with good prognosis in invasive breast cancer, especially ER+ luminal subtypes. Br J Cancer. 2018 Apr;118(8):1142-51.

36. Rakha EA, Tan PH, Varga Z, Tse GM, Shaaban AM, Climent F, van Deurzen CH, Purnell D, Dodwell D, Chan T, Ellis IO. Prognostic factors in metaplastic carcinoma of the breast: a multi-institutional study. Br J Cancer. 2015 Jan 20;112(2):283-9.
37. Mojarad S, Venturini B, Fulgenzi P, Papaleo R, Brisigotti M, Monti F, Canuti D, Ravaioli A, Woo L, Dlay S, Sherbet GV. Prediction of nodal metastasis and prognosis of breast cancer by ANN-based assessment of tumour size and p53, Ki-67 and steroid receptor expression. Anticancer Res. 2013 Sep;33(9):3925-33.
38. Saadatmand S, Bretveld R, Siesling S, Tilanus-Linthorst MM. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients.BMJ. 2015 Oct 6;351:h4901.
39. Cianfrocca M, Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. Oncologist. 2004;9(6):606-16.
40. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham prognostic index in primary breast cancer. Breast Cancer Res Treat. 1992;22(3):207-19.
41. Kurshumliu F, Gashi-Luci L, Kadare S, Alimehmeti M, Gozalan U. Classification of patients with breast cancer according to Nottingham prognostic index highlights significant differences in immunohistochemical marker expression. World Surg Oncol. 2014 Aug 1;12:243.
42. Weigelt B, Peterse JL, van 't Veer LJ. Breast cancer metastasis: markers and models. Nat Rev Cancer. 2005 Aug;5(8):591-602.
43. Abi-Raad R, Boutrus R, Wang R, Niemierko A, Macdonald S, Smith B, Taghian AG. Patterns and risk factors of locoregional recurrence in T1-T2 node negative breast cancer patients treated with mastectomy: implications for postmastectomy radiotherapy. Int J Radiat Oncol Biol Phys. 2011 Nov 1;81(3):e151-7.
44. Mohammed RA, Martin SG, Mahmmod AM, Macmillan RD, Green AR, Paish EC, Ellis IO. Objective assessment of lymphatic and blood vascular invasion in lymph node-negative breast carcinoma: findings from a large case series with long-term follow-up. J Pathol. 2011 Feb;223(3):358-65.
45. Blamey RW, Ellis IO, Pinder SE, Lee AH, Macmillan RD, Morgan DA, Robertson JF, Mitchell MJ, Ball GR, Haybittle JL, Elston CW. Survival of invasive breast cancer according to the Nottingham prognostic index in cases diagnosed in 1990–1999. Eur J Cancer. 2007 Jul;43(10):1548-55.
46. Rakha EA, Martin S, Lee AH, Morgan D, Pharoah PD, Hodi Z, Macmillan D, Ellis IO. The prognostic significance of lymphovascular invasion in invasive breast carcinoma. Cancer. 2012 Aug 1;118(15):3670-80.
47. Stacker SA, Williams SP, Karnezis T, Shayan R, Fox SB, Achen MG. Lymphangiogenesis and lymphatic vessel remodelling in cancer. Nat Rev Cancer. 2014 Mar;14(3):159-72.
48. Al-Rohil RN, Milton DR, Nagarajan P, Curry JL, Feldmeyer L, Torres-Cabala CA, Ivan D, Prieto VG, Tetzlaff MT, Aung PP. Intratumoral and peritumoral lymphovascular invasion detected by D2-40 immunohistochemistry correlates with metastasis in primary cutaneous Merkel cell carcinoma. Hum Pathol. 2018 Jul;77:98-107.
49. Zaorsky NG, Patil N, Freedman GM, Tuluc M. Differentiating lymphovascular invasion from retraction artifact on histological specimen of breast carcinoma and their implications on prognosis. J Breast Cancer. 2012 Dec;15(4):478-80.
50. Ellis IO, Coleman D, Wells C, Kodikara S, Paish EM, Moss S, Al-Sam S, Anderson N, Bobrow L, Buley I, Connolly CE, Dallimore NS, Hales S, Hanby A, Humphreys S, Knox F, Lowe J, Macartney J, Nash R, Parham D, Patnick J, Pinder SE, Quinn CM, Robertson AJ, Shrimankar J, Walker RA, Winder R. Impact of a national external quality assessment scheme for breast pathology in the UK. J Clin Pathol. 2006 Feb;59(2):138-45.
51. Mohammed RA, Martin SG, Gill MS, Green AR, Paish EC, Ellis IO. Improved methods of detection of lymphovascular invasion demonstrate that it is the predominant method of vascular invasion in breast cancer and has important clinical consequences. Am J Surg Pathol. 2007 Dec;31(12):1825-33.

52. Mohammed RA, Ellis IO, Elsheikh S, Paish EC, Martin SG. Lymphatic and angiogenic characteristics in breast cancer: morphometric analysis and prognostic implications. Breast Cancer Res Treat. 2009 Jan; 113(2):261-73.
53. Bono P, Wasenius VM, Heikkilä P, Lundin J, Jackson DG, Joensuu H. High LYVE-1–positive lymphatic vessel numbers are associated with poor outcome in breast cancer. Clin Cancer Res. 2004 Nov 1;10(21):7144-9.
54. Marinho VF, Metze K, Sanches FS, Rocha GF, Gobbi H. Lymph vascular invasion in invasive mammary carcinomas identified by the endothelial lymphatic marker D2-40 is associated with other indicators of poor prognosis. BMC Cancer. 2008 Feb 29;8:64.
55. Payne SJ, Bowen RL, Jones JL, Wells CA. Predictive markers in breast cancer – the present. Histopathology. 2008 Jan;52(1):82-90.
56. Hammond ME, Hayes DF, Wolff AC, Mangu PB, Temin S. American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Oncol Pract. 2010 Jul;6(4):195-7.
57. Lee ES, Han W, Kim MK, Kim J, Yoo TK, Lee MH, Lee KH, Kim TY, Moon HG, Im SA, Noh DY, Lee ES. Factors associated with late recurrence after completion of 5-year adjuvant tamoxifen in estrogen receptor positive breast cancer. BMC Cancer. 2016 Jul 7;16:430.
58. Rakha EA, Reis-Filho JS, Ellis IO. Combinatorial biomarker expression in breast cancer. Breast Cancer Res Treat. 2010 Apr;120(2):293-308.
59. Cancello G, Montagna E. Treatment of breast cancer in young women: do we need more aggressive therapies? Journal Thorac Dis. 2013 Jun;5 Suppl 1(Suppl 1):S47-54.
60. Zhou XL, Fan W, Yang G, Yu MX. The clinical significance of PR, ER, NF- kappa B, and TNF- alpha in breast cancer. Dis Markers. 2014;2014:494581.
61. Figueroa-Magalhães MC, Jelovac D, Connolly R, Wolff AC. Treatment of HER2-positive breast cancer. Breast. 2014 Apr;23(2):128-36.
62. Ingthorsson S, Andersen K, Hilmarsdottir B, Maelandsmo GM, Magnusson MK, Gudjonsson T. HER2 induced EMT and tumorigenicity in breast epithelial progenitor cells is inhibited by coexpression of EGFR. Oncogene. 2016 Aug 11;35(32):4244-55.
63. Mohammed ZM, McMillan DC, Edwards J, Mallon E, Doughty JC, Orange C, Going JJ. The relationship between lymphovascular invasion and angiogenesis, hormone receptors, cell proliferation and survival in patients with primary operable invasive ductal breast cancer. BMC Clin Pathol. 2013 Nov 25;13(1):31.
64. Feigin ME, Muthuswamy SK. Polarity proteins regulate mammalian cell-cell junctions and cancer pathogenesis. Curr Opin Cell Biol. 2009 Oct;21(5):694-700.
65. Ahmed S, Sami A, Xiang J. HER2-directed therapy: current treatment options for HER2-positive breast cancer. Breast Cancer. 2015 Mar;22(2):101-16.
66. Hamy AS, Lam GT, Laas E, Darrigues L, Balezeau T, Guerin J, Livartowski A, Sadacca B, Pierga JY, Vincent-Salomon A, Coussy F, Becette V, Bonsang-Kitzis H, Rouzier R, Feron JG, Benchimol G, Laé M, Reyal F. Lymphovascular invasion after neoadjuvant chemotherapy is strongly associated with poor prognosis in breast carcinoma. Breast Cancer Res Treat. 2018 Jun;169(2):295-304.
67. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lønning PE, Børresen-Dale AL. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001 Sep 11;98(19):10869-74.
68. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature. 2000 Aug 17;406(6797):747-52.
69. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale

AL, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003 Jul 8;100(14):8418-23.
70. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, Nobel A, Parker J, Ewend MG, Sawyer LR, Wu J, Liu Y, Nanda R, Tretiakova M, Ruiz Orrico A, Dreher D, Palazzo JP, Perreard L, Nelson E, Mone M, Hansen H, Mullins M, Quackenbush JF, Ellis MJ, Olopade OI, Bernard PS, Perou CM. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. 2006 Apr 27;7:96.
71. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, Nielsen TO, Gelmon K. Metastatic behavior of breast cancer subtypes. J Clin Oncol. 2010 Jul 10;28(20):3271-7.
72. Bastien RR, Rodríguez-Lescure Á, Ebbert MT, Prat A, Munárriz B, Rowe L, Miller P, Ruiz-Borrego M, Anderson D, Lyons B, Álvarez I, Dowell T, Wall D, Seguí MÁ, Barley L, Boucher KM, Alba E, Pappas L, Davis CA, Aranda I, Fauron C, Stijleman IJ, Palacios J, Antón A, Carrasco E, Caballero R, Ellis MJ, Nielsen TO, Perou CM, Astill M, Bernard PS, Martín M. PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. BMC Med Genomics. 2012 Oct 4;5:44.
73. Geyer FC, Lopez-Garcia MA, Lambros MB, Reis-Filho JS. Genetic characterization of breast cancer and implications for clinical management. J Cell Mol Med. 2009 Oct;13(10):4090-103.
74. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, Perou CM, Ellis MJ, Nielsen TO. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009 May 20;101(10):736-50.
75. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009 Mar 10;27(8):1160-7.
76. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, El-Sayed ME, Benhasouna A, Brunet JS, Akslen LA, Evans AJ, Blamey R, Reis-Filho JS, Foulkes WD, Ellis IO. Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. Clin Cancer Res. 2009 Apr 1;15(7):2302-10.
77. Weigelt B, Geyer FC, Reis-Filho JS. Histological types of breast cancer: how special are they? Mol Oncol. 2010 Jun;4(3):192-208.
78. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006 Jun 7;295(21):2492-502.
79. Sørlie T, Wang Y, Xiao C, Johnsen H, Naume B, Samaha RR, Børresen-Dale AL. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. BMC Genomics. 2006 May 26;7:127.
80. Feng Y, Spezia M, Huang S, Yuan C, Zeng Z, Zhang L, Ji X, Liu W, Huang B, Luo W, Liu B, Lei Y, Du S, Vuppalapati A, Luu HH, Haydon RC, He TC, Ren G. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes Dis. 2018 May 12;5(2):77-106.
81. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S; METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012 Apr 18;486(7403):346-52.

