



**University of
Nottingham**

UK | CHINA | MALAYSIA

**Exploring the Key Molecular drivers of lymphovascular
invasion in Invasive Breast Cancer.**

Sami Abbas M. Alsaeed

B.Sc., M.Sc.

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

January 2023

School of Medicine

ABSTRACT

BACKGROUND: Lymphovascular invasion (LVI) plays a crucial role in breast cancer (BC) metastasis and presents as a reliable indicator of poor outcome in terms of distant metastasis and shorter survival. Therefore, accurate and reproducible assessment of LVI and an understanding of its molecular mechanisms can help guide management decision-making. However, molecular mechanisms of LVI are complex and overlap with other biological processes, such as metastasis and BC progression in general. Therefore, deciphering the molecular mechanisms of LVI and identifying its driver genes remain challenging. This study aimed to utilise several well-defined BC datasets, evaluating whether the LVI mechanisms in BC are controlled by distinct transcriptome and proteomic profiles, and to assess the roles of potential target genes using various molecular techniques.

METHODS: In an attempt to discover new targets associated with LVI status, bioinformatics analysis was applied on multiple large cohorts of patients with BC, including the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort (n = 1,980) and The Cancer Genome Atlas Breast Carcinoma (TCGA-BRCA; n = 854) as discovery datasets and the Breast Cancer Gene-Expression Miner (bc-GenExMiner) dataset (Affymetrix; n = 4,904) cohorts as a validation dataset. The differential mRNA expression was performed to identify key genes with respect to LVI status, which was previously defined based on morphology and immunohistochemistry (IHC) for endothelial markers. In addition, mass spectrometry of a well-defined subset of BC with well-characterised LVI status was utilised to identify target genes at the protein level. The targeted genes that were strongly associated with LVI at the transcriptomic levels (*HSP90AA1*, *SLC9A3R1*, *NOP56* and *HMGB3*) were further investigated to confirm their prognostic value and association with LVI at the protein level utilising IHC and tissue microarrays of a large annotated early stage BC with long-term follow-up Nottingham cohort (n = 2,497). The protein expression of (*HSP90AA1*, *SLC9A3R1* and *NOP56*) was correlated with clinicopathological parameters, a large panel of relevant biomarkers, and patient outcomes to identify clinically reliable LVI-associated genes. Further functional experiments followed this to explore the mechanistic insights of the most likely target gene (*HMGB3*) for LVI in BC to identify novel BC prognostic and predictive value towards personalised therapy in patients with early stage BC.

RESULTS:

Using the METABRIC and TCGA-BRCA datasets for the identification of molecular signatures strongly correlated with identified LVI, a set of 99 genes demonstrated significantly different expression patterns between LVI+ve and LVI-ve cases. Clustering analysis with this gene set further divided the cases into two molecular subtypes (subtypes 1 and 2), which were significantly associated with LVI status and patient outcome. The 10-year overall survival of subtype 2 was significantly worse than that of subtype 1.

Expression of four individual genes strongly correlated with LVI. High expression of *HSP90AA1* at both the mRNA and protein levels was significantly associated with characteristics of poor BC prognosis, including high grade, LVI positivity, poor Nottingham Prognostic Index (NPI) and positive expression of p53 and phosphoinositide-3-kinase, catalytic, alpha polypeptide (PIK3CA). In addition, outcome analysis revealed that high *HSP90AA1* mRNA expression had significantly shorter BC-specific survival (BCSS) in the METABRIC and TCGA-BRCA datasets (hazard ratio (HR) = 1.38; 95% confidence interval [CI] = 1.16–1.65, $p = 0.0004$; and HR = 1.85; 95% CI = 1.318–2.61, $p < 0.0001$, respectively). The HSP90 α protein expression was an independent predictor of shorter BCSS (HR = 1.38; 95% CI = 1.06–1.80, $p = 0.018$) and distant metastasis-free survival (HR = 1.32; 95% CI = 1.03–1.69, $p = 0.027$).

High *SLC9A3R1* mRNA and/or *SLC9A3R1* protein expression was significantly correlated with aggressive tumour behaviour, LVI positivity, large tumour size, luminal B and human epidermal growth factor receptor 2 (HER2) positivity, high Ki67, and the NPI poor prognostic group. Both protein and mRNA expression of *SLC9A3R1* was significantly associated with PIK3CA and phosphorylated AKT. The *SLC9A3R1* mRNA expression was higher in those tumours showing a copy number gain in both METABRIC and TCGA-BRCA ($p < 0.0001$). In the outcome analysis, *SLC9A3R1* mRNA showed poor prognosis in bc-GenExMiner (HR = 1.59, 95% CI = 1.23–2.06, $p = 0.0005$), and high *SLC9A3R1* protein expression was prognostic of shorter survival (HR = 1.52; 95% CI = 1.14–2.04, $p = 0.005$).

High *NOP56* mRNA and/or protein expression was significantly correlated with aggressive tumour behaviour, including high histological grade, larger tumour size, positive nodal status, LVI positivity, hormone receptor negativity, triple-negative BC positivity and NPI poor prognostic group. High *NOP56* protein expression was associated with high Ki67 and significantly associated with PI3KCA and epidermal growth factor. A significant correlation

between NOP56 protein and *NOP56* mRNA was also observed ($p < 0.001$). High *NOP56* mRNA correlated with shorter survival BCSS in the METABRIC cohort ($p < 0.001$, HR = 1.3, 95% CI = 1.1–1.5) and bc-GenExMiner (HR = 1.38, 95% CI = 1.16–1.65, $p = 0.0003$), and NOP56 protein presented a poor prognosis of BCSS ($p = 0.003$, HR = 1.4, 95% CI = 1.1–2.0).

In both the METABRIC and TCGA-BRCA cohorts, high *HMGB3* mRNA expression was significantly associated with the poor BC prognosis characteristics, including LVI positivity, high tumour grade, NPI poor prognostic group, high Ki67, hormone receptor negativity, HER2 positivity and phosphorylated AKT positivity. In addition, high *HMGB3* mRNA expression was positively correlated with high level of PIK3CA. In the survival analysis, high *HMGB3* mRNA was prognostic for shorter BCSS in LVI+ tumours in METABRIC (HR = 1.447; 95% CI = 1.023–2.05, $p = 0.034$) and TCGA-BRCA (HR = 2.826; 95% CI = 1.222–6.539, $p = 0.015$). Consistent with the transcriptomic results, HMGB3 knockdown reduced proliferation activity and led to a significant decrease in tumour cell adhesion, migration and transmigration ability.

CONCLUSIONS: This investigation of independent large BC patient cohorts using several molecular techniques and bioinformatics analyses highlighted potential molecular and biological roles of LVI and their importance in BC progression, revealing the different aspects of BC metastasis and enhancing current knowledge. Moreover, the expression alteration of a group of genes and proteins may be responsible for the beginning or inhibition of LVI in invasive BC. This research showed that the clinicopathological index LVI in BC is linked to a unique transcriptome profile that may have prognostic significance. It identified a unique LVI profile with prognostic utility in BC. The proteins identified participate in various pathways manipulating the pre-LVI biological steps in BC, including adhesion, migration and invasion. In the future, the use of comprehensive and well-designed functional investigations with the help of several models and techniques may help further elucidate the roles played by each candidate target identified as controlling or inhibiting LVI status in patients with BC.

PUBLICATIONS AND SCIENTIFIC CONTRIBUTIONS

- 1- **Sami Alsaeed**, Michael S Toss, Mansour Alsaleem, Mohammed A Aleskandarany, Chitra Joseph , Sasagu Kurozumi , Graham Ball, Nigel P Mongan, Andrew R Green and Emad A Rakha. “Prognostic significance of heat shock protein 90AA1 (HSP90α) in invasive breast cancer”. *Journal of clinical pathology* (2021).
- 2- Sasagu Kurozumi, Chitra Joseph, Sultan Sonbul, **Sami Alsaeed**, Yousif Kariri, Abrar Aljohani, Sara Raafat, Mansour Alsaleem, Angela Ogden, Simon J Johnston, Mohammed A Aleskandarany, Takaaki Fujii, Ken Shirabe, Carlos Caldas, Ibraheem Ashankyty, Leslie Dalton, Ian O Ellis, Christine Desmedt, Andrew R Green, Nigel P Mongan, Emad A Rakha. “A key genomic subtype associated with lymphovascular invasion in invasive breast cancer”. *Breast Cancer Research and Treatment* (2019).
- 3- Sasagu Kurozumi, Chitra Joseph, Sultan Sonbul, Mohammed A Aleskandarany, Marian Pigera, Mansour Alsaleem, **Sami Alsaeed**, Yousif A. Kariri, Christopher C Nolan, Maria Diez-Rodriguez, Simon Johnston, Nigel P Mongan, Takaaki Fujii, Ken Shirabe, Stewart G Martin, Ian O Ellis, Andrew R Green, Emad A Rakha. “Clinicopathological and prognostic significance of Ras association and pleckstrin homology domains 1 (RAPH1) in breast cancer”. *Breast Cancer Research and Treatment* (2018).
- 4- Sasagu Kurozumi, **Sami Alsaeed**, Nnamdi Orah, Islam M Miligy, Chitra Joseph, Abrar Aljohani , Michael S Toss, Takaaki Fujii, Ken Shirabe, Andrew R Green, Mohammed A Aleskandarany, Emad A Rakha. “Clinicopathological significance of lipocalin 2 nuclear expression in invasive breast cancer”. *Breast Cancer Research and Treatment* (2019).
- 5- Sasagu Kurozumi, Chitra Joseph, Sara Raafat, Sultan Sonbul, Yousif A. Kariri, **Sami Alsaeed**, Marian Pigera, Mansour Alsaleem, Christopher C Nolan, Simon J Johnston, Mohammed A Aleskandarany, Angela Ogden, Takaaki Fujii, Ken Shirabe, Stewart G Martin, Ibraheem Alshankyty, Nigel P Mongan, Ian O Ellis, Andrew R Green, Emad A Rakha. “Utility of ankyrin 3 as a prognostic marker in androgen-receptor-positive breast cancer”. *Breast Cancer Research and Treatment* (2019).
- 6- Chitra Joseph, Mansour Alsaleem, Michael S. Toss, Yousif A. Kariri, Maryam Althobiti, **Sami Alsaeed**, Abrar I. Aljohani, Pavan L. Narasimha, Nigel P. Mongan, Andrew R. Green, Emad A. Rakha. “The ITIM-containing Receptor: Leukocyte-associated immunoglobulin-like receptor-1 (LAIR-1) modulates immune response and confers poor prognosis in invasive breast carcinoma”. *Cancers* (2020).

- 7- Yousif A Kariri, Mansour Alsaleem, Chitra Joseph, **Sami Alsaeed**, Abrar Aljohani, Sho Shiino, Omar J. Mohammed, Michael S Toss, Andrew R Green, Emad A Rakha. “The prognostic significance of Interferon Stimulated Gene 15 (ISG15) in invasive breast cancer”. *Breast Cancer Research and Treatment* (2020).
- 8- Yousif A Kariri, Chitra Joseph, Mansour Alsaleem, , Khloud A. Elsharawy, **Sami Alsaeed**, Michael S Toss, Nigel P Mongan, Andrew R Green, Emad A Rakha. “Mechanistic and Clinical Evidence Supports a Key Role for Cell Division Cycle Associated 5 (CDCA5) as an Independent Predictor of Outcome in Invasive Breast Cancer”’ *Cancers* (2022
- 9- **Sami Alsaeed**, Mansour Alsaleem, Andrew R Green and Emad A Rakha. “Clinicopathological characteristic of Solute carrier family 9, subfamily A, member 3 regulator 1 (SLC9A3R1) in breast cancer”’. To be submitted post-viva
- 10- **Sami Alsaeed**, Mansour Alsaleem, Yousif Kariri, Andrew R Green and Emad A Rakha. “Nucleolar protein 5 A (NOL5A /NOP56) is a potential participant in breast cancer’s poor prognosis”’. To be submitted post-viva.

CONFERENCE PRESENTATIONS

1. Poster presentation ‘‘*Clinical and Biological Role of Heat Shock Protein 90 Alpha Family Class A Member 1 (HSP90AA1) in Breast Cancer: A Marker of Lymphovascular Invasion and Poor Prognosis*’’ Cancer Research Nottingham 2018 Annual Symposium 19th September 2018, Medical School, Lecture Theatre 4 (LT4), Nottingham, United Kingdom.
2. Poster presentation ‘‘*HSP90AA1 significance as a target in breast cancer*’’ The 2nd UK Interdisciplinary Breast Cancer Symposium 2020, 27th-28th January 2020, International convention centre (ICC), Birmingham, United Kingdom.
3. Poster presentation (prize winner) ‘‘*Importance of Solute carrier family 9, subfamily A, member 3 regulator 1 (SLC9A3R1) in breast cancer progression*’’ The 13th Joint meeting of British Division of the International Academy of Pathology (BDIAP) and The Pathological Society, 6th-8th July 2021, Manchester Central Convention Centre, Manchester, United Kingdom.

ACKNOWLEDGEMENTS

I am grateful to Allah, the Almighty, who has provided me with the strength, wisdom and patience to complete my research and granted me this opportunity and the determination to see this research through.

Over the last four years, I have been privileged to study in the Department of Pathology, University of Nottingham, under the guidance and supervision of Professor Emad Rakha. First, I would like to offer my utmost respect and gratitude to Professor Rakha for his endless support with my PhD studies and research, for his motivation and immense knowledge of pathology and science, and for providing me with the opportunity to study under his tutelage. His patience and guidance have helped me greatly during my research and thesis writing.

I would also like to thank my supervisors, Professor Ian Ellis and Dr Andrew Green, for their instrumental advice, guidance and scientific contributions throughout this intense but thoroughly rewarding journey. There could not have been a superior combination of mentors and supervisors for me during my PhD studies. My gratitude to them is immense, and I hope to continue collaborating with these outstanding scholars.

My heartfelt appreciation goes again to Dr Andrew Green for his extreme support and enthusiasm and for always being available to offer his advice. His huge support during the writing process is similarly appreciated.

I am deeply indebted and grateful to Dr Mansour Alsaleem for his immeasurable knowledge of bioinformatics analysis. His graceful patience, guidance, instruction and friendship over these four years have encouraged me and been instrumental to this achievement and my progress. His consistent support and guidance during the writing of this thesis are very appreciated.

Special appreciation and thanks must be given to all of my PhD study friends and the members of the Cancer and Stem Cells Division and Histopathology Department for their energy, understanding and help throughout my project, especially Dr Chitra Joseph for their guidance throughout the lab work, Dr Sasagu Kurozomi for helping me in the analysis process and for encouragement, Dr Michael Toss for his scientific help, Dr Sultan Sonbul for sharing some analysis results with me. My time in this lab has been perfect.

Finally, my heartfelt appreciation goes to my sponsors in Saudi Arabia, the Northern Border University and the Saudi Arabian Cultural Bureau in the United Kingdom for their financial support during this period.

DEDICATION

Words cannot express my esteemed admiration and respect for the heroes in my life – my mother, Latifah, and my father, Abbas, who always supported me during my education journey. Mum and Dad, I keep pushing because of you, and you are the reason I can be brave through my struggles, hardships and pain. I love both of you so much!

I would also like to express my most profound appreciation to Nsreen, my wife. I'm deeply indebted to her. Her love, dedication and constant confidence in me have alleviated all of my burdens. Nsreen your presence and consistent encouragement have motivated us during our life abroad and the demanding process of completing my research. While you have struggled, too, it has been a relief to know you were willing to support me in any way you could. Many thanks to you and all of the women who have sacrificed their time to support their partners.

From the bottom of my heart, I would like to thank my beloved daughter, Eshraq, and my lovely son, Abdulaziz; with your presence, my life shines. You are my inspiration; you always give me strength during difficult times.

I also want to extend my sincere thanks to my beloved sisters and lovely brothers, their wonderful children, and my dear uncles Jabber, Yousef and Ahmed, who are supporting and encouraging me; your love and continued support have always propelled me forward.

I would also like to dedicate this to my mother-in-law, Maryam, who left us after struggling with metastatic breast cancer. She inspired me to engage in this project to help many women like her. My Mum, I hope you rest in peace; I love you so much.

To conclude, I cannot forget to thank my family and friends for all of their unconditional support during these very intense academic years.

ETHICAL APPROVAL

The Nottingham Research Ethics Committee 2 and Nottingham Health Science Biobank, reference number 15/NW/0685, approved this study with the title ‘Development of molecular genetic classification of BC’. This study utilised samples from Nottingham that were pseudo-anonymised and gathered before 2006. Thus, there was no requirement to have informed patient consent under the Human Tissue Act. Data release was also pseudo-anonymised in compliance with the Human Tissue Act regulations.

LIST OF ABBREVIATIONS

µl	Microliter
cm	Centimetre
g	Gram
M	Molar
ml	Millilitre
mm	Millimetre
W	Watt
µm	Micrometres

A

APOC1	Apolipoprotein C1
ATCC	American-Type Culture Collection
ATP	Adenosine Triphosphate

B

BC	Breast Cancer
BCA	Bicinchoninic Acid Assay
bc-GenExMiner v4.9	Breast Cancer Gene-Expression Miner v4.9
BCSS	Breast Cancer-Specific Survival
BRCA1	Breast Cancer Type-1 Susceptibility Protein
BRCA2	Breast Cancer Type-2 Susceptibility Protein
BV	Blood Vessels

C

CD31	Cluster of differentiation 31
CIs	Confidence Intervals
CN	Copy Number
CNA	Copy Number Aberrations
ICC	Correlation Coefficient

D

DAB	3,3' Diaminobenzidine
DEGs	Differentially Expressed Genes
DFS	Disease-Free Survival
DGE	Differential Genes Expression
dH ₂ O	Deionised water
DM	Distant Metastasis
DMEM	Dulbecco's Modified Eagle's Medium
DMFS	Distant Metastasis-Free Survival
DPBS	Dulbecco's Phosphate-Buffered Saline
DW	Distilled Water

E

EC	Endometrial Cancer
ECM	Extracellular Matrix
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial-Mesenchymal Transition
ERM	Ezrin, Radixin and Moesin

F

FAK	Focal Adhesion kinase
FBS	Fetal Bovine Serum
FFPE	Formalin-Fixed Paraffin-Embedded

G

GO	Gene Ontology
GPG	Good Prognostic Group
GSEA	Gene Set Enrichment Analysis
H	
H	Hour
H&E	Haematoxylin and Eosin
HCC	Hepatocellular Carcinoma
HER2	Human Epidermal Growth Factor Receptor-2
HIER	Heat-Induced Epitope Retrieval
HMG	High Mobility Group
HR	Hazard Ratio
HRE-	Hormonal Receptor-Negative
HRE+	Hormonal Receptor-Positive
HRT	Hormone Replacement Therapy
HSP90	Heat Shock Protein 90
HSP90αA1	Heat Shock Protein 90 Alpha family class A member 1
HSPs	Heat Shock Proteins
HUVECs	Human Umbilical Vein Endothelial Cells
I	
IBC	Invasive Breast Cancer
ICC	International Convention Centre
IDC	Invasive Ductal Carcinoma
IgG	Immunoglobulin G
IHC	Immunohistochemistry
IMS	Industrial Methylated Spirit
IntClusters	Integrative Clusters
K	
KD	Knockdown
KRAS	Kirsten Rat Sarcoma viral oncogene homolog
KSA	Kingdom of Saudi Arabia
L	
LECs	Lymphatic Endothelial Cells
LM	Lymph Node Metastasis
LN	Lymph Node
LVD	Lymphatic Microvascular Density
LVI	Lymphovascular Invasion
LVs	Lymphatic Vessels
M	
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MMPs	Matrix Metalloproteinases
MPG	Moderate Prognostic Group
mTOR	Mammalian target of rapamycin
MW	Molecular Weight
N	
NGS	Next-Generation Sequencing
NOL5A	Nucleolar Protein 5 A
NPI	Nottingham Prognostic Index
NPS	Nottingham Primary Series
O	
ER	Oestrogen Receptor

P

pAKT	Posphorylated AKT
PAM50	Prediction Analysis of Microarrays 50
PBS	Phosphate-Buffer Saline
PDPN	Podoplanin
PECAM-1	Platelet Endothelial Cell Adhesion Molecule-1
PH	Potential of Hydrogen
PIK3CA	Phosphoinositide-3-kinase, Catalytic, Alpha polypeptide
PMCA2	Plasma Membrane Calcium ATPase 2
PPG	Poor Prognostic Group
PPI	Protein-Protein Interactions
PR	Progesterone Receptor
PTEN	Phosphatase and Tensin homolog

R

RB	Running Buffer
RISC	RNA-Induced Silencing Complex
RNA-seq	RNA sequencing
RPMI	Roswell Park Memorial Institute Medium
RT-qPCR	Reverse Transcription-Polymerase Chain Reaction

S

SDS	Sodium Dodecyl Sulfate
SEM	Stander Error of the Mean
siRNA	Small double-stranded Interfering RNA
SLC9A3R1	Solute carrier family 9, subfamily A, member 3 regulator 1
SLN	Sentinel Lymph Nodes
snoRNPs	Small Nucleolar Ribonucleoproteins
SPSS	Statistical Package for Social Sciences

T

TB	Transfer Buffer
TBS	Tris Buffered Saline
TBS-T	Tris Buffered Saline Tween
TCGA-BRCA	The Cancer Genome Atlas Breast Carcinoma
TDLUs	Terminal Duct-Lobular Units
TMA	Tissue Microarray
TME	Tumour Microenvironment
TNBC	Triple-Negative Breast Cancer
TNFSF11	Tumor Necrosis Factor ligand Superfamily member 11
TNM	Tumour, Normal, and Metastatic Tissues

U

UALCAN	University of Alabama at Birmingham Cancer data analysis Portal
UK	United Kingdom
uPA	Urokinase Plasminogen Activator

V

VE-cadherin	Vascular Endothelial Cadherin
VEGFR3	Vascular Endothelial Growth Factor Receptor 3
VI	Vascular Invasion

W

WAD	Weighted Average Difference
WB	Western Blot

TABLE OF CONTENTS

ABSTRACT.....	i
PUBLICATIONS AND SCIENTIFIC CONTRIBUTIONS.....	iv
CONFERENCE PRESENTATIONS.....	vi
ACKNOWLEDGEMENTS.....	vii
DEDICATION.....	ix
ETHICAL APPROVAL.....	x
LIST OF ABBREVIATIONS.....	xi
TABLE OF CONTENTS.....	xiv
LIST OF TABLES.....	xix
LIST OF FIGURES.....	xxi
1 Chapter 1: Introduction.....	1
1.1 General Background.....	2
1.1.1 Cancer	2
1.2 Overview of Breast Cancer	4
1.2.1 Breast Cancer Epidemiology (Incidence and Mortality)	4
1.2.2 Breast Cancer Risk Factors	5
1.2.3 Breast Cancer Prognostic and Predictive Factors	8
1.3 The Molecular Pathways that Interact with Lymphovascular Invasion in Invasive Breast Cancer	14
1.3.1 Cancer Metastasis Cascade (Invasiveness, Tumour Microenvironment and Lymphovascular Invasion)	14
1.3.2 An overview of Lymphovascular Invasion.....	18
1.3.3 Lymphovascular Invasion Assessment	19
1.3.4 Lymphatic Vessels Structure, Function and LVI-related Markers	20
1.3.5 Conclusion	23
1.4 Hypothesis and Aims of the Study.....	24
1.4.1 Hypothesis.....	24
1.4.2 Aims.....	24
2 Chapter 2: General materials and methods.....	25
2.1 Study Cohorts	26
2.1.1 Transcriptomic Level	26
2.1.2 Proteomic Level	29
2.2 Optimisation of Antibodies Used in the Study.....	31
2.2.1 Cell Lines and Culture Media	31

2.2.2	Cell Passage	32
2.2.3	Preparation of Cell Lysates	32
2.2.4	Cryopreservation of Cell Lines	33
2.3	Antibody Specificity and Cross-Reactivity Assay: Western Blot.....	33
2.3.1	Protein Quantifications	34
2.3.2	Sample Preparation	34
2.3.3	Sample Loading and Protein Separation (Electrophoresis)	35
2.3.4	Protein Transferring	35
2.3.5	Membrane Blocking.....	35
2.3.6	Immunoblotting and Protein Reaction Detection	36
2.3.7	Protein Reaction Visualisation (the Bands)	36
2.4	Antibody Optimisation Assay: Immunohistochemistry	36
2.4.1	Staining Method.....	37
2.5	Immunoreactivity Evaluation before Scoring	40
2.6	Slide Imaging	41
2.7	Scoring of Tissue Microarray Cores	41
2.8	In-House Biomarker Consortium Data	41
2.9	Cut-Off Point Determination.....	43
2.10	Differential Gene Expression Analysis and Gene Selection for Study	43
2.10.1	Differential Gene Expression.....	43
2.10.2	Gene Selection for Study	44
2.11	Statistical Analysis	45
2.11.1	Introduction.....	45
2.11.2	Intra-Class Correlation Coefficient.....	45
2.11.3	Chi-Squared Test (Goodness of Fit)	45
2.11.4	Kaplan–Meier Survival Analysis and Log-Rank Test	46
2.11.5	Proportional Hazards Model (Cox Regression Model)	46
2.11.6	Student’s T-Test.....	46
2.11.7	Mann–Whitney U Test (Median Analysis).....	46
2.11.8	Spearman’s Rank Test (Model/Line of Best Fit).....	46
3	Chapter 3: Bioinformatics approach and differential gene expression.....	47
3.1	Overview	48
3.2	Aims and Hypothesis	49
3.3	Materials and Methods	49
3.3.1	Study Cohorts.....	49

3.3.2	Bioinformatics Analysis.....	51
3.3.3	Statistical Analysis.....	52
3.4	Results.....	52
3.4.1	Prognostic and Clinicopathological Significance of Lymphovascular Invasion Status	52
3.5	Discussion	62
4	Chapter 4: HSP90 α , a gene associated with LVI status selected from the DGE analysis and previous mass spectrometry analysis	66
4.1	Overview	67
4.2	Aim and Hypothesis	68
4.3	Materials and Methods	69
4.3.1	Study Cohort	69
4.3.2	Western Blotting and Immunohistochemistry	69
4.3.3	Statistical Analysis.....	71
4.4	Results.....	71
4.4.1	Western Blot Evaluation	71
4.4.2	Clinical Significance of <i>HSP90AA1</i> mRNA Expression.....	72
4.4.3	Outcome Analysis for <i>HSP90AA1</i> mRNA Expression.....	74
4.4.4	HSP90 α Protein Expression.....	76
4.4.5	HSP90 α Protein Expression and its Association with the Clinicopathological Parameters	77
4.4.6	Outcome Analysis for HSP90 α Protein Expression	79
4.5	Discussion	81
5	Chapter 5: SLC9A3R1, a biomarker that was chosen from the LVI-related gene signature set	84
5.1	Overview	85
5.2	Aim and Hypothesis	87
5.3	Materials and Methods	87
5.3.1	Study Cohorts.....	87
5.3.2	Western Blotting and Immunohistochemistry	87
5.3.3	Statistical Analysis.....	88
5.4	Results.....	88
5.4.1	Western Blot Evaluation	88
5.4.2	<i>SLC9A3R1</i> mRNA Expression and Clinicopathological Significance	89
5.4.3	Outcome Analysis for <i>SLC9A3R1</i> mRNA Expression	93
5.4.4	SLC9A3R1 Protein Expression	95

5.4.5	SLC9A3R1 Protein Expression and its Association with the Clinicopathological Parameters	95
5.4.6	Outcome Analysis for SLC9A3R1 Protein Expression	97
5.5	Discussion	98
6	Chapter 6: NOP56, a biomarker that was chosen from the LVI-related gene signature set	102
6.1	Overview	103
6.2	Aim and Hypothesis	105
6.3	Materials and Methods	105
6.3.1	Study Cohorts.....	105
6.3.2	Western Blotting and Immunohistochemistry	106
6.3.3	Statistical Analysis.....	106
6.4	Results	107
6.4.1	Western Blot Evaluation	107
6.4.2	<i>NOP56</i> Transcriptomic Expression and Clinicopathological Significance	107
6.4.3	Outcome Analysis for <i>NOP56</i> mRNA Expression	111
6.4.4	<i>NOP56</i> Protein Expression	113
6.4.5	<i>NOP56</i> Protein Expression and its Association with the Clinicopathological Parameters	113
6.4.6	Outcome Analysis for <i>NOP56</i> Protein Expression.....	115
6.5	Discussion	117
7	Chapter 7: Mechanistic insights of HMGB3, a biomarker that was chosen from the LVI-related gene signature set	120
7.1	Overview	121
7.2	Aim and Hypothesis	122
7.3	Materials and Methods	122
7.3.1	Selecting and Handling Breast Cancer Tumour Cells	122
7.3.2	Endothelial Cells.....	123
7.3.3	Silencing <i>HMGB3</i> using Transient Small Double-Stranded Interfering RNA	123
7.3.4	The Procedure of siRNA Transient Transfection	124
7.3.5	Cell Lysis for siRNA Samples.....	124
7.3.6	Assays of in vitro Studies (Preclinical).....	124
7.3.7	Clinical Methodology	128
7.3.8	Statistical Analysis.....	129
7.4	Results	129
7.4.1	Pre-Clinical Results	129

7.4.2	Clinical Results	136
7.5	Discussion	141
8	Chapter 8: General discussion	144
8.1	Overview	145
8.2	Hypothesis, Aim and Methodology	146
8.3	Chapter 3: Bioinformatics approach and differential gene expression	146
8.4	Chapter 4: HSP90 α , a gene associated with LVI status selected from the DGE analysis and previous mass spectrometry analysis.....	147
8.5	Chapter 5: SLC9A3R1, a biomarker that was chosen from the LVI-related gene signature set.....	147
8.6	Chapter 6: NOP56, a biomarker that was chosen from the LVI-related gene signature set.....	148
8.7	Chapter 7: Mechanistic insights of HMGB3, a biomarker that was chosen from the LVI-related gene signature set	149
8.8	Potential Strengths.....	150
8.9	Potential Limitations	151
8.10	Future Work	152
8.11	Final Conclusions.....	152
9	References	154
10	Appendices.....	182
10.1	Copyright approval for re-use Figure 1.1.....	182
10.2	Copyright approval for re-use “A key genomic subtype with lymphovascular invasion in invasive breast cancer” (223).	182
10.3	Copyright approval for the re-use “Prognostic significance of heat shock protein 90AA1 (HSP90 α in invasive breast cancer” (273).	183
10.4	List of top 350 genes significantly associated with lymphovascular invasion in the Nottingham cohort.....	184
10.5	List of top 350 genes significantly associated with lymphovascular invasion in the remaining METABRIC cases.	186
10.6	Genes full names list of the 99 genes significantly associated with lymphovascular invasion.	188
10.7	Mean value, SEM, subtraction and WAD ranking in the 99 genes significantly associated with lymphovascular invasion.	190

LIST OF TABLES

Table 1.1: Tumour grading system.	9
Table 1.2: NPI subgroups	10
Table 1.3: IntClusts association	12
Table 1.4: Molecular features for the integrative clusters IntCusts	12
Table 2.1: Characteristic of the METABRIC cohort (n = 1980).	27
Table 2.2: Characteristic of the TCGA-BRCA cohort (n = 854).	28
Table 2.3: Characteristic of the Nottingham IBC primary series (n = 2497).	30
Table 2.4: Cell lines used in this study.	31
Table 2.5: Western blotting details of antibodies used in this study	33
Table 2.6: Sample components	34
Table 2.7: Immunohistochemistry evaluation of the antibodies used in the study	37
Table 2.8: The contents of Novolink™ max polymer detection system kit (236).	39
Table 2.9: Available biological biomarkers from the Nottingham cohort that were used in this study.	42
Table 2.10: Antibodies Cut-off point and localisation.	43
Table 3.1: The association between lymphovascular invasion and other clinicopathological factors.	54
Table 3.2: Univariate and multivariate Cox regression analysis of lymphovascular invasion status.	55
Table 3.3: List of 99 genes significantly associated with lymphovascular invasion status.	56
Table 3.4: Gene ontology pathways significantly associated with 99 genes related to lymphovascular invasion.	57
Table 3.5: Clinicopathological and prognostic significance of genomic subtypes related to lymphovascular invasion.	60
Table 4.1: Association between <i>HSP90A41</i> mRNA expression and clinicopathological factors in the METABRIC (n = 1980) and TCGA-BRCA (n = 1097) invasive breast cancer datasets.	73
Table 4.2: Association between <i>HSP90A41</i> mRNA expression and other gene of interest in the METABRIC and TCGA-BRCA invasive breast cancer datasets.	74
Table 4.3: Associations between HSP90α protein expression and the clinicopathological factors of the Nottingham breast cancer cohort.	78

Table 4.4:Multivariate Cox regression analysis for predictors of BCSS and DMFS based on clinicopathological factors and protein expression of HSP90 α in the Nottingham breast cancer cohort.	81
Table 5.1: The <i>SLCA93R1</i> copy number alteration in the METABRIC cohort.	90
Table 5.2: The <i>SLCA93R1</i> copy number alteration in the TCGA-BRCA cohort.....	90
Table 5.3: Association of <i>SLC9A3R1</i> mRNA expression with clinicopathological characteristics in the METABRIC (n = 1980) and TCGA-BRCA (n = 854) cohorts.	91
Table 5.4: Association between <i>SLC9A3R1</i> mRNA expression and other gene of interest in the METABRIC and TCGA-BRCA invasive breast cancer datasets.	92
Table 5.5: Relationship of <i>SLC9A3R1</i> protein expression with clinicopathological characteristics in the Nottingham cohort.	96
Table 5.6: Multivariate Cox regression analysis of clinicopathological factors and protein expression of <i>SLC9A3R1</i> to predict breast cancer specific survival.....	98
Table 6.1: Relationship of <i>NOP56</i> mRNA expression with clinicopathological characteristics in the METABRIC (n = 1980) and TCGA-BRCA (n = 854) cohorts.	109
Table 6.2: Association between <i>NOP56</i> mRNA expression and other gene of interest in the METABRIC and TCGA-BRCA invasive breast cancer datasets.	110
Table 6.3: Relationship of <i>NOP56</i> protein expression with clinicopathological characteristics in the Nottingham cohort.	114
Table 6.4: Multivariate Cox regression analysis of clinicopathological factors and protein expression of <i>NOP56</i> to predict breast cancer specific survival.....	117
Table 7.1: Association of <i>HMGB3</i> mRNA expression with clinicopathological characteristics in the METABRIC (n = 1980) and TCGA-BRCA (n = 854) cohorts.	137
Table 7.2: Association between <i>HMGB3</i> mRNA expression and other genes of interest in the METABRIC and TCGA-BRCA invasive breast cancer datasets.	138

LIST OF FIGURES

Figure 1.1: Summary of the eight hallmarks capabilities and two enabling characteristics of cancer. from Hanahan and Weinberg (2022) (7). Copyright of re-use this figure in Appendix 10.1.....	4
Figure 1.2: Breast cancer classification according to a combination of Immunohistochemistry and microarray expression methods. Details in table adapted from (84).	12
Figure 1.3: Stages of tumour metastasis; the diagram shows the sequential steps utilise by tumour cells to metastasise from the primary site to the new organ.....	14
Figure 3.1: Lymphovascular invasion status-based stratification of the breast cancer patients' cumulative survival 10-year overall survival. In the METABRIC cases, the LVI+ group was noticeably poorer survival than the LVI– group (a); significant survival disparities between the LVI– and positive were observed in the TCGA-BRCA cases (b).	55
Figure 3.2: Gene ontology pathways significantly associated with genes related to lymphovascular invasion status, including ‘GO: 0031012 Extracellular matrix,’ ‘GO: 0005615 Extracellular space’ and ‘GO: 0072562 Blood micro-particle.	57
Figure 3.3: Cluster analysis and heat maps of the LVI-related genes group. Using the METABRIC cohort, a dendrogram of 99 LVI-associated genes was created, in which the arrangement of the branches shows how each gene is related to the others (a). Heat maps in line with LVI-related genes set for the METABRIC (b) and TCGA-BRCA (c) cohorts demonstrated that every case was categorically split into subtypes 1 and 2. Red colour indicates a high association, while green indicates a low association.	58
Figure 3.4: Dendrogram of the two subtypes 1 and 2 cases included in the METABRIC dataset.	59
Figure 3.5: Dendrogram and heat maps of the two subtypes 1 and 2 cases included in the TCGA-BRCA dataset. Red colour indicates a high association, while green indicates a low association.....	59
Figure 3.6: Lymphovascular invasion status-based stratification of the breast cancer patients' cumulative survival 10-year overall survival; In the METABRIC cases, the LVI+ group was noticeably poorer survival than the LVI– group (c); significant survival disparities between the LVI– and positive were observed in the TCGA-BRCA cases (d).	61
Figure 3.7: Survival analyses based on clinicopathological traits, including the LVI-related genetic subtype; multivariate survival analyses' hazard ratios and 95% CI are presented as forest plots in the METABRIC cohort (a) and TCGA-BRCA cohort (b).....	62

Figure 4.1: Western blotting validation showed mouse monoclonal anti-HSP90 α antibody ab13492; specificity with a single band observed at approximately 90 kDa in Hela, MCF7, SKBr3 and MDA-MB-231.	71
Figure 4.2: Cumulative survival of BC patients stratified by <i>HSP90AA1</i> mRNA expression. Breast Cancer Specific Survival in the METABRIC cases was significantly worse in the high <i>HSP90AA1</i> group than low <i>HSP90AA1</i> expression (a) in the TCGA cases, significant differences were noted in patient survival in the <i>HSP90AA1</i> high and low groups (b). Kaplan-Meier survival plots of <i>HSP90AA1</i> mRNA expression revealed that the high expression associated with poor patient outcome (c). BCOS, breast cancer overall survival; BCSS, breast cancer-specific survival.	75
Figure 4.3: HSP90 α immunohistochemical protein expression in tissue microarray cores of normal and breast cancer tissue. Invasive cancer cells (strong staining in the cytoplasm of cancer cells), (a) normal mammary gland (uniformly weak staining), (b) no staining (c), (images are at 20 $\times$ magnification).	76
Figure 4.4: Cumulative survival of BC patients stratified by HSP90 α protein expression. BCSS (a) and DMFS (b) plots of cytoplasmic HSP90 α revealed that high expression was associated with poor patient outcome. Breast Cancer Specific Survival (c) and distant metastasis-free survival (d) plots of cytoplasmic HSP90 α revealed that high expression was associated with poor patient outcome in Luminal B HER2/negative subtype.	80
Figure 5.1: Western blotting validation indicated SLC9A3R1 primary antibody specificity based on a single band observed at approximately 48 kDa.	89
Figure 5.2: The <i>SLC9A3R1</i> mRNA gene expression in breast cancer tissue based on sample type in RNA-Seq Data. (<i>p</i> -value is a comparison for multiple independent samples).	90
Figure 5.3: The <i>SLC9A3R1</i> mRNA gene expression on PAM50 subtypes in TCGA-BRCA data using bc-GenExMiner v4.9.	92
Figure 5.4: The Kaplan-Meier survival plots showing cumulative survival of patients with breast cancer stratified by <i>SLC9A3R1</i> mRNA expression. METABRIC (a), TCGA-BRCA (b) Breast Cancer Gene-Expression Miner (Affymetrix) (c), LVI- in METABRIC (d), LVI- in TCGA-BRCA (e), LVI+ in METABRIC (f) and LVI+ in TCGA-BRCA (g). OS, overall survival; BCSS, breast cancer-specific survival.	94
Figure 5.5: Immunohistochemically protein expression for SLC9A3R1 in tissue microarray cores of breast cancer tissue. Invasive cancer cells (intense staining in the cytoplasm of cancer cells) (a); and no staining (b) (images are at 20 $\times$ magnification).	95

Figure 5.6: Kaplan-Meier survival plots showing breast cancer specific survival of BC patients stratified by SLC9A3R1 expression in Nottingham cohort. SLC9A3R1 protein expression revealed that high expression was associated with poor patient outcome. Breast cancer specific survival in the Nottingham cohort (a), in the LVI– cases (b) and the LVI+ cases (c).....	97
Figure 6.1: The NOP56 primary antibody western blotting validation indicated specificity based on a single band observed at approximately 68 kDa.	107
Figure 6.2: The <i>NOP56</i> mRNA gene expression in breast cancer tissue based on sample type in RNA-Seq Data. (<i>p</i> -value is a comparison for multiple independent samples).....	108
Figure 6.3: The <i>NOP56</i> mRNA expression in breast cancer tissue based on breast cancer subclasses.	110
Figure 6.4: The Kaplan-Meier survival plots showing cumulative survival of patients with breast cancer stratified by <i>NOP56</i> mRNA expression. METABRIC (a), TCGA-BRCA (b) Breast Cancer Gene-Expression Miner (Affymetrix) (c), LVI– in METABRIC (d), LVI– in TCGA-BRCA (e), LVI+ in METABRIC (f) and LVI+ in T,CGA-BRCA (g). OS, overall survival; BCSS, breast cancer-specific survival.	112
Figure 6.5 Immunohistochemically protein expression for NOP56 in tissue microarray cores of breast cancer tissue. Invasive cancer cells (intense staining in the cytoplasm of cancer cells) (a); and no staining (b) (images are at 20× magnification).....	113
Figure 6.6: Kaplan-Meier survival plots showing breast cancer specific survival of BC patients stratified by NOP56 expression in Nottingham cohort. NOP56 protein expression revealed that high expression was associated with poor patient outcome. Breast cancer specific survival in the Nottingham cohort (a), in the LVI– cases (b) and the LVI+ cases(c).....	116
Figure 7.1: Western blotting validation indicated HMGB3 primary antibody specificity based on a single band observed at approximately 25 kDa.	130
Figure 7.2: The HMGB3 protein knockdown utilising various siRNA IDs in the BT-474 breast cancer cell line. Western blotting showed <i>HMGB3</i> knockdown utilising siRNA (IDs: 115646, 145247 and s66520) and negative control scrambled cells.	131
Figure 7.3: The HMGB3 protein expression assessed in the triple negative breast cancer (MDA-MB-231) cell line. The ability of the tumour cells to form colonies was reduced after <i>HMGB3</i> siRNA transfection. The <i>p</i> -value was interpreted as **** <i>p</i> < 0.0001.....	132
Figure 7.4: The HMGB3 protein expression in the triple negative breast cancer (MDA-MB-231) and luminal B–HER2 positive (BT-474) cell lines. Evaluation of the HMGB3 expression cell lines after transfection with <i>HMGB3</i> siRNA and negative control scrambled siRNA (a); the level of protein expression of HMGB3 in the KD cell compared to the negative control	

quantified by densitometry and normalised to β -actin levels showed almost complete loss of HMGB3 protein expression (b); the effect of KD *HMGB3* using siRNA in the proliferation ability of the cell line compared to the non-transfected negative control (c); the migration (wound closure) assay images showed the effect of *HMGB3* KD in MDA-MB-231 (d) and BT-474 (e) cell line migration; the migration effect in T0 and T48 in both the KD and negative control (f). The standard error of the mean was calculated in all experiments. The *p*-values were interpreted as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ and $****p < 0.0001$ 133

Figure 7.5: The photomicrograph of the triple negative breast cancer (MDA-MB-231) and luminal B–HER2 positive (BT-474) cell line adhesion with vascular and lymphatic endothelial cells (HUVECs and DLECs). The number of adhering cancer cells to the endothelial cells decreased after *HMGB3* knockdown in MDA-MB-231 (a) and BT-474 (b) BC with both the HUVECs and DLECs. The standard error of the mean was implicated in the experiments. The *p*-values were interpreted as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ and $****p < 0.0001$ 134

Figure 7.6: The photomicrograph of the triple negative breast cancer (MDA-MB-231) and luminal B–HER2 positive (BT-474) cell lines' transmigration across vascular and lymphatic endothelial cells (HUVECs and DLECs). The number of transmigrated cancer cells across the endothelial cells decreased after *HMGB3* knockdown in MDA-MB-231 (a) and BT-474 (b) BC with both the HUVECs and DLECs. The standard error of the mean was implicated in the experiments. The *p*-values were interpreted as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ and $****p < 0.0001$ 135

Figure 7.7: The *HMGB3* mRNA gene expression in breast cancer tissue based on sample type in RNA-Seq Data. (*p*-value is a comparison for multiple independent samples). 136

Figure 7.8: The *HMGB3* mRNA gene expression on PAM50 subtypes using bc-GenExMiner v4.9.....138

Figure 7.9: The Kaplan–Meier survival plots showing cumulative survival of patients with breast cancer stratified by *HMGB3* mRNA expression: METABRIC (a), TCGA-BRCA (b), Breast Cancer Gene-Expression Miner (Affymetrix) (c), LVI– in METABRIC (d), LVI– in TCGA-BRCA (e), LVI+ in METABRIC (f) and LVI+ in TCGA-BRCA (g). OS, overall survival; BCSS, breast cancer-specific survival. 140

1 Chapter 1: Introduction

1.1 General Background

1.1.1 Cancer

Cancer is a complex disease resulting from abnormal genetic and epigenetic alterations involved in how oncogenic cells behave, grow and differentiate (1). Worldwide, cancer is one of the leading causes of death, with 18.1 million new cases of cancer and 9.5 million cancer-related deaths globally in 2018 according to the International Agency for Research on Cancer (2). Medical research and advancement in cancer treatment have made considerable progress. However, by 2040, 29.5 million cancer cases are expected to be diagnosed yearly, and cancer-related deaths will take 16.4 million lives (2). Throughout carcinogenesis, mutated genes allow cells to develop traits to avoid cellular mechanisms that normally stop cell division and apoptosis. Therefore, these altered cells may gain uncontrolled cell division properties, migrating to distant sites of the body's tissues and organs. In addition, several molecular mechanisms must alter or stop for a tumour to develop, such as DNA repair processes and cellular systems that govern cellular division, differentiation, apoptosis and the cell cycle. Although cellular alteration occurs before cancer development, the specific nature and order of mechanisms that must occur before a cell turns into a malignant cell are yet to be understood (3-5).

1.1.1.1 *Cancer Hallmarks*

Carcinogenesis is the abnormal proliferation of cells in the body. Malignant tumours exhibit several distinct cellular characteristics (4). Hanahan and Weinberg's seminal framework defined the cellular traits and characteristics that all carcinogenic cells acquire (4). These include the ability to replicate limitlessly, resist apoptosis, sustain angiogenesis, create independent internal growth signals, remain insensitive to growth-inhibiting signals, and invade the proximal tissues and metastasise to other parts of the body (4). In 2011, Hanahan and Weinberg modified their framework to incorporate the ability to elude the immune system, thus evading discovery and death, as well as the cells' ability to self-modulate their metabolic activities (1). Moreover, in their 2011 review, Hanahan and Weinberg identified three tumourigenesis-supporting characteristics common in cancer that allow a tumour to develop the characteristics necessary for carcinogenesis (1). The characteristics include genomic instability and DNA mutation aggregation in the microenvironment around the tumour, tumour inflammation, and a high rates of endothelial cells, immune cells, pericytes, and stem cells. In addition, genomic and genetic functional techniques have helped identify genes that either guard against cancer (tumour suppressor), for example, *LKB*, *NF2*, *CDH1*, *TP53* and *RB1*, or boost cancer susceptibility (oncogenes), for example, *RAS*, *MYC* and *RAF* (6). During the cell cycle, carcinogenesis is impacted by oncogenes and suppressors because it affects the cell cycle process through either acquiring or losing functionality. Hanahan and Weinberg published a recent update in 2022 with additional core hallmarks of cancer that were previously considered 'emerging hallmarks', including deregulating cellular metabolism and avoiding immune destruction (7). Given that the mechanism of cancer pathogenesis is complex, Hanahan and Weinberg noted that the hallmarks, on their own, cannot decipher cancer pathogenesis at the level of molecular and cellular mechanisms that would facilitate tumour progression and malignancy through enabling preneoplastic cells to gain an aberrant phenotype. Therefore, Hanahan and Weinberg suggested a new concept, which they elaborated as enabling characteristics', resulting from the aberrant condition of neoplasia that provides a communication key between tumour cells and functional traits. In other words, the enabling characteristics affect molecular and cellular mechanisms by which hallmarks are acquired more than the aforementioned eight capabilities, as genomic instability and tumour-promoting inflammation are two examples of these (7) (Figure 1.1).

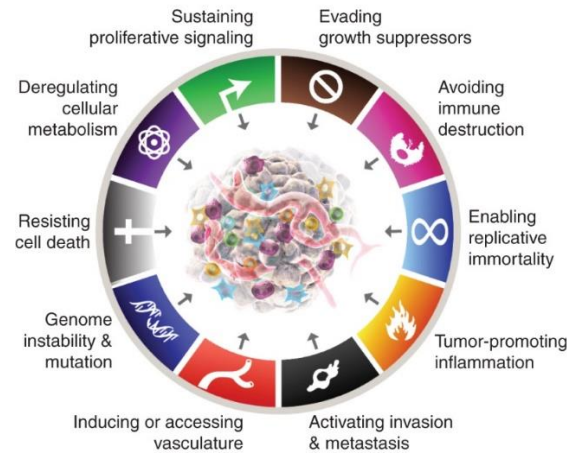


Figure 1.1: Summary of the eight hallmarks capabilities and two enabling characteristics of cancer. from Hanahan and Weinberg (2022) (7). Copyright of re-use this figure in Appendix 10.1.

1.2 Overview of Breast Cancer

1.2.1 Breast Cancer Epidemiology (Incidence and Mortality)

Breast cancer (BC) is a heterogeneous disease with varying clinical outcomes, histological presentations, and molecular and genetic profiles (8-10). Improved detection, screening strategies and personalised treatment regimens have improved the survival rate of patients with BC in England and Wales for over 20 years (11). However, BC is still the leading cause of cancer-related death in women all over the world. In 2018, the *World Cancer Report* classified BC as the most commonly diagnosed cancer type, with about 2.1 million new cases (~1 in 4) of women with BC causing 627,000 deaths worldwide(2). The incidence of BC in most countries is about 24.2% of registered cancer cases, and BC is the first cause of death in women in over 100 countries (2). In 2014, the *World Cancer Report* showed a drastic increase in BC among females worldwide, reaching a rate of 43.3 per 100,000. Comparing the number of women affected by BC globally in 2008 (1.4 million) with that of 2012 (1.67 million), indicates that the number has substantially increased (12). Breast cancer incidence also has overtaken other types of cancer as the most frequent malignant carcinoma, with 2.26 million new cases expected in 2020 globally (13). According to the World Health Organization, in 2020, in the United Kingdom (UK), BC was the second most diagnosed cancer, with 53,889 new cases each year, and the fourth leading cause of death, at 11,839 patients (14). Also, BC is still the most diagnosed cancer, with 3,954 new cases each year, and the second leading cause of death among all population in the Kingdom of Saudi Arabia (KSA) at 1,095 deaths. Thus, BC is regarded as having the highest rate of diagnosis and death in both the UK and KSA (14). This comes with a prediction that by 2025, the number of globally affected females will be in ~20

million patients, with most of them being in low-income countries (15). There are several ways BC could potentially be halted and the number of new cases controlled, such as via using early detection tools that could enhance the treatment options and educational campaigns that could raise awareness of how to avoid risk factors related to BC.

1.2.2 Breast Cancer Risk Factors

Genetic, environmental, and lifestyle variables may all predispose a person to the development or progression of BC (16). However, the exact mechanism of BC's occurrence is still unclear. The upcoming subsections will discuss BC incidence and progression risk factors.

1.2.2.1 Patient Age

The occurrence of BC and other human malignancies primarily depends on age (17). Moreover, it has been indicated that BC prognosis can be independently indicated by age at diagnosis. Females over the age of 50 contribute to 80% of the total BC cases. At age 40, the 10-year probability of developing invasive BC is less than 1.5%; at age 50, it is approximately 3%. At age 70, it is greater than 4%. Accordingly, an increased lifetime risk of 13.2% occurs, and women over 50 have a chance of developing cancer about nine times higher than those younger than them (18, 19). Within the UK, those aged 65–69 make up about 48% of all patients with BC; most of the incidence rates involve women between 85 and 89 (20). Furthermore, the tendency of BC to be in aggressive advanced-stage molecular subtypes, such as (human epidermal growth factor receptor 2 [HER2]) enriched tumours and triple-negative BC (TNBC), are more common in young women, which are highly correlated with poor prognosis. (21, 22). Thus, in addition to age, genetic abnormalities are one of the primary causes of BC.

1.2.2.2 Family History and Genetic Alteration

Of the total BC incidence, approximately 10% of cases are considered hereditary (23). The risk of developing BC increases in women with a close family member diagnosed with BC, such as a mother, daughter or sister, or with more than one relative previously affected by ovarian cancer. Furthermore, the risk of BC development is linked to inheriting multiple coding region and splice variants in BC type 1 (BRCA1), BC type 2 susceptibility protein (BRCA2), p53, phosphatase and tensin homolog (PTEN) and ataxia-telangiectasia mutated (24-26). A large number of total germline mutations are correlated with *BRCA1/BRCA2* mutation, which, according to a recent study, accounts for about 25% of familial risk (27). Interestingly, people with *BRCA1/BRCA2* mutations are not only at risk of BC but are also vulnerable to the ovarian, gastric, colon and pancreatic carcinomas (28-30). Nonetheless, BC as a disease is regarded as

multifactorial cancer that can be initiated by a polygenic mode of inheritance (31). For example, multiple studies have identified numerous loci associated with BC angiogenesis and metastasis (32). Polygenic factors are believed to increase BC risk factors by 18% in patients with high levels of polygenic risk (33). For instance, various variants which have been identified act through a wide range of critical molecular pathways and play essential roles in biology, such as maintaining the integrity of DNA repair processes (34). Polygenic testing has the potential to enhance health care by delivering personalised risk assessments to women at high risk for BC and within-population breast screening programmes (35, 36).

1.2.2.3 Lifestyle-Related Factors

Lifestyle factors, such as tobacco smoking, alcohol intake, obesity, , can be linked to BC initiation. Among smokers, both active and passive smokers are vulnerable to BC, with a high risk of invasive BC (IBC) (37). Moreover, in a previous study, BC development increased by <10% when individuals consumed ~10 g of alcohol daily (38). Increased incidence and mortality of BC are also linked to elevated body mass index and obesity (39-41), with a high risk of BC in post-menopausal women (42, 43). Several obesity-related mechanisms could account for the raised likelihood of developing BC. For instance, adipose tissue expresses aromatase, which is an enzyme that transforms androgen into oestrogen. In addition, obesity is associated with higher levels of circulating insulin, secreted inflammatory cytokines, and insulin-like growth factor I. All of these are associated with elevated probabilities of developing BC (44). In addition, long-term BC survivors may benefit from weight control for up to five years to improve survival and outcome (45).

1.2.2.4 The Effect of Breast Density

Breast tissue density can aid in indicating whether a woman is at higher risk of developing BC. Considering dense breast tissue, it is known that it is associated with an accumulation of glandular and supportive fibrous tissue compared to fatty tissue. To distinguish between the two, a mammogram can be used. Mammographic breast density is a strong indicator of BC risk. This shows that women with higher breast density risk having BC more than those with lower densities (46).

1.2.2.5 The Effect of Oestrogen and Hormonal Therapy

One of the leading causes of BC angiogenesis and development is hormonal factors. The body produces a sex hormone known as oestrogen involved in several biological functions. Thus, BC risk is increased with any manipulation of this hormone, such as exposure to hormonal therapy after menopause or taking oral contraceptives that may induce breast tissue growth (47). For example, lobule differentiation has a protective effect that could decrease IBC incidence, as well as by reducing the exposure to oestrogen through pregnancy and breastfeeding, the chance of protection will increase (48). A study showed that BC occurrence is significantly lower in women with a full-term pregnancy and those who are lactating than in those who are nulliparous or have never been pregnant (49).

Of note, premenopausal women may develop BC as well. They are at high risk of IBC when their ovulation is regulated with a form of progesterone that includes oestrogen and

medroxyprogesterone (50). Henceforth, it has been noted that the risk of BC increases with extended hormone replacement therapy (HRT) from 1.6–2 fold with a combined or single formula of HRT (51). For example, a study on a group of women aged 60–64 who had used HRT for at least five years showed that they had a 2-fold risk of BC compared to those with no history of HRT use (51). However, this high-risk factor for BC development is reduced five years after the termination of HRT use (51).

1.2.3 Breast Cancer Prognostic and Predictive Factors

The term prognostic factor is used to refer to any characteristic of a tumour that can be used to predict the patient's outcome based on BC-specific survival (BCSS), distant metastasis-free survival (DMFS) and disease-free survival (DFS). However, predictive factors are defined as any features depending on the clinicopathology of BC that can estimate the possibility of a patient's response to treatment. Prognostic and predictive qualities can be elucidated through definite clinicopathological features.

The severity and aggressiveness of the tumour and the chance of patient survival are usually estimated by certain traits known as prognostic factors, and this estimation is called prognosis. Thus, to obtain an overall idea of how BC cancer will affect a particular patient's survival, many prognostic factors should be used together. Likewise, predictive factors are tools used to guide clinicians in determining prognoses and patients' responses to treatment. Therefore, these factors hold the utmost importance for stratifying patients with BC for optimal treatment choices (52). For example, the establishment of multiple algorithmic tools such as Predict (<https://breast.predict.nhs.uk/tool>), Adjuvant Online and CancerMath® paved the way for clinicians to expand their knowledge of BC patients' survival by incorporating clinicopathological factors associated with BC (53). Pinpointing of precise prognostic and predictive parameters may guide proper therapeutic selection and improve survival. The following subsection will highlight the most important prognostic and predictive factors in BC.

1.2.3.1 Tumour Grade

Tumour grade is a fundamental prognostic indicator consisting of three tiers to assess the similarities between BC cells and ordinary breast cells. Furthermore, this prognosis indicator aids in showing how BC cells are structured in a malignant lesion. this study has embraced a tumour grading system consists of three categories with the parameters displayed in Table 1.1 (54). This tumour grading system has also been integrated into the Nottingham Prognostic Index (NPI) (55).

Table 1.1:Tumour grading system.

Tumour grade	Scale by definition
I	Breast cancer cells, like normal cells, well-differentiated and grow slowly.
II	Breast cancer cells are not like normal cells, normal differentiation and grow rapidly.
III	Breast cancer cells are extremely different to normal cells, poorly differentiated and have fast proliferation.

1.2.3.2 Tumour Size

Tumour size is important when predicting patient survival rates and BC behaviour and can aid in patient outcome prediction. Thus, important prognostic implications within IBC can be seen via the tumour size (56, 57). For instance, a small tumour size represents a better patient outcome, and vice versa. Furthermore, the size of the tumour has been seen to be a solid predictive factor of nodal metastasis in IBC (58). Within this study, the tumour size is divided into small or large tumour groups. The scale used to define whether a tumour is small or large is as follows: small tumour ≤ 2 cm; large tumour > 2 cm (55). The size of the tumour is incorporated into the NPI (55).

1.2.3.3 Lymph Node Stage

A global and most widely used tumour evaluation system consists of three major components; tumour, node and metastasis (TNM) (59). However, within IBC, lymph node (LN) stage is one of the most vital prognostic factors. In the early BC stage, the axillary LN stage shows a strong relationship with the DFS and BCSS (60, 61). In this study, three categories are used when categorising BC cases; the following groupings were obtained through the applied pathological NPI staging system for LN: 1) LN stage I (no LN is involved); 2) LN stage II (one, two or three LNs are involved); and finally, 3) LN stage III (the number of LNs involved is greater than three) (62, 63). There is a strong correlation between the LN stage status and the mortality rate among patients. A higher LN stage represents a poorer survival rate, and a lower LN stage shows a better survival rate. Over five years, the percentage of patients with BC who survive

while in LN stage I is approximately 83% (61). This LN staging system is incorporated into the NPI (55).

1.2.3.4 Nottingham Prognostic Index

For patients with IBC, the NPI is used as a prognostic stratification tool (63). The three parameters used within the NPI are tumour size, LN stage and tumour grade. This, in turn, enables enhanced predictive power to be provided for patients with IBC. The NPI has a unique formula, as follows:

$$\text{NPI} = (\text{tumour size in cm} \times 0.2) + \text{tumour grade (1–3)} + \text{LN stage (1–3)}$$

The first division of the NPI grouping allocates patients into three prognostic groups. Based on their calculated NPI, the three groups are <3.4, 3.4–5.4 and >5.4; the three groups are patients with good, moderate and poor prognosis, respectively (55). They can be further divided into six prognostic groups (Table 1.2) (64).

Table 1.2: NPI subgroups

Group 1	NPI range
Excellent prognostic group (EPG)	2.08–2.4
Good prognostic group (GPG)	2.41–3.40
Moderate prognostic group-I (MPG-I)	3.41–4.40
Moderate prognostic group-II (MPG-II)	4.41–5.40
Poor prognostic group (PPG)	5.41–6.40
Very poor prognostic group (VPG)	>6.40

1.2.3.5 Lymphovascular Invasion

An independent factor within patients with IBC is lymphovascular invasion (LVI), which is highly correlated with BCSS and DMFS (65). It refers to the presence of the tumour cells at the peritumoural area within blood vessels (BVs), lymphatic spaces or both. The morphological assessment of LVI is conducted by microscope with or without using endothelial-specific markers such as CD31, CD34 and D2-40 (immunohistochemistry [IHC] and haematoxylin and eosin [H&E]). The detection of LVI in the primary tumour of patients with BC is a known marker of metastatic potential. In addition, a number of independent studies have shown and demonstrated the prognostic value of LVI (65, 66). Therefore, this study focuses on identifying those molecular drivers for LVI within a large group of patients with IBC. Identification of LVI-related genes and proteins is needed to improve therapeutic strategies in the future. Thus,

if successful, these genes and proteins may provide independent prognostic information to pinpoint those groups of patients with IBC who are within the high-risk subgroup categories. The LVI detection methods and underlying molecular mechanisms are discussed further in this **Chapter (section 1.3)**.

1.2.3.6 Molecular Subtypes

Breast cancer can be classified according to histological grade and histological subtype (67) or biomarkers (68), which will further help in tailoring treatment regimens (69). In early 2001, the description of the gene expression patterns of BC was established (70). The classification of BC at that time started with two different groups of hormonal receptor tumours, that is, hormonal receptor–positive (HRE+) and hormonal receptor–negative (HRE–) tumours. Then, DNA microarray experiments with hierarchical clustering analysis identified a genes signature based on two sets from BC tumours using the Prediction Analysis of Microarrays of 50 genes (PAM50) panel. In addition, clustering analysis of HRE studies further classified BC into luminal A, luminal B, basal-like, HER2 and normal breast-like categories (70) (Figure 1.2). Using IHC, the positive expression of type II cytoskeletal 5 and 6 and/or epidermal growth factor receptor (EGFR+) helps to distinguish between the two triple-negative subtypes, that is, basal-like and non-basal BC (71). A previous study classified 10 integrative clusters (IntClusts) found in the studied cohort using genomic and transcriptomic profiling of 2,000 BC samples (72). The IntClusts displayed varied clinical, molecular and genetic features, and each had its own expression profiles of oestrogen receptor (ER), progesterone receptor (PR) and HER2 (72). The BCSS and IntClusts association is illustrated in (Table 1.3), and the molecular features for the IntClusts are shown in Table 1.4 (73).

According to their ER, PR and HER2 expression profiles, these phenotypes are denoted as luminal A tumours, ER and/or PR positive and HER2 negative, low cell proliferation phenotype and good prognosis (74). On the contrary, luminal B tumours have high proliferation rates, an aggressive histological type and a worse prognosis than luminal A tumours, with HER2 positive or negative (75). The basal subtype, also known as triple negative, presents ER, PR and HER2 negative (76) (Figure 1.2).

It is well documented that BC molecular subtypes are related to LVI and LN metastasis (77). The triple-negative molecular subtype is more mitotic (78), potently invasive and metastatic, as it expresses low level of claudin protein that participate in epithelial cell to mesenchymal transition (79, 80). Moreover, TNBC demonstrated significantly increased Ki-67 expression than

any BC subtypes (81). The HER2-positive subgroups behave better with targeted therapies (82). As reported, the BC molecular subtype determines the tumour properties (83).

Table 1.3: IntClusters association

Breast cancer specific survival, up to ten years after initial treatment	
BCSS	IntClust
Good	3,4,7 and 8
Moderate	1,6 and 9
Poor	2,5 and 10

Table 1.4: Molecular features for the integrative clusters IntCusts

IntClust	Defect	IntClust	Defect
1	17q23 amplification	6	8p12 amplification
2	11q13/14 amplification	7	16p gain, 17q loss, 8q amplification
3	a paucity of copy number changes	8	1q gain, 16q loss
4	copy number devoid	9	20q amplification
5	HER2 amplification	10	5q loss, 8q gain, 10p gain, 12p gain

	Luminal A	Luminal B	HER2+	Normal breast-like	Triple negative
Percentage at diagnosis	50-60%	10-20%	10-15%	5-10%	15-20%
Receptor expression	ER and/or PR +ve, Her-2 -ve	a) Her-2-ve, ER and/or PR +ve b) Her-2+ve, ER and/or PR +ve	Her-2 +ve ER -ve and PR -ve	ER -ve/+ve Her-2 -ve	ER -ve, PR -ve, and Her-2 -ve
Proliferation rate	Low	a) High b) Any	High	Low	High
Outcome	Excellent	Intermediate/ Bad	Bad	Intermediate/ good	Bad

Figure 1.2: Breast cancer classification according to a combination of Immunohistochemistry and microarray expression methods. Details in table adapted from (84).

1.2.3.7 Hormone Receptors, Oestrogen and Progesterone receptors

The recommended threshold of positive detection for ER and PR expression in patients with IBC is set at 1% of the tumour nuclei according to the College of American Pathologists and the American Society of Clinical Oncology (85). Oestrogen receptor is an important indicator for extended adjuvant endocrine treatment. Nevertheless, the duration for which the hormone thereby regimen is used for ER patients is still under discussion (86). The incidence of aggressive BC phenotypes in younger patients is common, considering the prognostic significance of ER-positive in these individuals is of limited utility. It is anticipated to be more beneficial for BC patients in post-menopausal. However, ER status could aid in clarifying the effect of patient age as an independent prognostic factor, which would allow a patient with BC to achieve a greater benefit from the chemotherapy treatment (87, 88).

Moreover, in routine clinical practice, co-expression of PR in ER+ patients is also measured. This status is well known for predicting response to adjuvant hormonal therapy in patients with BC. The hormonal treatment response is predicted to reach 80% when there is co-expression of ER and PR, whereas it is only 50% in ER+/PR- patients.(87).

1.2.3.8 Human Epidermal Growth Factor Receptor-2

The receptor tyrosine-protein kinase erbB-2 encoded by *ErbB2* (HER2) has a percentage of overexpression in patients with BC of around 15%. It has a strong correlation with an aggressive phenotype of IBC and unfavourable outcomes (89). Additionally, HER2 positivity is correlated with discohesive BC cells undergoing epithelial–mesenchymal transition (EMT), an indicator of oncogenic potentials (90). EMT is a biological process that refers to the transformation of polarised epithelial cells which form extensive cell-cell adhesions with one another, into mesenchymal cells (91). This change results in the functional promotion of cell migration, and invasion and evades programmed cell death.(92). Moreover, HER2 positivity within patients with IBC is linked to tumour invasion, abnormal growth, aberrant cellular polarity and poor survival (93). Interestingly, in a previous study of patients with IBC, HER2 positivity showed a strong relationship with LVI (94). Furthermore, in all IBC subtypes, the occurrence of LVI and HER2 positivity in BC is related to poor DFS, BCSS and DMFS (95).

In order to treat HER2 tumours, the monoclonal antibody (trastuzumab) is used to target the HER2 receptor in BC tumours (96).

1.3 The Molecular Pathways that Interact with Lymphovascular Invasion in Invasive Breast Cancer

1.3.1 Cancer Metastasis Cascade (Invasiveness, Tumour Microenvironment and Lymphovascular Invasion)

Metastasis is a complex series of events, starting from the detachment of tumour cells from the primary tumour mass with local invasion of surrounding host tissue and invasion or intravasation into the BV or LV through the extracellular matrix (ECM) (97, 98). Finally, the disseminated tumour cells invade distant organs through BVs and LVs (99). The distant metastasis process inhibits the cell cycle initially, and then the tumour cells adhere to the capillary bed inside the affected organ, thereby enhancing angiogenesis to prepare for extravasation into the new tissue/organ (98). At the same time, tumour cells attack the host's immune response and decreases apoptotic signals to survive and proliferate, forming new tumour masses (100). These sequential steps can be repeated to produce secondary metastases (Figure 1.3).