82. Dawson SJ, Rueda OM, Aparicio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. EMBO J. 2013 Mar 6;32(5):617-28.
83. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005 May 14;365(9472):1687-717.
84. Savci-Heijink CD, Halfwerk H, Hooijer GK, Horlings HM, Wesseling J, van de Vijver MJ. Retrospective analysis of metastatic behaviour of breast cancer subtypes. Breast Cancer Res Treat. 2015 Apr;150(3):547-57.
85. Sun Y, Wang X, Zhou Q, Lu Y, Zhang H, Chen Q, Zhao M, Su S. Inhibitory effect of emodin on migration, invasion and metastasis of human breast cancer MDA-MB-231 cells in vitro and in vivo. Oncol Rep. 2015 Jan;33(1):338-46.
86. Fazilaty H, Gardaneh M, Bahrami T, Salmaninejad A, Behnam B. Crosstalk between breast cancer stem cells and metastatic niche: emerging molecular metastasis pathway? Tumour Biol. 2013 Aug;34(4):2019-30.
87. González de Vega R, Clases D, Fernández-Sánchez ML, Eiró N, González LO, Vizoso FJ, Doble PA, Sanz-Medel A. MMP-11 as a biomarker for metastatic breast cancer by immunohistochemical-assisted imaging mass spectrometry. Anal Bioanal Chem. 2019 Jan;411(3):639-46.
88. Lee CC, Lin JC, Hwang WL, Kuo YJ, Chen HK, Tai SK, Lin CC, Yang MH. Macrophage-secreted interleukin-35 regulates cancer cell plasticity to facilitate metastatic colonization. Nat Commun. 2018 Sep 14;9(1):3763.
89. Insua-Rodríguez J, Pein M, Hongu T, Meier J, Descot A, Lowy CM, De Braekeleer E, Sinn HP, Spaich S, Sütterlin M, Schneeweiss A, Oskarsson T. Stress signaling in breast cancer cells induces matrix components that promote chemoresistant metastasis. EMBO Mol Med. 2018 Oct;10(10):e9003.
90. Shinde A, Wilmanski T, Chen H, Teegarden D, Wendt MK. Pyruvate carboxylase supports the pulmonary tropism of metastatic breast cancer. Breast Cancer Res. 2018 Jul 13;20(1):76.
91. Bray K, Gillette M, Young J, Loughran E, Hwang M, Sears JC, Vargo-Gogola T. Cdc42 overexpression induces hyperbranching in the developing mammary gland by enhancing cell migration. Breast Cancer Res. 2013;15(5):R91.
92. Lee JA, Bae JW, Woo SU, Kim H, Kim CH. D2-40, podoplanin, and CD31 as a prognostic predictor in invasive ductal carcinomas of the breast. J Breast Cancer. 2011 Jun;14(2):104-11.
93. Liang P, Nakada I, Hong JW, Tabuchi T, Motohashi G, Takemura A, Nakachi T, Kasuga T, Tabuchi T. Prognostic significance of immunohistochemically detected blood and lymphatic vessel invasion in colorectal carcinoma: its impact on prognosis. Ann Surg Oncol. 2007 Feb;14(2):470-7.
94. Kandemir NO, Barut F, Bektas S, Ozdamar SO. Can lymphatic vascular density be used in determining metastatic spreading potential of tumor in invasive ductal carcinomas? Pathol Oncol Res. 2012 Apr;18(2):253-62.
95. Hishida T, Yoshida J, Maeda R, Ishii G, Aokage K, Nishimura M, Nagai K. Prognostic impact of intratumoural microvascular invasion and microlymphatic permeation on node-negative non-small-cell lung cancer: which indicator is the stronger prognostic factor? Eur J Cardiothorac Surg. 2013 Apr;43(4):772-7.
96. Agarwal D, Pardhe N, Bajpai M, Gupta S, Mathur N, Vanaki SS, Puranik RS, Mittal M. Characterization, localization and patterning of lymphatics and blood vessels in oral squamous cell carcinoma: A comparative study using D2-40 and CD-34 IHC marker. J Clin Diagn Res. 2014 Oct;8(10):ZC86-9.
97. Rack B, Schindlbeck C, Jückstock J, Andergassen U, Hepp P, Zwingers T, Friedl TW, Lorenz R, Tesch H, Fasching PA, Fehm T, Schneeweiss A, Lichtenegger W, Beckmann MW, Friese K, Pantel K, Janni W; SUCCESS Study Group. Circulating tumor cells predict survival in early average-to-high risk breast cancer patients. J Natl Cancer Inst. 2014 May 15;106(5):dju066.

98. Kuo MC, Kothari AN, Kuo PC, Mi Z. Cancer stemness in bone marrow micrometastases of human breast cancer. Surgery. 2018 Feb;163(2):330-5.
99. Hartkopf AD, Taran FA, Wallwiener M, Hahn M, Becker S, Solomayer EF, Brucker SY, Fehm TN, Wallwiener D. Prognostic relevance of disseminated tumour cells from the bone marrow of early stage breast cancer patients – results from a large single-centre analysis. Eur J Cancer. 2014 Oct;50(15):2550-9.
100. Tang SL, Yang JQ, Du ZG, Tan QW, Zhou YT, Zhang D, Lv Q. Clinicopathologic study of invasive micropapillary carcinoma of the breast. Oncotarget. 2017 Jun 27;8(26):42455-65.
101. Gruel N, Benhamo V, Bhalshankar J, Popova T, Fréneaux P, Arnould L, Mariani O, Stern MH, Raynal V, Sastre-Garau X, Rouzier R, Delattre O, Vincent-Salomon A. Polarity gene alterations in pure invasive micropapillary carcinomas of the breast. Breast Cancer Res. 2014 May 8;16(3):R46.
102. Ni YB, Tse GM. Pathological criteria and practical issues in papillary lesions of the breast – a review. Histopathology. 2016 Jan;68(1):22-32.
103. Laser J, Cangiarella J, Singh B, Melamed J, Chiriboga L, Yee H, Darvishian F. Invasive lobular carcinoma of the breast: role of endothelial lymphatic marker D2-40. Ann Clin Lab Sci, 2008 Spring;38(2):99-104.
104. Fernández B, Paish EC, Green AR, Lee AH, Macmillan RD, Ellis IO, Rakha EA. Lymph-node metastases in invasive lobular carcinoma are different from those in ductal carcinoma of the breast. J Clin Pathol. 2011 Nov;64(11):995-1000.
105. Bentz JS, Yassa N, Clayton F. Pleomorphic lobular carcinoma of the breast: clinicopathologic features of 12 cases. Mod Pathol. 1998 Sep;11(9):814-22.
106. Buchanan CL, Flynn LW, Murray MP, Darvishian F, Cranor ML, Fey JV, King TA, Tan LK, Sclafani LM. Is pleomorphic lobular carcinoma really a distinct clinical entity? J Surg Oncol. 2008 Oct 1;98(5):314-7.
107. Sahin S, Karatas F, Erdem GU, Hacioglu B, Altundag K. Invasive pleomorphic lobular histology is an adverse prognostic factor on survival in patients with breast cancer. Am Surg. 2017 Apr;83(4):359-64.
108. Ueng SH, Mezzetti T, Tavassoli FA. Papillary neoplasms of the breast: A review. Arch Pathol Lab Med. 2009 Jun;133(6):893-907.
109. Song YJ, Shin SH, Cho JS, Park MH, Yoon JH, Jegal YJ. The role of lymphovascular invasion as a prognostic factor in patients with lymph node-positive operable invasive breast cancer. J Breast Cancer. 2011 Sep;14(3):198-203.
110. Truong PT, Yong CM, Abnousi F, Lee J, Kader HA, Hayashi A, Olivotto IA. Lymphovascular invasion is associated with reduced locoregional control and survival in women with node-negative breast cancer treated with mastectomy and systemic therapy. J Am Coll Surg. 2005 Jun;200(6):912-21.
111. Guarnieri A, Neri A, Correale PP, Lottini M, Testa M, Mariani F, Tucci E, Megha T, Cintorino M, Carli A. Prediction of lymph node status by analysis of prognostic factors and possible indications for elective axillary dissection in T1 breast cancers. Eur J Surg. 2001 Apr;167(4):255-9.
112. Choi SH, Barsky SH, Chang HR. Clinicopathologic analysis of sentinel lymph node mapping in early breast cancer. Breast J. 2003 May-Jun;9(3):153-62.
113. Chen HL, Ding A. Comparison of invasive micropapillary and triple negative invasive ductal carcinoma of the breast. Breast. 2015 Dec;24(6):723-31.
114. Ugras S, Stempel M, Patil S, Morrow M. Estrogen receptor, progesterone receptor, and HER2 status predict lymphovascular invasion and lymph node involvement. Ann Surg Oncol. 2014 Nov;21(12):3780-6.
115. Sabatier R, Finetti P, Guille A, Adelaide J, Chaffanet M, Viens P, Birnbaum D, Bertucci F. Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. Mol Cancer. 2014 Oct 2;13:228.
116. Holliday DL, Speirs V. Choosing the right cell line for breast cancer research. Breast Cancer Res. 2011 Aug 12;13(4):215.
117. Sato M, Taguchi A, Fukui Y, Kawata A, Taguchi S, Kashiyama T, Eguchi S, Inoue T, Tomio K, Tanikawa M, Sone K, Mori M, Nagasaka K, Adachi K, Ikemura M, Domoto Y, Arimoto T, Oda K, Fukayama M, Osuga Y, Fujii T. Blood vessel invasion is a

strong predictor of postoperative recurrence in endometrial cancer. Int J Gynecol Cancer. 2018 Jun;28(5):875-81.
118. Wu JK, Kitajewski C, Reiley M, Keung CH, Monteagudo J, Andrews JP, Liou P, Thirumoorthi A, Wong A, Kandel JJ, Shawber CJ. Aberrant lymphatic endothelial progenitors in lymphatic malformation development. PLoS One. 2015 Feb 26;10(2):e0117352.
119. Dadiani M, Kalchenko V, Yosepovich A, Margalit R, Hassid Y, Degani H, Seger D. Real-time imaging of lymphogenic metastasis in orthotopic human breast cancer. Cancer Res. 2006 Aug 15;66(16):8037-41.
120. Shields JD, Fleury ME, Yong C, Tomei AA, Randolph GJ, Swartz MA. Autologous chemotaxis as a mechanism of tumor cell homing to lymphatics via interstitial flow and autocrine CCR7 signaling. Cancer Cell, 2007 Jun;11(6):526-38.
121. Wong SY, Hynes RO. Lymphatic or hematogenous dissemination: how does a metastatic tumor cell decide? Cell Cycle. 2006 Apr;5(8):812-7.
122. Steinskog ES, Sagstad SJ, Wagner M, Karlsen TV, Yang N, Markhus CE, Yndestad S, Wiig H, Eikesdal HP. Impaired lymphatic function accelerates cancer growth. Oncotarget. 2016 Jul 19;7(29):45789-802.
123. Ran S, Volk L, Hall K, Flister MJ. Lymphangiogenesis and lymphatic metastasis in breast cancer. Pathophysiology, 2010 Sep;17(4):229-51.
124. Shi SJ, Wang LJ, Yu B, Li YH, Jin Y, Bai XZ. LncRNA-ATB promotes trastuzumab resistance and invasion-metastasis cascade in breast cancer. Oncotarget. 2015 May 10;6(13):11652-63.
125. Bhat-Nakshatri P, Appaiah H, Ballas C, Pick-Franke P, Goulet R Jr, Badve S, Srour EF, Nakshatri H. SLUG/SNAI2 and tumor necrosis factor generate breast cells with CD44+/CD24- phenotype. BMC Cancer. 2010 Aug 6;10:411.
126. Balkwill FR, Capasso M, Hagemann T. The tumor microenvironment at a glance. J Cell Sci. 2012 Dec 1;125(Pt 23):5591-6.
127. Yousef EM, Tahir MR, St-Pierre Y, Gaboury LA. MMP-9 expression varies according to molecular subtypes of breast cancer. BMC Cancer. 2014 Aug 23;14:609.
128. Wu QW, Yang QM, Huang YF, She HQ, Liang J, Yang QL, Zhang ZM. Expression and clinical significance of matrix metalloproteinase-9 in lymphatic invasiveness and metastasis of breast cancer. PLoS One. 2014 May 20;9(5):e97804.
129. Fidalgo F, Rodrigues TC, Pinilla M, Silva AG, Maciel Mdo S, Rosenberg C, de Andrade VP, Carraro DM, Krepischi AC. Lymphovascular invasion and histologic grade are associated with specific genomic profiles in invasive carcinomas of the breast. Tumour Biol. 2015 Mar;36(3):1835-48.
130. Hotary K, Li XY, Allen E, Stevens SL, Weiss SJ. A cancer cell metalloprotease triad regulates the basement membrane transmigration program. Genes Dev. 2006 Oct 1;20(19):2673-86.
131. Perentes JY, Kirkpatrick ND, Nagano S, Smith EY, Shaver CM, Sgroi D, Garkavtsev I, Munn LL, Jain RK, Boucher Y. Cancer cell–associated MT1-MMP promotes blood vessel invasion and distant metastasis in triple-negative mammary tumors. Cancer Res. 2011 Jul 1;71(13):4527-38.
132. Pula B, Wojnar A, Werynska B, Ambicka A, Kruczak A, Witkiewicz W, Ugorski M, Podhorska-Okolow M, Dziegiel P. Impact of different tumour stroma assessment methods regarding podoplanin expression on clinical outcome in patients with invasive ductal breast carcinoma. Anticancer Res. 2013 Apr;33(4):1447-55.
133. Pula B, Wojnar A, Witkiewicz W, Dziegiel P, Podhorska-Okolow M. Podoplanin expression in cancer-associated fibroblasts correlates with VEGF-C expression in cancer cells of invasive ductal breast carcinoma. Neoplasma. 2013;60(5):516-24.
134. Issa A, Le TX, Shoushtari AN, Shields JD, Swartz MA. Vascular endothelial growth factor-C and C-C chemokine receptor 7 in tumor cell–lymphatic cross-talk promote invasive phenotype. Cancer Res. 2009 Jan 1;69(1):349-57.
135. Wyckoff J, Wang W, Lin EY, Wang Y, Pixley F, Stanley ER, Graf T, Pollard JW, Segall J, Condeelis J. A paracrine loop between tumor cells and macrophages is required for tumor cell migration in mammary tumors. Cancer Res. 2004 Oct 1;64(19):7022-9.

136. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, Perou CM, Nielsen TO. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res. 2008 Mar 1;14(5):1368-76.
137. Huo Z, Gao Y, Zuo W, Zheng G, Kong R. Metastases of basal-like breast invasive ductal carcinoma to the endometrium: A case report and review of the literature. Thorac Cancer. 2015 Jul;6(4):548-52.
138. Joimel U, Gest C, Soria J, Pritchard LL, Alexandre J, Laurent M, Blot E, Cazin L, Vannier JP, Varin R, Li H, Soria C. Stimulation of angiogenesis resulting from cooperation between macrophages and MDA-MB-231 breast cancer cells: proposed molecular mechanism and effect of tetrathiomolybdate. BMC Cancer. 2010 Jul 17;10:375.
139. Cursiefen C, Chen L, Borges LP, Jackson D, Cao J, Radziejewski C, D'Amore PA, Dana MR, Wiegand SJ, Streilein JW. VEGF-A stimulates lymphangiogenesis and hemangiogenesis in inflammatory neovascularization via macrophage recruitment. J Clin Invest. 2004 Apr;113(7):1040-50.
140. Robinson BD, Sica GL, Liu YF, Rohan TE, Gertler FB, Condeelis JS, Jones JG. Tumor microenvironment of metastasis in human breast carcinoma: a potential prognostic marker linked to hematogenous dissemination. Clin Cancer Res. 2009 Apr 1;15(7):2433-41.
141. Cheung KJ, Ewald AJ. Illuminating breast cancer invasion: diverse roles for cell-cell interactions. Curr Opin Cell Biol. 2014 Oct;30:99-111.
142. Alizadeh D, Trad M, Hanke NT, Larmonier CB, Janikashvili N, Bonnotte B, Katsanis E, Larmonier N. Doxorubicin eliminates myeloid-derived suppressor cells and enhances the efficacy of adoptive T-cell transfer in breast cancer. Cancer Res. 2014 Jan 1;74(1):104-18.
143. Maniotis AJ, Folberg R, Hess A, Seftor EA, Gardner LM, Pe'er J, Trent JM, Meltzer PS, Hendrix MJ. Vascular channel formation by human melanoma cells in vivo and in vitro: vasculogenic mimicry. Am J Pathol. 1999 Sep;155(3):739-52.
144. Kirschmann DA, Seftor EA, Hardy KM, Seftor RE, Hendrix MJ. Molecular pathways: vasculogenic mimicry in tumor cells: diagnostic and therapeutic implications. Clin Cancer Res. 2012 May 15;18(10):2726-32.
145. Shirakawa K, Tsuda H, Heike Y, Kato K, Asada R, Inomata M, Sasaki H, Kasumi F, Yoshimoto M, Iwanaga T, Konishi F, Terada M, Wakasugi H. Absence of endothelial cells, central necrosis, and fibrosis are associated with aggressive inflammatory breast cancer. Cancer Res. 2001 Jan 15;61(2):445-51.
146. Shirakawa K, Kobayashi H, Heike Y, Kawamoto S, Brechbiel MW, Kasumi F, Iwanaga T, Konishi F, Terada M, Wakasugi H. Hemodynamics in vasculogenic mimicry and angiogenesis of inflammatory breast cancer xenograft. Cancer Res. 2002 Jan 15;62(2):560-6.
147. Liu T, Sun B, Zhao X, Li Y, Gu Q, Dong X, Liu F. OCT4 expression and vasculogenic mimicry formation positively correlate with poor prognosis in human breast cancer. Int J Mol Sci. 2014 Oct 28;15(11):19634-49.
148. Paulis YW, Soetekouw PM, Verheul HM, Tjan-Heijnen VC, Griffioen AW. Signalling pathways in vasculogenic mimicry. Biochim Biophys Acta. 2010 Aug;1806(1):18-28.
149. Qiao L, Liang N, Zhang J, Xie J, Liu F, Xu D, Yu X, Tian Y. Advanced research on vasculogenic mimicry in cancer. J Cell Mol Med. 2015 Feb;19(2):315-26.
150. Rakha EA, El-Sayed ME, Reis-Filho JS, Ellis IO. Expression profiling technology: its contribution to our understanding of breast cancer. Histopathology. 2008 Jan; 52(1):67-81.
151. Mannelqvist M, Stefansson IM, Bredholt G, Hellem Bø T, Oyan AM, Jonassen I, Kalland KH, Salvesen HB, Akslen LA. Gene expression patterns related to vascular invasion and aggressive features in endometrial cancer. Am J Pathol. 2011 Feb;178(2):861-71.
152. Mannelqvist M, Wik E, Stefansson IM, Akslen LA. An 18-gene signature for vascular invasion is associated with aggressive features and reduced survival in breast cancer. PLoS One. 2014 Jun 6;9(6):e98787.

153. Mínguez B, Hoshida Y, Villanueva A, Toffanin S, Cabellos L, Thung S, Mandeli J, Sia D, April C, Fan JB, Lachenmayer A, Savic R, Roayaie S, Mazzaferro V, Bruix J, Schwartz M, Friedman SL, Llovet JM. Gene-expression signature of vascular invasion in hepatocellular carcinoma. J Hepatol. 2011 Dec;55(6):1325-31.
154. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, Ménard S, Palazzo JP, Rosenberg A, Musiani P, Volinia S, Nenci I, Calin GA, Querzoli P, Negrini M, Croce CM. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 2005 Aug;65(16):7065-70.
155. Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature. 2007 Oct 11;449(7163):682-8.
156. Fiannaca A, La Rosa M, La Paglia L, Rizzo R, Urso A. Analysis of miRNA expression profiles in breast cancer using biclustering. BMC Bioinformatics. 2015;16 Suppl 4(Suppl 4):S7.
157. Zhao YC, Ni XJ, Wang MH, Zha XM, Zhao Y, Wang S. Tumor-derived VEGF-C, but not VEGF-D, promotes sentinel lymph node lymphangiogenesis prior to metastasis in breast cancer patients. Med Oncol. 2012 Dec;29(4):2594-600.
158. Zhang XH, Huang DP, Guo GL, Chen GR, Zhang HX, Wan L, Chen SY. Coexpression of VEGF-C and COX-2 and its association with lymphangiogenesis in human breast cancer. BMC Cancer. 2008 Jan 13;8:4.
159. Chiavenna SM, Jaworski JP, Vendrell A. State of the art in anti-cancer mAbs. J Biomed Sci. 2017 Feb 20;24(1):15.
160. Storr SJ, Mohammed RA, Woolston CM, Green AR, Parr T, Spiteri I, Caldas C, Ball GR, Ellis IO, Martin SG. Calpastatin is associated with lymphovascular invasion in breast cancer. Breast. 2011 Oct;20(5):413-8.
161. Storr SJ, Carragher NO, Frame MC, Parr T, Martin SG. The calpain system and cancer. Nat Rev Cancer. 2011 May;11(5):364-74.
162. Jerjees DA, Alabdullah M, Alkaabi M, Abduljabbar R, Muftah A, Nolan C, Green AR, Ellis IO, Rakha EA. ERK1/2 is related to oestrogen receptor and predicts outcome in hormone-treated breast cancer. Breast Cancer Res Treat. 2014 Aug;147(1):25-37.
163. Alshareeda AT, Negm OH, Albarakati N, Green AR, Nolan C, Sultana R, Madhusudan S, Benhasouna A, Tighe P, Ellis IO, Rakha EA. Clinicopathological significance of KU70/KU80, a key DNA damage repair protein in breast cancer. Breast Cancer Res Treat. 2013 Jun;139(2):301-10.
164. Alshareeda AT, Negm OH, Green AR, Nolan C, Tighe P, Albarakati N, Sultana R, Madhusudan S, Ellis IO, Rakha EA. SUMOylation proteins in breast cancer. Breast Cancer Res Treat. 2014 Apr;144(3):519-30.
165. Green AR, Caracappa D, Benhasouna AA, Alshareeda A, Nolan CC, Macmillan RD, Madhusudan S, Ellis IO, Rakha EA. Biological and clinical significance of PARP1 protein expression in breast cancer. Breast Cancer Res Treat. 2015 Jan;149(2):353-62.
166. Al-Kaabi MM, Alshareeda AT, Jerjees DA, Muftah AA, Green AR, Alsubhi NH, Nolan CC, Chan S, Cornford E, Madhusudan S, Ellis IO, Rakha EA. Checkpoint kinase1 (CHK1) is an important biomarker in breast cancer having a role in chemotherapy response. Br J Cancer. 2015 Mar 3;112(5):901-11.
167. Nguyen-Ngoc KV, Cheung KJ, Brenot A, Shamir ER, Gray RS, Hines WC, Yaswen P, Werb Z, Ewald AJ. ECM microenvironment regulates collective migration and local dissemination in normal and malignant mammary epithelium. Proc Natl Acad Sci U S A. 2012 Sep 25;109(39):E2595-604.
168. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012 May;2(5):401-4.
169. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013 Apr 2;6(269):pl1.

170. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C, Bowlby R, Shen H, Hayat S, Fieldhouse R, Lester SC, Tse GM, Factor RE, Collins LC, Allison KH, Chen YY, Jensen K, Johnson NB, Oesterreich S, Mills GB, Cherniack AD, Robertson G, Benz C, Sander C, Laird PW, Hoadley KA, King TA; TCGA Research Network, Perou CM. Comprehensive molecular portraits of invasive lobular breast cancer. Cell. 2015 Oct 8;163(2):506-19.
171. Senkus E, Kyriakides S, Penault-Llorca F, Poortmans P, Thompson A, Zackrisson S, Cardoso F; ESMO Guidelines Working Group. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2013 Oct;24 Suppl 6:vi7-23.
172. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, Macmillan D, Blamey RW, Ellis IO. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer. 2005 Sep 1;116(3):340-50.
173. Barros FF, Abdel-Fatah TM, Moseley P, Nolan CC, Durham AC, Rakha EA, Chan S, Ellis IO, Green AR. Characterisation of HER heterodimers in breast cancer using in situ proximity ligation assay. Breast Cancer Res Treat. 2014 Apr;144(2):273-85.
174. Rakha EA, Abd El Rehim D, Pinder SE, Lewis SA, Ellis IO. E-cadherin expression in invasive non-lobular carcinoma of the breast and its prognostic significance. Histopathology. 2005 Jun;46(6):685-93.
175. Abd El-Rehim DM, Pinder SE, Paish CE, Bell JA, Rampaul RS, Blamey RW, Robertson JF, Nicholson RI, Ellis IO. Expression and co-expression of the members of the epidermal growth factor receptor (EGFR) family in invasive breast carcinoma. Br J Cancer. 2004 Oct 18;91(8):1532-42.
176. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, Perez EA, Press MF, Rhodes A, Sturgeon C, Taube SE, Tubbs R, Vance GH, van de Vijver M, Wheeler TM, Hayes DF; American Society of Clinical Oncology/College of American Pathologists. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch Pathol Lab Med. 2007;131(1):18-43.
177. Towbin H, Staehelin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A. 1979 Sep;76(9):4350-4.
178. McCarty KS Jr, McCarty KS Sr. Histochemical approaches to steroid receptor analyses. Semin Diagn Pathol. 1984 Nov;1(4):297-308.
179. Aleskandarany MA, Sonbul SN, Mukherjee A, Rakha EA. Molecular mechanisms underlying lymphovascular invasion in invasive breast cancer. Pathobiology. 2015 Sep;82(3-4):113-23.
180. Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, Kwei KA, Hernandez-Boussard T, Wang P, Gazdar AF, Minna JD, Pollack JR. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One. 2009 Jul 3;4(7):e6146.
181. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. World J Clin Oncol. 2014 Aug 10;5(3):412-24.
182. Engel LW, Young NA, Tralka TS, Lippman ME, O'Brien SJ, Joyce MJ. Establishment and characterization of three new continuous cell lines derived from human breast carcinomas. Cancer Res. 1978 Oct;38(10):3352-64.
183. Trempe GL. Human breast cancer in culture. Recent Results Cancer Res. 1976(57):33-41.
184. Petrova NM, Zenkova MA, Chernolovskaya E. Structure - functions relations in small interfering RNAs. In: Andrade AO, Alves Pereira A, Naves ELM, Soares AB, editors. Practical Applications in Biomedical Engineering. Intech Open; 2013. p. 187-228.
185. Cory AH, Owen TC, Barltrop JA, Cory JG. Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. Cancer Commun. 1991 Jul;3(7):207-12.