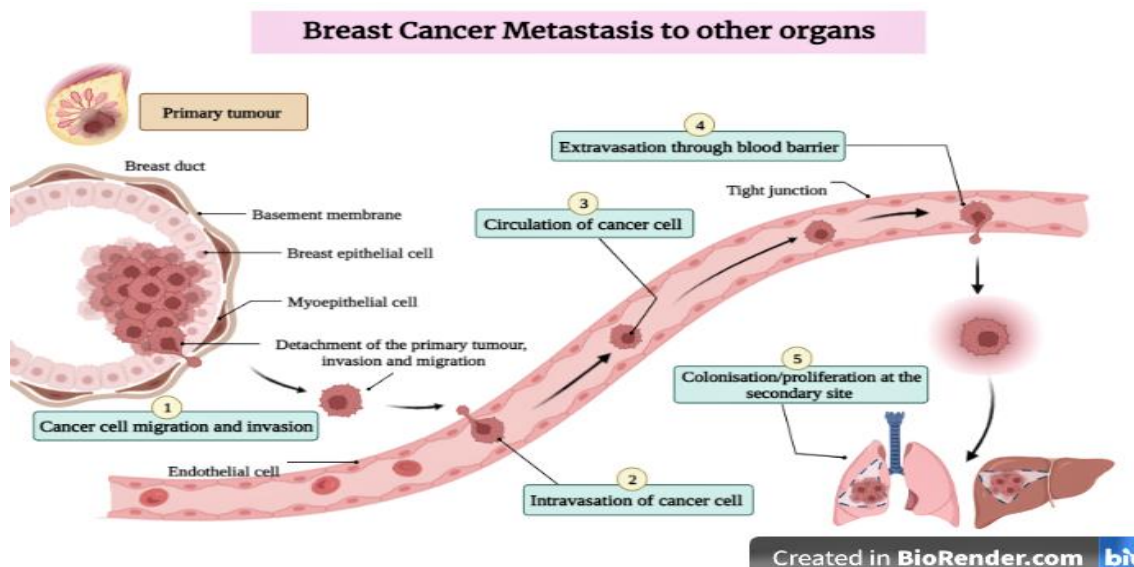


Figure 1.3: Stages of tumour metastasis; the diagram shows the sequential steps utilise by tumour cells to metastasise from the primary site to the new organ.

According to theories, metastasis occurs via two methods. The first method is the late-stage or clonal selection theory; only tiny parts of tumour cells with selective advantage induce distant metastasis via the aggregation of genetic modifications, which occurs as the tumour advances (101). In contrast to the late-stage theory, new DNA microarray-based techniques have revealed that cancer cells' potential to metastasise is a pre-programmed process that gives them an inherited general tendency for metastasis (102). Thus, each type of cancer has different preferred metastatic organs. For example, BC tends to metastasise to the bones, lungs and liver more than other organs (103). The role of particular organs in the metastatic dissemination of tumours has been the subject of many different hypotheses. In 1889, Stephen Paget introduced the seed and soil theory postulating that cancer metastases by means of an organ-specific process in which particular organs serve as an ideal environment (soil) for the growth of cancer cells (seeds) (104, 105). Local growth factors, specific endothelial cell surface receptors and specific chemotactic agents differentially expressed by various organs may play a role in the tendency for some cancers to selectively develop in particular organs (106-108), indicating that these factors are behind metastasis distribution patterns. In contrast to the seed and soil theory, in 1928, James Ewing suggested that the vascular system's anatomical structure primarily defines the pattern of metastasis distribution (109). According to his prediction, when cancer emboli detach from the primary cancer site and intravasation into the circulation became mechanically obstructed in the vessel of the initial organ being attacked, this increases the chance of metastasis there (109). In 2016, an investigation on metastasis initiation done by Massagué and Obenauf, suggested that a well-understanding of the effect of tumour microenvironment (TME) obstacles, which is control metastasis colonisation, is the key factor in preventing tumour metastasis and assisting in exploring metastasis treatment (110).

Furthermore, it has been suggested that BC metastasis could occur via both LVs and BVs, indicating aggressive diseases that have poor outcomes (111). Dissemination through LVs and BVs is considered the primary reason behind BC mortality, by which lymphangiogenesis and angiogenesis initiate preluding for cancer invasiveness and metastasis (112-114). Lymphovascular invasion is the mechanism by which tumour emboli find a way to penetrate vascular/lymphatic channels at the beginning of the metastatic process, while angiogenesis is the creation of neovasculature to support cancer growth and enhance invasion. It has been shown that LVs are involved more often in BC initial metastasis than BVs; this might be due to the affinity of intrinsic tropism of tumour cells to lymphatic channels (115). Even though it has long been suggested that cancer cells spread via the lymphatics by following the lymphatic

drainage paths, the exact mechanisms by which tumour cells enter lymphatic capillaries remain unclear (113, 116-118). Thus, lymphangiogenic factor expression might impact the path of metastatic dissemination via LVs due to their unique structure, promoting cancer progression (119, 120). Despite the significant correlation between LVI and both angiogenesis and lymphangiogenesis, none is comparable to LVI, regardless of the type of vessels engaged, whether BVs or LVs, native or newly created (121). Although the vasculature is crucial in creating an ideal environment for LVI to flourish, it is not equivalent with LVI.

Moreover, malignant epithelial cells of the breast use the blood and lymph vessels to spread to other surrounding tissues and distal organs of the body by recruiting the cellular parts of the nearby stroma TME (122, 123). Stromal cells, including BV and LV endothelial cells, tumour-related fibroblasts and inflammatory cells that are originate from bone marrow, operate as stimulators of tumour cell migration and invasion and as path controllers in the ECM in many cases (124).

The intracellular communication of malignant cells with stromal parts of the tumour, termed the TME metastasis, could enhance migration, adhesion and invasion of the adjacent tissues. Afterwards, the penetration of blood and lymph vessels becomes more likely to happen (125). The controlled genome organises the biological and situated functions of the confined epithelial cells. For instance, the alteration of the normal DNA coding of the epithelial cell due to the oncogenic changes allows the initiation and angiogenesis of malignant epithelial cells. The previously mentioned process could initiate embryogenesis transformation of the new malignant cells, which plays a significant role in invading various tissues and effectively accomplishing the resulting metastasis events (126). Epithelial–mesenchymal transition is another term for this cellular renovation.

The cell-to-cell adhesion and cell adhesion to the ECM are changed, causing the ECM invasion. Before BC cells penetrate into the vascular lumina, they must first separate from the developing tumour cell mass and then invade and migrate via the ECM to the vascular walls. As a result, these processes are regarded as prerequisites for the LVI process. According to a previous report, the mediation of cell-to-cell adhesion and BC metastasis dramatically depends on the cadherin family (127). The metastatic sprout BC cells result from insufficient E-cadherin, which maintains cell–cell junctions (128). A report has also suggested a relationship between the downregulation of E-cadherin and progression alongside poor patient outcomes (129). There is also a strong relationship between EMT and N-cadherin (130).

N-cadherin is associated with malignant tumour progression elements in humans, implying that it might be a potential therapeutic target in case of tumour invasion and metastasis (131). It is increasingly being proven that EMT, which alters the ECM, dramatically contributes to tumour progression (132, 133). According to Kotb, mice get cysts and fibrosis in their mammary glands because of the elevated N-cadherin switch over E-cadherin (130). In stromal invasion, tumour cells get adhered to stromal cells and consequently invade the stroma. Moreover, stromal invasion is also often shown to upregulate N-cadherin and downregulate E-cadherin (134, 135). Transcription factors are proteins with a significant biological role in the upregulation or downregulation of the transcription of tumorigenesis-associated genes. Increased motility is achieved through the downregulation of desmosomes and by diminishing cell-to-cell adhesion proteins. As far as invasive breast carcinoma is concerned, the malignant cells on the tumour boundaries upregulate other genes, such as *SLUG*, *SNAIL* and *ZEB1*, speeding up their potential transformation to the mesenchymal phase (136, 137). However, they downregulate their production of adhesion molecules such as E-cadherin (138). Apparently, E-cadherin expression levels in tumour cells could affect invasion via different processes. Therefore, downregulation or loss of E-cadherin expression might promote the migration of cancer cells to BVs or LVs. With an increase in the high probability of a malignant cell invading the vascular system, the metastasis, recurrence and mortality rates will aggressively elevate (77).

It has been suggested that tumour cells can interact with the ECM, mainly due to transmembrane receptors known as integrins (139, 140), which are found associated with ECM components such as vitronectin (141), fibrinogen, collagen, laminin and fibronectin (142). Integrin receptors for cell adhesion to the ECM play a key role in tumour development and growth. They promote ECM-degrading enzymes, such as matrix metalloproteinases (MMPs), contributing to the modulation of tumour motility (143). This can be seen in integrins $\alpha 5 \beta 1$ and $\alpha 3 \beta 1$, both of which result in the upregulation of *MMP-9* (143). Another activator that enhances the ECM process through MMPs is urokinase plasminogen activator (uPA). After the invasion, the degradation process of ECM is escalated due to uPA, and the penetration of MMPs and tissue boundaries becomes possible as the ECM degrades (144). In patients with BC, uPA has prognostic significance as a risk factor for distant metastases (145). Tumour tissue with higher uPA concentrations progresses more aggressively, and patients are at greater risk of recurrence and have lower survival rates (146). A recent study that transfected uPA to human umbilical vein endothelial cells (HUVECs) showed that uPA could promote proliferation and fibrinolytic activity, consequently improving the invasion and angiogenesis ability of HUVECs via

promoting their expression of MMP-2, MMP-9, VEGF (147) and MMP-1 (148). This might explain the LVI phenomenon since studies have shown a strong positive correlation between LVI and MMP-9 (149, 150).

Metastatic progression through BVs and LVs occurs through various biochemical and biophysical interactions (151). Properties of malignant cells include elevated survival, resistance to apoptosis, increased protein expression activity, higher cellular motility and proliferation rates, adhesion, invasion and migration (152). These properties represent the internal and external characteristics of the tumour, which regulates genes. Henceforth, finding new ways to reduce the breast epithelium's tumorigenicity will depend on the activation/inhibition of these signalling pathways (153). Therefore, an urgent investigation is needed to improve personalised treatment regimens to uncover the molecular mechanisms underlying LVI and its driver genes.

1.3.2 An overview of Lymphovascular Invasion

Despite the fact that fewer than 10% of patients with BC present with metastatic tumours, up to 30% of those with early stage disease continue to develop distant metastasis over time (154-156). Distant metastasis is a TME indicator of poor outcome in cancer since. It develops through particular and coordinated processes and interacting pathways. These processes involve various mechanisms, such as proliferation, interactions with the cell matrix, stromal invasion, migration of cells and circumvention of the host immune system (157-159).

Breast cancer cells undergo several biological interactions at the molecular level to propagate and metastasise through lymphovascular channels (151, 160, 161). Further investigation of these signalling cascades is essential to recognise the specific molecular entities that govern the critical stages of LVI. Hanahan and Weinberg discussed several mechanisms for cancer initiation and progression, indicating that there are dominant intrinsic and extrinsic characteristics of cancer cells that control their genetic regulation (1). However, these properties are not linked directly to the incidence of LVI and can be recognised in LVI- and LVI+ tumours. Nevertheless, they are critical characteristics of LVI development, and the controlling molecular mechanisms behind them must remain active, especially in some cellular clones, during the stage of carcinogenesis from initiation to metastasis via vascular channels.

1.3.3 Lymphovascular Invasion Assessment

Traditionally, LVI is evaluated via morphological examination of the peritumoural area, which is identified as areas outside the tumoural compartment. Intratumoural the areas are areas inside the tumoural compartment. It's noteworthy that pure intratumoural LVI is very rare and its value as a prognostic factor for BC falls behind the peritumoural LVI (115, 162-164).

Nearly 75% of LVI+ cases develop peritumoural LVI, while (~25%) develop intratumoural LVI, of which approximately 4% have only intratumoural compartment LVI (115). The low prevalence of intratumoural LVI means that the evaluation regarding its implications is unclear and generates the dilemma of whether intratumoural LVI is significant. Moreover, intratumoural LVI might explain metastatic cascades when LN positivity is present in intratumoural LVI cases only (115).

Nevertheless, intratumoural LVI should still be considered assiduously in the absence of cancerous cells in LN (121). Therefore, it may be pragmatic to acknowledge that the tumour has developed mechanisms for dissemination into the vasculature and should be placed under close clinical observation.

Various studies indicate that LVI occurs in BC tumours despite the tumours lacking morphological signs of LVI, such as the presence of circulating tumour cells (165), bone marrow micrometastases (166) and LN positivity (167). The reason behind this is hypothesised to be linked with BV and lymphatic vessel (LV) cancer cell migration. It has been documented that cancer cells utilise several biochemical and biophysical interactions to facilitate their metastatic progression through vascular and lymphatic channels (121). Therefore, understanding of the molecular processes that underlie LVI in BC remains sparse, as determining the likelihood of metastases in BC tumours lacking LVI remains a challenge.

Furthermore, molecular changes identified in tumours containing LVI compared to tumours lacking LVI may indicate a number of underlying biological changes rather than the histological presentation. Thus, in BC, LVI may reflect a biological spectrum comparable to alternative morphological markers of cancer biology, such as the grade of malignancy, where cases without LVI are on one end of the range while cases with substantial LVI are on the other end. Moreover, the complicated nature of the molecular pathways underpinning LVI, their intersection with the metastatic tumour cascade involved in tumour dissemination and the absence of apposite models to simulate BC cells' intravasation have increased the complexity

of recognising the governing molecular entities, pathways and possible therapeutic targets related to LVI in BC.

1.3.4 Lymphatic Vessels Structure, Function and LVI-related Markers

Structurally, BVs have tight junctions in contrast with LVs (119), making LVs more susceptible to cancer cell migration than BVs; consequently, cancer cells can easily surpass intravasation. With that in mind, peritumoural LV intravasation might be considered an essential path for IBC cells failing to bypass the vascular endothelial cells' physical restraints (168).

The lymphatic system is vital for maintaining tissue homeostasis. Through a combination of osmotic and hydrostatic pressure, the blind ending vessels of lymphatics collect intercellular fluid from the interstitial area and return it to the blood circulation, keeping the body's fluid balance normal (169). In response to increased interstitial tissue pressure, overlapping intracellular junctions can open, allowing fluid entry into the lymph vessels. When interstitial fluid enters the lumen, it is called lymph. As the lymph fluid circulates through the lymphatic system, it eventually enters the right lymphatic duct and thoracic duct. Lastly, it is returned to the bloodstream via the left and right subclavian. Lymphatic vessels are also implicated in spreading cancer cells throughout the body.

Metastatic malignant cells are also believed to migrate towards lymphatics near the tumour periphery due to the pressure in the interstitial fluid, increasing the possibility of passive intravasation into lymphatic channels (168, 170). Despite that, recent studies have suggested that lymphatic endothelium may attract tumour cells. Endothelial cells release chemokines, small molecules that attract cancer cells, to lymphatic capillaries during this process (171, 172). The relationship between chemokine receptors and their ligands is essential for controlling the movement of leukocytes (173). This association has been hypothesised to be utilised by cancer cells to modify access to the lymphatic system (173). Nonetheless, active molecular processes can attract cancer cells to both kinds of capillaries (174). Lymphatic research has not developed as swiftly as vascular research due to a lack of histological and immunohistochemical markers that compare blood and lymphatic endothelial cells (LECs). However, in recent years, the identification of LVs has indeed been implemented by use of specific lymphatic markers, such as D2-40/podoplanin (PDPN) (175, 176), vascular endothelial growth factor receptor-3 (VEGFR-3) (177) and cluster of differentiation 31(CD31)/platelet endothelial cell adhesion molecule-1 (PECAM-1) (178).

The antibody D2-40 and lymphatic microvascular density (LVD) parameters are widely used in research for detecting LVI. To start, D2-40 is a monoclonal antibody that detects PDPN, a mucin-like transmembrane glycoprotein found at the endothelium of LVs (179). It has high sensitivity in differentiating lymphatics in BC tissue (180-182). Also, in both intratumoural and peritumoural LECs, D2-40 exhibits the highest immunoreactivity (180). In addition, several studies have found that the D2-40 IHC marker helps diagnose and verify LVI in IBC (183) and increases the detection accuracy of LVI (184). Moreover, a higher density of LVs around a tumour is associated with lymphatic metastasis in BC (185), which eventually will lead to local recurrence (186) and shorter patient survival (187). For example, a high LVD might indicate that the risk of tumour cell intravasation is elevated and that the tumour cells communicate with LECs more intensively, which is essential for tumour progression (188). According to a review of 28 studies involving 2,920 patients with BC, peritumoural LVD is more robustly correlated with LVI and LN metastasis (LNM) than intratumoural LVD, suggesting that the peritumoural lymphatic system may be the main path of dissemination for BC cells. As a result, it is recommended that patients detected to have high LVD or LVI presence, especially with peritumoural high LVD, should be treated more aggressively to prevent BC progression and metastasis (185). Britto et al. (189) reported that despite improvement of diagnostic accuracy of LVD and LVI using D2-40, neither can be used to predict sentinel LN (SLN) macrometastasis. Interestingly, progressive metastasis and poor prognosis are indicated if the SNLs have malignant breast carcinoma cells (190). Britto et al. did not find that a correlation between LVD and SLN status might be due to the mechanistic complexity of circulation and nesting of neoplastic cells, which could explain this dilemma.

Podoplanin is a transmembrane mucin that is highly O-glycosylated and expressed by LECs. It is employed as a selective lymphatic endothelium marker (191). Additionally, the PDPN cytoplasmic tail interacts with ezrin and/or moesin, which are the members of the ezrin, radixin and moesin (ERM) protein family, and with MMP- 9 and 2. They induce EMT, or restructuring of the actin cytoskeleton and cellular protrusions; consequently, this transmembrane glycoprotein enhances cancer cell motility and invasion (192-194). According to Neri et al. (195), tumours with high PDPN expression have increased local migration and invasion compared to tumours with low PDPN expression. This implies that PDPN may be a viable therapeutic target via migration suppression using a PDPN monoclonal antibody (NZ-1) and maackia amurensis seed lectin inhibition (196). Positive PDPN expression is more common in

LVI than CD31 or CD34 positivity and is related to the development of nodal metastases and poor patient outcomes (197, 198).

Sometimes referred to as FMS-related receptor tyrosine kinase, VEGFR-3 is a tyrosine kinase receptor expressed only on LECs in normal adult tissues (199, 200). It is believed to contribute to metastatic spread through lymph vessels (114). When malignant cells overexpress VEGFR-3, it affects the process of lymphangiogenesis (201). Moreover, VEGF-C and VEGF-D are new VEGF family members that can control lymphatic channel formation through VEGFR-3, a receptor expressed mainly by LV during growth (202-204). As a result of activated VEGFR-3, LECs proliferate, enhancing the ability of cancer cells to metastasise into lymphadenoids and lymph vessels. However, several studies have found no significant relationship between VEGF-C and LVI or nodal metastases, which might be explained by the possible involvement of many variables in addition to VEGF-C and VEGF-D and the heterogeneity of research samples (205).

Lymphatic-specific adhesion molecules are crucial in migrating and transmigrating cancer cells into LVs. Initially, they must form an adhesive bond with LECs. Several adhesion molecules have been found to be expressed on LECs especially. For example, PECAM-1, also known as CD31, is a cell-cell adhesion protein (206) encoded by the *PECAM-1* gene that has been primarily utilised to identify endothelial cells in several research works (207). Several adhesion molecules have also been detected at intercellular connections in LN sinus-specialised ‘endothelial’ cells (208). Vascular endothelial cadherin (VE-cadherin) is the main element of endothelial cell adherens junctions and plays a critical role in controlling vascular permeability and angiogenesis (125, 209). Notably, CD31 colocalises partly with VE-cadherin at zipper-like junctions and at the tip of overlapping flaps where there are less button-like junctions (210, 211). Vascular endothelial cadherin is necessary for junctional integrity preservation in C57BL/6 mice, while PECAM-1 is not (211). This proves that gene profiling data are used by LECs to create junctional proteins that express CD31, junctional adhesion molecule-A and occludin in cultured cells (212, 213).

1.3.5 Conclusion

Researchers have comprehensively characterised the molecular profiles of tumour differentiation based on histological grade and the development of nodal dissemination (214-216). Nevertheless, the molecular processes of LVI and driver genes, which may be used as potential therapeutic targets or biomarkers, have yet to be fully define. Determining genes that are differently expressed in LVI+ and LVI– BC might lead to the discovery of essential genes and pathways that trigger LVI. However, the nature of the analysed cohorts is one of the most significant challenges in identifying LVI-associated genes in genome-wide gene profiling studies. Thus, identifying molecular profiles associated with LVI status in BC is challenging given the lack of LVI status in available large-scale molecular studies combined with the inherent subjectivity of morphological assessment.

. Although LVI is well-known as a pre-requested step for BC metastasis, the key molecular factor/s driving LVI still needs to be detected. Bioinformatics research is required to discover the target genes that play a crucial role in LVI mechanisms, which might help to understand the pathways underlying BC-LVI; thus, identifying suspected biomarkers could be used as a therapeutic target.

1.4 Hypothesis and Aims of the Study**1.4.1 Hypothesis**

This study hypothesises that specific genes and proteins can regulate LVI progression in patients with BC through different convergent and/or parallel molecular and biological pathways.

1.4.2 Aims

1. To investigate the molecular characterisation of LVI in BC by creating a gene predictor of LVI.
2. To investigate the role of highly significant LVI-associated gene candidates in a large, well-characterised series of BC at the protein level, including their relevance to molecular subtypes, response to therapy and patient outcome.
3. To build a predictive model for LVI to form a panel of markers that could be conjointly used as a reliable tool in patient risk stratification.
4. To investigate the mechanistic and phenotypic functions of significant LVI-associated genes in BC cells.

2 Chapter 2: General materials and methods

2.1 Study Cohorts

2.1.1 Transcriptomic Level

2.1.1.1 *Molecular Taxonomy of Breast Cancer International Consortium*

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) series (n = 1,980) is a microarray dataset with a large set of data consisting of well-defined genomic clinical characteristics and patient outcome follow-ups and was used to identify the LVI mechanisms in this study. A subset of the METABRIC cohort was originally derived from the Nottingham Primary Series (NPS), which is composed of patients with IBC. A total of 296 primary IBC fresh frozen biopsy specimens were previously sent from Nottingham City Hospital (UK), part of Nottingham University Hospitals, and received at Addenbrooke's Hospital, which is affiliated with Cambridge University Hospitals. The Nottingham contribution was essential to build the complete METABRIC cohort of 1,980 patients and to integrate gene expression and copy number analyses to uncover new BC subgroups based on genomic and transcriptomic data (72, 78, 217). The decrypted and reviewed data file was received on 5 September 2016, provided by Dr Oscar M. Rueda, senior research associate at the Cambridge Institute, University of Cambridge. At the transcriptomic level, the differential gene expressions (DGEs) related to the molecular regulation of LVI were determined for a large number of patients via independent and mutual confirmation approaches.

In the METABRIC cohort, 1,980 female patients with primary operable IBC were included in the clinicopathological profile (Table 2.1). For each patient, the BCSS was calculated in months based on the start date of the primary surgical treatment and ending at death from BC. By the last follow-up, 505 patients (25.5%) had died from BC, while 1,071 patients (54.1%) had survived. Additionally, 404 (21%) patients had died because of other causes. The histological phenotype characteristics of all patients were described as invasive ductal carcinoma (IDC) at 84%, followed by invasive lobular carcinoma at 8%. The remaining patients were diagnosed with a special type of BC, including tubular, medullary and mucinous carcinoma. Moreover, PAM50 (218), a technique based on reverse transcription–quantitative polymerase chain reaction (RT-qPCR), was used to assess the intrinsic subtype distributions of BC in the METABRIC cohorts. Ten IntClusts were further classified according to gene expression, somatic aberration and integration germline variants. The clusters identified new heterogeneous gene expression landscapes for molecular subtypes classified by hormone receptor expression. More details about the clinical characteristics of the METABRIC cohort are listed in Table 2.1.

Table 2.1: Characteristic of the METABRIC cohort (n = 1980).

Parameter	Frequency (%)	Parameter	Frequency (%)
Age at diagnosis		Menopausal status	
≤50 years	424 (21)	Pre	436 (22)
>50 years	1556 (79)	Post	1533 (78)
Tumour size		Tumour grade	
≤2.0 cm	622 (32)	I	170 (9)
>2.0 cm	1338 (68)	II	770 (41)
		III	952 (50)
Nottingham prognostic index		Lymph node-stage	
Good prognostic group	680 (34)	I	987 (51)
Moderate prognostic group	1101 (56)	II	622 (32)
Poor prognostic group	199 (10)	III	316 (17)
Lymphovascular invasion		Oestrogen receptor	
Negative	930 (59)	Negative	474 (24)
Positive	635 (41)	Positive	1506 (76)
Progesterone receptor		Human epidermal growth factor receptor-2	
Negative	940 (48)	Negative	1733 (87)
Positive	1040 (52)	Positive	247 (13)
Triple-negative breast cancer		Follow-up status	
No	1660 (84)	Alive	1071 (54)
Yes	320 (16)	Died from BC	505 (26)
		Died from other causes	384 (19)
		Lost follow-up	20 (1)
Tumour type		PAM50	
Invasive ductal	1544 (80)	Normal-like	199 (10)
Invasive lobular	146 (8)	Luminal-A	718 (36)
Medullary	32 (2)	Luminal-B	488 (25)
Special type	206 (10)	HER2-enriched	240 (12)
		Basal-like	330 (17)

2.1.1.2 The Cancer Genome Atlas Breast Carcinoma

The Cancer Genome Atlas Breast Carcinoma (TCGA-BRCA) is restricted on female primary BC patients with available H&E slides (n = 854) cohort retrieved from the Genomic Data Commons Data Portal and cBioPortal was one of the genomic and transcriptomic datasets in which mRNA expression was comprehensively analysed in a large number of early stage BC cases (219-222). The TCGA-BRCA cohort contained 676 ER+/PR+ cases and 164 HER2 positive BC (223) cases; more details on the TCGA-BRCA cohort are provided in Table 2.2. Also, normalised gene expression of chosen biomarkers was retrieved from the UCSC Xena browser (<https://xenabrowser.net/datapages/>) to access the GDC TCGA BC RNA-sequencing dataset to determine the normalised expression of chosen biomarkers in normal BC tissue (n=113), primary tumour (n=1097) and metastatic samples (n=7). The RNA sequencing (RNA-seq) expression level read counts were normalised via the calculation of fragments per kilo base of transcript per million mapped reads upper quartile.

Table 2.2: Characteristic of the TCGA-BRCA cohort (n = 854).

Parameter	Frequency (%)	Parameter	Frequency (%)
Age at diagnosis		Tumour size	
≤50 years	231 (27)	≤2.0 cm	239 (28)
>50 years	607 (73)	>2.0 cm	615 (72)
Tumour grade		Lymph node-stage	
I	79 (21)	I	162 (19)
II	145 (38)	II	499 (58)
III	155 (41)	III	193 (23)
Lymphovascular invasion		Oestrogen receptor	
Negative	559 (65)	Negative	185 (22)
Positive	295 (35)	Positive	639 (78)
Progesterone receptor		Human epidermal growth factor receptor-2	
Negative	272 (33)	Negative	567 (81)
Positive	546 (67)	Positive	133 (19)
Triple-negative breast cancer		Follow-up status	
No	706 (86)	Alive	676 (79)
Yes	119 (14)	Died from BC	88 (10)
		Lost follow-up	90 (11)
PAM50		Tumour type	
Normal-like	209 (17)	No special types (NST)	600 (70)
Luminal-A	415 (48)	lobular	177 (20)
Luminal-B	141 (15)	NST& Medullary	49 (6)
HER2-enriched	133 (14)	NST & good prognostic special type	28 (4)
Basal-like	56 (6)		

2.1.1.3 External Validation of the Data with Online Dataset

2.1.1.3.1 Breast Cancer Gene Expression Miner

External validation was performed using Breast Cancer Gene-Expression Miner (bc-GenExMiner v4.9; <http://bcgenex.ico.unicancer.fr/BC-GEM/GEM-Accueil.php?js=1>) to examine mRNA expression with the exemption of the METABRIC and TCGA-BRCA datasets (224). It contained three modules, that is, a correlation module, a prognostic module and an expression module. The prognostic effects of a gene on the patient outcome being tested were assessed with the prognostic module. The outcomes analysis for some of the markers in this thesis was conducted using Affymetrix (n = 4,904), which is a platform for both gene expression analysis (mRNA) and genetic analysis (DNA) using GeneChip technology (225).

2.1.1.3.2 Tumour, Normal, and Metastatic Tissues and the University of Alabama at Birmingham Cancer Data Analysis Portal

The Tumour, Normal, and Metastatic tissues (TNM) (<https://www.tnmplot.com/>) and University of Alabama at Birmingham Cancer data analysis portal (UALCAN) datasets (<http://ualcan.path.uab.edu/index.html>) together contained 1,097 primary cancers, 7 metastatic cancers and 113 normal tissue samples (226-228).

2.1.2 Proteomic Level

2.1.2.1 Nottingham Invasive Breast Cancer Primary Series

This study was conducted on a large, well-characterised cohort (n = 2,497) of primary IBC cases presented to Nottingham City Hospital between 1987 and 1998 with available clinicopathological data (Table 2.3). A tissue microarray (TMA) using formalin-fixed paraffin-embedded (FFPE) wax blocks was conducted on all biopsy samples of the cohort. A FFPE paraffin wax block contained 15 cores representing patients' specimens with 0.6-mm tissue thickness and was used to examine a variety of biological markers (229). Moreover, to avoid using samples with heterogeneous reactions, a portion of the FFPE full-face tissue sections that included the morphological characteristic of IBC was used. Each sample cohort was made available along with the most important clinicopathological indicators, including the tumour size and histological grade, nodal stage, nodal status, LVI, NPI, molecular and histological subtypes, and distant metastasis. Treatment and outcome data included the duration, in months, from primary surgical treatment to the time of death due to BC, BCSS, the duration (in months) from primary surgical treatment to the occurrence of first distant metastasis, DMFS, and regional and local metastasis were also reported (230).

Table 2.3: Characteristic of the Nottingham IBC primary series (n = 2497).

Parameter	Frequency (%)	Parameter	Frequency (%)
Age at diagnosis		Menopausal status	
≤50 years	669 (34)	Pre	916 (37)
>50 years	1285 (66)	Post	1545 (63)
Tumour size		Tumour grade	
≤2.0 cm	1009 (52)	I	440 (20)
>2.0 cm	932 (48)	II	757 (34)
		III	1017 (46)
Nottingham prognostic index		Lymph node-stage	
Good prognostic group	789 (36)	I	1502 (66)
Moderate prognostic group	1072 (49)	II	595 (26)
Poor prognostic group	332 (15)	III	177 (8)
Lymphovascular invasion (haematoxylin and eosin: Podoplanin/CD34)		Oestrogen receptor	
Negative	1327 (69)	Negative	549 (26)
Positive	602 (31)	Positive	1601 (74)
Progesterone receptor		Human epidermal growth factor receptor-2	
Negative	761 (41)	Negative	1653 (88)
Positive	1111 (59)	Positive	231 (12)
Triple-negative breast cancer		Follow-up status	
No	1719 (84)	Alive	1356 (57)
Yes	320 (16)	Died from BC	698 (29)
		Died from other causes	314 (13)
		Lost follow-up	35 (1)
Distant metastasis		Recurrence	
No	1682 (69)	No	1483 (62)
Yes	770 (31)	Yes	926 (38)
Tumour type			
Invasive ductal/Not Special Type	1559 (80)	Special type, Tubular/Mucinous	113 (6)
Invasive lobular	215 (11)	Miscellaneous	10 (1)
Medullary-like	44 (2)		

2.2 Optimisation of Antibodies Used in the Study

2.2.1 Cell Lines and Culture Media

Various human cell lysates of the BC cell lines listed in (Table 2.4) were chosen for evaluating and confirming the specificity of biomarkers used in this study. These included TNBC MDA-MB-231 (epithelial cells obtained from the breast mammary gland were established from a pleural effusion of a 51-year-old Caucasian female with metastatic mammary adenocarcinoma) (231), MCF7 (derived from the pleural effusion of a 69-year-old Caucasian with BC [adenocarcinoma]) and SKBR3 (obtained from a patient who underwent radiation). Then, HeLa epithelial cells derived from a patient with human cervical carcinoma (232), BT-474, were obtained from a 60-year-old woman who suffered from advance-stage IDC of the breast (233). The American Type Culture Collection (Rockville, MD, USA) was the source of all of the cell lines used. All cells were incubated at 37°C in a controlled humidified atmosphere of 5% CO₂.

Mycoplasma contamination tests were conducted regularly using the MycoAlert Detection Kit (CUL001B, R&D Systems, UK), and short tandem repeat profiling was used to determine the integrity of all cell lines (Source BioSciences, UK).

Table 2.4: Cell lines used in this study.

Cell line type and origin		Type of media
TNBC (ER-, PR+/-, HER2-)	MDA-MB-231	RPMI + 10% FBS
Luminal A (ER+, PR+, HER2- / low Ki67)	MCF7	RPMI + 10% FBS
HER2-enriched (ER-, PR-, HER2+)	SKBR3	McCoy's 5A + 10% FBS
Cervical carcinoma cell line	HeLa	DMEM + 10% FBS
Luminal B (ER+, PR+, HER2+)	BT-474	DMEM + 10% FBS
• FBS: fetal bovine serum (F9665, Sigma, Brazil). • RPMI: Roswell park memorial institute medium (RPMI-1640, Cytiva, UK). • McCoy's: M8403, Sigma, UK. • DMEM: Dulbecco's modified eagle's medium (D5796, Sigma, UK).		

2.2.2 Cell Passage

2.2.2.1 Procedure

All cell lines were observed visually by phase-contrast microscope every 2–4 days to check viability, and the media was replaced every three days. Once the confluence of the cells was between 70% and 80%, the consumed cell culture medium was discarded from the culture flask. Next, Dulbecco's phosphate-buffered saline (DPBS) was used at a measurement of approximately 2 ml/10 cm² from the culture surface area to wash the cells.

As a harvesting solution, 5 ml of room-temperature trypsin dissociation reagent (0.25% phenol red, 100 ml: 15050065; Thermo Fisher Scientific) was added to cover a cell layer, which was then incubated for approximately 3–5 minutes at 37°C or as required to detach a cell line from the flask. Once the cells detached, a room-temperature complete growth medium was added under a sterilised environment to the flasks to neutralise the trypsin solution. Afterwards, the cells were moved to sterile tubes and centrifuged for 5–10 minutes at a rate of 1,000 rpm. Finally, the cell pellet was re-suspended in a small amount of preheated complete growth medium, with a small sample of 10 µl taken for counting.

To determine the total number of cells and the percentage of viable cells, the following equation was used: number of cells/ml = mean cell count × dilution factor (10) × 10⁴. The cell suspensions were diluted under the seeding density required for each cell line. Then, the desired and appropriate amounts were pipetted into new cell culture flasks containing a proper amount of fresh growth media and cultured at 37°C in an adjusted humidity atmosphere at 5% CO₂ in an incubator for further use or passaging.

2.2.3 Preparation of Cell Lysates

To ensure no disturbance to the adherent cells, the cell culture medium was detached from the flasks carefully. The adherent cells were washed in cold DPBS. The cells were covered using 2 ml of trypsin and incubated for five minutes at 37°C. As previously mentioned, the cells were re-suspended and moved to sterile tubes, before being centrifuged for 5–10 minutes at 1,000 rpm. The supernatant was then disposed of, and lysis of the pellets containing cells was determined by including a radioimmunoprecipitation assay buffer (25 mM Tris, 150 mM NaCl, 1% Nonidet P-40, 1% sodium deoxycholate and 0.1% sodium dodecyl sulphate [SDS]) to a protease inhibitor and phosphate mixture (Pierce, Thermo Fisher Scientific, UK). The cell lysates were transferred to a microfuge tube and incubated for 30 minutes on ice, with the incubator shaken approximately every two minutes. To ensure all cell debris was eliminated,

the cell lysates underwent centrifugation for 20 minutes at 13,000 rpm at a temperature of 4°C. The supernatant layer was stored away at a temperature of – 20°C until required.

2.2.4 Cryopreservation of Cell Lines

Cells were harvested, as mentioned previously, and the cryovials were labelled. Once a percentage of 70% or more cell viability was obtained, the cells were equipped for freezing. The freezing medium was then prepared (appropriate medium + 10% dimethyl sulfoxide; 276855, Sigma-Aldrich).

After the medium was prepared, the supernatant layer was removed to enable the freezing medium to be added at the needed cell density of 1×10^6 cells per ml. Once this was complete, 1 ml of aliquots was moved into the cryovials and stored in a freezer at a temperature of – 80°C for around 24 h. These cryovials were then transferred to a liquid nitrogen container to be stored long term. Next, cell thawing was conducted by placing the cells into a 37°C water bath; after that, complete fresh media was added to a 15-ml tube. Then, the tube was centrifuged at 1,000 rpm for five minutes, and the pellet was re-suspended in the fresh complete media, transferred to a flask and incubated at 37°C.

2.3 Antibody Specificity and Cross-Reactivity Assay: Western Blot

To evaluate the antibody specificity, western blotting (WB) was done. This technique is based on separating the protein mixtures according to their molecular weights (MW) under electrical current in acrylamide gel and then transferring these separated proteins onto a nitrocellulose membrane, as described by Towbin et al. (234). Detailed information about the antibodies, including their manufacturers, sources, clones, concentrations and expected MWs, is presented in Table 2.5.

Table 2.5: Western blotting details of antibodies used in this study

Antibody	Supplier	Species	Clone	Western blotting		Expected molecular weight as kDa
				Concentration	Incubation time	
HSP90AA1	Abcam	Mouse monoclonal	AC88	1:700	Overnight	90
SLC9A3R1	Abcam	Rabbit monoclonal	EPR5562	1:500	Overnight	48
NOP56	Sigma-Aldrich	Mouse monoclonal	CL2603	1:3000	Overnight	66
HMGB3	Sigma-Aldrich	Rabbit polyclonal	HPA062583	1:1000	Overnight	22

2.3.1 Protein Quantifications

Each sample was tested for protein quantity equivalence using a Thermo Fisher Scientific PierceTM Bicinchoninic Acid Assay (BCA) Protein Assay Kit (catalogue number: 23225, Invitrogen, UK). To prepare the samples, distilled water (DW) was used as a calibrator and solvent solution and to dilute the samples at a ratio of 1 part DW (16 μ l) to 8 parts of the sample (4 μ l). The working reagent BCA contained two reagents (A and B) and was performed with a dilution ratio of 50 (40 μ l) parts reagent A to 1 (1,960 μ l) part reagent B. A 96-well plate (Invitrogen, Thermo Fisher Scientific, UK) was used, and 8 μ l of the diluted samples and DW was added to each well in duplicate. Then, the working solution was added to each well at 200 μ l. To prevent exposure to light, the plate was covered with foil and incubated for 30 minutes at 37°C. To measure the absorbance, an Infinite F50 spectrophotometer (Tecan, Switzerland) was used at a 570-nm wavelength.

2.3.2 Sample Preparation

All cell lysates were thawed on ice and prepared with sample preparation reagents in chemical-resistant polypropylene microfuge tubes (Table 2.6). The samples were spun gently using a vortex mixer before being digested on a dry heating block at 100°C for five minutes and then kept on ice until use.

Table 2.6: Sample components

Cell lysate and reagents	Volume
Cell lysate	13 μ l
NOVEX Bolt TM LDS Sample Buffer	5 μ l
NuPAGE Reducing Agent (10x)	2 μ l
Total volume	20 μl

2.3.3 Sample Loading and Protein Separation (Electrophoresis)

The Mini Gel Tank (Thermo Fisher Scientific, UK) was used to run the gels with a previously prepared running buffer (RB) solution that contained 3-(N-morpholino) propanesulfonic acid SDS Running Buffer (20×) (Invitrogen by Life Technologies, UK) and DW at a dilution of 1:10 (50 ml RB to 950 ml DW). Bolt™ Anti-Oxidant (to protect the protein from possible oxidation; Invitrogen by Life Technologies) was added to the RB.

A precast gel, Bolt 3–8% Tris-Acetate Gel, 1.0 mm × 12 well (Invitrogen by Thermo Fisher Scientific, UK), was inserted into the tank. The gel helps in denatured and reduced protein separations. The gel wells were loaded with a protein ladder 10–180 kDa (Thermo Fisher Scientific, Baltics, UAB) (7 µl) at the first well and lysates (13 µl each) in the following wells. After the lid was placed on the tank, the electrode cords were connected to the electrical power supply (PowerEase 90W PS0091; Invitrogen, Life Technologies, UK). The power supply was set to run for 50 minutes at 180 volts or for 60 minutes at 150 volts.

2.3.4 Protein Transferring

The blot transfer module (Thermo Fisher Scientific, USA) was prepared by making a sandwich with sponges (Invitrogen Sponge Pad, Thermo Fisher Scientific, USA) and filter papers (Hybond Blotting Paper, GE Healthcare, UK) in between the gel and the nitrocellulose membrane. A cylindrical roller was passed over the layered gel in filter paper to eliminate air bubbles. Sponges were added to maintain the pressure of the gel against the membrane. The sponges, filter paper and membrane were pre-soaked with transfer buffer (TB) (50 ml NuPAGE Blot 20× TB (Thermo Fisher Scientific, USA), 100 ml of 10% methanol (Sigma Aldrich, USA), 849 ml deionised water (dH₂O) and 1 ml NuPAGE Bolt Antioxidant. Transfer of the protein was done at 10 volts for one hour. Ponceau stain (catalogue number: 3761-53-3, Sigma Aldrich, UK) was used to evaluate the efficiency of the transfer before the blocking step.

2.3.5 Membrane Blocking

To reduce non-specific binding of the primary antibody, the membrane was blocked using 5% milk (Marvel Original Dried Skimmed Milk, Premier Food Groups Ltd, UK) in DPBS and 0.1% Tween (DPBS Tween). A rocker (Stuart See-Saw Rocker, Bibby Scientific Limited, UK) was used at 30 oscillations/min for 60 minutes at room temperature to ensure complete blocking.

2.3.6 Immunoblotting and Protein Reaction Detection

After blocking, the membranes were incubated with primary antibody diluted in the blocking solution overnight at 4°C or for one hour according to the manufacturer's instructions or by moderating the recommended dilutions to acquire ideal results (a single band with a clean background on the expected MW; for example, in the case of recommended 1:2,000 dilutions, 1:1,500, 1:2,500 and 1:3,000 were tried). Each antibody that was used in the study for the different BC cell lysates had its own binding specificity. This specificity was measured qualitatively (Table 2.5). Beta-actin (Anti-beta-actin antibody-Loading Control was used as either anti-beta-actin antibodies were produced in mouse (catalogue number: A5441, Sigma-Aldrich: Clone AC-15; Sigma, UK) or rabbit (catalogue number: MABT523; Sigma-Aldrich; Clone RM112; Sigma, UK). The anti-beta-actin was diluted in a blocking solution (milk) with a dilution ratio of 1:5,000. The membranes were washed for 10 × 3 minutes with Thermo Scientific Pierce DPBS Tween with continuous rocking. Then, they were incubated with fluorescent horseradish peroxidase conjugated secondary antibody IRDye® 680RD donkey anti-mouse immunoglobulin G (IgG) (LI-COR, Bioscience, Lincoln, NE, USA) and IRDye® 800CW donkey anti-rabbit IgG (LI-COR, Bioscience, Lincoln, NE, USA) antibodies in order to detect housekeeping protein and AB. The dilution was prepared as 1/15,000 fluorescent secondary antibodies in a blocking solution (5% Marvel milk). To avoid direct exposure to light, the membrane was covered with foil and then incubated for 60 minutes on a continuous rocker. Again, the membranes were washed with DPBS Tween for 10 × 3 minutes.

2.3.7 Protein Reaction Visualisation (the Bands)

The images were developed via chemiluminescence using an Odyssey Fc (LI-COR Biosciences, Lincoln, NE, USA). The final image showed red and green fluorescent bands visualised via the Image Studio™ software (Java, UK). To obtain images of the detected protein band, the IRDye® 680RD and IRDye® 800CW secondary antibodies with wavelengths of 694 nm and 795 nm, respectively, were used. A secure hard drive was used to store all labelled images.

2.4 Antibody Optimisation Assay: Immunohistochemistry

Since its first description in the 1940s, IHC has become the most robust technique examining BC and easing the prediction of targeted treatment applications. Immunohistochemistry is involved in evaluating molecular subtyping, assessing protein morphological patterns in tumour tissue and identifying prognostic biomarkers, which adds more reliability to pathology reports in many oncogenic diseases, including IBC (235).

Between 10 and 15 full-face sections, randomly representing different histological and molecular subtypes, were stained with IHC for (HSP90AA1, SLC9A3R1 and NOP56), which enabled the exploration of the heterogeneity of the indicated target before the TMA staining using 4- μ m-thick TMA sections of IBC. Positive and negative controls were used for each primary antibody as a confirmation step of antibody specificity. A positive IHC reactivity marker (the rabbit polyclonal β 2-microglobulin marker [1:2,000; Dako, USA]), with a dilution ratio of 1:1,000 to antibody diluent (SLCB3189; Leica Bond, UK) was used as a positive control. However, the antibody diluent was added without the antibody to represent the negative control. Slide sections were incubated at the temperature and time recommended by the manufacturer (Table 2.7).

Table 2.7: Immunohistochemistry evaluation of the antibodies used in the study

Antibody	Immunohistochemistry dilution and incubation	Pre-treatment	Expression localisation	Control tissue
HSP90AA1	1:200 overnight	Microwave, citrate pH 6, 20 minutes	Cytoplasm	Kidney
SLC9A3R1	1:250 1 hour room temperature	Microwave, citrate pH 6, 20 minutes	Cytoplasm	Kidney
NOP56	1:250 overnight	Microwave, citrate pH 6, 20 minutes	Cytoplasm	Kidney
pH potential hydrogen.				

2.4.1 Staining Method

2.4.1.1 Pre- Staining Tissue Section Slides Preparation

The labelled tissue section slides were placed on a hotplate (Leica HI 1220, Germany) at 60°C for 10 minutes to remove the wax before being cooled for 5 minutes at room temperature. Next, the slides were dewaxed/rehydrated by immersion in sequential xylene baths ($\times 2$, five minutes) (Fisher Scientific, UK) and, subsequently, by immersion in 100% and 70% industrial methylated spirit (IMS) ethanol (ethyl alcohol) ($\times 2$, three minutes) (Sigma-Aldrich, Poland) baths, followed by a wash ($\times 3$, five minutes) in dH₂O wash. The above steps were performed by an ST5010 Autostainer XL (Leica Biosystems, Germany).

2.4.1.2 Heat-Induced Epitope Retrieval: Citrate Buffer (pH = 6.00)

Due to their nature and storage time, surgical biopsies preserved in formalin or aldehyde and embedded in paraffin might undergo significant physical and chemical changes. Having poor epitope exposure may adversely affect the expression strength. Thus, multiple heat-induced epitope retrieval (HIER) procedures were conducted under different conditions to stimulate the FFPE tissue's antigenic sites. Among the HIER systems, the citrate buffer antigen retrieval procedure provided the best outcomes in stimulating the antigenic site of the FFPE in the IBC tissues.

To prepare the 10× sodium citrate buffers, 2.94 g of trisodium citrate was dissolved in 1,000 ml of DW, and the potential of hydrogen (pH) was adjusted to pH 6.0. Heat-mediated/ citrate antigen retrieval was performed using 1.0 M sodium citrate buffer (one part 10× citrate buffer solution into nine parts DW) at pH 6.0 (Russel RL060P Portable pH Meter, ThermoElect Co., Colorado, USA) and microwaved (Whirlpool, Jetstream, JT359) at 1,000 W (240°C) for 20 minutes. Later, it was slowly cooled under tap water for five minutes.

For staining full-face and TMA sections, sequential and flatbed methods were used, respectively. Water-resistant boundaries (ImmEdge Hydrophobic Pen, Vector Laboratories, Canada) were created in a suitable manner surrounding the tissue sector, which enabled the reagents to be confined within the TMA cores. For the sequential method and after antigen retrieval, Tris-buffered saline (TBS) with a pH of 7.6 was used to fill the sample slides to ensure a humid environment that helped preserve the tissue from dryness and damage. Tris-buffered saline Tween (TBS-T) (Trizma base, Sigma-Aldrich; NaCl and HCl diluted at 1:10 with dH₂O, pH 7.6) with 0.1% Tween (Tween 20, Dako, Denmark) was used in the TMA cores (×2, five minutes) after antigen retrieval and (×3, five minutes) for all other followed washing steps. The washing steps were beneficial for removing all previous reagents and ensuring reactions had stopped.

2.4.1.3 Staining With Novolink Max Polymer Detection System

The NovolinkTM Max Polymer Detection System (RE7280-K, Leica Biosystems, UK) was used to stain the slides with mouse and rabbit IgG (Table 2.8). The quantification of the specific binding percentage of the primary antibody to the targeted biomarker was measured by the development of brown colouration of the chromogenic 3,3' diaminobenzidine (DAB) against the counterstaining haematoxylin, which gave a blue background for the non-stained untargeted cellular component.

Table 2.8: The contents of Novolink™ max polymer detection system kit (236).

Reagent	Composition
Peroxidase block	3–4% (v/v) Hydrogen peroxide
Protein block	0.4% Casein in DPBS, with stabilisers, surfactant, and 0.2% Bronidox L as a
Post primary	Rabbit anti-mouse IgG (<10 jig/ml) in 10% (v/v) animal serum in TBS/0.09% ProClin™ 950
Novolink™ polymer	Anti-rabbit Poly-HRP-IgG (<25jig/ml) containing 10% (v/v) animal serum in TBS /0.09% ProClin™
DAB chromogen	1.74% w/v 3,3' - diaminobenzidine, in a stabilizer
DAB substrate Buffer	A buffered solution containing ≤0.1% hydrogen peroxide and preservative
Haematoxylin	<0.1% Haematoxylin

2.4.1.3.1 Endogenous Blocking Peroxidase and Protein Block

Endogenous peroxidase activity that remained active within the formalin-fixed tissue was blocked by a peroxidase block for five minutes, followed by a wash with TBS-T. This step ensured the quality of intensity of the colour produced by the antigen–antibody reactions. Non-specific binding of the primary antibody (non-specific band) was reduced by adding Novolink Protein Block for five minutes, which enhanced the quality reflection for the specific reaction between the antibody and the antigen, followed by a wash with TBS-T.

Negative and positive procedures were used with different staining for the various antibodies. There was no primary antibody encountered in the negative control. Therefore, tissue samples were incubated in antibody diluent only. Each antibody had a known reactivity as described in the manufacture date sheet, enabling it to react with positive control tissue when incubated together.

2.4.1.3.2 Antibody Optimisation

All antibodies in this study were stained following the manufacturer's instructions, including those for the antibody dilution concentration and incubation time. Thus, according to the manufacturer supplier's recommendation of antibody optimisation, the concentration was applied with four various dilutions, two higher and two lower (e.g. dilutions of 1:300, 1:400, 1:500 and 1:600), which were also used in addition to the recommended concentration of 1:500. This procedure ensured that the best antibody concentration was met and the antibody was selected according to their specificity, where the most specific value of the antibody is achieved when the targeted cellular component is stained without affecting immune cells and the background. The slides were incubated with the primary antibody diluted in protein stabiliser and surfactant (Bond Primary Antibody Diluent, AR9352, Leica Biosystems, UK) for 30 minutes, one hour or overnight as per the manufacturer's instructions and adjusted in accordance with the staining results (Table 2.7).

2.4.1.3.3 Post-Primary Antibody and Novolink™ Polymer

The slides were incubated with post-primary for 30 minutes, which is used to detect mouse antibodies, followed by washing with TBS-T. Then Novolink™ Polymer was added after that to detect both the post-primary and any tissue-bound rabbit antibodies. The incubation time was set for 30 minutes, and then the slides were washed with TBS-T.

2.4.1.3.4 3,3' Diaminobenzidine

A DAB working solution was prepared by the 1:20 dilution Novolink™ DAB Chromogen and a Novolink™ DAB Substrate Buffer and applied for five minutes; then, the slides were washed with TBS-T. The solution was always freshly made, and due to its sensitivity to light, it was preserved in the dark during incubation until use. The chromogen reacted with the enzymes, forming a visible brown precipitate in the presence of water, indicating the primary Ab expression sites.

2.4.1.3.5 Haematoxylin Counter Stain

The slides were counterstained with haematoxylin (less than 1% concentration) for six minutes. The haematoxylin stained the nuclei and aided in the IHC staining protein assessment. A plastic container filled with water was used to remove any excess haematoxylin counterstain from the slides.

2.4.1.4 Tissue Dehydration

The slides were dehydrated with a gradual concentration of IMS (3 × 2 min), followed by clearing with xylene twice for five minutes using Leica Autostainer XL with program 2. In the end, the slides were mounted with dibutylphthalate polystyrene xylene (BDH, Leica Microsystems, UK), coverslipped (Fisherbrand Coverslips, 22 mm × 40 mm, UK) and dried for 24 h at room temperature.

2.5 Immunoreactivity Evaluation before Scoring

The expression of proteins in normal breast tissue sections and the level of heterogeneity in tumour tissue were analysed using full-face breast tissue sections taken from 10–15 different patient cases representing various histological grades, stages and molecular subtypes; prior to the TMA staining, IHC staining was done on these to manage and minimise the number of consumed TMAs. All biomarkers in this study were selected based on their homogeneous distribution throughout all slides and were used to stain patients' TMA sections later on.

2.6 Slide Imaging

Slides were converted into high-resolution images through digital scanning (NanoZoomer, Hamamatsu Photonics, Queens Medical Centre, UK). Images were observed (Aperio ImageScope, Aperio Technologies) at 20× magnification. This helped to generate a library for the slides to be available for editing and sharing. This was also very helpful when the quality and reflection resolution of IHC or H&E staining slides were affected by long storage duration, shrinking of tissue caused by light or any loss of the histological architecture within FFPE specimens.

2.7 Scoring of Tissue Microarray Cores

Only the cores representing $\geq 15\%$ of the invasive tumour were scored, excluding carcinoma in situ, breast stroma, inflammatory infiltrates, and normal acini and ducts. The percentage of the malignant cells that were stained (0–100) and the intensity of staining (0: negative, 1: weak, 2: moderate and 3: strong) were considered in terms of their semi-quantitative evaluation H-score. The formula used for calculating the final H-score was as follows: $(0 \times \% \text{ stained negative}) + (1 \times \% \text{ stained 'weakly'}) + (2 \times \% \text{ stained moderately}) + (3 \times \% \text{ stained strongly})$. Although only one observer carried out all scoring, a more experienced external observer (pathologist) performed scoring of a subset of 10%–30% of cases for added reliability.

2.8 In-House Biomarker Consortium Data

A wide spectrum of biological biomarkers with clinical relevance to BC progression was used to examine their interaction with a chosen set of biomarkers discovered in this study and their relation with LVI incidence and disease progression. These biomarkers play an essential role in BC, such as EGFR, phosphoinositide-3-kinase, catalytic, alpha polypeptide (PIK3CA), AKT, Ki67 and p53. EGFR regulates EMT and cancer invasion through its downstream pathway (237). The high expression of EGFR is considered an independent predictor of poor outcomes in IBC (237). PI3KCA mutation occurs in about 20–30% of BC, which enhances cancer growth and decreases patient survival (238, 239). AKT is an essential component of the PI3K/AKT/mTOR signalling pathway, as it regulates cell development, proliferation and survival. This pathway is frequently dysregulated in BC (240). In BC, the rate of tumour cell proliferation is one of the most significant prognostic markers, and Ki67 is consequently useful for prognosis (241). P53 mutation is linked with more aggressive malignancy and poorer overall prognosis in BC (242). In addition, the p53 mutation occurs in 70–80% of cases of the most aggressive subtype of BC, the TNBC (243).

These data are illustrated in Table 2.9. For the dichotomisation of biomarkers, several methods were used to determine the cut-off value. The median or mean was used depending on the distribution of the data.

Table 2.9: Available biological biomarkers from the Nottingham cohort that were used in this study.

Proteins marker	Expression location	Cut-off point	Reference
EGFR	membrane	100	(244)
PIK3CA	cytoplasm	100	(245)
AKT	cytoplasm	65	(246)
Ki67	cytoplasm	14	(247)
P53	nucleus	10	(248)

2.9 Cut-Off Point Determination

The cut-off points for each biomarker were determined individually. According to the patient follow-up time, output events (both uncensored and censored) were used for this sub-classification. In the case of patients who developed distant metastasis (DM) or died of IBC, the results were considered negative and uncensored. Censored patients either had not developed DM or their data were forfeited during follow-up and they were alive. The cut-off point was recognised as the median contained most patients, which presented as the peak pixel in Statistical Package for Social Sciences (SPSS) version 24 (SPSS Inc., Chicago, IL, USA). To determine the statistical significance, these cut-off points were used in conjunction with the SPSS software, in which *p*-values were calculated and relative against DMFS. The localisations and cut-off points for the biomarkers are presented in Table 2.10.

Table 2.10: Antibodies Cut-off point and localisation

Biomarker	Expression Localisation	Cut-off point
HSP90AA1	Cytoplasm	H-score 50 using the median, with a score >50 was considered high expression.
SLC9A3R1	Mainly in Cytoplasm	H-score 0 using the median, with a score > 0 was considered high expression.
NOP56	Mainly in Cytoplasm	H-score 35 H-score using the median, with a score >35 was considered high expression.

2.10 Differential Gene Expression Analysis and Gene Selection for Study

2.10.1 Differential Gene Expression

A DGE experiment is a statistically quantifying measurement of possible modifications in a gene in terms of reading counts or expression levels performed between two experimental conditions. Converting the information in the gene into a functional gene product, such as a protein with microarray technology, offers robust data for prognostic candidates that could ultimately contribute to BC studies which shed light on a new approach in BC personalised treatment. In addition, such a technique improves patient survival and outcomes, consequently helping to enhance clinical management. Based on tissue availability and integrity, the cases involved in the analysis were used in two cohorts: Nottingham (*n* = 285/1,980 cases) and Addenbrooke's Hospital (*n* = 1,280/1,980 cases). The critical determination of the statistical distribution properties of mRNA LVI status formed part of the foundation of this study. The LVI status of the subgroup was assessed via IHC, utilising CD31, CD34 and D2-40 (115). This enabled the end LVI status to be determined using IHC along with several H&E tumour

sections. Of the 1,565 samples in the mRNA microarrays, 930 were LVI–, whereas the remainder were LVI+ based on a linear model of DGE.

As part of the TCGA-BRCA cohort, the availability of LVI information also became a critical consideration when analysing the RNA sequence data. Therefore, identifying gene candidates with already given prognoses using the method mentioned above was beneficial. The method used to identify DGEs between both LVI positive and negative cases was the weighted average difference (WAD) method. The DGEs were nominated with respect to their determined WAD ranking (249, 250). In addition to the WAD method, the Edge-R software was also considered in the DGE analysis. Fold changes ($> \pm 2$) were utilised to identify DGEs for filtering, using $p < 0.05$ as the maximum significance threshold. In the WAD assay results, there were 350 genes associated with LVI status in both the Nottingham cases ($n = 285$) and the remaining METABRIC cases ($n = 1,280$), as summarised in **Appendices 10.4** and **10.5**, respectively. There was an overlap between the DGEs of both groups, and they overlapped genes have been made available (note that the set that is associated with LVI is in **Appendix 10.6**). More details about the DGEs will be presented in **Chapter 3**.

2.10.2 Gene Selection for Study

Several approaches were followed for gene evaluation and selection for this study. First, the markers (HSP90AA1, SLC9A3R1, NOP56 and HMGB3) were chosen according to their strong correlation with LVI status that resulted from DGE analysis and validated by various statistical analyses. While All of the selected genes were obtained by DGE analysis. However, one of these genes was subsequently pursued from a previous mass spectrometry study by Nottingham BC group member (Dr Sultan Sonbul) and validated by DGE. For example, to verify the association between the targeted gene from the previously mentioned analysis sources and the LVI status, the t-test method (chi-squared) was utilised. In addition, the univariate method (Mann–Whitney test) was also used to confirm the gene expression with LVI status. Nonetheless, chosen genes were individually examined to expand understanding of their molecular function and biological properties.

The positive association of LVI and these genes and the correlation with clinicopathological parameters and patient outcomes at the transcriptomic level using METABRIC and TCGA-BRCA datasets were considered as a basis of the gene selection. The online dataset bc-GenExMiner v4.9 (Affymetrix dataset) was also utilised to verify the correlation of each gene with patient survival. Pearson's correlation was used to validate the association of identified

oncogenic and BC metastasis-related pathways and proteins, such as p53, Ki67, AKT, PIK3CA, MYC and EGFR, with the targeted genes to investigate any relationship between them and BC LVI. In conclusion, the genes used in this study were obtained from various resources, including DGEs and mass spectrometry (*HSP90AA1*), and the 99 genes in the LVI-associated gene signature (*SLC9A3R1*, *NOP56* and *HMGB3*).

2.11 Statistical Analysis

2.11.1 Introduction

A number of different statistical methods were used to determine and test the significance of the results obtained. Other nominal variables, such as ordinal and categorical, independent or dependent, and continuous scale data, epitomised this variety. The statistical software package SPSS was used to perform the analyses (IBM SPSS Statistics, versions 24–28, SPSS Inc., Chicago, IL, USA). A value of $p < 0.05$ (two-tailed) was set as the optimal level of significance for all statistical tests. A histogram was produced for normality testing of the staining distribution to determine the biomarkers' dichotomising cut-offs. Independent analyses were carried out regarding cytoplasmic H-scores.

2.11.2 Intra-Class Correlation Coefficient

A pathologist re-evaluated 10-30% of the whole cohort that the author previously evaluated to confirm the reliability of the H-scoring data. This was conducted using an intra-class correlation coefficient (ICC) test. The pathologist and the author independently assessed sections, ensuring the individual scores were assessed blindly. A sufficient correlation of the ICC was evaluated if the H-score threshold agreement was equal to or greater than 75%. Visualisation differences sometimes put the ICC percentage less than 75%; in this case, and must rescored. If any discrepancies occurred between the TMA scores (the correlation level was less than <50), each slide would be re-assessed and scored, with each difference being evaluated until a better level of agreement was reached. To ensure a high level of quality control for the H-scoring level, no patient data were available throughout all scoring sessions.

2.11.3 Chi-Squared Test (Goodness of Fit)

The chi-squared non-parametric statistical test was used to enable statistical relationships to be shown and identified between numerous nominal independent variables. The reasoning behind using the chi-squared test was to enable a thorough mathematical evaluation of the differences in rates regarding distribution for categorised and selected biomarkers. The assessment was against the verified and valid equivalent data, which belonged to other nominal variables, such

as tumour size, tumour grade, patients' age, LVI, chemotherapy, LN stage or other stained biomarkers that had previously been noted.

2.11.4 Kaplan–Meier Survival Analysis and Log-Rank Test

The log-rank test was used to evaluate the association between biomarker expression and patient survival. The analysis depended on the distribution of two populations of BC patients using the median expression of examined biomarkers to ensure censored observations. Kaplan–Meier curves were used to plot the patients' associated evaluation with a log-rank test for significance.

2.11.5 Proportional Hazards Model (Cox Regression Model)

The Cox proportional hazards model was robust to evaluate standardised prognostic factors against the prognostic independence of both proteins and differentially expressed genes (DEGs). The ability of this test to measure the prognostic or predictive independence of encountered genes from common prognostic/predictive factors, such as tumour grade, Ki67 expression, tumour size and LN stage, stands out the value of this test.

2.11.6 Student's T-Test

Using the student's t-test, an evaluation of the statistical difference between the two normally distributed variables was achieved.

2.11.7 Mann–Whitney U Test (Median Analysis)

The Wilcoxon–Mann–Whitney non-parametric test is an additional test that was conducted on selected biomarkers to undertake a univariate analysis of those biomarkers. With this test being conducted, statistical relations were identified and shown between the distributed median and the nominal data of the continuous variables.

2.11.8 Spearman's Rank Test (Model/Line of Best Fit)

Using the Spearman's rank correlation coefficient, the relation of strength within the linear association between two non-normalised continuous variables, such as protein/gene expressions, was determined. The range of R-values went from -1 to $+1$, with -1 being a strong inverse correlation, 0 indicating no relation or association between the two variables and $+1$ showing a solid relation/association in the same direction. Other values which fell within that range also showed different strengths in relation/association; for example, values from -0.1 to 0.1 were regarded to not correlate; those from 0.1 to 0.3 or -0.1 to -0.3 were considered to show a weak correlation; those from 0.3 to 0.5 or -0.5 to -0.3 were moderate; and finally, values from 0.5 to 1 or -1 to -0.5 were regarded to have a strong correlation.

3 Chapter 3: Bioinformatics approach and differential gene expression

This chapter comprises data published in the *British Journal of Cancer* (2019) under the title ‘A key genomic subtype with lymphovascular invasion in invasive breast cancer’ with the corresponding DOI: [10.1038/s41416-019-0486-6](https://doi.org/10.1038/s41416-019-0486-6) (223). The copyright approval for re-use of these material is provided in **Appendix 10.2**.

3.1 Overview

As an introduction to this research, BC is a highly malignant, heterogeneous disease with differences across patients in terms of staging, grading and the severity of the tumour which are caused by inter-tumour heterogeneity. Early stage BC in individuals is characterised by morphological and genomic, transcriptomic and proteomic expression complexity, and this has been validated (251). Due to the continuous molecular alteration in patients with BC, it is imperative to refine the present molecular categorisation systems and develop tailored therapy approaches for patient subgroups more susceptible to invasion mechanisms. The precise identification of crucial genetic characteristics, such as abnormal gene, mRNA or protein expressions, could enhance BC therapeutic application and management. Moreover, discovery of various atypically expressed biomarkers and pathways that drive BC progression and might affect the survival of patients with BC has been made easier by recent developments in high-throughput sequencing technologies. Concerning various clinicopathological characteristics of BC, LVI could be accurately identified using H&E and IHC methods at the histological and molecular levels.

Recent research has used a combinational approach of integrating DGE and protein expression with bioinformatics to identify clusters of upregulated DEGs in cancer-related pathways. A recently published work examined the most highly expressed and repressed genes between 113 BC and 38 in cells neighbouring normal-like tissues by importing the gene expression profiles from an online resource called Gene Expression Omnibus (252, 253). Their findings revealed the differential expression of 82 genes, whose roles may be connected to cellular division and structural component receptor interaction of ECM as well as proteinaceous ECM (252). While this study's results seemed promising for further investigation, however, this was not applied to the association of LVI-positive BC cases. Thus, applying the same methodological approaches to a larger and well-characterised cohort of BC to identify more reliable prognostic and predictive biomarkers seemed promising to improve patients' outcomes. Additionally, another investigation evaluated the DEGs across patients with BC with both unfavourable and favourable outcomes in

luminal A BC using the TCGA-BRCA online database (254). The study found that a group of 18 prognostic biomarkers was linked to nine separate pathways (254). In a recent study, DEGs related to LVI were discovered in 79 LVI+ and 29 LVI– patients with BC using the significance analysis of the microarrays and the R limma package. According to the study’s findings, the protein–protein interactions, gene ontology (GO), gene set enrichment analysis (GSEA) and pathways analysis suggested that the cytokine–receptor interaction pathway and the imbalance between the components of the ECM and TME may be a reason behind the triggering of underlying LVI mechanisms in BC (255). Using LVI implications in BC prognosis, it may be possible to identify high-risk subgroups of patients with BC. The decisive gene/protein signatures have yet to be uncovered entirely, and the molecular code for LVI is still unclear. This is attributed to various factors that may contest the precise molecular signature of LVI in BC, leading to a lack of understanding of LVI status. The sample size of an LVI representative cohort, the various characteristic molecular expressions and clinicopathological traits of an LVI prospective cohort, potential false LVI– despite LNM+ detection, and the inaccurate processing and staining of tissue samples are just examples of the challenges that could hinder the precise detection of potential LVI drivers.