186. Munshi A, Hobbs M, Meyn RE. Clonogenic cell survival assay. Methods Mol Med. 2005;110:21-8.
187. Grada A, Otero-Vinas M, Prieto-Castrillo F, Obagi Z, Falanga V. Research techniques made simple: Analysis of collective cell migration using the wound healing assay. J Invest Dermatol. 2017 Feb;137(2):e11-6.
188. Ibidi: cells in focus. Creating the gap: Different approaches [Internet]. c2021 [cited 2021 Oct 25]. Available from: https://ibidi.com/content/282-creating-the-gap.
189. Koopman G, Reutelingsperger CP, Kuijten GA, Keehnen RM, Pals ST, van Oers MH. Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. Blood. 1994 Sep 1;84(5):1415-20.
190. Riccardi C, Nicoletti I. Analysis of apoptosis by propidium iodide staining and flow cytometry. Nat Protoc. 2006;1(3):1458-61.
191. Safuan S. The assessment of lymphatic vessel invasion in cancer. LAP Lambert Academic Publishing; 2014.
192. Marshall DC, Webb TE, Hall RA, Salciccioli JD, Ali R, Maruthappu M. Trends in UK regional cancer mortality 1991-2007. Br J Cancer. 2016 Feb 2;114(3):340-7.
193. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, Symmans WF, Gonzalez-Angulo AM, Hennessy B, Green M, Cristofanilli M, Hortobagyi GN, Pusztai L. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol. 2008 Mar 10;26(8):1275-81.
194. Wo JY, Chen K, Neville BA, Lin NU, Punglia RS. Effect of very small tumor size on cancer-specific mortality in node-positive breast cancer. J Clin Oncol. 2011 Jul 1;29(19):2619-27.
195. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, Fox SB, Ichihara S, Jacquemier J, Lakhani SR, Palacios J, Richardson AL, Schnitt SJ, Schmitt FC, Tan PH, Tse GM, Badve S, Ellis IO. Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res, 2010;12(4):207.
196. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Larsimont D, Cardoso F, Peterse H, Nuyten D, Buyse M, Van de Vijver MJ, Bergh J, Piccart M, Delorenzi M. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst. 2006 Feb 15;98(4):262-72.
197. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, Brat DJ, Cooper LAD. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci U S A. 2018 Mar 27;115(13):e2970-9.
198. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S; METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012 Apr 18;486(7403):346-52.
199. Kadota K, Nakai Y, Shimizu K. A weighted average difference method for detecting differentially expressed genes from microarray data. Algorithms Mol Biol. 2008 Jun 26;3:8. doi: 10.1186/1748-7188-3-8.
200. Kurozumi S, Joseph C, Sonbul S, Alsaeed S, Kariri Y, Aljohani A, Raafat S, Alsaleem M, Ogden A, Johnston SJ, Aleskandarany MA, Fujii T, Shirabe K, Caldas C, Ashankyty I, Dalton L, Ellis IO, Desmedt C, Green AR, Mongan NP, Rakha EA. A key genomic subtype associated with lymphovascular invasion in invasive breast cancer. Br J Cancer. 2019 Jun;120(12):1129-36.
201. Zafeiris D, Rutella S, Ball GR. An artificial neural network integrated pipeline for biomarker discovery using alzheimer's disease as a case study. Comput Struct Biotechnol J. 2018 Feb 21;16:77-87.
202. Nevin L; PLOS Medicine Editors. Advancing the beneficial use of machine learning in health care and medicine: Toward a community understanding. PLoS Med. 2018 Nov 30;15(11):e1002708.

203. Lancashire LJ, Rees RC, Ball GR. Identification of gene transcript signatures predictive for estrogen receptor and lymph node status using a stepwise forward selection artificial neural network modelling approach. Artif Intell Med. 2008 Jun;43(2):99-111.
204. Swan AL, Stekel DJ, Hodgman C, Allaway D, Alqahtani MH, Mobasheri A, Bacardit J. A machine learning heuristic to identify biologically relevant and minimal biomarker panels from omics data. BMC Genomics, 2015;16 Suppl 1(Suppl 1):S2.
205. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet. 2015 Jun;16(6):321-32.
206. Abdel-Fatah TMA, Agarwal D, Liu DX, Russell R, Rueda OM, Liu K, Xu B, Moseley PM, Green AR, Pockley AG, Rees RC, Caldas C, Ellis IO, Ball GR, Chan SYT. SPAG5 as a prognostic biomarker and chemotherapy sensitivity predictor in breast cancer: a retrospective, integrated genomic, transcriptomic, and protein analysis. Lancet Oncol. 2016 Jul;17(7):1004-18.
207. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019 Jul 2;47(W1):W199-205.
208. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res. 2017 Jul 3;45(W1):W130-7.
209. Rakha EA, Abbas A, Pinto Ahumada P, ElSayed ME, Colman D, Pinder SE, Ellis IO. Diagnostic concordance of reporting lymphovascular invasion in breast cancer. J Clin Pathol. 2018 Sep;71(9):802-5.
210. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One. 2014 Jan 16;9(1):e78644.
211. Wolff A, Bayerlová M, Gaedcke J, Kube D, Beißbarth T. A comparative study of RNA-Seq and microarray data analysis on the two examples of rectal-cancer patients and Burkitt Lymphoma cells. PLoS One. 2018 May 16;13(5):e0197162.
212. Merker JD, Oxnard GR, Compton C, Diehn M, Hurley P, Lazar AJ, Lindeman N, Lockwood CM, Rai AJ, Schilsky RL, Tsimberidou AM, Vasalos P, Billman BL, Oliver TK, Bruinooge SS, Hayes DF, Turner NC. Circulating tumor DNA analysis in patients with cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review. J Clin Oncol. 2018 Jun 1;36(16):1631-41.
213. Cedolini C, Bertozzi S, Londero AP, Bernardi S, Seriau L, Concina S, Cattin F, Risaliti A. Type of breast cancer diagnosis, screening, and survival. Clin Breast Cancer. 2014 Aug;14(4):235-40.
214. Verhelst J, Hulpiau P, Saelens X. Mx proteins: antiviral gatekeepers that restrain the uninvited. Microbiol Mol Biol Rev. 2013 Dec;77(4):551-66.
215. Haller O, Kochs G. Interferon-induced mx proteins: dynamin-like GTPases with antiviral activity. Traffic. 2002 Oct;3(10):710-7.
216. Boerner JL, Nechiporchik N, Mueller KL, Polin L, Heilbrun L, Boerner SA, Zoratti GL, Stark K, LoRusso PM, Burger A. Protein expression of DNA damage repair proteins dictates response to topoisomerase and PARP inhibitors in triple-negative breast cancer. PLoS One. 2015 Mar 16;10(3):e0119614.
217. Borden EC. Interferons. In: Mendelsohn J, Howley PM, Israel MA, Gray J, Thompson C, editors. The molecular basis of cancer. Fourth Edition. Atlanta: Elsevier Inc.; 2014. p. 739-752.
218. Calmon MF, Rodrigues RV, Kaneto CM, Moura RP, Silva SD, Mota LD, Pinheiro DG, Torres C, de Carvalho AF, Cury PM, Nunes FD, Nishimoto IN, Soares FA, da Silva AM, Kowalski LP, Brentani H, Zanelli CF, Silva WA Jr, Rahal P, Tajara EH, Carraro DM, Camargo AA, Valentini SR. Epigenetic silencing of CRABP2 and MX1 in head and neck tumors. Neoplasia. 2009 Dec;11(12):1329-39.
219. Croner RS, Stürzl M, Rau TT, Metodieva G, Geppert CI, Naschberger E, Lausen B, Metodiev MV. Quantitative proteome profiling of lymph node-positive vs. -negative colorectal carcinomas pinpoints MX1 as a marker for lymph node metastasis. Int J Cancer. 2014 Dec 15;135(12):2878-86.

220. Lee SJ, Hwang CS, Kim YK, Lee HJ, Ahn SJ, Shin N, Lee JH, Shin DH, Choi KU, Park DY, Lee CH, Huh GY, Sol MY, Lee HJ, Gong G, Kim JY, Kim A. Expression of myxovirus resistance A (MxA) is associated with tumor-infiltrating lymphocytes in human epidermal growth factor receptor 2 (HER2)-positive breast cancers. Cancer Res Treat. 2017 Apr;49(2):313-21.
221. Zitvogel L, Galluzzi L, Kepp O, Smyth MJ, Kroemer G. Type I interferons in anticancer immunity. Nat Rev Immunol. 2015 Jul;15(7):405-14.
222. Han YY, Liu HY, Han DJ, Zong XC, Zhang SQ, Chen YQ. Role of glycosylation in the anticancer activity of antibacterial peptides against breast cancer cells. Biochem Pharmacol. 2013 Nov 1;86(9):1254-62.
223. Sistigu A, Yamazaki T, Vacchelli E, Chaba K, Enot DP, Adam J, Vitale I, Goubar A, Baracco EE, Remédios C, Fend L, Hannani D, Aymeric L, Ma Y, Niso-Santano M, Kepp O, Schultze JL, Tüting T, Belardelli F, Bracci L, La Sorsa V, Ziccheddu G, Sestili P, Urbani F, Delorenzi M, Lacroix-Triki M, Quidville V, Conforti R, Spano JP, Pusztai L, Poirier-Colame V, Delaloge S, Penault-Llorca F, Ladoire S, Arnould L, Cyrta J, Dessoliers MC, Eggermont A, Bianchi ME, Pittet M, Engblom C, Pfirschke C, Préville X, Uzè G, Schreiber RD, Chow MT, Smyth MJ, Proietti E, André F, Kroemer G, Zitvogel L. Cancer cell–autonomous contribution of type I interferon signaling to the efficacy of chemotherapy. Nat Med. 2014 Nov;20(11):1301-9.
224. Smolková K, Ježek P. The Role of mitochondrial NADPH-dependent isocitrate dehydrogenase in cancer cells. Int J Cell Biol. 2012;2012:273947.
225. Teoh ST, Lunt SY. Metabolism in cancer metastasis: bioenergetics, biosynthesis, and beyond. Wiley Interdiscip Rev Syst Biol Med. 2018 Mar;10(2).
226. Bergaggio E, Piva R. Wild-type IDH enzymes as actionable targets for cancer therapy. Cancers (Basel). 2019 Apr 19;11(4):563.
227. Raineri S, Mellor J. *IDH1*: linking metabolism and epigenetics. Front Genet. 2018 Oct 23;9:493.
228. Chiang S, Weigelt B, Wen HC, Pareja F, Raghavendra A, Martelotto LG, Burke KA, Basili T, Li A, Geyer FC, Piscuoglio S, Ng CK, Jungbluth AA, Balss J, Pusch S, Baker GM, Cole KS, von Deimling A, Batten JM, Marotti JD, Soh HC, McCalip BL, Serrano J, Lim RS, Siziopikou KP, Lu S, Liu X, Hammour T, Brogi E, Snuderl M, Iafrate AJ, Reis-Filho JS, Schnitt SJ. IDH2 mutations define a unique subtype of breast cancer with altered nuclear polarity. Cancer Res. 2016 Dec 15;76(24):7118-29.
229. Li J, He Y, Tan Z, Lu J, Li L, Song X, Shi F, Xie L, You S, Luo X, Li N, Li Y, Liu X, Tang M, Weng X, Yi W, Fan J, Zhou J, Qiang G, Qiu S, Wu W, Bode AM, Cao Y. Wild-type IDH2 promotes the Warburg effect and tumor growth through HIF1α in lung cancer. Theranostics. 2018 Jul 16;8(15):4050-61.
230. Jiang CY, Gao Y, Wang XJ, Ruan Y, Bei XY, Wang XH, Jing YF, Zhao W, Jiang Q, Li J, Han BM, Xia SJ, Zhao FJ. Long non-coding RNA lnc-MX1-1 is associated with poor clinical features and promotes cellular proliferation and invasiveness in prostate cancer. Biochem Biophys Res Commun. 2016 Feb 12;470(3):721-7.
231. Hu JL, Hua YJ, Chen Y, Yu B, Gao S. Structural analysis of tumor-related single amino acid mutations in human MxA protein. Chin J Cancer. 2015 Sep 28;34(12):583-93.
232. Horisberger MA. Interferon-induced human protein MxA is a GTPase which binds transiently to cellular proteins. J Virol. 1992 Aug;66(8):4705-9.
233. Nagano K, Masters JR, Akpan A, Yang A, Corless S, Wood C, Hastie C, Zvelebil M, Cramer R, Naaby-Hansen S. Differential protein synthesis and expression levels in normal and neoplastic human prostate cells and their regulation by type I and II interferons. Oncogene. 2004 Mar 4;23(9):1693-703.
234. Jonasch E, Haluska FG. Interferon in oncological practice: review of interferon biology, clinical applications, and toxicities. Oncologist. 2001;6(1):34-55.
235. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett. 2009 Dec 17;583(24):3966-73.
236. Rizzardi AE, Johnson AT, Vogel RI, Pambuccian SE, Henriksen J, Skubitz AP, Metzger GJ, Schmechel SC. Quantitative comparison of immunohistochemical

staining measured by digital image analysis versus pathologist visual scoring. Diagn Pathol. 2012 Jun 20;7:42.
237. Heerboth S, Housman G, Leary M, Longacre M, Byler S, Lapinska K, Willbanks A, Sarkar S. EMT and tumor metastasis. Clin Transl Med. 2015 Feb 26;4:6.
238. Huang H. Matrix metalloproteinase-9 (MMP-9) as a cancer biomarker and MMP-9 biosensors: Recent advances. Sensors (Basel). 2018 Sep 27;18(10):3249.
239. Gluz O, Liedtke C, Gottschalk N, Pusztai L, Nitz U, Harbeck N. Triple-negative breast cancer--current status and future directions. Ann Oncol. 2009 Dec;20(12):1913-27.
240. Liu YL, Saraf A, Lee SM, Zhong X, Hibshoosh H, Kalinsky K, Connolly EP. Lymphovascular invasion is an independent predictor of survival in breast cancer after neoadjuvant chemotherapy. Breast Cancer Res Treat. 2016 Jun;157(3):555-64.
241. Weigelt B, Lo AT, Park CC, Gray JW, Bissell MJ. HER2 signaling pathway activation and response of breast cancer cells to HER2-targeting agents is dependent strongly on the 3D microenvironment. Breast Cancer Res Treat. 2010 Jul;122(1):35-43.
242. Garrett M, Fujii Y, Osaka N, Ito D, Hirota Y, Sasaki A, Debinski W. Emerging roles of wild-type and mutant IDH1 in growth, metabolism and therapeutics of glioma. Gliomas [Internet]. 2021 Apr 30. Chapter 4.
243. Terunuma A, Putluri N, Mishra P, Mathé EA, Dorsey TH, Yi M, Wallace TA, Issaq HJ, Zhou M, Killian JK, Stevenson HS, Karoly ED, Chan K, Samanta S, Prieto D, Hsu TY, Kurley SJ, Putluri V, Sonavane R, Edelman DC, Wulff J, Starks AM, Yang Y, Kittles RA, Yfantis HG, Lee DH, Ioffe OB, Schiff R, Stephens RM, Meltzer PS, Veenstra TD, Westbrook TF, Sreekumar A, Ambs S. MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. J Clin Invest. 2014 Jan;124(1):398-412.
244. Fang X, Cai Y, Liu J, Wang Z, Wu Q, Zhang Z, Yang CJ, Yuan L, Ouyang G. Twist2 contributes to breast cancer progression by promoting an epithelial–mesenchymal transition and cancer stem-like cell self-renewal. Oncogene. 2011 Nov 24;30:4707.
245. Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. Cancer Cell. 2012 Mar 20;21(3):297-308.
246. Cantor JR, Sabatini DM. Cancer cell metabolism: one hallmark, many faces. Cancer Discov. 2012 Oct;2(10):881-98.
247. El Ansari R, Craze ML, Miligy I, Diez-Rodriguez M, Nolan CC, Ellis IO, Rakha EA, Green AR. The amino acid transporter SLC7A5 confers a poor prognosis in the highly proliferative breast cancer subtypes and is a key therapeutic target in luminal B tumours. Breast Cancer Res. 2018 Mar 22;20(1):21.
248. Lv Q, Xing S, Li Z, Li J, Gong P, Xu X, Chang L, Jin X, Gao F, Li W, Zhang G, Yang J, Zhang X. Altered expression levels of IDH2 are involved in the development of colon cancer. Exp Ther Med. 2012 Nov;4(5):801-806.
249. Hunter KW, Crawford NP, Alsarraj J. Mechanisms of metastasis. Breast Cancer Res. 2008;10 Suppl 1(Suppl 1):S2.
250. Fang J, Ge X, Xu W, Xie J, Qin Z, Shi L, Yin W, Bian M, Wang H. Bioinformatics analysis of the prognosis and biological significance of HMGB1, HMGB2, and HMGB3 in gastric cancer. J Cell Physiol. 2020 Apr;235(4):3438-46.
251. Agresti A, Bianchi ME. HMGB proteins and gene expression. Curr Opin Genet Dev. 2003 Apr;13(2):170-8.
252. Štros M. HMGB proteins: Interactions with DNA and chromatin. Biochim Biophys Acta. 2010 Jan-Feb;1799(1-2):101-13.
253. Guo ZS, Liu Z, Bartlett DL, Tang D, Lotze MT. Life after death: targeting high mobility group box 1 in emergent cancer therapies. Am J Cancer Res. 2013;3(1):1-20.
254. Zhang Z, Chang Y, Zhang J, Lu Y, Zheng L, Hu Y, Zhang F, Li X, Zhang W, Li X. HMGB3 promotes growth and migration in colorectal cancer by regulating WNT/β-catenin pathway. PloS One. 2017 Jul 5;12(7):e0179741.

255. Zheng W, Yang J, Dong Z, Wang L, Fang M, Wu W, Yao D, Yao M. High mobility group box 3 as an emerging biomarker in diagnosis and prognosis of hepatocellular carcinoma. Cancer Manag Res. 2018 Nov 20;10:5979-89.
256. Takeda T, Izumi H, Kitada S, Uramoto H, Tasaki T, Zhi L, Guo X, Kawatsu Y, Kimura T, Horie S, Nabeshima A, Noguchi H, Wang KY, Sasaguri Y, Kohno K, Yamada S. The combination of a nuclear HMGB1-positive and HMGB2-negative expression is potentially associated with a shortened survival in patients with pancreatic ductal adenocarcinoma. Tumour Biol. 2014 Oct;35(10):10555-69.
257. Ueda M, Takahashi Y, Shinden Y, Sakimura S, Hirata H, Uchi R, Takano Y, Kurashige J, Iguchi T, Eguchi H, Sugimachi K, Yamamoto H, Doki Y, Mori M, Mimori K. Prognostic significance of high mobility group box 1 (HMGB1) expression in patients with colorectal cancer. Anticancer Res. 2014 Oct;34(10):5357-62.
258. Guo S, Wang Y, Gao Y, Zhang Y, Chen M, Xu M, Hu L, Jing Y, Jing F, Li C, Wang Q, Zhu Z. Knockdown of high mobility group-box 3 (HMGB3) expression inhibits proliferation, reduces migration, and affects chemosensitivity in gastric cancer cells. Med Sci Monit. 2016 Oct 24;22:3951-60.
259. Wang K, Yin YH, Yang ZQ, Yu HF, Wang YS, Guo B, Yue ZP. Hmgb3 induces the differentiation of uterine stromal cells through targeting Ptn. Reprod Sci. 2019 Jul;26(7):891-9.
260. Wen B, Wei YT, Zhao K. The role of high mobility group protein B3 (HMGB3) in tumor proliferation and drug resistance. Mol Cell Biochem. 2021 Apr;476(4):1729-39.
261. Xie Y, Wang L, Yang D. CircEPSTI1 promotes the progression of non-small cell lung cancer through miR-145/HMGB3 axis. Cancer Manag Res. 2020 Aug 4;12:6827-36.
262. Li M, Cai Y, Zhao H, Xu Z, Sun Q, Luo M, Gu L, Meng M, Han X, Sun H. Overexpression of HMGB3 protein promotes cell proliferation, migration and is associated with poor prognosis in urinary bladder cancer patients. Tumour Biol. 2015 Jun;36(6):4785-92.
263. Yamada Y, Nishikawa R, Kato M, Okato A, Arai T, Kojima S, Yamazaki K, Naya Y, Ichikawa T, Seki N. Regulation of HMGB3 by antitumor miR-205-5p inhibits cancer cell aggressiveness and is involved in prostate cancer pathogenesis. J Hum Genet. 2018 Feb;63(2):195-205.
264. Zheng WJ, Yao M, Fang M, Wang L, Dong ZZ, Yao DF. Abnormal expression of *HMGB-3* is significantly associated with malignant transformation of hepatocytes. World J Gastroenterol. 2018 Aug 28;24(32):3650-62.
265. Gong Y, Cao Y, Song L, Zhou J, Wang C, Wu B. HMGB3 characterization in gastric cancer. Genet Mol Res. 2013 Dec 2;12(4):6032-9.
266. Song N, Liu B, Wu JL, Zhang RF, Duan L, He WS, Zhang CM. Prognostic value of HMGB3 expression in patients with non-small cell lung cancer. Tumour Biol. 2013 Oct;34(5):2599-603.
267. Liu J, Wang L, Li X. HMGB3 promotes the proliferation and metastasis of glioblastoma and is negatively regulated by miR-200b-3p and miR-200c-3p. Cell Biochem Funct. 2018 Oct;36(7):357-65.
268. Kruse JP, Gu W. Modes of p53 regulation. Cell. 2009 May 15;137(4):609-22.
269. Armstrong MJ, Stang MT, Liu Y, Gao J, Ren B, Zuckerbraun BS, Mahidhara RS, Xing Q, Pizzoferrato E, Yim JH. Interferon regulatory factor 1 (IRF-1) induces p21(WAF1/CIP1) dependent cell cycle arrest and p21(WAF1/CIP1) independent modulation of survivin in cancer cells. Cancer Lett. 2012 Jun 1;319(1):56-65.
270. Xu C, Wu A, Zhu H, Fang H, Xu L, Ye J, Shen J. Melatonin is involved in the apoptosis and necrosis of pancreatic cancer cell line SW-1990 via modulating of Bcl-2/Bax balance. Biomed Pharmacother. 2013 Mar;67(2):133-9.
271. Mrozik KM, Blaschuk OW, Cheong CM, Zannettino ACW, Vandyke K. N-cadherin in cancer metastasis, its emerging role in haematological malignancies and potential as a therapeutic target in cancer. BMC Cancer. 2018 Oct 1;18(1):939.

272. Ribeiro AS, Paredes J. P-cadherin linking breast cancer stem cells and invasion: A promising marker to identify an "intermediate/metastable" EMT state. Front Oncol. 2015 Jan 5;4:371.
273. Akrami M, Meshksar A, Ghoddusi JM, Safarpour MM, Tahmasebi S, Zangouri V, Talei A. Prognostic role of lymphovascular invasion in patients with early breast cancer. Indian J Surg Oncol. 2021 Dec;12(4):671-77.
274. Thelander M, Gräslund A, Thelander L. Subunit M2 of mammalian ribonucleotide reductase. Characterization of a homogeneous protein isolated from M2-overproducing mouse cells. J Biol Chem. 1985 Mar 10;260(5):2737-41.
275. Lin ZP, Belcourt MF, Cory JG, Sartorelli AC. Stable suppression of the R2 subunit of ribonucleotide reductase by R2-targeted short interference RNA sensitizes p53(-/-) HCT-116 colon cancer cells to DNA-damaging agents and ribonucleotide reductase inhibitors. J Biol Chem. 2004 Jun 25;279(26):27030-8.
276. Yang Y, Lin J, Guo S, Xue X, Wang Y, Qiu S, Cui J, Ma L, Zhang X, Wang J. RRM2 protects against ferroptosis and is a tumor biomarker for liver cancer. Cancer Cell Int. 2020 Dec 7;20(1):587.
277. Chabes A, Thelander L. Controlled protein degradation regulates ribonucleotide reductase activity in proliferating mammalian cells during the normal cell cycle and in response to DNA damage and replication blocks. J Biol Chem. 2000 Jun 9;275(23):17747-53.
278. Torrents E, Aloy P, Gibert I, Rodríguez-Trelles F. Ribonucleotide reductases: Divergent evolution of an ancient enzyme. J Mol Evol. 2002 Aug;55(2):138-52.
279. Nordlund P, Reichard P. Ribonucleotide reductases. Annu Rev Biochem. 2006;75:681-706.
280. D'Angiolella V, Donato V, Forrester FM, Jeong YT, Pellacani C, Kudo Y, Saraf A, Florens L, Washburn MP, Pagano M. Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair. Cell. 2012 May 25;149(5):1023-34.
281. Chen WX, Yang LG, Xu LY, Cheng L, Qian Q, Sun L, Zhu YL. Bioinformatics analysis revealing prognostic significance of RRM2 gene in breast cancer. Biosci Rep. 2019 Apr 9;39(4):BSR20182062.
282. Yoshida Y, Tsunoda T, Doi K, Tanaka Y, Fujimoto T, Machida T, Ota T, Koyanagi M, Takashima Y, Sasazuki T, Kuroki M, Iwasaki A, Shirasawa S. KRAS-mediated up-regulation of RRM2 expression is essential for the proliferation of colorectal cancer cell lines. Anticancer Res. 2011 Jul;31(7):2535-9.
283. Morikawa T, Maeda D, Kume H, Homma Y, Fukayama M. Ribonucleotide reductase M2 subunit is a novel diagnostic marker and a potential therapeutic target in bladder cancer. Histopathology. 2010 Dec. 57(6):885-92.
284. Li J, Pang J, Liu Y, Zhang J, Zhang C, Shen G, Song L. Suppression of RRM2 inhibits cell proliferation, causes cell cycle arrest and promotes the apoptosis of human neuroblastoma cells and in human neuroblastoma RRM2 is suppressed following chemotherapy. Oncol Rep. 2018 Jul;40(1):355-60.
285. Mazzu YZ, Armenia J, Chakraborty G, Yoshikawa Y, Coggins SA, Nandakumar S, Gerke TA, Pomerantz MM, Qiu X, Zhao H, Atiq M, Khan N, Komura K, Lee GM, Fine SW, Bell C, O'Connor E, Long HW, Freedman ML, Kim B, Kantoff PW. A novel mechanism driving poor-prognosis prostate cancer: Overexpression of the DNA repair gene, ribonucleotide reductase small subunit M2 (RRM2). Clin Cancer Res. 2019 Jul 15;25(14):4480-92.
286. Shin JU, Lee CH, Lee KT, Lee JK, Lee KH, Kim KM, Kim KM, Park SM, Rhee JC. Prognostic significance of ATM and cyclin B1 in pancreatic neuroendocrine tumor. Tumour Biol. 2012 Oct;33(5):1645-51.
287. Morgan DO. Principles of CDK regulation. Nature. 1995 Mar 9;374(6518):131-4.
288. Fang Y, Yu H, Liang X, Xu J, Cai X. Chk1-induced CCNB1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer. Cancer Biol Ther. 2014 Sep;15(9):1268-79.
289. Kakino S, Sasaki K, Kurose A, Ito H. Intracellular localization of cyclin B1 during the cell cycle in glioma cells. Cytometry. 1996 May1;24(1):49-54.