3.2 Aims and Hypothesis

This work aimed to explore novel DEGs influencing LVI processes in patients with BC to designate promising prognostic and targeted treatment in transcriptomic level, as well as perhaps proteins, for advanced investigation. In this study, several cohorts of patients with BC were utilised, including the METABRIC and TCGA-BRCA cohorts, characterized by sufficient clinical follow-up, and other well-defined and characterised BC cohort, to investigate the unidentified molecular stages that may underpin LVI mechanisms in patients with BC.

3.3 Materials and Methods

3.3.1 Study Cohorts

3.3.1.1 *The Molecular Taxonomy of Breast Cancer International Consortium Cohort*

In total, 285 patients from Nottingham in the METABRIC cohort and the additional METABRIC cases (n = 1,280) presented a significant number of transcriptome and copy number (CN) alteration profiles, as discussed in **Chapter 2 (section 2.1.1.1)**.

The METABRIC dataset, accompanied by validated clinicopathological features and several transcriptome and CN aberration (CNA) profiles, was used as a discovery cohort in this investigation. This cohort’s key applications included the following: 1) verifying the Linear

Models for Microarray Data (limma) result for the DEG experiment conducted on the LVI+/- cohort (Limma is a popular statistical software for analysing gene expression data from microarray experiments) (256); 2) providing this study with definite LVI+/- and LN+/- transcriptomic profiles of subsets of patients with BC and CNA data; and 3) conducting a thorough correlation analysis of gene expression on a chosen cluster of DGEs using a large-scale cohort of patients with BC.

In the METBARIC cohort, 55.4% of all patients had LVI- status, whereas LVI+ patients made up 44.6%. Moreover, in 78.3% of patients, the age at diagnosis was > 50 years. Concerning grade, 50.3% of the sample size was made up of patients with grade III tumour IBC. While slightly more than 50% of the cases in this group received a moderate NPI rating, 47.5% of patients had positive nodal involvement. In this group, invasive ductal BC was the most common histological type observed in 82.6% of patients. Each patient's BCSS was measured in months, up to a maximum of 300 months, with the recording of the date starting on the day when the initial surgical treatment began and continuing until the patient passed away due to BC, as described previously in **Chapter 2 (section 2.1.2.1)**. According to the most recent follow-up data, IBC was the cause of death for 25.5% of all patients.

With regard to molecular subtypes, 75.7% of patients had the ER+/IHC tumours. Moreover, 14.5% of patients had elevated HER2 expression, while 3.3% were considered borderline. According to the intrinsic subtyping classification of the samples, 10% of the patients in this cohort had the normal-like subtype, 36% had the luminal A type, 25% had the luminal B type, 12% had the HER2-enriched subtype, and 17% had the basal-like tumour subtype'. According to the IntClusts distribution of the METABRIC dataset, 7.0% of all patients fell into IntClust-1, IntClust-2 contained 3.6%, 14.6% were categorised into IntClust-3, IntClust-4 had 17.3%, 9.6% were included in IntClust-5, IntClust-6 had 4.3%, 9.6% were put into IntClust-7, IntClust-8 contained 15.1%, and 7.4% of all patients were put into IntClust-9, and finally data distribution revealed 11.4% of patients as IntClust-10.

3.3.1.2 The Cancer Genome Atlas Breast Carcinoma Cohort

In **Chapter 2 (section 2.1.1.2)**, the TCGA-BRCA cohort was described in detail. In this study, it served as a validation cohort to confirm the genes associated with LVI status identified by the METABRIC cohort. Even though 1,089 patients with BC formed the total TCGA-BRCA cohort, only 854 patients, were deemed IBC cases. Overall, 35% of patients in this cohort were classified as LVI+, whereas 65% were classified as LVI-. LVI status was assessed by an expert

breast pathologist (Leslie Dalton). Also, in this cohort, 70% of patients had a high histopathological grade. Furthermore, in 28% of cases, significant tumour size was seen. Patients older than 50 were found to account for two-thirds of cases. In all cases, 28% had nodal staging positive. The patient rate of survival in the TCGA-BRCA cohort was calculated to be 10 years of BCSS, and the time of survival was represented in months (for a total of up to 120 months). In addition, the BCSS time recording was from the day of collection of the primary surgical sample until the death day due to BC. In 75% of patient samples, the ER+ status was identified at the molecular level. In 65% of cases, increased PR expression was found. Additionally, 15% of cases were found to have increased HER2 expression. Moreover, 14% of cases displayed TNBC status. The following molecular intrinsic subtypes of the patient specimens were identified: 7% basal-like BC, 31% normal-like BC, 36% luminal A BC, 17% luminal B BC and 16% HER2-enriched BC.

3.3.2 Bioinformatics Analysis

Using the WAD approach, the analysis was carried out a former team member (Dr Sasagu) by counting on a reliable histological approach for analysing LVI tissues through identifying CD34 or D2-40 on a subset of Nottingham BC cases included in METABRIC as previously discovered (115). This subset of LVI cases from Nottingham was examined using stringent morphological and immunohistochemical criteria. However, only H&E-stained slides from the cBioPortal dataset were used to determine LVI status in the TCGA-BRCA cohort. Thus, DEGs in LVI+ and LVI- patients was performed, and DEGs were chosen based on their WAD ranking. Lists of the top 350 genes correlated with LVI as determined by the WAD analysis are summarised in **Appendices 10.4** and **10.5**. These lists were generated from a subset of the Nottingham cases in the METABRIC cohort (n = 285) and other METABRIC cases (n = 1,280). Both Nottingham and METABRIC cases were analysed independently, and then, overlapping DEGs between the two groups were included in the final gene set associated with LVI.

The clustering and heat map construction was performed through the Cluster 3.0 package (257). The clustering analysis was conducted using the METABRIC dataset as the discovery and the TCGA-BRCA dataset as the validation. Before the clustering analysis was conducted, the TCGA-BRCA count reads dataset was log2 transformed by a former team member (Dr. Sasagu). Furthermore, to determine significantly enriched GOs and pathways linked with these genes, the WEB-based Gene SeT AnaLysis Toolkit (WebGestalt) was utilised (258, 259). To manage the

level of false discovery, the Benjamini–Hochberg procedure was adopted within WebGestalt and consisted of an adjusted p -value <0.05 . This level was considered to be statistically significant.

3.3.3 Statistical Analysis

To assess the patient survival in both the METABRIC and TCGA-BRCA cohorts, the Kaplan–Meier survival curves of 10-year survival data were used to investigate the correlation between targeted genes and patient outcomes. Several clinicopathological parameters, including LVI and LN status, tumour size, tumour grade, hormonal status (ER and PR), HER2 status and the LVI-associated genomic subtypes and prognoses, were analysed against each other using the Cox proportional hazards regression model with 95% confidence intervals (CIs) in both univariate and multivariate analyses.

3.4 Results

3.4.1 Prognostic and Clinicopathological Significance of Lymphovascular Invasion Status

A subset of 635/1,565 patients, who constituted around 41% of the METABRIC dataset, displayed LVI positivity, while the remaining 930 (59%) were classified as LVI–. In the Nottingham group, 41.1%, or 117/285, of the total patients were LVI+, while 40.5% (518/1,280) were LVI+ cases, as observed in the METABRIC cohort. Meanwhile, LVI+ patients represented 35% (295) of the total cases (854), while LVI– cases represented 65% (559) in the TCGA-BRCA cohort. Moreover, LVI positivity was strongly associated with a high histological grade in both cohorts (both METABRIC and TCGA-BRCA cases representing $p < 0.0001$; TCGA-BRCA, $p = 0.00055$), positive nodal status (both METABRIC and TCGA-BRCA also representing $p < 0.0001$) and large tumour size (METABRIC: $p < 0.0001$; TCGA-BRCA: $p = 0.00055$) (Table 3.1).

In both the METABRIC and TCGA-BRCA cohorts, the LVI+ BC patient survival rate was considerably poorer than that of the patients with LVI– BC (hazard ratio (HR) = 1.70, 95% = CI 1.45–2.01, $p < 0.0001$ [Figure 3.1]; HR = 2.2, 95% CI = 1.46–3.38, $p = 0.00019$ [Figure 3.1], respectively). The univariate and multivariate analyses of the TCGA-BRCA and METABRIC datasets are listed in Table 3.2. Univariate analyses revealed poor prognostic indicators of different factors in both cohorts using the Cox proportional hazards regression model, including LVI+ status, large tumour size (METABRIC: HR = 1.82, 95% CI = 1.49–2.21, $p < 0.0001$; TCGA-BRCA: HR = 1.81, 95% CI = 1.08–3.04, $p = 0.025$), positive nodal

stage (METABRIC: HR = 2.06, 95% CI = 1.74–2.44, $p < 0.0001$; TCGA-BRCA: HR = 1.85, 95% CI = 1.20–2.85, $p = 0.0056$), negative ER status (METABRIC: HR = 1.66, 95% CI = 1.38–1.99, $p < 0.0001$; TCGA-BRCA: HR = 1.89, 95% CI = 1.19–2.98, $p = 0.0065$) and negative PR status (METABRIC: HR = 1.67, 95% CI = 1.42–1.98, $p < 0.0001$; TCGA-BRCA: HR = 1.68, 95% CI = 1.08–2.61, $p = 0.020$). Moreover, positive HER2 status (HR = 1.92, 95% CI = 1.54–2.38, $p < 0.0001$) and high histological grade (HR = 1.63, 95% CI = 1.37–1.93, $p < 0.0001$) were also important prognostic variables in the METABRIC cohort. Multivariate analysis also revealed LVI positivity to be a poor and independent prognostic factor (METABRIC: HR = 1.29, 95% CI = 1.07–1.56, $p = 0.0073$; TCGA-BRCA: HR = 2.19, 95% CI = 1.32–3.62, $p = 0.0023$) (Table 3.2).

Table 3.1: The association between lymphovascular invasion and other clinicopathological factors.

Factors	METABRIC cohort			TCGA-BRCA cohort		
	Lymphovascular invasion status		p value	Lymphovascular invasion status		p value
	Positive N (%)	Negative N (%)		Positive N (%)	Negative N (%)	
Tumour size						
≤2 cm	149 (23.5)	332 (36.3)	<0.0001	61 (20.7)	178 (31.8)	0.00055
>2 cm	485 (76.5)	582 (63.7)		234 (79.3)	381 (68.2)	
Nodal stage						
Positive	430 (67.8)	305 (32.9)	<0.0001	226 (77.1)	197 (35.4)	<0.0001
Negative	204 (32.2)	621 (67.1)		67 (22.9)	359 (64.6)	
Histological grade						
Grade 3	374 (60.7)	399 (44.6)	<0.0001	155 (55.0)	197 (36.9)	<0.0001
Grade 1,2	242 (39.3)	496 (55.4)		127 (45.0)	337 (63.1)	
Oestrogen receptor						
Positive	473 (74.5)	731 (78.6)	0.058	219 (76.6)	420 (78.1)	0.63
Negative	162 (25.5)	199 (21.4)		67(23.4)	118 (21.9)	
Progesterone receptor						
Positive	319 (50.2)	509 (54.7)	0.080	194 (68.6)	352 (65.8)	0.43
Negative	316 (49.8)	421 (45.3)		89 (31.4)	183 (34.2)	
Human epidermal growth factor receptor-2						
Positive	105 (16.5)	83 (8.9)	<0.0001	50 (21.2)	83 (17.9)	0.29
Negative	530 (83.5)	847 (91.1)		186 (78.8)	381 (82.1)	
Molecular Subtypes						
Luminal A	225 (35.5)	368 (39.7)	0.0021	286 (71)	119 (29)	<0.0001
Luminal B	178 (28.1)	215 (23.2)		68 (48)	73 (52)	
HER2 enriched BC	87 (13.7)	94 (10.1)		30 (54)	26 (46)	
Basal-like	99 (15.6)	147 (15.8)		94 (71)	39 (29)	
Normal-like	44 (7.0)	104 (11.2)		18 (60)	12 (40)	
Values in bold are statistically significant.						

Table 3.2: Univariate and multivariate Cox regression analysis of lymphovascular invasion status.

Factors	METABRIC cohort					
	Univariate Analysis			Multivariate analysis		
	Hazard ratio	95% CI	p-value	Hazard ratio	95% CI	p-value
Lymphovascular invasion	1.70	1.45-2.01	<0.001	1.29	1.07-1.56	0.0073
Tumour size	1.82	1.49-2.21	<0.001	1.48	1.21-1.83	<0.001
Nodal stage	2.06	1.74-2.44	<0.001	1.63	1.35-1.97	<0.001
Histological grade	1.63	1.37-1.93	<0.001	1.16	0.96-1.40	0.13
Oestrogen receptor	1.66	1.38-1.99	<0.001	1.14	0.91-1.43	0.25
Progesterone receptor	1.67	1.42-1.98	<0.001	1.38	1.13-1.69	0.002
Human epidermal growth factor receptor-2	1.92	1.54-2.38	<0.001	1.45	1.15-1.83	0.0019
	TCGA cohort					
	Univariate Analysis			Multivariate analysis		
	Hazard ratio	95% CI	p-value	Hazard ratio	95% CI	p-value
Lymphovascular invasion	2.22	1.46-3.38	0.00019	2.19	1.32-3.62	0.0023
Tumour size	1.81	1.08-3.04	0.025	1.33	0.77-2.31	0.302
Nodal stage	1.85	1.20-2.85	0.0056	1.13	0.67-1.92	0.651
Histological grade	1.46	0.94-2.25	0.092	0.95	0.52-1.80	0.884
Oestrogen receptor	1.89	1.19-2.98	0.0065	1.70	0.82-3.50	0.15
Progesterone receptor	1.68	1.08-2.61	0.020	1.21	0.60-2.42	0.603
Human epidermal growth factor receptor-2	1.51	0.83-2.77	0.18	1.42	0.73-2.61	0.330

Values in bold are statistically significant.

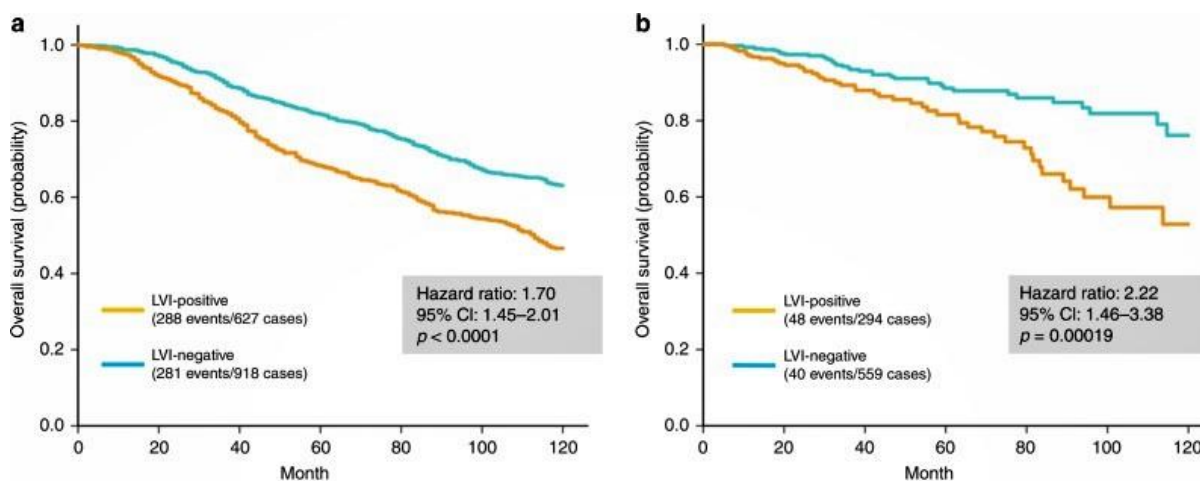


Figure 3.1: Lymphovascular invasion status-based stratification of the breast cancer patients' cumulative survival 10-year overall survival. In the METABRIC cases, the LVI+ group was noticeably poorer survival than the LVI- group (a); significant survival disparities between the LVI- and positive were observed in the TCGA-BRCA cases (b).

3.4.1.1 Genes Associated with Lymphovascular Invasion status

This analysis revealed 99 genes significantly associated with LVI status. The complete names of the targeted genes identified in this study are listed in **Appendix 10.6**. Moreover, the statistically relevant details, such as the mean value, standard error of the mean (SEM), subtraction and WAD ranking, are available in **Appendix 10.7**. Of these genes, 42 were significantly upregulated, and 57 were underexpressed genes (Table 3.3) found in the overlapping DEGs between the Nottingham cases in the METABRIC cohort (n = 285) and the remaining METABRIC cases (n = 1,280) (**Appendices 10.4 and 10.5**, respectively).

Table 3.3: List of 99 genes significantly associated with lymphovascular invasion status.

Upregulated genes					
APOC1	ELF3	ISG15	LY6E	S100P	UCP2
APOE	ERBB2	KRT18	MMP11	SCD	YWHAZ
CALML5	GNAS	KRT18P55	MX1	SLC52A2	
CCNB2	HMGA1	KRT19	NME1	SLC9A3R1	
CDC45	HMGB3	KRT7	NOP56	SPDEF	
COX6C	HSPB1	KRT8	PGAP3	TM7SF2	
DNAJA4	IDH2	LAPTM4B	PITX1	UBE2C	
EEF1A2	IFI27	LRRC26	PTTG1	UBE2S	
Downregulated genes					
ACTG2	CXCL14	FCGBP	MAOA	S100A4	TNS3
ANG	CYBRD1	FGD3	MFAP4	SELENOM	TPM2
ANXA1	CYP4X1	FOS	MGP	SERPINA3	TXNIP
C1S	DCN	FST	MT1E	SERPINE2	UBD
CDC42EP4	DKK3	GAS1	NDP	SGCE	VIM
CEBPD	DPYSL2	GSTP1	NINJ1	SLC40A1	VTCN1
CFB	DUSP1	HBA2	PDGFRL	SLC44A1	ZBTB20
CFD	EEF1B2	HBB	PLGRKT	SRPX	
CLIC6	FBLN1	HLA-DQA1	PYCARD	STC2	
CXCL12	FCER1A	IL17RB	RPL3	SUSD3	

To explore the possible pathways linked to the 99 targeted genes, the GSEA online tool was utilised to apply the ‘molecular component’ GO pathway, and the analysis revealed various GO pathways, including ‘GO: 0031012 Extracellular matrix’, ‘GO: 0005615 Extracellular space’ and ‘GO: 0072562 Blood micro-particle’ (Table 3.4). All significant pathways were associated with the ‘Cellular component’ of GO (Figure 3.2).

Table 3.4: Gene ontology pathways significantly associated with 99 genes related to lymphovascular invasion.

Ontology	Name	Genes in ontology	Observed	Expected	Enrichment	p-value	Genes
GO:0072562	Blood microparticle	110	7	0.52	13.51	0.00043	<i>SERPINA3, HBB, APOE, CFB, C1S, ACTG2, YWHAZ</i>
GO:0031012	Extracellular matrix	503	11	2.37	4.64	0.0079	<i>DCN, FBLN1, ANG, HSPB1, APOE, MFAP4, MGP, MMP11, NDP, SERPINE2, VIM</i>
GO:0005615	Extracellular space	1385	23	6.52	3.53	< 0.0001	<i>SERPINA3, DCN, CFD, FBLN1, DKK3, ANG, GSTP1, ANXA1, HBB, HSPB1, APOC1, APOE, MFAP4, NDP, SERPINE2, S100A4, CFB, CXCL12, C1S, ACTG2, YWHAZ, STC2, CXCL14</i>

Values in bold are statistically significant.

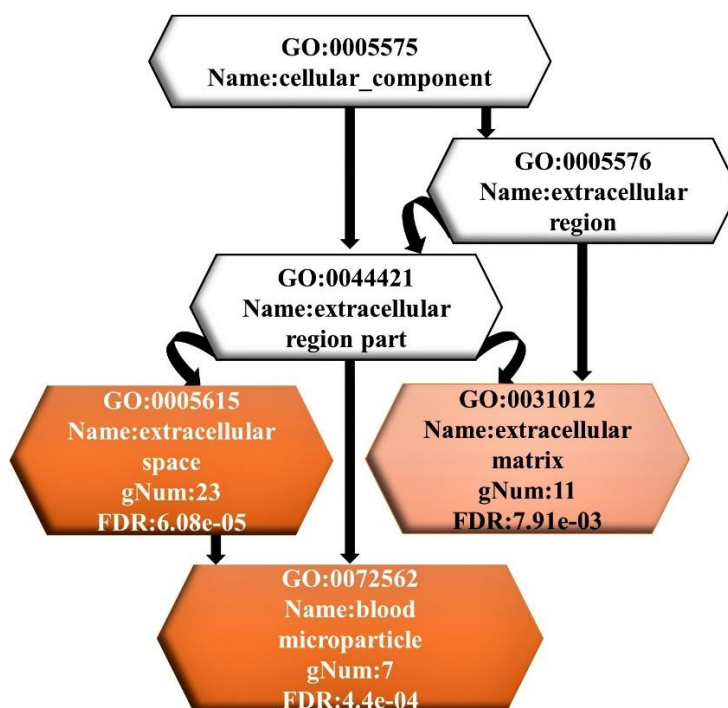


Figure 3.2: Gene ontology pathways significantly associated with genes related to lymphovascular invasion status, including ‘GO: 0031012 Extracellular matrix,’ ‘GO: 0005615 Extracellular space’ and ‘GO: 0072562 Blood micro-particle.

On the basis of similarity in expression, hierarchical clustering was performed to better examine these 99 genes (Figure 3.3a). Cases were divided into two subtypes by clustering in the discovery (METABRIC) cohort, as follow: subtypes 1 (n = 738 cases; 45%) and 2 (n = 827 cases; 55%) (Figure 3.3b). The METABRIC case dendrogram is depicted in Figure 3.4, where the branches' arrangement reveals the association for every case. Furthermore, hierarchical clustering was done on the TCGA-BRCA cohort using the 99 genes previously mentioned to confirm these findings. A dendrogram was used to categorise these 854 cases (Figure 3.5), which once more divided the cases into subtypes 1 (n = 263 cases; 31%) and 2 (n = 591 cases; 69%) (Figure 3.3c). The LVI+ cases were considerably more common in subtype 2 tumours versus subtype 1 tumours in both cohorts (TCGA-BRCA and METABRIC: $p < 0.0001$) (Table 3.5).

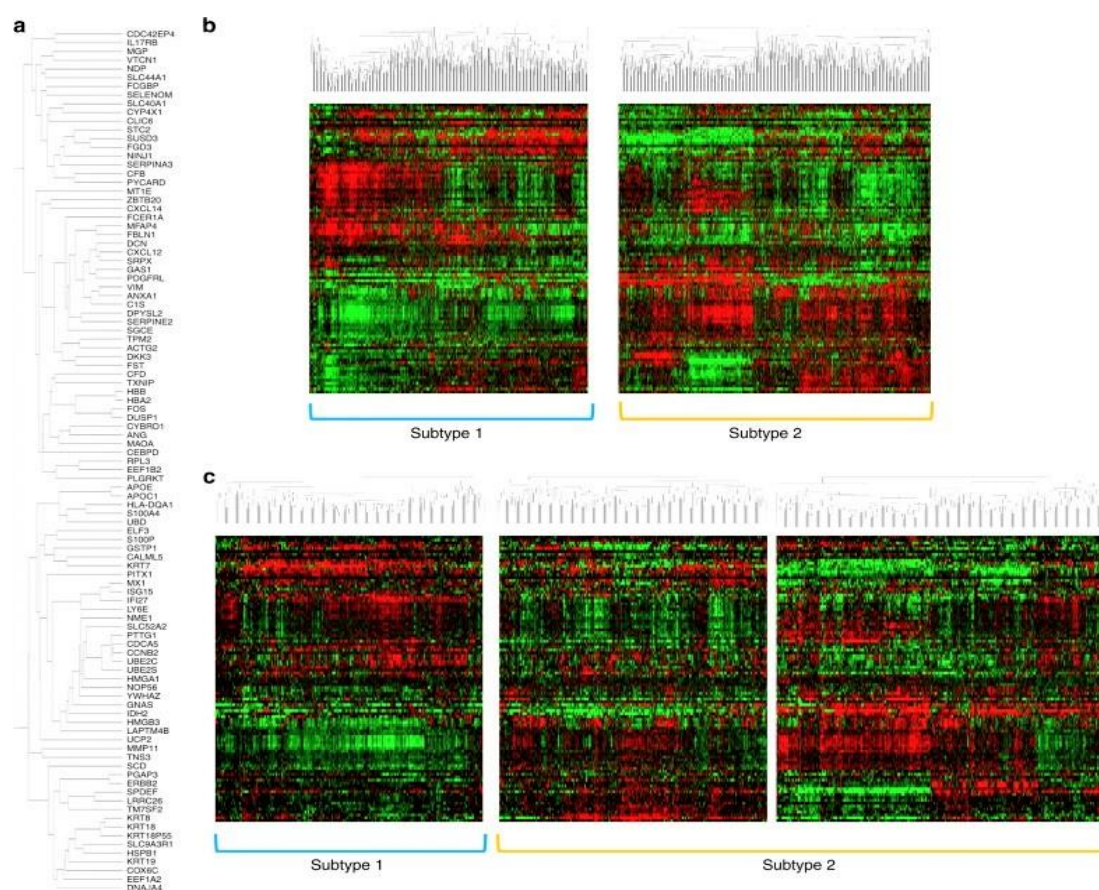


Figure 3.3: Cluster analysis and heat maps of the LVI-related genes group. Using the METABRIC cohort, a dendrogram of 99 LVI-associated genes was created, in which the arrangement of the branches shows how each gene is related to the others (a). Heat maps in line with LVI-related genes set for the METABRIC (b) and TCGA-BRCA (c) cohorts demonstrated that every case was categorically split into subtypes 1 and 2. Red colour indicates a high association, while green indicates a low association.

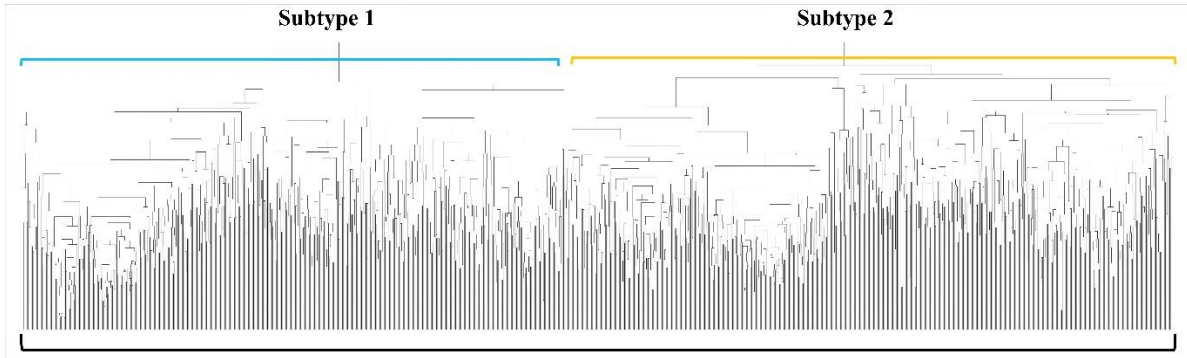


Figure 3.4: Dendrogram of the two subtypes 1 and 2 cases included in the METABRIC dataset.

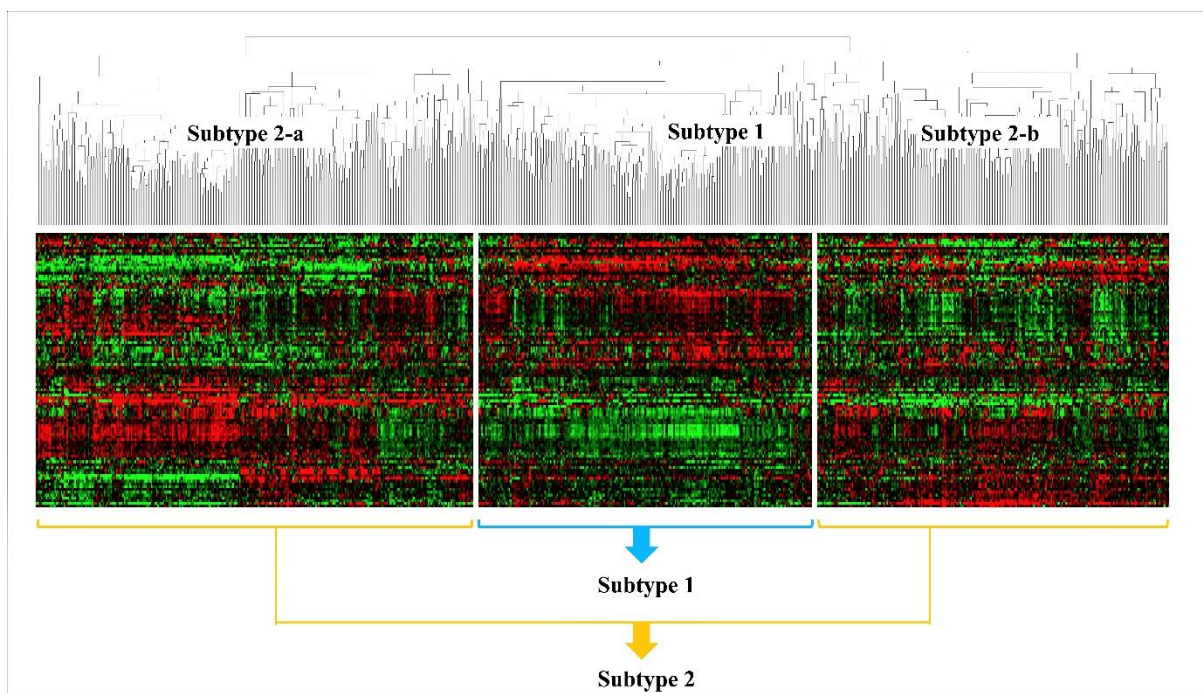


Figure 3.5: Dendrogram and heat maps of the two subtypes 1 and 2 cases included in the TCGA-BRCA dataset. Red colour indicates a high association, while green indicates a low association.

Table 3.5: Clinicopathological and prognostic significance of genomic subtypes related to lymphovascular invasion.

Factors	METABRIC cohort			TCGA-BRCA cohort		
	LVI-associated genomic subtypes		p value	LVI-associated genomic subtypes		p value
	Subtype 1 N (%)	Subtype 2 N (%)		Subtype 1 N (%)	Subtype 2 N (%)	
Lymphovascular invasion						
Positive	262 (35.5)	373 (45.1)	<0.0001	61 (23.2)	234 (39.6)	<0.0001
Negative	476 (64.5)	454 (54.9)		202 (76.8)	357 (60.4)	
Tumour size						
≤2cm	279 (38.1)	202 (24.8)	<0.0001	99 (37.6)	140 (23.7)	<0.0001
>2cm	454 (61.9)	613 (75.2)		164 (62.4)	451 (76.3)	
Nodal stage						
Positive	307 (41.7)	428 (51.9)	<0.0001	128 (48.9)	295 (50.3)	0.71
Negative	429 (58.3)	396 (48.1)		134 (51.1)	292 (49.7)	
Histological grade						
Grade 3	187 (26.5)	586 (72.8)	<0.0001	28 (11.3)	324 (56.9)	<0.0001
Grade 1,2	519 (73.5)	219 (27.2)		219 (88.7)	245 (43.1)	
Oestrogen receptor						
Positive	707 (95.8)	497 (60.1)	<0.0001	246 (97.6)	393 (68.7)	<0.0001
Negative	31 (4.2)	330 (39.9)		6 (2.4)	179 (31.3)	
Progesterone receptor						
Positive	533 (72.2)	295 (35.7)	<0.0001	235 (94.0)	311 (54.8)	<0.0001
Negative	205 (27.8)	532 (64.3)		15 (6.0)	257 (45.2)	
Human epidermal growth factor receptor-2						
Positive	20 (2.7)	168 (20.3)	<0.0001	20 (9.6)	113 (23.0)	<0.0001
Negative	718 (97.3)	659 (79.7)		189 (90.4)	378 (77.0)	
Molecular Subtypes						
Luminal A	467 (63.5)	126 (15.3)	<0.0001	336 (83)	69 (17)	<0.0001
Luminal B	121 (16.5)	272 (32.9)		48 (34)	93 (66)	
HER2 enriched BC	10 (1.4)	171 (20.7)		10 (18)	46 (82)	
Basal-like	24 (3.3)	222 (26.9)		34 (26)	99 (74)	
Normal-like	113 (15.4)	35 (4.2)		22 (73)	8 (27)	
Values in bold are statistically significant.						

3.4.1.2 Prognostic and Clinicopathological Significance of the LVI-Related Genes

Subtype 2 showed a statistically significant association with the following tumour phenotypes and genotypes in both the TCGA-BRCA and METABRIC cohorts, including large tumour size (>2cm), high histological grade (grade 3), negative ER and PR, and HER2 positivity ($p < 0.0001$ for both cohorts). Notably, subtype 2 was assigned to (the METABRIC: 95%; TCGA-BRCA: 82%) for HER2-enriched BC, (METABRIC: 90%; TCGA-BRCA: 74%) for basal-like BC and (METABRIC: 69%; TCGA-BRCA: 66%) for the luminal B BC subtype (Table 3.5).

In both cohorts, patients who had LVI-related subtype 2 tumours had a considerably poorer prognosis than those who did not (METABRIC: HR = 1.78, 95% CI = 1.50–2.12, $p < 0.0001$; TCGA-BRCA: HR = 2.32, 95% CI = 1.35–3.99, $p = 0.0023$ [Figure 3.6 c and d respectively]). Multivariate survival analysis in both cohorts revealed that the LVI-related genomic subtype was independent poor prognostic factor (METABRIC: HR = 1.32, 95% CI = 1.07–1.63, $p = 0.0098$; TCGA-BRCA: HR = 2.76, 95% CI = 1.19–6.38, $p = 0.018$ [Table 3.2 and Figure 3.7]).