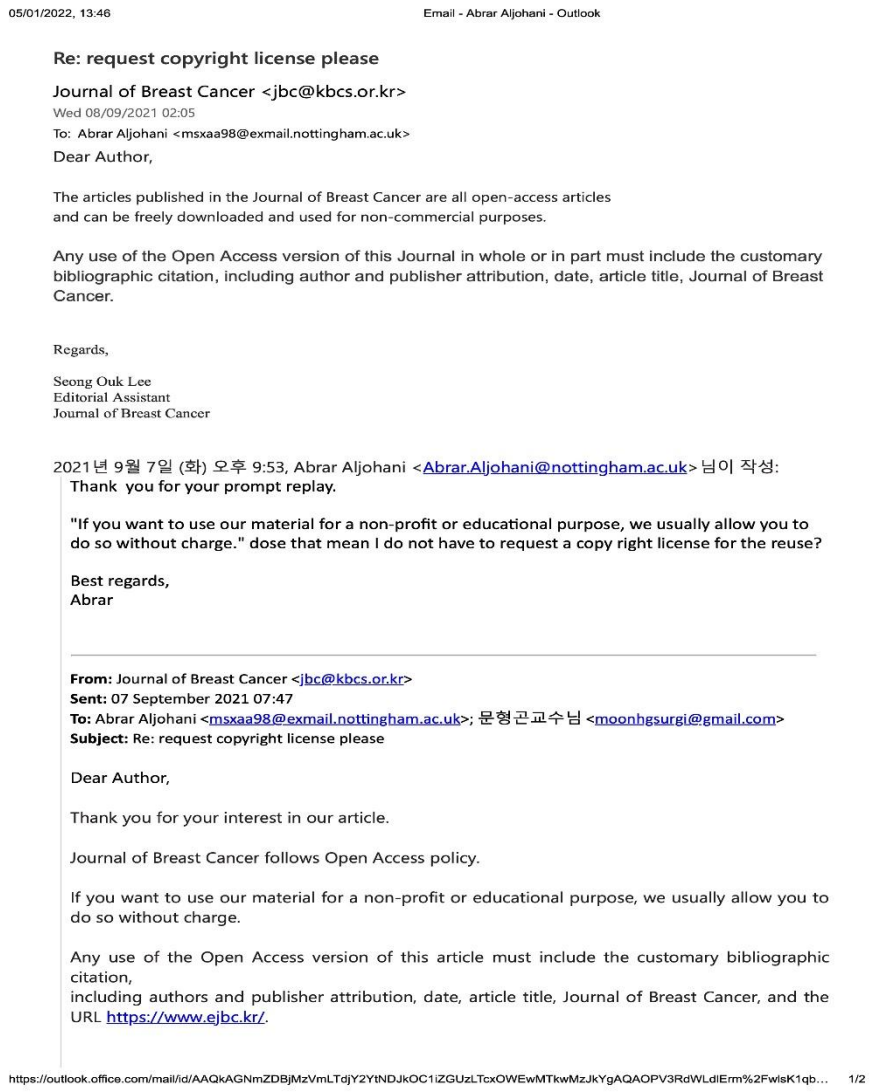
290. Winters ZE, Hunt NC, Bradburn MJ, Royds JA, Turley H, Harris AL, Norbury CJ. Subcellular localisation of cyclin B, Cdc2 and p21(WAF1/CIP1) in breast cancer. association with prognosis. Eur J Cancer. 2001 Dec;37(18):2405-12.
291. Kedinger V, Meulle A, Zounib O, Bonnet ME, Gossart JB, Benoit E, Messmer M, Shankaranarayanan P, Behr JP, Erbacher P, Bolcato-Bellemin AL. Sticky siRNAs targeting survivin and cyclin B1 exert an antitumoral effect on melanoma subcutaneous xenografts and lung metastases. BMC Cancer. 2013 Jul;13:338.
292. Innocente SA, Abrahamson JL, Cogswell JP, Lee JM. p53 regulates a G2 checkpoint through cyclin B1. Proc Natl Acad Sci U S A. 1999 Mar 2;96(5):2147-52.
293. Niméus-Malmström E, Koliadi A, Ahlin C, Holmqvist M, Holmberg L, Amini RM, Jirström K, Wärnberg F, Blomqvist C, Fernö M, Fjällskog ML. Cyclin B1 is a prognostic proliferation marker with a high reproducibility in a population-based lymph node negative breast cancer cohort. Int J Cancer. 2010 Aug 15;127(4):961-7.
294. Zhou L, Li J, Zhao YP, Cui QC, Zhou WX, Guo JC, You L, Wu WM, Zhang TP. The prognostic value of Cyclin B1 in pancreatic cancer. Med Oncol. 2014 Sep;31(9):107.
295. Song Y, Zhao C, Dong L, Fu M, Xue L, Huang Z, Tong T, Zhou Z, Chen A, Yang Z, Lu N, Zhan Q. Overexpression of cyclin B1 in human esophageal squamous cell carcinoma cells induces tumor cell invasive growth and metastasis. Carcinogenesis. 2008 Feb;29(2):307-15.
296. Matthess Y, Raab M, Sanhaji M, Lavrik IN, Strebhardt K. Cdk1/cyclin B1 controls Fas-mediated apoptosis by regulating caspase-8 activity. Mol Cell Biol. 2010 Dec;30(24):5726-40.
297. Gomez LA, de Las Pozas A, Reiner T, Burnstein K, Perez-Stable C. Increased expression of cyclin B1 sensitizes prostate cancer cells to apoptosis induced by chemotherapy. Mol Cancer Ther. 2007 May;6(5):1534-43.
298. Stamatakos M, Palla V, Karaiskos I, Xiromeritis K, Alexiou I, Pateras I, Kontzoglou K. Cell cyclins: triggering elements of cancer or not? World J Surg Oncol. 2010 Dec 22;8:111.
299. Li R, Jiang X, Zhang Y, Wang S, Chen X, Yu X, Ma J, Huang X. Cyclin B2 overexpression in human hepatocellular carcinoma is associated with poor prognosis. Arch Med Res. 2019 Jan;50(1):10-17.
300. Shubbar E, Kovács A, Hajizadeh S, Parris TZ, Nemes S, Gunnarsdóttir K, Einbeigi Z, Karlsson P, Helou K. Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. BMC Cancer. 2013 Jan 2;13:1.
301. Petri ET, Errico A, Escobedo L, Hunt T, Basavappa R. The crystal structure of human cyclin B. Cell Cycle. 2007 Jun 1;6(11):1342-9.
302. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer. 2009 Mar;9(3):153-66.
303. Bellanger S, de Gramont A, Sobczak-Thépot J. Cyclin B2 suppresses mitotic failure and DNA re-replication in human somatic cells knocked down for both cyclins B1 and B2. Oncogene. 2007 Nov 8;26(51):7175-84.
304. Wu T, Zhang X, Huang X, Yang Y, Hua X. Regulation of cyclin B2 expression and cell cycle G2/M transition by menin. J Biol Chem. 2010 Jun 11;285(24):18291-300.
305. Qian X, Song X, He Y, Yang Z, Sun T, Wang J, Zhu G, Xing W, You C. CCNB2 overexpression is a poor prognostic biomarker in Chinese NSCLC patients. Biomed Pharmacother. 2015 Aug;74:222-7.
306. Lei CY, Wang W, Zhu YT, Fang WY, Tan WL. The decrease of cyclin B2 expression inhibits invasion and metastasis of bladder cancer. Urol Oncol. 2016 May;34(5):237.e1-10.
307. Bolognese F, Wasner M, Dohna CL, Gurtner A, Ronchi A, Muller H, Manni I, Mossner J, Piaggio G, Mantovani R, Engeland K. The cyclin B2 promoter depends on NF-Y, a trimer whose CCAAT-binding activity is cell-cycle regulated. Oncogene. 1999 Mar 11;18(10):1845-53.

308. Kariri YA, Aleskandarany MA, Joseph C, Kurozumi S, Mohammed OJ, Toss MS, Green AR, Rakha EA. Molecular complexity of lymphovascular invasion: The role of cell migration in breast cancer as a prototype. Pathobiology. 2020;87(4):218-31.
309. Liu X, Peng J, Zhou Y, Xie B, Wang J. Silencing RRM2 inhibits multiple myeloma by targeting the Wnt/β-catenin signaling pathway. Mol Med Rep. 2019 Sep;20(3):2159-66.
310. Li C, Zheng J, Chen S, Huang B, Li G, Feng Z, Wang J, Xu S. RRM2 promotes the progression of human glioblastoma. J Cell Physiol. 2018 Oct;233(10):6759-67.
311. Zhuang S, Li L, Zang Y, Li G, Wang F. RRM2 elicits the metastatic potential of breast cancer cells by regulating cell invasion, migration and VEGF expression via the PI3K/AKT signaling. Oncol Lett. 2020 Apr;19(4):3349-55.
312. Liu Y, Beyer A, Aebersold R. On the dependency of cellular protein levels on mRNA abundance. Cell. 2016 Apr 21;165(3):535-50.
313. Liang WH, Li N, Yuan ZQ, Qian XL, Wang ZH. DSCAM-AS1 promotes tumor growth of breast cancer by reducing miR-204-5p and up-regulating RRM2. Mol Carcinog. 2019 Apr;58(4):461-73.
314. de Klerk E, 't Hoen PA. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. Trends Genet. 2015 Mar;31(3):128-39.
315. Venkov C, Plieth D, Ni T, Karmaker A, Bian A, George AL Jr, Neilson EG. Transcriptional networks in epithelial-mesenchymal transition. PloS One. 2011;6(9):e25354.
316. Zhou BS, Tsai P, Ker R, Tsai J, Ho R, Yu J, Shih J, Yen Y. Overexpression of transfected human ribonucleotide reductase M2 subunit in human cancer cells enhances their invasive potential. Clin Exp Metastasis. 1998 Jan;16(1):43-9.
317. Souglakos J, Boukovinas I, Taron M, Mendez P, Mavroudis D, Tripaki M, Hatzidaki D, Koutsopoulos A, Stathopoulos E, Georgoulias V, Rosell R. Ribonucleotide reductase subunits M1 and M2 mRNA expression levels and clinical outcome of lung adenocarcinoma patients treated with docetaxel/gemcitabine. Br J Cancer. 2008 May 20;98(10):1710-5.
318. Han P, Chen RH, Wang F, Zeng JY, Yu ST, Xu LH, Cai Q, Liang FY, Xia TL, Lin ZR, Zhong Q, Huang XM. Novel chimeric transcript RRM2-c2orf48 promotes metastasis in nasopharyngeal carcinoma. Cell Death Dis. 2017 Sep;8(9):e3047.
319. Jin CY, Du L, Nuerlan AH, Wang XL, Yang YW, Guo R. High expression of RRM2 as an independent predictive factor of poor prognosis in patients with lung adenocarcinoma. Aging (Albany NY). 2020 Dec 19;13(3):3518-35.
320. Spaderna S, Schmalhofer O, Wahlbuhl M, Dimmler A, Bauer K, Sultan A, Hlubek F, Jung A, Strand D, Eger A, Kirchner T, Behrens J, Brabletz T. The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. Cancer Res. 2008 Jan 15;68(2):537-44.
321. Zheng H, Kang Y. Multilayer control of the EMT master regulators. Oncogene. 2014 Apr 3;33(14):1755-63.
322. Park M, Chae HD, Yun J, Jung M, Kim YS, Kim SH, Han MH, Shin DY. Constitutive activation of cyclin B1-associated cdc2 kinase overrides p53-mediated G2-M arrest. Cancer Res. 2000 Feb 1;60(3):542-5.
323. Guardavaccaro D, Pagano M. Stabilizers and destabilizers controlling cell cycle oscillators. Mol Cell. 2006 Apr 7;22(1):1-4.
324. Sherr CJ. Cancer cell cycles. Science. 1996 Dec 6;274(5293):1672-7.
325. Ye C, Wang J, Wu P, Li X, Chai Y. Prognostic role of cyclin B1 in solid tumors: a meta-analysis. Oncotarget. 2017 Jan 10;8(2):2224-32.
326. Soria JC, Jang SJ, Khuri FR, Hassan K, Liu D, Hong WK, Mao L. Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication. Cancer Res. 2000 Aug 1;60(15):4000-4.
327. Weng L, Du J, Zhou Q, Cheng B, Li J, Zhang D, Ling C. Identification of cyclin B1 and Sec62 as biomarkers for recurrence in patients with HBV-related hepatocellular carcinoma after surgical resection. Mol Cancer. 2012 Jun 8;11:39.
328. Kawamoto H, Koizumi H, Uchikoshi T. Expression of the G2-M checkpoint regulators cyclin B1 and cdc2 in nonmalignant and malignant human breast lesions:

immunocytochemical and quantitative image analyses. Am J Pathol. 1997 Jan;150(1):15-23.
329. Pines J, Hunter T. Human cyclins A and B1 are differentially located in the cell and undergo cell cycle-dependent nuclear transport. J Cell Biol. 1991 Oct;115(1):1-17.
330. Toyoshima F, Moriguchi T, Wada A, Fukuda M, Nishida E. Nuclear export of cyclin B1 and its possible role in the DNA damage-induced G2 checkpoint. EMBO J. 1998 May 15;17(10):2728-35.
331. Taylor WR, DePrimo SE, Agarwal A, Agarwal ML, Schönthal AH, Katula KS, Stark GR. Mechanisms of G2 arrest in response to overexpression of p53. Mol Biol Cell. 1999 Nov;10(11):3607-22.
332. Grabsch H, Lickvers K, Hansen O, Takeno S, Willers R, Stock W, Gabbert HE, Mueller W. Prognostic value of cyclin B1 protein expression in colorectal cancer. Am J Clin Pathol. 2004 Oct;122(4):511-6.
333. Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nat Rev Cancer. 2003 Jun;3(6):453-8.
334. Pantel K, Brakenhoff RH. Dissecting the metastatic cascade. Nat Rev Cancer. 2004 Jun;4(6):448-56.
335. Christofori G. New signals from the invasive front. Nature. 2006 May 25;441(7092):444-50.
336. Melzer C, von der Ohe J, Hass R. Breast carcinoma: From initial tumor cell detachment to settlement at secondary sites. Biomed Res Int. 2017;2017:8534371.
337. Li B, Zhu HB, Song GD, Cheng JH, Li CZ, Zhang YZ, Zhao P. Regulating the CCNB1 gene can affect cell proliferation and apoptosis in pituitary adenomas and activate epithelial-to-mesenchymal transition. Oncol Lett. 2019 Nov;18(5):4651-8.
338. Begnami MD, Fregnani JH, Nonogaki S, Soares FA. Evaluation of cell cycle protein expression in gastric cancer: cyclin B1 expression and its prognostic implication. Hum Pathol. 2010 Aug;41(8):1120-7.
339. Debled M, de Mascarel I, Brouste V, Mauriac L, MacGrogan G. Re: Population-based study of peritumoral lymphovascular invasion and outcome among patients with operable breast cancer. J Natl Cancer Inst. 2010 Feb 24;102(4):275-6.
340. Tang J, Kong D, Cui Q, Wang K, Zhang D, Gong Y, Wu G. Prognostic genes of breast cancer identified by gene co-expression network analysis. Front Oncol. 2018 Sep 11;8:374.
341. Peng R, Wang Y, Mao L, Fang F, Guan H. Identification of core genes involved in the metastasis of clear cell renal cell carcinoma. Cancer Manag Res. 2020 Dec 30;12:13437-13449.
342. Tang D, Lotze MT, Kang R, Zeh HJ. Apoptosis promotes early tumorigenesis. Oncogene. 2011 Apr 21;30(16):1851-4.
343. Ghelli Luserna di Rora' A, Iacobucci I, Martinelli G. The cell cycle checkpoint inhibitors in the treatment of leukemias. J Hematol Oncol. 2017 Mar 29;10(1):77.
344. Kozłowski J, Kozłowska A, Kocki J. Breast cancer metastasis - insight into selected molecular mechanisms of the phenomenon. Postepy Hig Med Dosw (Online). 2015 Apr 8;69:447-51.
345. Vieira AF, Paredes J. P-cadherin and the journey to cancer metastasis. Mol Cancer. 2015 Oct 6;14:178.
346. Han YH, Mun JG, Jeon HD, Kee JY, Hong SH. Betulin inhibits lung metastasis by inducing cell cycle arrest, autophagy, and apoptosis of metastatic colorectal cancer cells. Nutrients. 2019 Dec 26;12(1):66.
347. Lindström LS, Yau C, Czene K, Thompson CK, Hoadley KA, Van't Veer LJ, Balassanian R, Bishop JW, Carpenter PM, Chen YY, Datnow B, Hasteh F, Krings G, Lin F, Zhang Y, Nordenskjöld B, Stål O, Benz CC, Fornander T, Borowsky AD, Esserman LJ; STO Trialists Group. Intratumor heterogeneity of the estrogen receptor and the long-term risk of fatal breast cancer. J Natl Cancer Inst. 2018 Jul 1;110(7):726-33.
348. Hong D, Fritz AJ, Zaidi SK, van Wijnen AJ, Nickerson JA, Imbalzano AN, Lian JB, Stein JL, Stein GS. Epithelial-to-mesenchymal transition and cancer stem cells

contribute to breast cancer heterogeneity. J Cell Physiol. 2018 Dec;233(12):9136-9144.
349. García-Fernández A, Barco I, Fraile M, Lain JM, Carmona A, Gonzalez S, Pessarrodona A, Giménez N, García-Font M. Factors predictive of mortality in a cohort of women surgically treated for breast cancer from 1997 to 2014. Int J Gynaecol Obstet. 2016 Aug;134(2):212-6.

Appendix A



05/01/2022, 13:46 Email - Abrar Aljohani - Outlook

Re: request copyright license please

Journal of Breast Cancer <jbc@kbcs.or.kr>
Wed 08/09/2021 02:05
To: Abrar Aljohani <msxaa98@exmail.nottingham.ac.uk>

Dear Author,

The articles published in the Journal of Breast Cancer are all open-access articles and can be freely downloaded and used for non-commercial purposes.

Any use of the Open Access version of this Journal in whole or in part must include the customary bibliographic citation, including author and publisher attribution, date, article title, Journal of Breast Cancer.

Regards,

Seong Ouk Lee
Editorial Assistant
Journal of Breast Cancer

2021년 9월 7일 (화) 오후 9:53, Abrar Aljohani <Abrar.Aljohani@nottingham.ac.uk>님이 작성:

Thank you for your prompt replay.

"If you want to use our material for a non-profit or educational purpose, we usually allow you to do so without charge." dose that mean I do not have to request a copy right license for the reuse?

Best regards,
Abrar

From: Journal of Breast Cancer <jbc@kbcs.or.kr>
Sent: 07 September 2021 07:47
To: Abrar Aljohani <msxaa98@exmail.nottingham.ac.uk>; 문형곤교수님 <moonhgsurgi@gmail.com>
Subject: Re: request copyright license please

Dear Author,

Thank you for your interest in our article.

Journal of Breast Cancer follows Open Access policy.

If you want to use our material for a non-profit or educational purpose, we usually allow you to do so without charge.

Any use of the Open Access version of this article must include the customary bibliographic citation,
including authors and publisher attribution, date, article title, Journal of Breast Cancer, and the URL https://www.ejbc.kr/.

https://outlook.office.com/mail/id/AAQkAGNmZDBjMzVmLTdjY2YtNDJkOC1iZGUzLTcxOWEwMTkwMzJkYgAQAOPV3RdWLdIErm%2FwisK1qb… 1/2

Appendix B

WG: Message From Contact Form: Other

Susanne Seifert <sseifert@ibidi.de>

Mon 25/10/2021 08:30

To: Abrar Aljohani <msxaa98@exmail.nottingham.ac.uk>

Dear Abrar,

Thank you for your message.

Of course, we agree that you may use the image from our website.
Please quote it with the credit: Courtesy of ibidi

If you let me know exactly which image from the website you would like to use, I can send you the high-resolution version.

Best regards,
Susanne

Greeting, I would like to use an image from your website on my PhD thesis, and I would like to get a permission from you to use it. Can you please allow me to use it? Here is the link for the image https://ibidi.com/content/282-creating-the-gap Many thanks, Abrar

Dr. Susanne Seifert | Head of Marketing

E: sseifert@ibidi.de **T:** +49 89 520 46 17-34 **F:** +49 89 520 46 17-59
ibidi GmbH, Lochhamer Schlag 11, 82166 Gräfelfing, Germany

cells in focus ibidi

www.ibidi.com Follow us:



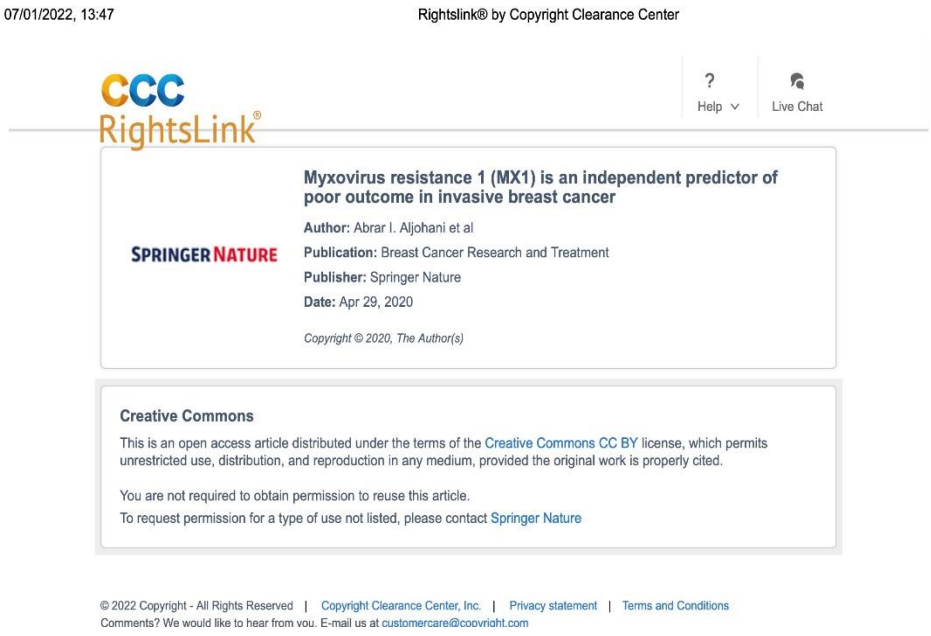
Geschäftsführer: Dr. Valentin Kahl, Dr. Roman Zantl | Registergericht: Amtsgericht München | HRB-Nr.: 135 854

Please refrain from sending Microsoft Office attachments to ibidi. Please share only PDF or Office documents by secured platforms.

GEHEIMHALTUNGSPFLICHT:
Dieses E-Mail und alle damit verbundenen Anlagen sind vertraulich und dürfen nur bestimmten Personen zugänglich gemacht werden. Sofern Sie nicht zu den angegebenen Empfängern gehören, benachrichtigen Sie bitte unverzüglich den Absender. Der Inhalt darf weder an Dritte weitergegeben noch zu anderen Zwecken verwendet werden. Die Informationen dürfen auch nicht auf einem Datenträger gespeichert oder auf einen Datenträger kopiert werden.

CONFIDENTIALITY:
This e-mail and any attachments are confidential and may be privileged. If you are not a named recipient, please notify the sender immediately and do not disclose the contents to another person, use it for any purpose or store or copy the information in any medium.