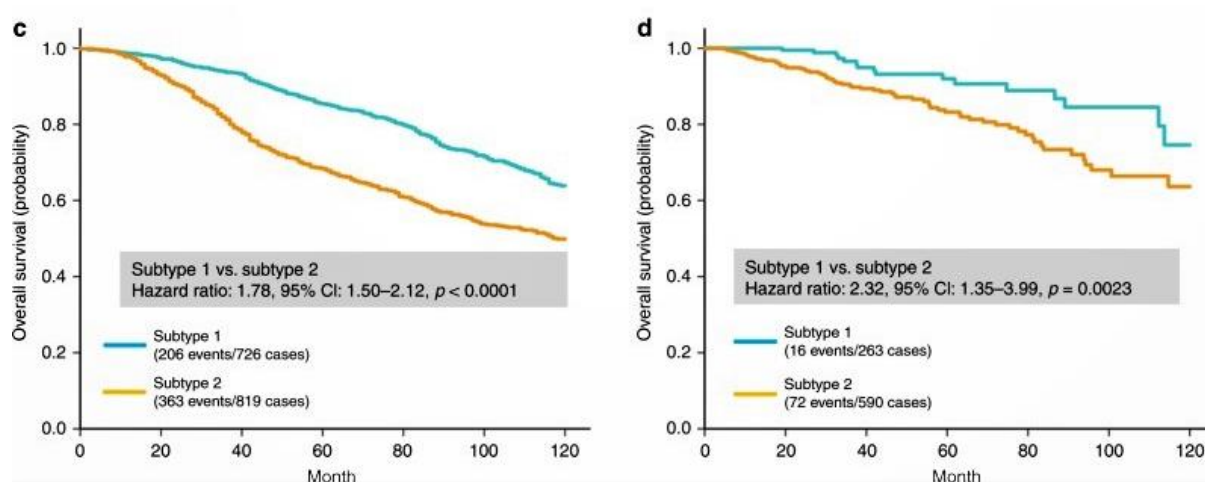


Figure 3.6: Lymphovascular invasion status-based stratification of the breast cancer patients' cumulative survival 10-year overall survival; In the METABRIC cases, the LVI+ group was noticeably poorer survival than the LVI– group (c); significant survival disparities between the LVI– and positive were observed in the TCGA-BRCA cases (d).

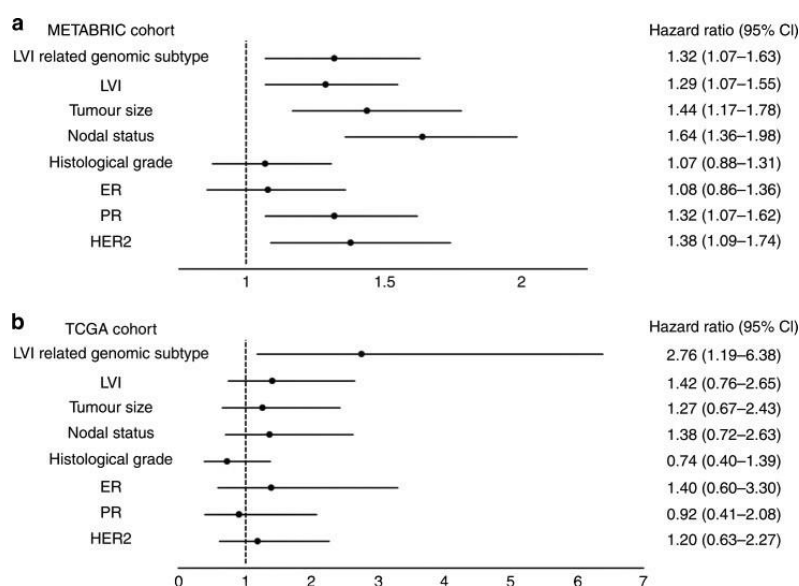


Figure 3.7: Survival analyses based on clinicopathological traits, including the LVI-related genomic subtype; multivariate survival analyses' hazard ratios and 95% CI are presented as forest plots in the METABRIC cohort **(a)** and TCGA-BRCA cohort **(b)**.

3.5 Discussion

This study, based on the METABRIC dataset, used WAD analysis to identify a set of 99 genes that are significantly correlated with LVI status. The results from the discovery cohort were confirmed using the TCGA-BRCA cohort. In the aggressive form of BC, the LVI is used as a prognostic tool that predicts the likelihood of metastasis (121). Even though similar research (255, 260-262) has been conducted to discover target genes that play a crucial role in LVI mechanisms, few overlapping genes have been determined due to the nature of the tumour and the approach and cohorts used. Previous studies have revealed gene sets related to vascular invasion (VI) in various kinds of cancer, such as endometrial cancer (EC) (260) and hepatocellular carcinoma (HCC) (261). In their study, Mannelqvist et al. (262) found a set of 18 genes linked with VI status in patients with EC (260). The verification of these genes was further evaluated in the BC patient cohort, which revealed a significant association with ER and PR hormonal receptor negativity, HER2 positivity, basal-like phenotype and unfavourable patient survival (262). In agreement with Mannelqvist and colleagues' results, this analysis of the METABRIC cohort indicated that the subtype 2 cases included in this cohort contained 90% basal-like, 95% HER2-enriched, 69% luminal B and 95% BC subtypes. Mínguez et al. studied 214 HCC cases for a gene signature that might predict the occurrence of LVI in patients

with surgically treated HCC (261). From their set, *CD24*, *CDKN3*, *UBE2C* (found in the Nottingham 99-gene set) and *PIK3R1* were some genes that have a role in the development and progression of BC (261). A recent study investigated DNA CNAs and gene expression signatures related to LVI in patients with IBC using array comparative genomic hybridisation and RT-qPCR microarrays (148). The analysis identified possible candidate genes for LVI+ grade 3 tumours, including interstitial collagenase, the mediator of RNA polymerase II transcription subunit-1 and protein S100-A8 (148). Despite this, it is worth noting that the BC cohort was small ($n = 57$) and that frozen tissue was used without additional verification of LVI status.

Despite the fact that LVI might be used as a prospective treatment target, the complex molecular processes that drive LVI in BC have not yet been discovered. The findings of this research showed that strong associations exist between extracellular pathways and the 99 genes that make up the LVI-related gene signature. A finding in line with a previous study by Klahan et al. (255) using analysed microarray data from 108 patients with BC found that the ECM component such as collagens was linked to their LVI gene set. Also, as part of their gene set, both Endothelial PAS domain protein 1, a VEGF production stimulator, and tumour necrosis factor ligand superfamily member 11 (TNFSF11), a receptor activator of nuclear factor kappa B ligand, contribute to BC cell invasion, migration and dissemination. In addition, TNFSF11, alongside interleukin 6 cytokine family signal transducer, contributes to cytokine–receptor interaction, the most frequent enriched pathway linked to LVI (255). In addition, genes involved in EMT in prostate cancer have also been linked to extracellular space–related pathways (263). In the process of BC development and progression, many reconfigurations and reorganisations take place on the protein structural network component of the ECM (264). In addition, EMT plays a major role in tumour dissemination to distant organs and tissue, and some EMT genes are thought to be correlated with LVI status in BC (115). Among the 99 genes in the LVI signature, several have been linked to extracellular pathways, which play a fundamental role in BC development. For instance, heat shock protein 27 and heat shock protein family B member 1 (HSPB1) have been previously reported for their association with BC dissemination (265). Interestingly, the early upregulation of HSPB1 expression during cell differentiation suggests a potential function for HSPB1 in regulating the proliferation and migration of cancer stem-like cells (266). The next example is the apolipoprotein C1 (APOC1) biomarker, which is used to predict patient prognosis in TNBC (267). Research performed by Ko et al. proposed a role for APOC1 in controlling the inflammatory response of tumour tissue

(268), which may have a close relationship to the elimination of proliferating tumour cells (269). Moreover, MMPs are well known as a poor prognostic factor in BC; their high expression is associated with regulating tumour motility, invasion and EMT (270). For instance, MMP-11 facilitates cancer progression through multiple processes, including inhibition of the immune response and gradual reduction of tumour surveillance activities, blocking apoptotic activities and, consequently, facilitating BC cells' migration and invasion (271, 272). More functional investigations of these markers are required to investigate the link between genetic aberration function and LVI-related proteins in BC.

Associating the METABRIC and TCGA-BRCA cohorts was a limitation of this research due to the varied methodologies utilised to measure and statistically analyse transcriptional activity and assess LVI. A reliable histological approach for analysing LVI tissues by identifying CD34 or D2-40 was previously discovered (115). The subset of LVI cases from Nottingham included in METABRIC was examined using stringent morphological and immunohistochemical criteria. However, only H&E-stained slides from the cBioPortal dataset were used to determine LVI status in the TCGA-BRCA cohort. Though it is possible to evaluate LVI with only one H&E slide, doing so may make it difficult to reliably detect LVI negativity (273). In the current analysis, the Nottingham cases, the remaining METABRIC cases and the TCGA-BRCA cases all had nearly identical LVI+ rates regardless of the methods of LVI evaluation used. While these findings might imply that a single H&E-stained slide is sufficient for LVI assessment, studies with a larger cohort assessing LVI status with both H&E and IHC slides are essential for reliable reporting of LVI status.

The data were analysed differently between the METABRIC and TCGA-BRCA datasets. In the METABRIC dataset, transcriptomic expression of the genes was evaluated utilising microarrays. However, RNA-seq employing NGS was utilised to analyse genes in the TCGA. Microarray systems have been utilised and verified for about 20 years, and this method has seen widespread use for the purpose of assessing multiple gene expression profiles. However, analysis of all identified transcripts, the discovery of new transcripts, splice junctions and non-coding RNAs are specifications of an unbiased RNA-seq approach. Variations in the technical and methodological levels may support the known difficulties in connecting microarray and RNA sequencing results among different research works. For instance, various methods may cover different parts of the genome or have different minimum detectable thresholds. As a result, conducting extensive cross-assay comparisons is required since it would be erroneous to presume that one approach may be substituted for another. While statistical concordance

was shown across this study's various cohorts, further investigation using the same methodologies (microarray and/or NGS analyses) could give more definitive confirmation of the LVI gene signature.

In summary, the METABRIC and TCGA-BRCA datasets are suitable for use in the LVI evaluation method, and this approach has significant prognostic implications. As a result of this research, a set of genetic subtypes connected with LVI status and patient outcomes in BC that enhance the aggressiveness of BC has been identified. This provides an experimental tool that might help untangle underlining gene mechanisms associated with clinically significant LVI in BC. Using the same criteria and assessment methodologies for LVI status could potentially increase consistency among clinical cohorts classified according to the LVI gene signature.

**4 Chapter 4: HSP90 α , a gene
associated with LVI status
selected from the DGE analysis
and previous mass spectrometry
analysis**

This chapter comprises data published in the *Journal of Clinical Pathology* (2021) under the title ‘Prognostic significance of heat shock protein 90AA1 (HSP90 α in invasive breast cancer’ with the corresponding DOI: [10.1136/jclinpath-2020-207106](https://doi.org/10.1136/jclinpath-2020-207106) (274). The copyright approval for the re-use of this material is provided in **Appendix 10.3**.

4.1 Overview

In the BC overview conducted in **Chapter 1**, it was illustrated how IBC is a heterogeneous group of diseases with various presentations, morphological features, outcomes and responses to therapy (9). According to several studies, genomic and transcriptomic alterations in different individuals lead to high invasive potential, diverse proteomic expressions (251, 275), and including poor prognosis (276, 277). However, a number of studies have linked invasiveness of BC to certain molecular intrinsic subtypes that regulate hormonal signals in the invasion process (278). Younger patients with BC have a greater likelihood of having aggressive phenotypic features of IBC, such as luminal B (279). One study identified a strong association between younger patients with luminal B IBC and LVI (280). Considering the apparent molecular alterations in patients with IBC, further molecular stratification to improve targeted treatment methodologies must be applied for BC patients. In other words, medical treatments and BC control could be enhanced if critical genetic characteristics were identified, such as genetic alteration, mRNA and protein expression. Thus, one of the main focuses of research is to decipher the critical role of LVI prognostic marker association in promoting tumour aggressiveness in various cancers. An evaluation of previous in-house and external LVI studies looking at multiple genes involved in other types of cancer was conducted to determine which genes may be linked to BC LVI.

Although many factors associated with BC incidence and progression have been identified, current understanding of the diverse molecular and cellular mechanisms contributing to BC progression remains limited. In addition, BC is considered one of the cancers that most commonly develop distant metastasis and enhance the spreading of tumour cells to different human tissues and organs (281), leading to an increase in the patient mortality rate. Better knowledge of the molecular factors involved in BC progression will help in tailoring personalised therapy and improving outcomes (73). Thus, identifying these factors, which initiate BC progression with detachment of the primary tumour, invasion and migration until proliferating at the secondary site, is considered essential to avoid and manage BC LVI and metastatic cascade.

Existing data indicate that the cellular stress response is important in cancer progression and cancer therapy (282). Expression of heat shock proteins (HSPs) increases in response to cellular stress (283, 284) and can enable tumour cells to survive cellular stress by evading cell death and senescence mechanisms (285). Consistent with this, it was demonstrated that HSPs have an essential role in multiple pro-oncogenic signalling pathways, and it is an adenosine triphosphate (ATP) dependent protein chaperone (286, 287). Subsequently, the heat shock protein 90 (HSP90) family has attracted considerable attention as a potential therapeutic target (288) as these proteins are the key players involved in protein modification and have been implicated in BC progression (289). The human HSP90 alpha family class A member 1 (HSP90 α), encoded by the *HSP90AA1* gene, which is located at chromosome 14q32.31, appears to be critically important in the prevention of cell death (290) and evasion of apoptosis (291). Previous studies have indicated that HSP90 α overexpression inhibits major apoptosis-associated proteins in leukaemia, enabling the malignant cells to escape apoptosis and undergo proliferation (292, 293). Consistent with this, mechanistic evidence from a lung cancer model indicates HSP90 α role in tumour growth (294).

Autophagy can induce genomic instability, malignant transformations and may help metastatic cancer cells to escape from anoikis (295, 296). *In vitro* studies in osteosarcoma have also shown that *HSP90AA1* knockdown increased chemo-sensitivity by decreasing autophagy activity of cancer cells (297), which raises the potential that HSP90 α also contributes to pro-invasive mechanisms.

4.2 Aim and Hypothesis

Through its varied interactions, HSP90 α could play a vital role in various biological processes such as maintaining the proper conformations of protein molecules in cells to ensure they remain stable (298). HSP90 α regulates protein synthesis and has demonstrated extracellular functions in wound repair and malignancy (299, 300). Consequently, the roles of HSP90AA1 may confer a favourable strategy utilised by tumour cells to promote tumour invasion and migration, which are sequential steps of LVI leading to cancer dissemination (301, 302). *HSP90AA1* was one of the 350 genes identified by our previous DGE analysis in LVI+ and LVI- patients using the WAD approach; more details are provided in Chapter 3 (section 3.3.2). Its importance was shown also when a set of FFPE TMA blocks (n = 352, LVI+ = 167, LVI- = 185) were examined by high-performance liquid chromatography/mass spectrometry as a high-throughput screening technique carried out by Dr Sultan Sonbul (Pathology, School of

Medicine, University Hospital, Nottingham, UK) as a collaborative project to identify genes associated with LVI status in IBC.

This resulted in the identification of HSP90 α as one of the top genes associated with LVI status in the examined cohort. As a result, it was hypothesised that HSP90 α is involved in pathways contributing to the invasion and migration, development and metastasis of cancer cells and that it could be used as a novel therapeutic target in IBC.

This study aimed to investigate the clinical and prognostic value of HSP90 α expression at both the mRNA and protein levels in BC LVI, relating these findings to key clinicopathological factors and the expression of other proteins relevant to HSP90 α function.

4.3 Materials and Methods

4.3.1 Study Cohort

The clinical significance of *HSP90AA1* mRNA expression was assessed using the METABRIC dataset of IBC (n = 1,980). The clinicopathological associations and prognostic value of *HSP90AA1* mRNA expression were further evaluated using the whole BC cohort of TCGA-BRCA (n = 1,097) to examine its association with outcome and clinicopathological parameters, whereas the other chosen biomarkers in this study were evaluated only on a restricted cohort of TCGA-BRCA (n = 854) due to not being able to retrieve all H&E slides for re-evaluation of vascular invasion in the cohort. Nonetheless, the bc-GenExMiner v4.5 datasets (n = 5,056) were also incorporated for examining *HSP90AA1* mRNA expression and association with patients' outcomes. This study included a well-characterised cohort of IBC (n = 911) to validate the prognostic significance of HSP90 α protein expression using IHC. The clinicopathological features and outcomes of patients were obtained from the patients' medical records in **Chapter 2 (section 2.1.2.1)**.

4.3.2 Western Blotting and Immunohistochemistry

To validate the antibody specificity, WB was performed using mouse monoclonal anti-HSP90 α antibody (ab13492; Abcam, Cambridge, UK) with a dilution of 1:700. The housekeeping protein rabbit β -actin antibody (MABT523, clone RM112, Sigma-Aldrich, Gillingham, UK) was used at a dilution of 1:5,000. For WB, cell lysates from HeLa, MCF7, SKBr3 and MDA-MB231 cell lines were used. This procedure was performed in keeping with the protocol of WB validation for specific Ab-Ab binding mentioned in **Chapter 2 (section 2.3)**. The TMA slides were stained using the IHC procedure previously described in **Chapter 2 (section 2.4.1)**.

For interobserver concordance, a subset of cores (25%) was independently scored by another scorer (M. T) in addition to the author, and ICC was performed ($ICC = 0.7$). Scoring was performed on anonymised samples of clinical variables at the time of scoring. Discordant cases were re-scored by both observers, and a final score was agreed upon.

4.3.3 Statistical Analysis

The SPSS 24 program (IBM, Chicago, IL, USA) was used to perform the statistical analysis. For *HSP90A1* mRNA expression, the median value was used to dichotomise the data. For the HSP90 α protein expression, the cut-off point was determined using the median. The association between HSP90 α expression and the clinicopathological factors was evaluated using the chi-squared test. The univariate survival analysis was conducted using Kaplan–Meier plots, and multivariate survival analyses were conducted using the Cox regression model with adjustment of covariates fitted to test independence from standard prognostic factors in BC. A value of $p < 0.05$ (two-tailed) was considered significant. Further statistical data are detailed in **Chapter 2 (section 2.11)**.

4.4 Results

4.4.1 Western Blot Evaluation

The anti-HSP90 α primary antibody's WB image revealed a single specific band at the predicted MW (90 kDa; Figure 4.1), confirming the specificity of the antibody for each cell lysate.

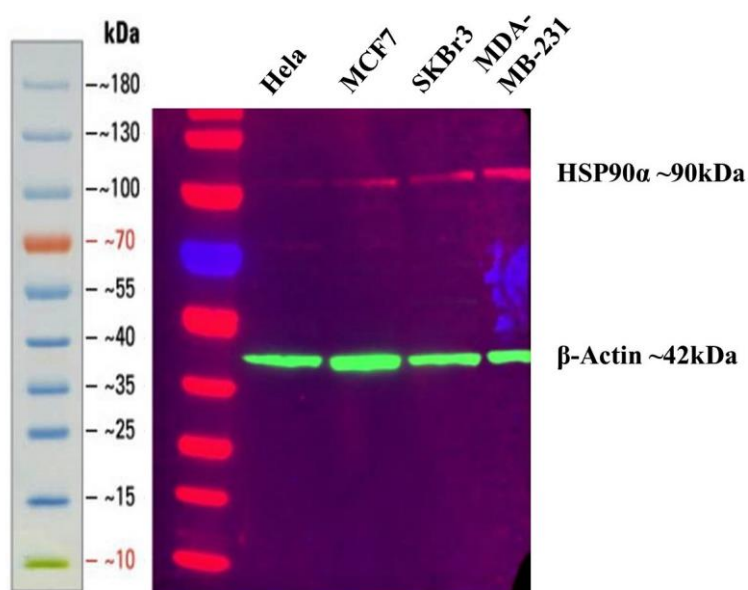


Figure 4.1: Western blotting validation showed mouse monoclonal anti-HSP90 α antibody ab13492; specificity with a single band observed at approximately 90 kDa in HeLa, MCF7, SKBr3 and MDA-MB-231.

4.4.2 Clinical Significance of *HSP90AA1* mRNA Expression

High expression of *HSP90AA1* mRNA, and the median was used to determine the cut-off point, was seen in 990 of 1,980 cases in METABRIC. A similar frequency of positivity was seen in the TCGA-BRCA (549 of 1,097 cases) and the bc-GenExMiner v4.5 datasets (2,537 of 5,056 cases). High expression of *HSP90AA1* mRNA was significantly associated with high grade, LVI positivity, ER negativity, and HER2 positivity. The *HSP90AA1* mRNA expression was significantly higher in luminal B, HER2 positive and TNBC tumours than in luminal A tumours (Table 4.1), and it was associated with high PIK3CA and p53 (Table 4.2). These results were seen in both METABRIC and TCGA-BRCA.

Table 4.1: Association between *HSP90AA1* mRNA expression and clinicopathological factors in the METABRIC (n = 1980) and TCGA-BRCA (n = 1097) invasive breast cancer datasets.

Factors	METABRIC			TCGA-BRCA		
	Low <i>HSP90AA1</i>	High <i>HSP90AA1</i>	<i>p</i> -value	Low <i>HSP90AA1</i>	High <i>HSP90AA1</i>	<i>p</i> -value
	N (%)	N (%)		N (%)	N (%)	
Lymphovascular invasion						
Negative	505 (54.3)	425 (45.7)	0.017	324 (58.0)	235 (42.0)	<0.001
Positive	306 (48.2)	329 (51.8)		127 (43.1)	168 (56.9)	
Tumour size						
≤2.0 cm	451 (52.6)	407 (47.4)	0.041	99 (55.9)	78 (44.1)	0.01
>2.0 cm	528 (47.9)	574 (52.1)		273 (45.7)	325 (54.3)	
Age						
≤50 years	197(46.5)	227 (53.5)	0.10	149 (45.2)	181 (54.8)	0.037
>50 years	793(51)	763 (49)		397 (52.0)	366 (48.0)	
Nodal Status						
Negative	542 (52.4)	493 (47.6)	0.026	423 (47.0)	477 (53.0)	0.564
Positive	444 (47.3)	494 (52.7)		10 (47.6)	11 (52.4)	
Histological grade						
Grade 1 and 2	534 (56.8)	406 (43.2)	<0.001	158 (59.0)	110 (41.0)	<0.001
Grade 3	405 (42.5)	547 (57.5)		67 (33.7)	132 (66.3)	
Oestrogen receptor						
Negative	215 (45.4)	259 (54.6)	0.02	100 (42.0)	138 (58.0)	0.001
Positive	775 (51.5)	731 (48.5)		431 (53.5)	375 (46.5)	
Progesterone receptor						
Negative	459 (48.8)	481 (51.2)	0.32	148 (43.0)	196 (57.0)	<0.001
Positive	531 (51.1)	509 (48.9)		380 (54.5)	317 (45.5)	
Human epidermal growth factor receptor 2						
Negative	897 (51.8)	836 (48.2)	<0.001	311 (53.2)	274 (46.8)	<0.001
Positive	93 (37.7)	154 (62.3)		61 (34.5)	116 (65.5)	
Nottingham prognostic index						
Good prognostic group	390 (57.4)	290 (42.6)	<0.001			
Moderate prognostic group	510 (46.3)	591 (53.7)				
Poor prognostic group	90 (45.2)	109 (54.8)				
Molecular subtypes						
Luminal A	430 (59.9)	288 (40.1)	<0.001	301 (60.3)	198 (39.7)	<0.001
Luminal B	169 (34.6)	319 (65.4)		59 (29.9)	138 (70.1)	
HER2 enriched	105 (43.8)	135 (56.3)		16 (20.5)	62 (79.5)	
Basal-like	148 (44.8)	181 (55.2)		74 (43.3)	97 (56.7)	
Normal-like	134 (67.3)	65 (32.7)		30 (83.3)	6 (16.7)	
Values in bold are statistically significant						

Table 4.2: Association between *HSP90AA1* mRNA expression and other gene of interest in the METABRIC and TCGA-BRCA invasive breast cancer datasets.

Gene of Interest	HSP90AA1 mRNA Level (METABRIC)		Mean Rank		Z-score	p-value	HSP90AA1 mRNA Level (TCGA-BRCA)		Mean Rank		Z-score	p-value
	Low (%)	High (%)	Low	High			Low (%)	High (%)	Low	High		
Ki67	990 (50)	990 (50)	920	1060	-5.4	<0.001	548(49.9)	549 (50)	443	654	-11	<0.001
PIK3CA			982	998	-3.15	0.002			483	614	-6.9	<0.001
P53			941	1039	-3.8	<0.001			550	547	-0.1	0.889
AKT			981	999	-0.72	0.471			506	591	-4.4	<0.001
EGFR			1011	969	-1.66	0.096			554	543	-0.6	0.558
Values in bold are statistically significant.												

4.4.3 Outcome Analysis for *HSP90AA1* mRNA Expression

The outcome analysis revealed that patients who had BC with higher *HSP90AA1* mRNA expression had significantly shorter BCSS than those who had low *HSP90AA1* expression (METABRIC: HR = 1.38; 95% CI = 1.16–1.65, $p = 0.0004$; TCGA-BRCA: HR = 1.85; 95% CI = 1.318–2.61, $p < 0.0001$; Figure 4.2a and b). Using the prognostic module of bc-GenExMiner v4.5, *HSP90AA1* mRNA was validated as a significant prognostic factor for shorter survival (HR = 1.17; 95% CI = 1.07–1.28, $p = 0.0004$; Figure 4.2c).

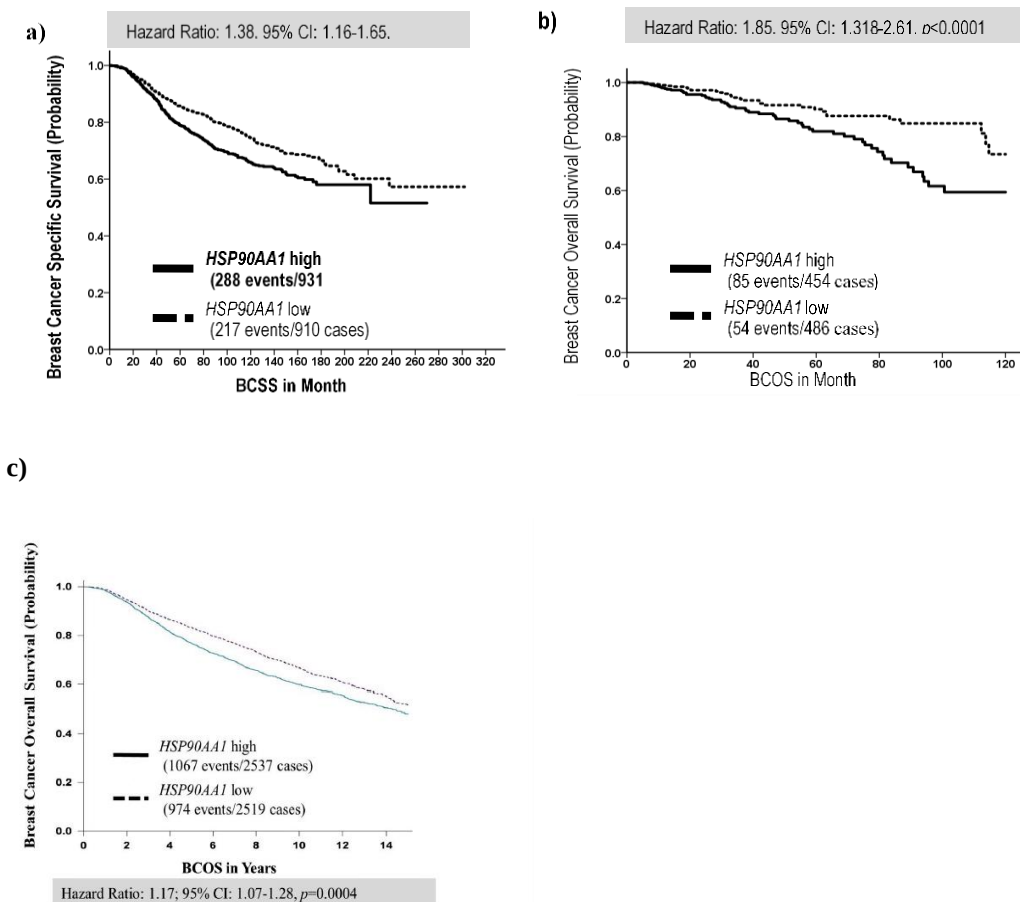


Figure 4.2: Cumulative survival of BC patients stratified by *HSP90AA1* mRNA expression. Breast Cancer Specific Survival in the METABRIC cases was significantly worse in the high *HSP90AA1* group than low *HSP90AA1* expression (a) in the TCGA cases, significant differences were noted in patient survival in the *HSP90AA1* high and low groups (b). Kaplan-Meier survival plots of *HSP90AA1* mRNA expression revealed that the high expression associated with poor patient outcome (c). BCOS, breast cancer overall survival; BCSS, breast cancer-specific survival.

4.4.4 HSP90 α Protein Expression

To validate the antibody protein expression, IHC was performed using the anti-HSP90 α antibody with a dilution of 1:200 and showed cytoplasmic immunoreactivity. Full-face BC tissue sections (n = 14) that were selected based on LVI status, tumour grade and histological type were stained using IHC to evaluate the pattern of HSP90 α protein expression as a mandatory prerequisite step for using the Nottingham cohort. The HSP90 α staining was homogeneously distributed within the invasive tumour cells in the full-face tissue sections. The normal terminal duct lobular units (TDLUs) adjacent to the tumour showed uniformly weak HSP90 α staining compared to the invasive cancer cells. The HSP90 α expression in IBC and adjacent the normal TDLUs is illustrated in Figure 4.3 a–b. The cut-off point for the HSP90 α protein expression was determined as an H-score of 50 using the median, with a score >50 considered a high expression (Table 2.10). High HSP90 α protein expression was observed in 525/911 (58%) of cases, while low HSP90 α protein expression was performed in 386/911 (42%) of the cases (Figure 4.3a and 4.3b, respectively).

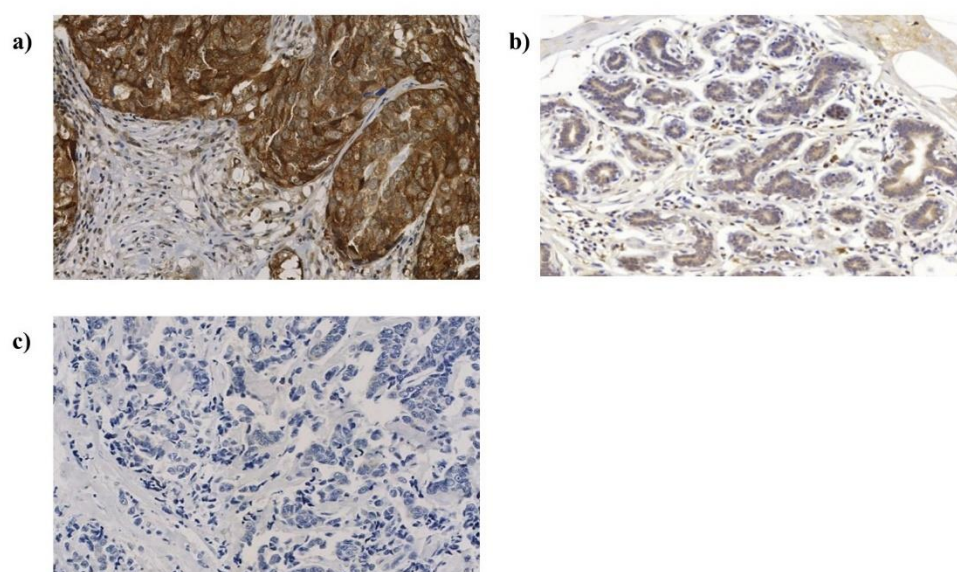


Figure 4.3: HSP90 α immunohistochemical protein expression in tissue microarray cores of normal and breast cancer tissue. Invasive cancer cells (strong staining in the cytoplasm of cancer cells), **(a)** normal mammary gland (uniformly weak staining), **(b)** no staining **(c)**, (images are at 20 $\times$ magnification).

4.4.5 HSP90 α Protein Expression and its Association with the Clinicopathological Parameters

Here, HSP90 α showed a positive linear correlation with *HSP90AA1* mRNA (Spearman's coefficient 0.3; $p = 0.001$). High HSP90 α protein expression was significantly associated with high tumour grade ($p < 0.0001$), positive nodal status ($p = 0.0016$), positive LVI ($p = 0.014$), high PIK3CA ($p < 0.0001$) and positive p53 ($p < 0.002$) (Table 4.3). Moreover, HSP90 α was highly expressed in the luminal B/HER2-negative, TNBC, luminal B/HER2-positive and HER2-enriched subtypes, while luminal A tumours showed the lowest levels of expression ($p = 0.002$) (Table 4.3).

Table 4.3: Associations between HSP90 α protein expression and the clinicopathological factors of the Nottingham breast cancer cohort.

Factors	Expression of HSP90α			Factors	Expression of HSP90α		
	Low	High	p-value		Low	High	p-value
	N (%)	N (%)			N (%)	N (%)	
Tumour size				Age			
≤2 cm	190 (43.9)	243 (56.1)	0.435	≤50 years	125 (39.6)	191 (60.4)	0.210
>2 cm	195 (41.3)	277 (58.7)		>50 years	261 (43.9)	334 (56.1)	
Histological grade				Nodal status			
1,2	234 (50.6)	228 (49.4)	< 0.0001	Negative	255 (45.6)	304 (54.4)	0.016
3	150 (34.1)	290 (65.9)		Positive	130 (37.5)	217 (62.5)	
Lymphovascular invasion				Oestrogen receptor			
Negative	271 (45.5)	325 (54.5)	0.014	Negative	85 (37.8)	140 (62.2)	0.126
Positive	114 (37.0)	194 (63.0)		Positive	296 (43.6)	383 (56.4)	
Progesterone receptor				Human epidermal growth factor receptor-2			
Negative	144 (39.9)	217 (60.1)	0.534	Negative	312 (41.8)	434 (58.2)	0.848
Positive	215 (42.0)	297 (58.0)		Positive	53 (40.8)	77 (59.2)	
Ki67				Phosphoinositide-3-kinase, Catalytic, Alpha polypeptide			
Low	107 (41.6)	150 (58.4)	0.352	Negative	95 (55.2)	77 (44.8)	< 0.0001
High	163 (45.4)	196 (54.6)		Positive	191 (36.1)	338 (63.9)	
Posphorylated AKT				P53			
Negative	59 (39.9)	89 (60.1)	0.861	Negative	278 (44.8)	342 (55.2)	0.002
Positive	182 (39.1)	284 (60.9)		Positive	83 (33.6)	164 (66.4)	
Epidermal growth factor receptor							
Negative	300 (43.2)	395 (56.8)	0.083				
Positive	66 (36.1)	117 (63.9)					
Intrinsic Subtypes				Nottingham prognostic index			
Luminal A	115 (54.0)	98 (46.0)	0.002	Good prognostic group	149 (53.8)	128 (46.2)	< 0.0001
Luminal B HER2 negative	98 (36.3)	172 (63.7)		Moderate prognostic group	182 (38.6)	290 (61.4)	
Luminal B/HER2 positive	16 (39.0)	25 (61.0)		Poor prognostic group	51 (34.5)	97 (65.5)	
HER2 enriched BC	57 (36.5)	99 (63.5)					
TNBC	37 (42.0)	51 (58.0)					
Values in bold are statistically significant.							