Appendix C



07/01/2022, 13:47 Rightslink® by Copyright Clearance Center

CCC RightsLink®

Help Live Chat

SPRINGER NATURE

Myxovirus resistance 1 (MX1) is an independent predictor of poor outcome in invasive breast cancer

Author: Abrar I. Aljohani et al

Publication: Breast Cancer Research and Treatment

Publisher: Springer Nature

Date: Apr 29, 2020

Copyright © 2020, The Author(s)

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

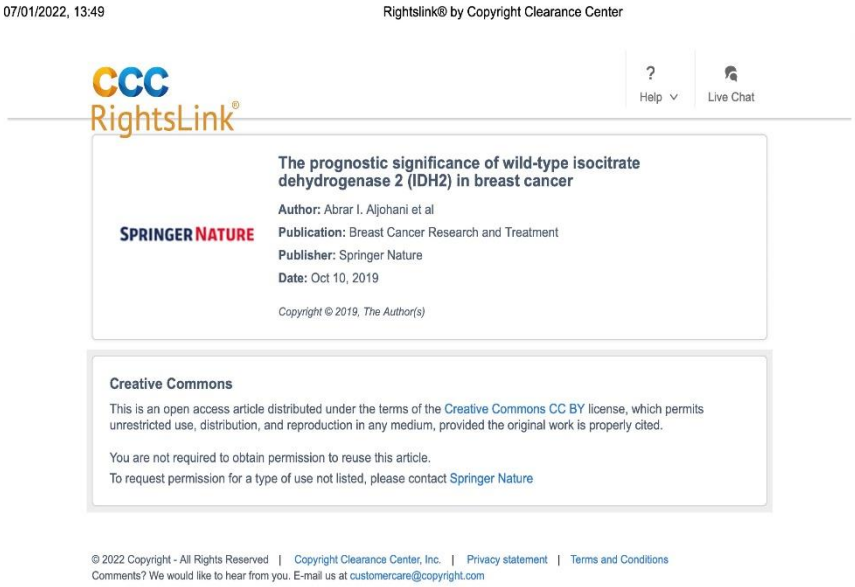
You are not required to obtain permission to reuse this article.

To request permission for a type of use not listed, please contact Springer Nature

© 2022 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions

Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Appendix D



07/01/2022, 13:49 Rightslink® by Copyright Clearance Center

CCC RightsLink®

? Help Live Chat

SPRINGER NATURE

The prognostic significance of wild-type isocitrate dehydrogenase 2 (IDH2) in breast cancer

Author: Abrar I. Aljohani et al

Publication: Breast Cancer Research and Treatment

Publisher: Springer Nature

Date: Oct 10, 2019

Copyright © 2019, The Author(s)

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

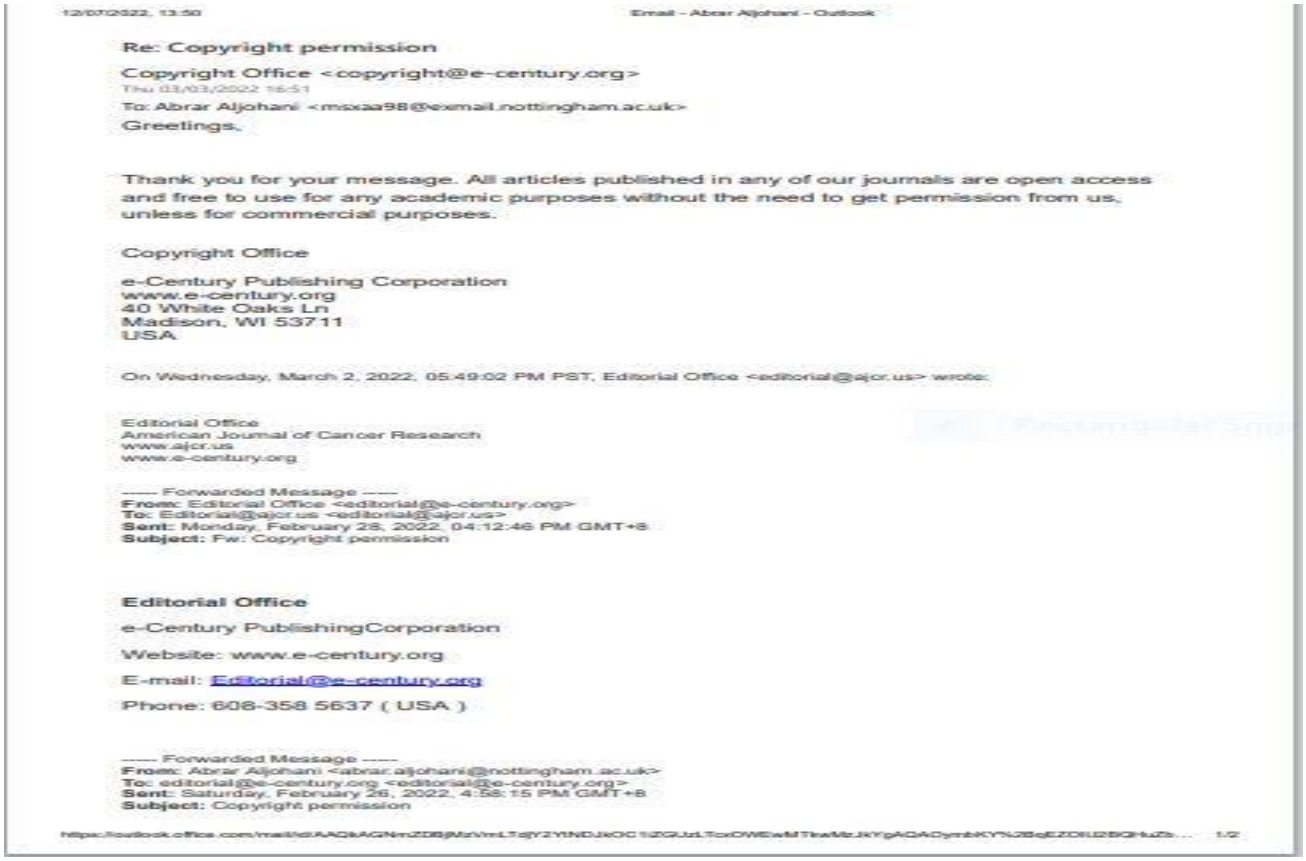
You are not required to obtain permission to reuse this article.

To request permission for a type of use not listed, please contact Springer Nature

© 2022 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions

Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Appendix E



12/07/2022, 13:50 Email - Abrar Aljohani - Outlook

Re: Copyright permission

Copyright Office <copyright@e-century.org>
Thu 03/03/2022 16:51
To: Abrar Aljohani <mxaa98@exmail.nottingham.ac.uk>

Greetings,

Thank you for your message. All articles published in any of our journals are open access and free to use for any academic purposes without the need to get permission from us, unless for commercial purposes.

Copyright Office

e-Century Publishing Corporation
www.e-century.org
40 White Oaks Ln
Madison, WI 53711
USA

On Wednesday, March 2, 2022, 05:49:02 PM PST, Editorial Office <editorial@ajcr.us> wrote:

Editorial Office
American Journal of Cancer Research
www.ajcr.us
www.e-century.org

----- Forwarded Message -----
From: Editorial Office <editorial@e-century.org>
To: Editorial@ajcr.us <editorial@ajcr.us>
Sent: Monday, February 28, 2022, 04:12:46 PM GMT+8
Subject: Fw: Copyright permission

Editorial Office

e-Century PublishingCorporation

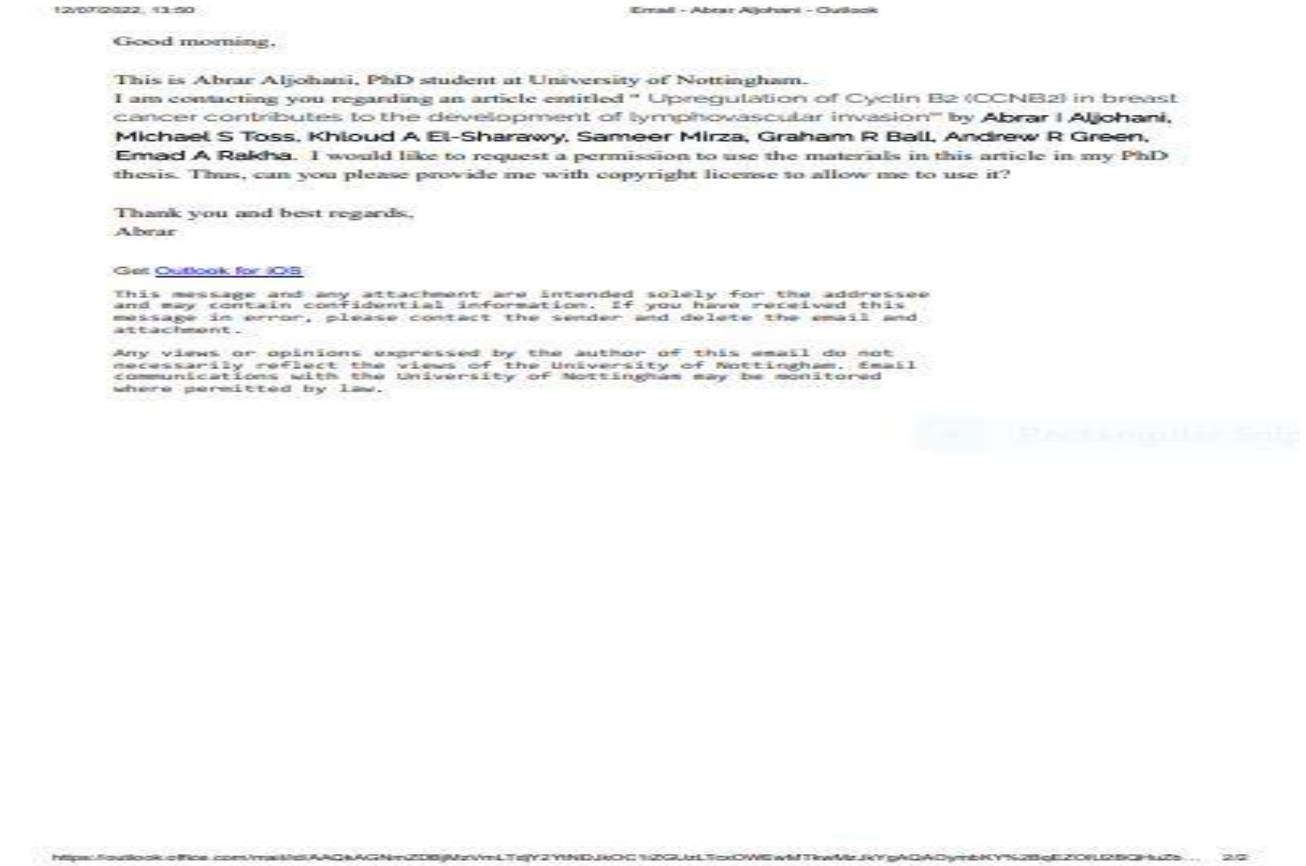
Website: www.e-century.org

E-mail: Editorial@e-century.org

Phone: 608-358 5637 (USA)

----- Forwarded Message -----
From: Abrar Aljohani <abrar.aljohani@nottingham.ac.uk>
To: editorial@e-century.org <editorial@e-century.org>
Sent: Saturday, February 26, 2022, 4:58:15 PM GMT+8
Subject: Copyright permission

https://outlook.office.com/mail/id/AAQkAGNmY2I8ZjMzLWmLTqjY2YtNDJkOC1iZGUzLTcxOWEwMTkwMzJkYgAQAOymbKYNJ0BqEZOEUZBQHuZb... 1/2



12/07/2022, 13:50 Email - Abrar Aljohani - Outlook

Good morning,

This is Abrar Aljohani, PhD student at University of Nottingham.
I am contacting you regarding an article entitled " Upregulation of Cyclin B2 (CCNB2) in breast cancer contributes to the development of lymphovascular invasion" by **Abrar I Aljohani, Michael S Toss, Khloud A El-Sharawy, Sameer Mirza, Graham R Ball, Andrew R Green, Emad A Rakha**. I would like to request a permission to use the materials in this article in my PhD thesis. Thus, can you please provide me with copyright license to allow me to use it?

Thank you and best regards,
Abrar

Get Outlook for iOS

This message and any attachment are intended solely for the addressee and may contain confidential information. If you have received this message in error, please contact the sender and delete the email and attachment.

Any views or opinions expressed by the author of this email do not necessarily reflect the views of the University of Nottingham. Email communications with the University of Nottingham may be monitored where permitted by law.

https://outlook.office.com/mail/id/AAQkAGNmY2I8ZjMzLWmLTqjY2YtNDJkOC1iZGUzLTcxOWEwMTkwMzJkYgAQAOymbKYNJ0BqEZOEUZBQHuZb... 2/2