4.4.6 Outcome Analysis for HSP90 α Protein Expression

Cases with high HSP90 α protein expression had significantly shorter BCSS than those with low HSP90 α expression (HR = 1.45; 95% CI = 1.17–1.89, p = 0.005; Figure 4.4a). In addition, cases with high HSP90 α expression displayed significantly shorter DMFS (HR = 1.35; 95% CI 1.65–1.74, p = 0.013; Figure 4.4b). High expression of HSP90 α in the luminal B/HER2-negative cohort was associated with shorter BCSS and DMFS (p = 0.002 and p = 0.009, respectively; Figure 4.4c–d). There was a similar trend observed in the TNBC cases, while there was no association with outcomes in the other molecular classes. The multivariate Cox regression analysis for HSP90 α , adjusting for standard prognostic parameters, including tumour size, nodal stage and LVI, showed that high HSP90 α protein expression was an independent poor prognostic factor for shorter BCSS (HR = 1.38; 95% CI = 1.06–1.80, p = 0.018) and DMFS (HR = 1.32; 95% CI = 1.03–1.69, p = 0.027; Table 4.4).

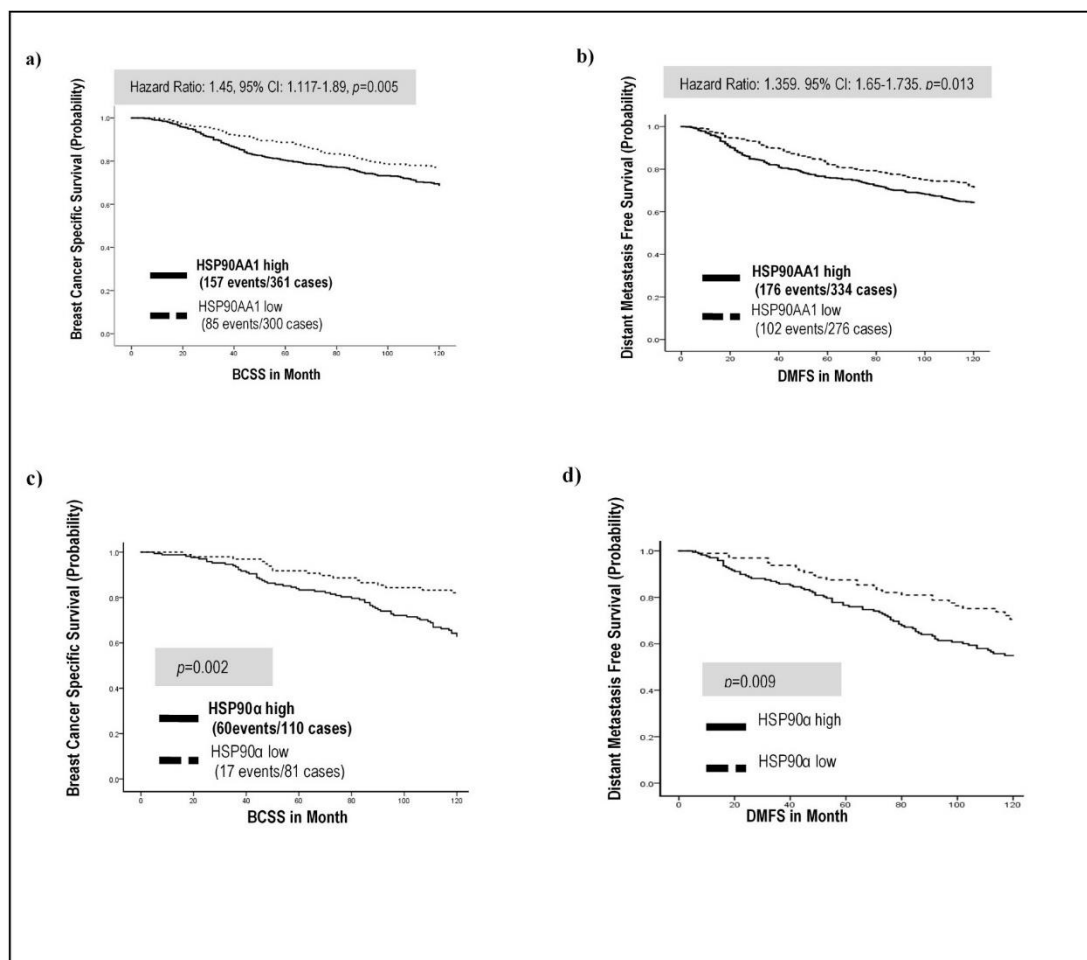


Figure 4.4: Cumulative survival of BC patients stratified by HSP90α protein expression. BCSS (a) and DMFS (b) plots of cytoplasmic HSP90α revealed that high expression was associated with poor patient outcome. Breast Cancer Specific Survival (c) and distant metastasis-free survival (d) plots of cytoplasmic HSP90α revealed that high expression was associated with poor patient outcome in Luminal B HER2/negative subtype.

Table 4.4: Multivariate Cox regression analysis for predictors of BCSS and DMFS based on clinicopathological factors and protein expression of HSP90 α in the Nottingham breast cancer cohort.

Factors	Multivariate analysis BCSS			Multivariate analysis DMFS		
	Hazard Ratio	95% CI	<i>p</i> -value	Hazard Ratio	95% CI	<i>p</i> -value
HSP90α expression	1.378	1.056-1.798	0.018	1.321	1.032-1.690	0.027
Tumour size	1.62	1.225-2.136	0.001	1.725	1.338-2.224	<0.0001
Nodal stage	1.914	1.579-2.320	<0.0001	2.723	1.919-3.864	<0.0001
Lymphovascular invasion	1.46	1.106-1.905	0.007	0.647	0.387-1.081	0.096
Values in bold are statistically significant						

4.5 Discussion

Breast cancer is a group of heterogeneous diseases that differ in morphological, molecular, and clinical behaviour. This heterogeneity poses challenges in understanding the biology and the mechanisms underlying BC progression (73). BC risk stratification and prediction of the usefulness of different therapeutic protocols currently depend on the tumour clinicopathological criteria in addition to clinically validated biomarkers such as ER, PR and HER2. However, substantial proportions of patients with apparently similar clinicopathological features can show variable outcomes and different responses to the same therapeutic regimens. The HSP family of proteins have essential roles in multiple oncogenic signalling pathways through acting to maintain cellular proteins in their correct configurations to ensure their stability (286, 287). Various previous approaches identified Hsp90 as a possible tumour target oncogenic behaviour, and targeted cancer treatments for Hsp90 have been developed (303-306). To date, clinical studies have investigated at least 18 potential inhibitors with favourable pharmacological features, notably ganetespib (STA-9090); however, none has demonstrated sufficient effectiveness to be authorised by the FDA (307). HSP90 α controls protein assembly with proven extracellular roles in wound healing and cancer (299, 300). Therefore, the functions of HSP90 α may confer a favourable mechanism used by cancer cells to facilitate cancer cell migration and metastasis (301, 302). Thus, the expression of HSP90 α in multiple BC datasets at mRNA and protein levels was explored to correlate with clinicopathological parameters and prognostic utility. The current study shows that high expression of HSP90 α is associated with poor prognostic parameters and an independent prognostic marker of poor clinical outcome.

The current study results showed that high expression of *HSP90AA1* mRNA was significantly associated with clinicopathological variable characteristics of poor prognosis, including high histological grade, higher NPI prognostic scores, and loss of hormone receptor expression. Importantly, the expression was associated with shorter patients' survival, independent of other variables. These results are in line with the findings from a recent study investigating the mRNA expression of six HSPs in BC and reported an association between *HSP90AA1* and poor prognosis (285). Patients with non-small cell lung cancer and lung adenocarcinoma showed that high mRNA expression of *HSP90AA1* is associated with poor outcome (308).

The positive correlation of HSP90 α with PIK3CA and p53 expression, (309, 310) indicates that high HSP90 α expression is associated with aggressive BC. PIK3CA is an EMT-promoting protein that contributes to the cell cycle control and interacts with HSP90 leading to apoptosis inhibition (283, 311). During the cell cycle, the active form of PIK3CA can phosphorylate the regular component of the membrane PIP2 to PIP3, which in turn activates AKT protein kinase B that interacts with HSP90 leading to inhibition of apoptosis. *in vivo* studies have reported the increased sensitivity of the cells to an apoptosis-inducing stimulus by inhibiting AKT-HSP90 binding (312, 313). These results were supported by another study that showed HSP90 inhibition enhances AKT dephosphorylation process (314). *In vitro* and *in vivo* experiments have also concluded that PI3K/AKT is a major contributing signalling pathway in Burkitt lymphoma (309). In addition, numerous cellular processes are dependent on the PI3K signalling pathway, which is frequently altered in BC, resulting in increased tumour development and decreased patients survivability (238). In BC, the PI3K/Akt, the target of the rapamycin mTOR pathway, is linked to developing resistance to endocrine treatment, cytotoxic treatment, and HER2-directed treatment (315). Wild-type p53 is well known as a tumour suppressor, whereas mutant p53 seems to acquire an oncogenic phenomenon (316, 317). Interestingly, the activation of HSP90 expression induces stabilisation of mutant p53, consequently promoting tumour progression and metastases through the Hsp90-p53 co-binding mechanism (310, 318).

Lymph node positivity and LVI in BC are strong prognostic factors in predicting patient outcome (319). In this study, HSP90AA1 mRNA and protein expression was strongly associated with LVI positivity and positive nodal status. This association indicate a certain role of HSP90AA1 in LVI, and to help us understand this role then must understand the long journey of tumour cell during the invasion. The tumour cell journey starts when detaching from

the origin site, migrating, and invading the neighbour tissue, and invasion into LV through the ECM to reach the distant organ where the extravasation process occurs and establishes metastasis (97-99). In addition, invasive tumours are known for harbouring the degradation of ECM using MMPs. This degradation permits tumour dissemination by permitting entry to the vascular and lymphatic systems (320). MMP expression and activation are elevated in nearly all human malignancies. (321). Among the MMPs family, MMP-2 and MMP-9 play a critical role in the development and progression of tumours (322, 323). Interestingly, HSP90AA1 was found to interact with MMP-2 and 9 in BC through localising and enhancing their proteolytic activity in several studies (324, 325). A significant positive correlation between LVI and MMP-9 has also been reported (149, 150). A study demonstrated that mAb 4c5 prevents MMP-2 and MMP-9 activation by inhibiting their communication with extracellular HSP90AA1 leading to preventing metastatic BC cells aggregation (325). Considering the role of MMP-2 and 9 in the invasion and the strong association between them and HSP90AA1 and LVI, we could assume that MMPs are the key link between HSP90AA1 and LVI in the metastatic cascade. Therefore, inhibiting HSP90AA1 activities could aid in preventing LVI occurrence in BC patients and improve outcomes.

Luminal B, TNBC and HER2-enriched BC molecular subtypes are biologically and clinically aggressive subtypes compared to luminal A tumours (326-328). This study results showed that HSP90 α is less expressed in the Luminal-A BC molecular subtype than the Luminal-B / HER2-negative, Luminal-B / HER2-positive, HER2 enriched and TNBC classes. This could reflect the potential interaction between HSP90 α expression, tumour proliferation, and HER2 signalling pathway, which may trigger invasion and LVI resulting in BC metastasis (329). In line with current study findings, clinical trials of HSP90 inhibitors have shown promising results in the Luminal B and HER2-positive tumours (330).

In conclusion, the current results highlighted the significant association of HSP90 α with poor prognosis in BC patients and especially in the intrinsic subtypes of aggressive behaviour, including Luminal B. Functional studies are required to determine the influence of aberrant expression of HSP90 α and to explore its utility as a potential therapeutic target for precision medicine.

**5 Chapter 5: SLC9A3R1, a
biomarker that was chosen from
the LVI-related gene signature set**

5.1 Overview

The BC overview in this study illustrated how IBC is a heterogeneous disease encompassing different morphological and molecular characteristics, leading to a wide range of patient prognoses, outcomes and responses to therapy (9). In addition, death from BC is often a result of metastasis. Thus, several attempts were carried out to explore the underlying molecular mechanisms involved in BC progression. However, a key factor responsible remains to be elucidated (56, 331, 332): LVI occurs when the tumour emboli appear in LVs around the primary tumour (121). Several studies have demonstrated a prognostic value of LVI, suggesting that LVI presents a risk of axillary LNM and distant metastasis (65, 115, 333). Therefore, LVI could act as a potential therapeutic target in BC; however, such targets associated with LVI status in BC have yet to be identified.

Based on mRNA expression data, our previous study identified a novel DEG signature in both LVI+ and LVI- cases (223). Solute carrier family 9, subfamily A, member 3 regulator 1 (SLC9A3R1) was observed as one of the genes that is positively correlated with LVI status. It is located on chromosome 17q25.1, consists of six exons and encodes Na⁺/H⁺ exchanger regulatory factor-1. Moreover, SLC9A3R1 functions as an interface between plasma membrane proteins and the ERM membrane-binding phosphoprotein 50 family, connecting them to the actin cytoskeleton controlling their surface expression (334).

Wnt signalling is a critical pathway for the progression of various biological mechanisms such as cell polarity, cell migration, and patterning throughout embryonic development, adult tissue homeostasis and regeneration of diverse tissues and organs. (335). Genome-wide sequencing and gene expression profile analysis have shown that Wnt signalling is primarily implicated in the progression and dissemination of BC (336). In addition, it is reported that this pathway is also essential for the modulation of the BC immune microenvironment, stem cell maintenance, treatment resistance and phenotype formation (336, 337). β -catenin is a multifunctional protein considered an adhesion molecule that plays a fundamental role in the normal physiological microenvironment (338). It is well known as a transcriptional factor in Wnt signalling, which plays a critical role in cell development, differentiation, stem cell renewal and apoptosis (338). Abnormal β -catenin expression is linked to enhancing carcinogenesis, and its aberrant activity has been seen in various cancer including BC (339, 340). In normal conditions, β -catenin is expressed in the cytoplasm at a low level (341). However, the activation of Wnt signalling dysregulated the β -catenin degradation components leading to an elevation of β -catenin level in the cytoplasm (337, 342).

Existing evidence indicates that SLC9A3R1 is involved in the Wnt signalling pathway, which triggers the adherens junction complex constituent β -catenin to aggregate in the cytoplasm, subsequently translocation into nucleolus (339), promoting to carcinogenesis (343). In addition, plasma membrane calcium ATPase-2 (PMCA2) which maintains mammary epithelial cell survival and controls cell development and death, is also influenced by SLC9A3R1 overexpression in BC (344). Moreover, SLC9A3R1 has been linked to the activity of cellular proteins, autophagy and modulation of BC proliferation (345). A correlation also exists between SLC9A3R1 and growth factor tyrosine kinase receptors participating in cancer progression (346-348). However, the biological and prognostic significance of SLC9A3R1 expression in BC cases is still unclear.

5.2 Aim and Hypothesis

Several reports have suggested that SLC9A3R1 overexpression modulates tumour growth, migration and metastasis in breast, endometrial, lung, prostate and colon cancers (346, 349-353). However, there is a lack of evidence in all these studies regarding the combined prognostic value of SLC9A3R1 and LVI positivity modulating BC invasiveness. Thus, the hypothesis is that SLC9A3R1 plays an essential role in BC LVI progression through its contribution to several vital intrinsic subtypes and pathways, consequently leading to poor patient outcomes. Therefore, this study aimed to assess SLC9A3R1 expression, identified by the previously mentioned gene signature (**chapter 3**), and evaluate its clinical significance with related mRNA and protein biomarkers in BC and its association with LVI.

5.3 Materials and Methods

5.3.1 Study Cohorts

To explore *SLC9A3R1* mRNA expression in breast tissue, gene expression TPM values from the TNM plot database were used. The clinical significance, clinicopathological associations and CNA of *SLC9A3R1* mRNA expression were assessed using the METABRIC dataset of IBC (n = 1,980) and the BC cohort retrieved from the cBioPortal for Cancer Genomics utilising the TCGA-BRCA cohort (n = 854) for which LVI data were available. Both cohorts were used as the discovery. In addition, the bc-GenExMiner v4.9 dataset (Affymetrix; n = 4,904) was investigated to explore the role of SLC9A3R1 and patient outcomes for further validation. More details about the methods and clinical materials used in this study are given in **Chapter 2 (section 2.1.1)**. Furthermore, SLC9A3R1 protein expression was immunohistochemically evaluated in a well-characterised BC cohort (n = 696), which was used as a validation cohort. An overview of the clinicopathological parameters of patients from Nottingham cohort is summarised in **Chapter 2 (section 2.1.2.1)**.

5.3.2 Western Blotting and Immunohistochemistry

The specificity of SLC9A3R1 protein (rabbit monoclonal ab109430; Abcam, Cambridge, UK) was validated using WB on human BC cell lysates harvested from the MCF7 and SKBr3 cell lines. The cell lines were chosen according to the manufacturer's recommendation, protein atlas and availability. The housekeeping protein mouse β -actin antibody (A5441, clone AC-15; Sigma-Aldrich, Gillingham, UK) was used at a dilution of 1:5,000. This procedure was done in keeping with the protocol of WB validation for specific Ab–Ag binding given in **Chapter 2 (section 2.3)**.

The TMA slides were stained according to the IHC procedure previously described in **Chapter 2 (section 2.4.1)**. Two scorers (the author and Y. K) calculated the inter-observer concordance in 25% of the cores, and they scored the data blindly (ICC = 0.74), indicating an excellent concordance between scorers.

5.3.3 Statistical Analysis

The statistical analysis was conducted via SPSS 28.0 (IBM, Chicago, IL, USA). The data dichotomisation used for *SLC9A3R1* mRNA and protein expression was based on the median. The chi-squared test was used to evaluate the associations between *SLC9A3R1* expression and clinicopathological parameters. Moreover, Kaplan–Meier plots were employed to undertake the univariate survival analysis with adjustment of covariates to investigate the independence of *SLC9A3R1* expression with standard prognostic factors in BC. Cox regression and log-rank test were used in the multivariate analysis. Statistical significance was reflected by a *p*-value <0.05 (two-tailed). Additionally, a subset of the Nottingham series incorporated in METABRIC (n = 150/1,980) was used to examine the correlation between *SLC9A3R1* mRNA and *SLC9A3R1* protein expression using Pearson’s correlation test. This test was also used to validate the correlation between *SLC9A3R1* mRNA expression and genes of interest. Further statistical data are detailed in **Chapter 2 (section 2.11)**.

5.4 Results

5.4.1 Western Blot Evaluation

The anti-*SLC9A3R1* primary antibody’s WB images showed a single specific band at the anticipated MW (48 kDa; Figure 5.1). Using housekeeping protein mouse β -Actin was sufficient to provide evidence of the antibody specificity for each cell lysate.

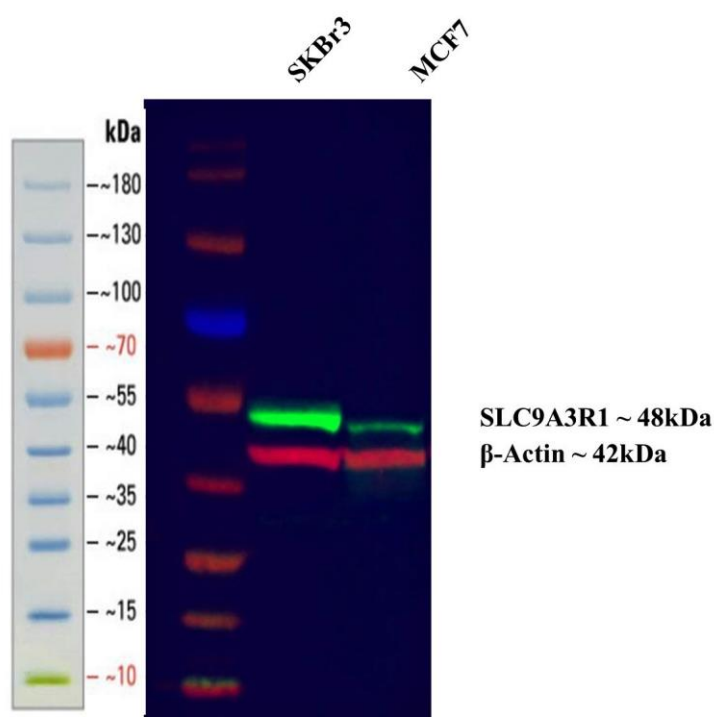


Figure 5.1: Western blotting validation indicated SLC9A3R1 primary antibody specificity based on a single band observed at approximately 48 kDa.

5.4.2 *SLC9A3R1* mRNA Expression and Clinicopathological Significance

To examine the prognostic value of *SLC9A3R1* mRNA expression in breast tissue, this study carried out an in-depth investigation using gene expression data from the TNM plot portal, and high *SLC9A3R1* mRNA expression was observed in tumoural and metastatic samples of BC when compared to adjacent normal breast tissue samples ($n = 403$) (Figure 5.2). Furthermore, when investigating CNA data within the METABRIC and TCGA-BRCA cohorts, *SLC9A3R1* CN gain in the METABRIC cohort was observed in 214/1,891 (11.3%) cases, TCGA-BRCA 207/830 (24.9%) than in neutral and those showing a loss ($p < 0.0001$) (Table 5.1 and Table 5.2). High *SLC9A3R1* mRNA expression in the METABRIC and TCGA-BRCA cohorts was significantly associated with LVI+, ER+, PR+ and HER2 positive tumours. In addition, using the PAM50 classification of BC tumours, *SLC9A3R1* mRNA was significantly associated with the luminal B subtype in the METABRIC cohort (Table 5.3). Also, high *SLC9A3R1* expression was associated with high Ki67 and AKT levels (Table 5.4). High *SLC9A3R1* mRNA also showed a significant correlation with luminal B in the TCGA-BRCA dataset ($p < 0.0001$) using bc-GenExMiner v4.9 (Figure 5.3).

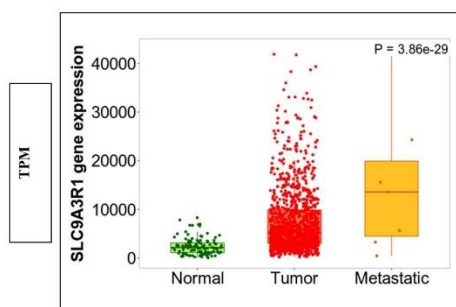


Figure 5.2: The *SLC9A3R1* mRNA gene expression in breast cancer tissue based on sample type in RNA-Seq Data. (*p*-value is a comparison for multiple independent samples).

Table 5.1: The *SLCA93R1* copy number alteration in the METABRIC cohort.

CNA region for <i>SLCA93R1</i>	CNA	Cases with reduced <i>SLCA93R1</i> mRNA expression (%)	Cases with High <i>SLCA93R1</i> mRNA Expression (%)	Total (%)	p-value
Chromosome 17q25.1	Loss	29 (79)	8 (21)	37 (100)	<0.0001
	Gain	48 (18)	214 (82)	262 (100)	
	Neutral	830 (52)	762 (48)	1592 (100)	
	Total	907 (48)	984 (52)	1891 (100)	

Table 5.2: The *SLCA93R1* copy number alteration in the TCGA-BRCA cohort.

CNA region for <i>SLCA93R1</i>	CNA	Cases with reduced <i>SLCA93R1</i> mRNA expression (%)	Cases with High <i>SLCA93R1</i> mRNA Expression (%)	Total (%)	p-value
Chromosome 17q25.1	Loss	107 (83)	22 (17)	129 (100)	<0.0001
	Gain	91 (31)	207 (69)	262 (100)	
	Neutral	219 (54)	184 (46)	403 (100)	
	Total	417 (50)	413 (50)	830 (100)	

Table 5.3: Association of *SLC9A3R1* mRNA expression with clinicopathological characteristics in the METABRIC (n = 1980) and TCGA-BRCA (n = 854) cohorts.

Factors	METABRIC			TCGA-BRCA		
	Low <i>SLC9A3R1</i>	High <i>SLC9A3R1</i>	<i>p</i> -value	Low <i>SLC9A3R1</i>	High <i>SLC9A3R1</i>	<i>p</i> -value
	N (%)	N (%)		N (%)	N (%)	
Lymphovascular invasion						
Negative	494 (53.1)	436 (46.9)	0.010	296 (53.0)	263 (47.0)	0.018
Positive	295 (46.5)	340 (53.5)		131 (44.4)	164 (55.6)	
Tumour size						
≤2.0 cm	354 (56.9)	268 (43.1)	<0.001	128 (53.6)	111 (46.4)	0.195
>2.0 cm	625 (46.7)	713 (53.3)		299 (48.6)	316 (51.4)	
Age						
≤50 years	253 (59.7)	171 (40.3)	<0.001	142 (54.8)	117 (45.2)	0.063
>50 years	737 (47.4)	817 (52.6)		285 (47.9)	310 (52.1)	
Nodal Status						
Negative	532 (51.4)	503 (48.6)	0.154	230 (54.0)	196 (46.0)	0.026
Positive	452 (48.2)	486 (51.8)		196 (46.3)	227 (53.7)	
Histological grade						
Grade 1 and 2	472 (49.68)	468 (49.68)	0.011	200 (43.1)	264 (56.9)	<0.001
Grade 3	478 (50.2)	474 (49.8)		205 (58.2)	147 (41.8)	
Oestrogen receptor						
Negative	357 (75.3)	117 (24.7)	<0.001	157 (84.9)	28 (15.1)	<0.001
Positive	633 (42.0)	873 (58.0)		260 (40.7)	379 (59.3)	
Progesterone receptor						
Negative	551 (58.6)	389 (41.4)	<0.001	195 (71.7)	77 (28.3)	<0.001
Positive	739 (42.2)	601 (57.8)		221 (40.5)	325 (59.5)	
Human epidermal growth factor receptor 2						
Negative	886 (51.1)	847 (48.9)	0.008	306 (54.0)	261 (46.0)	<0.001
Positive	104 (42.1)	143 (57.9)		47 (35.3)	86 (64.7)	
Nottingham prognostic index						
Good prognostic group	350 (51.5)	330 (48.5)	0.419			
Moderate prognostic group	548 (49.8)	553 (50.2)				
Poor prognostic group	92 (46.2)	107 (53.8)				
Molecular subtypes						
Luminal A	336 (46.8)	382 (53.2)	<0.001	178 (44)	227 (56)	<0.001
Luminal B	150 (30.7)	338 (69.3)		32 (22.7)	109 (77.3)	
HER2 enriched	94 (39.2)	146 (60.8)		27 (48.2)	29 (51.8)	
Basal-like	281 (85.2)	49 (14.8)		124 (93.2)	9 (6.8)	
Normal-like	128 (64.3)	71 (35.7)		24 (80)	6 (20)	
Values in bold are statistically significant, GPG, good prognostic group; MPG, moderate prognostic group; PPG, poor prognostic group.						

Table 5.4: Association between *SLC9A3R1* mRNA expression and other gene of interest in the METABRIC and TCGA-BRCA invasive breast cancer datasets.

Gene of Interest	<i>SLC9A3R1</i> mRNA Level (METABRIC)		Mean Rank		Z-score	p-value	<i>SLC9A3R1</i> mRNA Level (TCGA-BRCA)		Mean Rank		Z-score	p-value
	Low (%)	High (%)	Low	High			Low (%)	High (%)	Low	High		
<i>Ki67</i>	990 (50)	990 (50)	933	1047	-4.5	<0.001	427 (50)	427 (50)	442	413	-1.757	0.079
<i>PIK3CA</i>			997	995	-2.4	0.018			427	401	-3.143	0.002
<i>AKT</i>			865	1115	-10	<0.001			506	474	-5.471	<0.001

Values in bold are statistically significant .

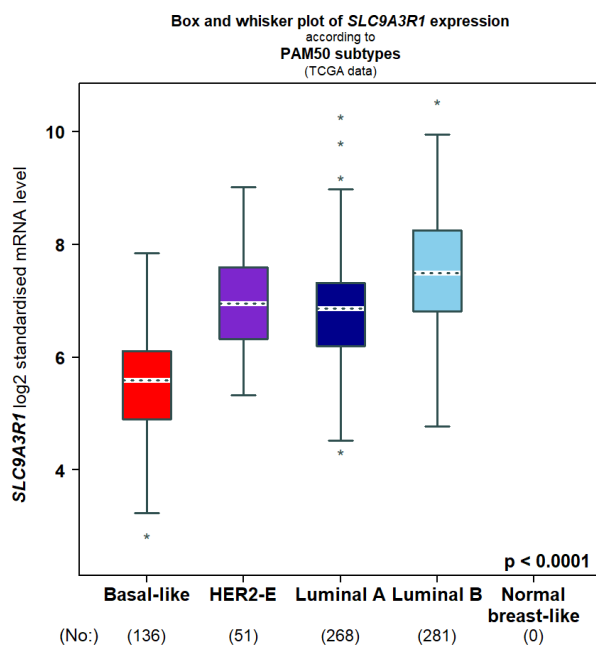


Figure 5.3: The *SLC9A3R1* mRNA gene expression on PAM50 subtypes in TCGA-BRCA data using bc-GenExMiner v4.9.

5.4.3 Outcome Analysis for *SLC9A3R1* mRNA Expression

In the outcome analysis, high expression of *SLC9A3R1* mRNA indicated a trend of poor prognosis features in the prognostic module, High *SLC9A3R1* mRNA expression cases showed a trend of worse BCSS than low *SLC9A3R1* mRNA expression cases in the METABRIC cohort (HR = 1.053, 95% CI = 0.883–1.257, $p = 0.564$; Figure 5.4a) and the TCGA-BRCA cohort (HR = 1.046, 95% CI = 0.686–1.595, $p = 0.835$; Figure 5.4b). Even though all datasets showed poor prognosis, bc-GenExMiner v4.9 (Affymetrix) was the only dataset with a significant association (HR = 1.23, 95% CI = 1.02–1.48, $p = 0.0279$; Figure 5.4c).

While the stratification of *SLC9A3R1* mRNA cases based on LVI status revealed that those with high *SLC9A3R1* mRNA expression favoured poor prognosis implications for the BC patients. However, this finding did not reach statistical significance in the LVI– cases for the METABRIC (HR = 1.269, 95% CI = 0.938–1.718, $p = 0.122$; Figure 5.4d) or TCGA-BRCA cohorts (HR = 0.819, 95% CI = 0.492–1.754, $p = 0.819$; Figure 5.4e). The LVI+ cases showed a similar association in the METABRIC dataset (HR = 0.956, 95% CI = 0.730–1.251, $p = 0.743$; Figure 5.4f) and TCGA-BRCA dataset (HR = 0.995, 95% CI = 0.564–1.756, $p = 0.986$; Figure 5.4g).

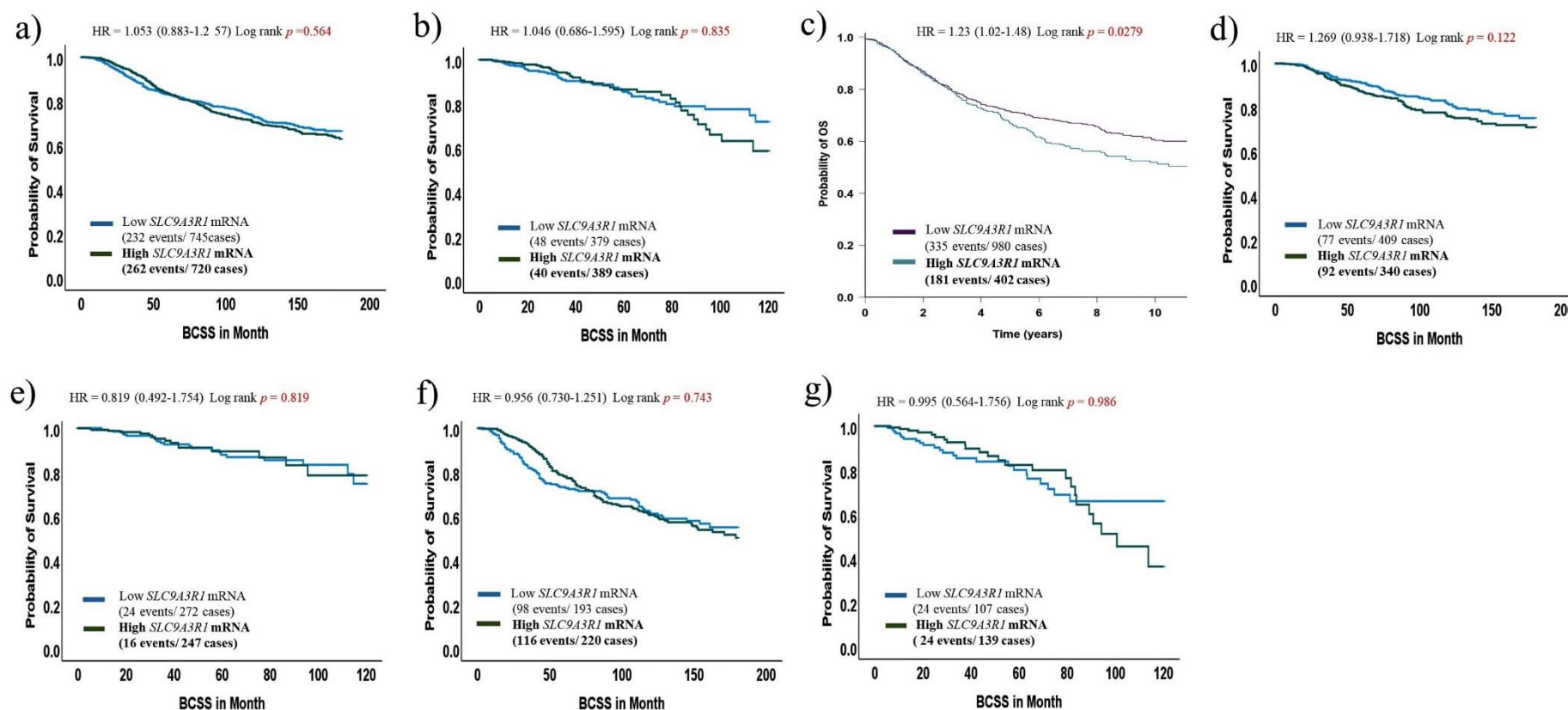


Figure 5.4: The Kaplan-Meier survival plots showing cumulative survival of patients with breast cancer stratified by *SLC9A3R1* mRNA expression. METABRIC (a), TCGA-BRCA (b) Breast Cancer Gene-Expression Miner (Affymetrix) (c), LVI- in METABRIC (d), LVI- in TCGA-BRCA (e), LVI+ in METABRIC (f) and LVI+ in TCGA-BRCA (g). OS, overall survival; BCSS, breast cancer-specific survival.

5.4.4 SLC9A3R1 Protein Expression

The full-face invasive tumour tissue sections showed homogenously distributed staining of SLC9A3R1. The IHC staining was conducted on the Nottingham BC cohort TMAs using anti-SLC9A3R1 primary antibody, as previously described in **Chapter 2 (section 2.4.1)**. The immunoreactivity of SLC9A3R1 expression was mainly observed in the cytoplasm. The median H-score (0) (range 0-160) was used as the basis for dichotomising SLC9A3R1 protein expression (n = 696) (Figure 5.5a and b, respectively).

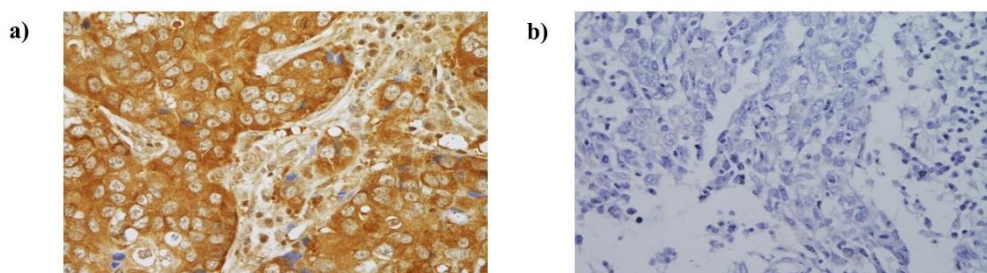


Figure 5.5: Immunohistochemically protein expression for SLC9A3R1 in tissue microarray cores of breast cancer tissue. Invasive cancer cells (intense staining in the cytoplasm of cancer cells) **(a)**; and no staining **(b)** (images are at 20× magnification).

5.4.5 SLC9A3R1 Protein Expression and its Association with the Clinicopathological Parameters

There was a positive linear correlation between SLC9A3R1 protein and *SLC9A3R1* mRNA expression (Spearman's coefficient: 0.5; $p < 0.001$). High SLC9A3R1 protein expression was significantly associated with LVI positivity ($p = 0.001$), larger tumour size ($p = 0.032$), ER positivity ($p < 0.001$), PR positivity ($p = 0.030$), HER2 positivity ($p = 0.015$) and high Ki67 levels ($p = 0.017$). High SLC9A3R1 protein expression was also associated with high levels of PIK3CA ($p = 0.046$) and pAKT ($p = 0.018$). High SLC9A3R1 expression showed a negative correlation with triple-negative BC ($p < 0.001$) (Table 5.5) and showed a significant association with intrinsic molecular subtypes, particularly with luminal B and HER2 ($p < 0.001$) (Table 5.5).

Table 5.5: Relationship of SLC9A3R1 protein expression with clinicopathological characteristics in the Nottingham cohort.

Factors	Expression of SLC9A3R1			Factors	Expression of SLC9A3R1		
	Low	High	p-value		Low	High	p-value
	N (%)	N (%)			N (%)	N (%)	
Tumour size				Age			
≤2 cm	204 (59.3)	140 (40.7)	0.032	≤50 years	140 (54.3)	118 (45.7)	0.654
>2 cm	176 (51.2)	168 (48.8)		>50 years	242 (56)	192 (44)	
Histological grade				Nodal status			
1,2	185 (57.3)	138 (42.7)	0.256	Negative	238 (57.9)	173 (42.1)	0.086
3	196 (53.0)	174 (47.0)		Positive	142 (51.3)	135 (48.7)	
Lymphovascular invasion				Oestrogen receptor			
Negative	269 (61.3)	177 (39.7)	0.001	Negative	120 (67.4)	58 (32.6)	<0.001
Positive	111 (46.3)	129 (53.8)		Positive	262 (50.9)	253 (49.1)	
Progesterone receptor				Human epidermal growth factor receptor-2			
Negative	176 (60.3)	116 (39.7)	0.030	Negative	337 (57.6)	248 (42.4)	0.015
Positive	202 (51.9)	187 (48.1)		Positive	38 (43.7)	49 (56.3)	
Ki67				Phosphoinositide-3-kinase, Catalytic, Alpha polypeptide			
Low	119 (63.3)	69 (36.7)	0.017	Negative	82 (63.1)	48 (36.9)	0.046
High	141 (52.0)	130 (48.0)		Positive	209 (53.0)	185 (47.0)	
Phosphorylated AKT				Triple-negative			
Negative	70 (61.9)	43 (38.1)	0.018	Non-TNBC	278 (50.1)	277 (49.9)	<0.001
Positive	190 (49.4)	195 (50.6)		TNBC	98 (78.4)	27 (21.6)	
Intrinsic Subtypes				Nottingham prognostic index			
Luminal A	55 (53.4)	48 (46.6)	<0.001	Good prognostic group	118 (61.1)	75 (38.9)	0.065
Luminal B	66 (48.9)	69 (51.1)		Moderate prognostic group	201 (54.5)	168 (45.5)	
HER2 enriched BC	26 (44.1)	33 (55.9)		Poor prognostic group	61 (48.0)	66 (52.0)	
TNBC	98 (78.4)	27 (21.6)					
Values in bold are statistically significant.							

5.4.6 Outcome Analysis for SLC9A3R1 Protein Expression

For the BCSS outcome analysis, cases with high SLC9A3R1 expression showed significantly shorter survival than those with low expression (HR = 1.523; 95% CI = 1.137–2.040, $p = 0.004$; Figure 5.6a). The categorisation of SLC9A3R1 protein expression based on LVI status revealed a results that favoured a poor prognosis association with high SLC9A3R1 expression (LVI–, HR = 1.227; 95% CI = 0.799–1.883, $p = 0.348$; Figure 5.6b); however did not reach statistical significance and a significant association with survival considering LVI+ tumours (HR = 1.593; 95% CI = 1.052–2.412, $p = 0.026$; Figure 5.6c).

The SLC9A3R1 multivariate Cox regression analysis, correlated with standard prognostic parameters, including LVI, tumour size and age, revealed that high SLC9A3R1 protein expression was a significant independent predictor for shorter BCSS (HR = 2.047; 95% CI = 1.516–2.763, $p = 0.042$; Table 5.6).

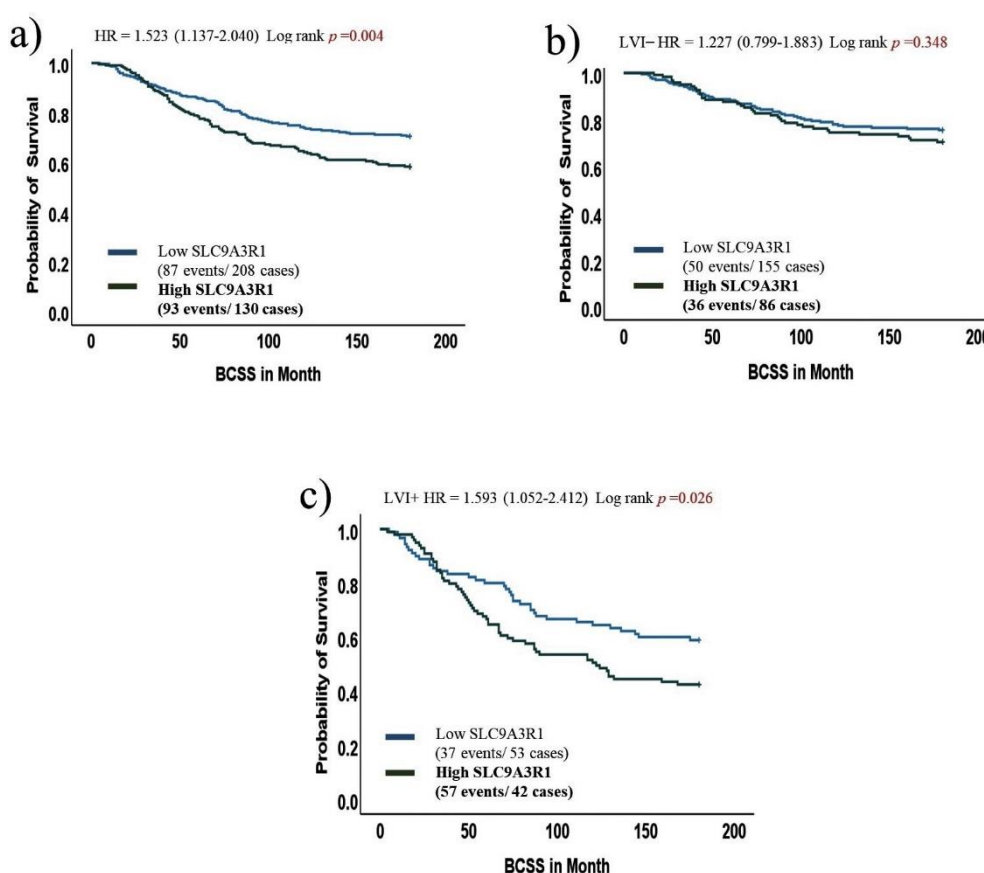


Figure 5.6: Kaplan-Meier survival plots showing breast cancer specific survival of BC patients stratified by SLC9A3R1 expression in Nottingham cohort. SLC9A3R1 protein expression revealed that high expression was associated with poor patient outcome. Breast cancer specific survival in the Nottingham cohort (a), in the LVI– cases (b) and the LVI+ cases (c).

Table 5.6: Multivariate Cox regression analysis of clinicopathological factors and protein expression of SLC9A3R1 to predict breast cancer specific survival.

Factors	Multivariate analysis BCSS		
	Hazard Ratio	95% CI	<i>p</i> -value
SLC9A3R1 expression	1.359	1.012-1.824	0.042
Lymphovascular invasion	2.047	1.516-2.763	0.001
Tumour size	1.822	1.334-2.489	<0.001
Age	1.320	1.012-1.824	0.072

5.5 Discussion

Breast cancer is a series of molecularly and histopathologically distinct conditions with different clinical outcomes and treatment responses (68). Owing to this heterogeneity, it is challenging to elucidate the biological processes and mechanisms that underpin the progression of BC (73). The LVI is a critical step in BC invasiveness as it presents a risk of axillary LNM in patients with BC (65, 115, 333). However, the molecular mechanisms associated with LVI in BC and the genes behind this phenomenon are still poorly understood.

Based on the previous gene signature study (223), *SLC9A3R1* has been indicated to be significantly associated with LVI positivity. Moreover, *SLC9A3R1* overexpression has been linked to aggressive tumour behaviours and poor outcomes in different cancers (346, 349-353). However, data on *SLC9A3R1*'s prognostic value in BC have yet to be defined. Previously, increased *SLC9A3R1* expression was observed more commonly in BC as compared to normal breast tissue using immunohistochemical evaluation, suggesting that *SLC9A3R1* might contribute to BC progression (354). However, knowledge remains limited about the correlation between *SLC9A3R1* and LVI. Therefore, the present work aimed to evaluate expression of *SLC9A3R1* and association with established LVI+ status in a large BC cohort to assess the viability of *SLC9A3R1* for prognostic purposes.

One of the primary features of oncogenic transformation is DNA amplification; *SLC9A3R1* is located on chromosome 17q25.1. This study observed more CN gain than CN loss in the METABRIC and TCGA-BRCA cohorts. Furthermore, the *SLC9A3R1* CN gain showed a strong positive correlation with its corresponding mRNA and protein. Thus, amplification may be a common mechanism leading to increased expression of *SLC9A3R1* in BC, and detection of *SLC9A3R1* gene amplification may also be used as a clinical marker for cancer metastasis.

In agreement with earlier studies on BC (350, 355), this study found that higher *SLC9A3R1* expression in the cytoplasm was correlated not only with poor prognosis, aggressive clinical

parameters, such as high Ki67 levels, and aggressive types of BC, such as luminal B and HER2-enriched subtypes, but also with evidence of LVI at both the mRNA and protein levels, which in turn is related to BC progression. Consistent with this, studies have shown an increased reduction of proliferation and migration capacities in prostate cancer cells and HCC after the SLC9A3R1 knockdown (352, 356). Moreover, according to findings at the in vitro level, a correlation exists between SLC9A3R1 overexpression and BC aggressiveness in the MDA-MB-435 and MCDF-7 cell lines, which can promote invasion of BC cells and activate an invasive phenotype, either on its own or alongside changes in the tumour metabolic microenvironment, (355) which may trigger invasion and LVI in accordance with current study findings at both the transcriptomic and protein levels.

Regarding *SLC9A3R1* mRNA and protein levels, the ER–ve and TNBC subtypes had the lowest levels, whereas the ER+ and HER2 positive subtypes had higher levels. Evidence supports a close correlation between the HER2 signalling pathway and the tumour microenvironment (326). High-proliferating ER+ve and HER2 positive tumours is the main characteristic of the luminal B BC subtype, associated with poor prognosis (327). Furthermore, high SLC9A3R1 expression has been associated with high-proliferation Ki67, which is encountered mainly in the aggressive forms of luminal B tumours with a lower sensitivity rate to endocrine therapy (357) – a relationship that is in accordance with current study findings. These results are compatible with other studies, which have shown that SLC9A3R1 is expressed in luminal B (358) and HER2 positive BC tumours (350). Interestingly, SLC9A3R1 is located in the same chromosome as *ErbB2*, and this may explain the positive correlation of SLC9A3R1 with HER2-enriched in this study. It has been suggested that overexpression of *ErbB2* in BC cells can elevate VEGF levels and increase MMP-9 activity, promoting angiogenesis (359, 360). Furthermore, a correlation has been found between *ErbB2* overexpression and the number of LN metastases in patients with BC with positive nodal status (361). The overexpression of *ErbB2* could manipulate MMP-9 and MMP-2 protease activities, which are correlated with LVs invasiveness, LN dissemination and tumour recurrence (359, 362-364). Clinically, detectable LVI in the primary BC correlates dramatically with LN and DM (365). Together, these results indicate the role of SLC9A3R1 in tumour development and progression, which is in agreement with this study's observations.

Additionally, HER2/neu-overexpressing breast carcinoma displays co-localisation of SLC9A3R1 and the oncogenic receptor HER2/neu (366). One study noted that PMCA2 and HER2 are expressed in lower levels after *SLC9A3R1* knockdown, which also suppresses HER2

signalling and causes HER2 detachment from HSP90 as well as HER2 internalisation and ubiquitination (344), a mechanism that has been identified as playing a critical role in endocytosis and workability of the anti-HER2 antibody-drug conjugate (367). It can thus be deduced that HER2 signalling modulation and membrane preservation in BC likely rely on the integrated action of SLC9A3R1 and PMCA2 (368). Furthermore, another study suggested an association between HER2 and Wnt signalling pathways in promoting the proliferation of mammary tumours (369) pathways influenced by SLC3A3R1 overexpression, as recently mentioned. Moreover, β -catenin, a cell adhesion molecule, seems to be altered by Wnt, which leads to β -catenin translocation to the nucleus (339) to function as an oncogene in breast tissue (340).

Therefore, these indirect interactions between SLC9A3R1 and BC subtypes may alter cell–cell adhesion and stimulate breast tumour development, promoting LVI and invasion. This connection could potentially be exploited by BC treatments.

Notably, the PI3K/AKT pathway is associated with BC cancer progression and metastasis (370, 371). Therefore, PI3K/AKT deregulation is involved in tumorigenesis and maintains cellular functions including, metabolism, autophagy, apoptosis, and pAKT and *PIK3CA* mutations (372-374). The current study demonstrated that high expression of SLC9A3R1 protein is associated with high levels of PI3K and pAKT. Furthermore, a previous study on BC showed that SLC9A3R1 PTEN functional interaction inhibits the PI3K/AKT pathway (375). The SLC9A3R1 interactions above imply its vital role in manipulating the cell cycle and inducing genomic instability to facilitate the LVI mechanism.

While this study uncovered promising results indicating that SLC9A3R1 may have a possible role in BC LVI progression depending on the data at the transcriptomic and proteomic levels, it possesses some limitations. While this study was based on retrospective cohort data with long clinical follow-up time, multiple data resources are needed to validate these results. Some TMA cores were missing or had barely detectable tumours in the TMA sections used in this study. More experimental studies should be performed in the future to uncover the molecular mechanisms associated with SLC9A3R1 to understand its prognostic value in BC LVI+ tumours.

In conclusion, our study demonstrated the prognostic value of SLC9A3R1 BC LVI+ tumours. It interacts with most aggressive intrinsic subtypes, including the luminal B, HER2 positive and poor patient outcome subtypes. It contributes to BC-related pathways leading to tumour

proliferation, adhesion, invasion, migration and, consequently, BC progression and metastasis. Efficient studies are needed to determine the correlation of SLC9A3R1 with LVI and validate its therapeutic efficiency.

6 Chapter 6: NOP56, a biomarker that was chosen from the LVI- related gene signature set

6.1 Overview

Breast cancer cells are controlled by highly complex mechanisms that may govern their proliferation and viability. They exhibit multi-factorial characteristics, escaping growth suppressors and the immune system and activating inflammation, invasion and metastasis pathways as well as genomic instability (1). Lymphovascular invasion is a highly unfavourable histological feature for patients with IBC, as it reduces their chances of survival. Moreover, BC is a complex, heterogeneous disease with distinct molecular characteristics (68), which adds more complexity to understanding its mechanisms and histological features, such as LVI. Therefore, obtaining a deep understanding of BC's underlying genetic pathways involved in LVI and BC metastasis through examining the transcriptomic and proteomic expression may be critical to uncovering BC's molecular complexity and improving its clinical outcomes.

Nucleolar protein 5A (*NOL5A/NOP56*) is a protein-coding gene located on chromosome 20.p13 and engaged in pre-ribosomal RNA (rRNA) processing. Furthermore, NOP56 is considered one of the ribonuclear proteins accompanied by NOP58, non-histone chromosome protein 2-like 1 and fibrillarin to form and activate the core protein set of the box C/D small nucleolar ribonucleoprotein complex (376, 377). The complex plays a fundamental role in the ribosome assembly process by increasing the methylation process of rRNA at the 2'-O-ribose and modulating rRNA processing (376, 377). The rRNA processing indicated extreme influence by the small nucleolar RNAs (snoRNAs) that are considered a key factor role in the pre-rRNAs (378). Previous studies had identified that NOP56p was accompanied by NOP77p/NOP4p mutant strains, which were assigned to a primary ribosome assembly defect and a secondary pre-rRNA processing defect (379, 380). Moreover, the genes mediating ribosome synthesis are the only genes in cells that have conserved Myc/Max binding sites from early metazoans through humans (381). A recent study made by Ryu H et al., had suggested that the deubiquitinase USP36 promotes SUMOylation of NOP56 and NOP58 and promotes binding of NOP58 to snoRNA, and knocking out of deubiquitinase USP36 significantly inhibits rRNA processing (382). Interestingly, MYC directly affects the human transcription of RNA Pol I along with other nucleolar proteins, including NOP56 (383). MYC proteins were detected to play a pivotal role in ribosome biogenesis and protein synthesis by controlling the expression of large numbers of proteins involved in this process (384, 385), including the expression of the target gene NOP56 (386, 387).

High expression of NOP56 has been observed in several types of cancers, including lung cancer (388), Burkitt lymphoma (387) and hepatic carcinoma (389). It has been reported that Burkitt

lymphoma associated Myc mutants demands NOP56 to induce cell transformation and tumour growth (387). Cells rely on ribosomes to synthesise proteins, making them perhaps the most important macromolecular machine, and small nucleolar ribonucleoproteins (snoRNPs) are among the most important (390). The snoRNPs play an essential role in cancer transformation, tumorigenesis and metastasis and represent potential therapeutic targets (391, 392). Interestingly, NOP56 may have extra-ribosomal functions, but these have yet to be determined (393).

6.2 Aim and Hypothesis

Based on the results from **Chapter 3** (223), NOP56 is one of the genes in the signature for predicting the presence of LVI. Although NOP56 has been indicated to play a substantial role in ribosomal biology in several types of cancers, its role in BC has yet to be defined. Thus, the hypothesis is that NOP56 plays an essential role in BC LVI progression through its contribution to several vital intrinsic subtypes and pathways, consequently leading to poor patient outcomes. Therefore, this study aimed to evaluate the expression of NOP56 in BC at both the transcriptomic and proteomic levels in various well-characterised BC cohorts with long-term follow-up and to investigate its correlation with patient outcomes and several clinicopathological characteristics, including LVI.

6.3 Materials and Methods

6.3.1 Study Cohorts

As described in **Chapter 2 (section 2.1.1)**, gene expression data were obtained from a TNM plot to evaluate the prognostic significance of *NOP56* mRNA expression in BC as well as in the UALCAN dataset. Moreover, two transcriptomic datasets, METABRIC (n = 1,980) and TCGA-BRCA (n = 854), were used as discovery cohorts to evaluate and investigate the prognostic value of *NOP56* mRNA expression. In addition, the prognostic significance of *NOP56* mRNA expression was assessed using bc-GenExMiner v4.9 (Affymetrix; n = 4,904) online dataset. The expression of the NOP56 protein was also evaluated by IHC in a well-characterised BC cohort (n = 803). The data related to clinicopathological features of the cohort series were previously outlined in **Chapter 2 (section 2.1.2.1)**.

6.3.2 Western Blotting and Immunohistochemistry

The specificity validation of the NOP56 protein (mouse monoclonal AMAb91013, Sigma-Aldrich, Gillingham, UK) was carried out using immunoblotting in human BC lysates obtained from MDA-MB-231 and BT-474, both from the American Type Culture Collection, Rockville, MD, USA. The housekeeping protein rabbit β -actin antibody (MABT523, clone RM112, Sigma-Aldrich, Gillingham, UK) was used at a dilution of 1:5,000. Details regarding the WB validation procedure are available in **Chapter 2 (section 2.3)**.

The TMAs were stained using the IHC procedure previously described in **Chapter 2 (section 2.4.1)**. Two researchers (the author and Y. K) scored 25% of the cores independently in order to calculate the inter-observer concordance. There was a reasonable concordance rate in double-scored cases (ICC = 0.7, $p = 0.019$).

6.3.3 Statistical Analysis

The data were analysed manually and statistically using SPSS 28.0 (IBM, Chicago, IL, USA). The NOP56 mRNA and NOP56 protein expression were classified into two distinct groups based on the median. The chi-squared test was used to study the relationship between protein, mRNA expression and other categorical parameters. Pearson's correlation test was used to evaluate the relationship between NOP56 transcriptomic and protein expression using a portion of the Nottingham series ($n = 224/286$). Then, NOP56 and patient survival correlation were investigated using Kaplan–Meier plots. In the multivariate analysis, Cox regression and log-rank tests were employed. A p -value <0.05 (two-tailed) was used to indicate statistical significance. All statistical data are discussed in detail in **Chapter 2 (section 2.11)**.

6.4 Results

6.4.1 Western Blot Evaluation

The anti-NOP56 antibody showed a specific band at the expected MW of 68 KDa (Figure 6.1), confirming the specificity of the antibody for each cell lysate using housekeeping protein rabbit β -actin.

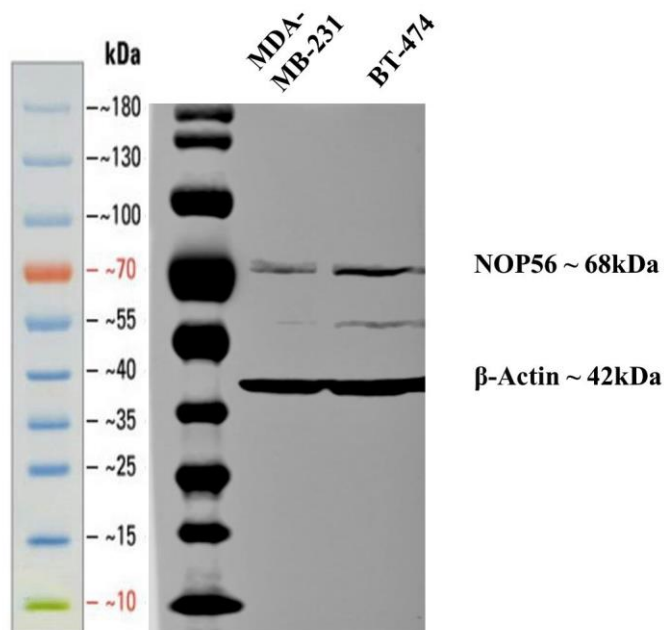


Figure 6.1: The NOP56 primary antibody western blotting validation indicated specificity based on a single band observed at approximately 68 kDa.

6.4.2 *NOP56* Transcriptomic Expression and Clinicopathological Significance

High *NOP56* mRNA expression was observed more in tumoural and metastatic BC as compared to normal breast tissue samples in the TNM plot portal and UALCAN datasets (Figure 6.2). High *NOP56* mRNA expression was significantly associated with LVI+ status in both the METABRIC and TCGA-BRCA cohorts ($p < 0.001$) and with positive nodal stage in the METABRIC cohort ($p = 0.001$). The significant association of high *NOP56* mRNA expression was also observed with other clinicopathological parameters associated with poor patient outcome, including larger tumour size, high grade, and ER- and PR-negative tumours ($p < 0.001$), HER2 positive status in the METABRIC ($p < 0.001$) and TCGA-BRCA cohorts ($p = 0.002$) (Table 6.1) and with high levels of gene of interest (Ki67: $p < 0.001$) and (MYC: $p < 0.001$) (Table 6.2). The expression of *NOP56* mRNA was higher in TNBC and HER2-enriched BC than in other subtypes (Table 6.1; Figure 6.3).

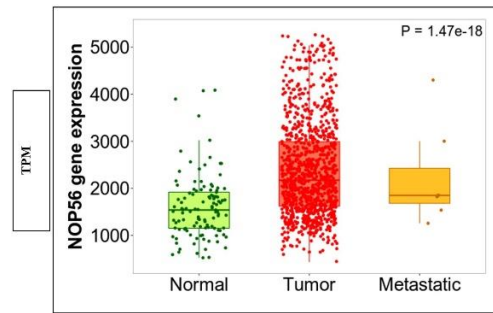


Figure 6.2: The *NOP56* mRNA gene expression in breast cancer tissue based on sample type in RNA-Seq Data. (p -value is a comparison for multiple independent samples).

Table 6.1: Relationship of NOP56 mRNA expression with clinicopathological characteristics in the METABRIC (n = 1980) and TCGA-BRCA (n = 854) cohorts.

Factors	METABRIC			TCGA-BRCA		
	Low NOP56	High NOP56	p-value	Low NOP56	High NOP56	p-value
	N (%)	N (%)		N (%)	N (%)	
Lymphovascular invasion						
Negative	499 (53.8)	428 (46.2)	<0.001	303 (54.2)	256 (45.8)	<0.001
Positive	278 (44.1)	352 (55.9)		124 (42.0)	171 (58.0)	
Tumour size						
≤2.0 cm	358 (57.8)	261 (42.2)	<0.001	137 (57.3)	102 (42.7)	0.008
>2.0 cm	602 (45.3)	728 (54.7)		290 (47.2)	325 (52.8)	
Age						
≤50 years	253 (59.7)	171 (40.3)	<0.001	115 (44.4)	144 (55.6)	0.031
>50 years	737 (47.4)	817 (52.6)		312 (52.4)	283 (47.6)	
Nodal Stage						
Negative	542 (52.6)	489 (47.4)	0.001	218 (51.2)	208 (48.8)	0.560
Positive	422 (45.3)	509 (54.7)		208 (49.2)	215 (50.8)	
Histological grade						
Grades 1 and 2	580 (63.0)	357 (37.0)	<0.001	298 (64.2)	166 (35.8)	<0.001
Grade 3	337 (35.7)	607 (64.3)		105 (29.8)	247 (70.2)	
Oestrogen receptor						
Negative	152 (32.2)	320 (67.8)	<0.001	52 (28.1)	133 (71.9)	<0.001
Positive	815 (54.4)	682 (45.6)		364 (57.0)	275 (43.0)	
Progesterone receptor						
Negative	386 (41.3)	548 (58.7)	<0.001	92 (33.8)	180 (66.2)	<0.001
Positive	581 (56.1)	454 (43.9)		320 (58.6)	226 (41.)	
Human epidermal growth factor receptor 2						
Negative	885 (51.4)	838 (48.6)	<0.001	296 (52.2)	271 (47.8)	0.002
Positive	82 (33.3)	164 (66.7)		50 (37.6)	83 (62.4)	
Nottingham prognostic index						
Good prognostic group	424 (62.4)	255 (37.6)	<0.001			
Moderate prognostic group	479 (43.8)	615 (56.2)				
Poor prognostic group	64 (32.7)	132 (67.3)				
Molecular subtypes						
Luminal A	468 (65.5)	246 (34.5)	< 0.001	263 (64.9)	142 (35.1)	< 0.001
Luminal B	213 (44.1)	270 (55.9)		56 (39.7)	85 (60.3)	
HER2 enriched	84 (35.0)	156 (65.0)		16 (28.6)	40 (71.4)	
Basal-like	85 (25.9)	243 (74.1)		28 (21.1)	105 (78.9)	
Normal-like	116 (58.3)	83 (41.7)		17 (56.7)	13 (43.3)	
Values in bold are statistically significant.						

Table 6.2: Association between *NOP56* mRNA expression and other gene of interest in the METABRIC and TCGA-BRCA invasive breast cancer datasets.

Gene of Interest	<i>NOP56</i> mRNA Level (METABRIC)		Mean Rank		Z-score	p-value	<i>NOP56</i> mRNA Level (TCGA-BRCA)		Mean Rank		Z-score	p-value
	Low (%)	High (%)	Low	High			Low (%)	High (%)	Low	High		
<i>Ki67</i>	967 (49)	1002 (51)	797	1167	-14.4	<0.001	427(50)	427 (50)	362	493	-7.79	<0.001
<i>PIK3CA</i>			988	982	-1.3	0.202			458	397	-3.59	<0.001
<i>AKT</i>			935	1033	-3	<0.001			430	425	-0.284	0.776
<i>MYC</i>			906	1061	-6.04	<0.001			357	498	-8.4	<0.001
<i>EGFR</i>			972	997	-0.97	0.332			449	406	-2.5	0.011

Values in bold are statistically significant; PIK3CA, Phosphoinositide-3-kinase, Catalytic, Alpha polypeptide, epidermal growth factor receptor (EGFR).

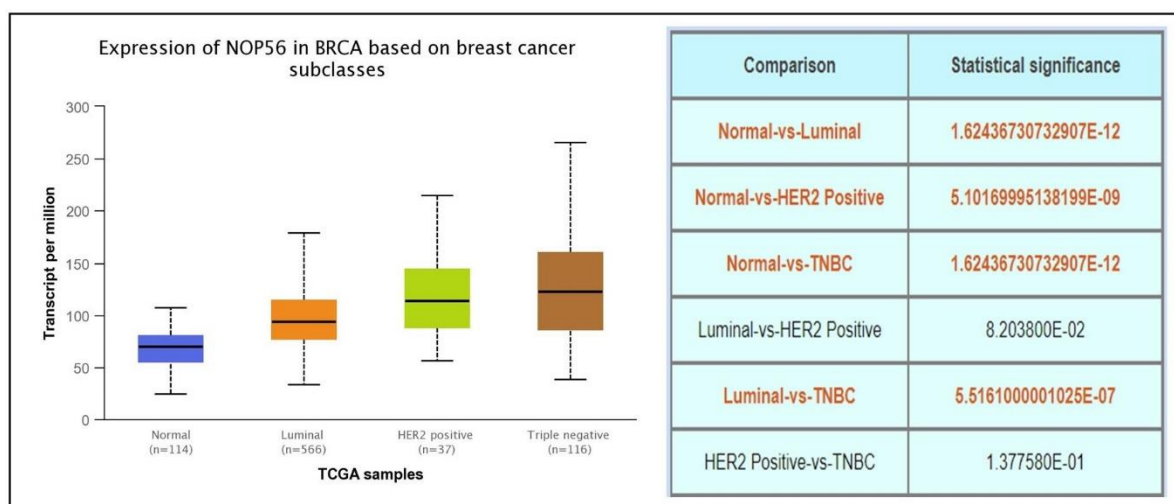


Figure 6.3: The *NOP56* mRNA expression in breast cancer tissue based on breast cancer subclasses.

6.4.3 Outcome Analysis for *NOP56* mRNA Expression

High *NOP56* mRNA expression was significantly correlated with shorter survival BCSS in the METABRIC cohort (HR = 1.894, 95% CI = 1.491–2.405, $p < 0.001$; Figure 6.4a). However, in the TCGA-BRCA cohort, high *NOP56* mRNA expression showed a poor prognosis but did not reach statistical significance (HR = 1.439, 95% CI = 0.874–2.370, $p = 0.150$; Figure 6.4b). Moreover, bc-GenExMiner v4.9 (Affymetrix) showed a similar significant association as the METABRIC cohort (HR = 1.38, 95% CI = 1.16–1.65, $p = 0.0003$; Figure 6.4c).

The stratification of *NOP56* mRNA cases based on LVI status revealed that high *NOP56* mRNA expression was significantly associated with shorter survival in LVI– cases (METABRIC: HR = 1.527, 95% CI = 1.006–2.317, $p = 0.045$ Figure 6.4d). In the TCGA-BRCA LVI– cases, no significant association was observed (HR = 1.517, 95% CI = 0.754–3.050, $p = 0.239$; Figure 6.4e). The LVI+ cases showed a highly significant association with poor patient prognosis in the METABRIC dataset (HR = 1.979, 95% CI = 1.371–2.856, $p < 0.001$; Figure 6.4f), whereas the LVI+ cases in the TCGA-BRCA dataset showed no significant association (HR = 1.176, 95% CI = 0.575–2.407, $p = 0.656$; Figure 6.4g).

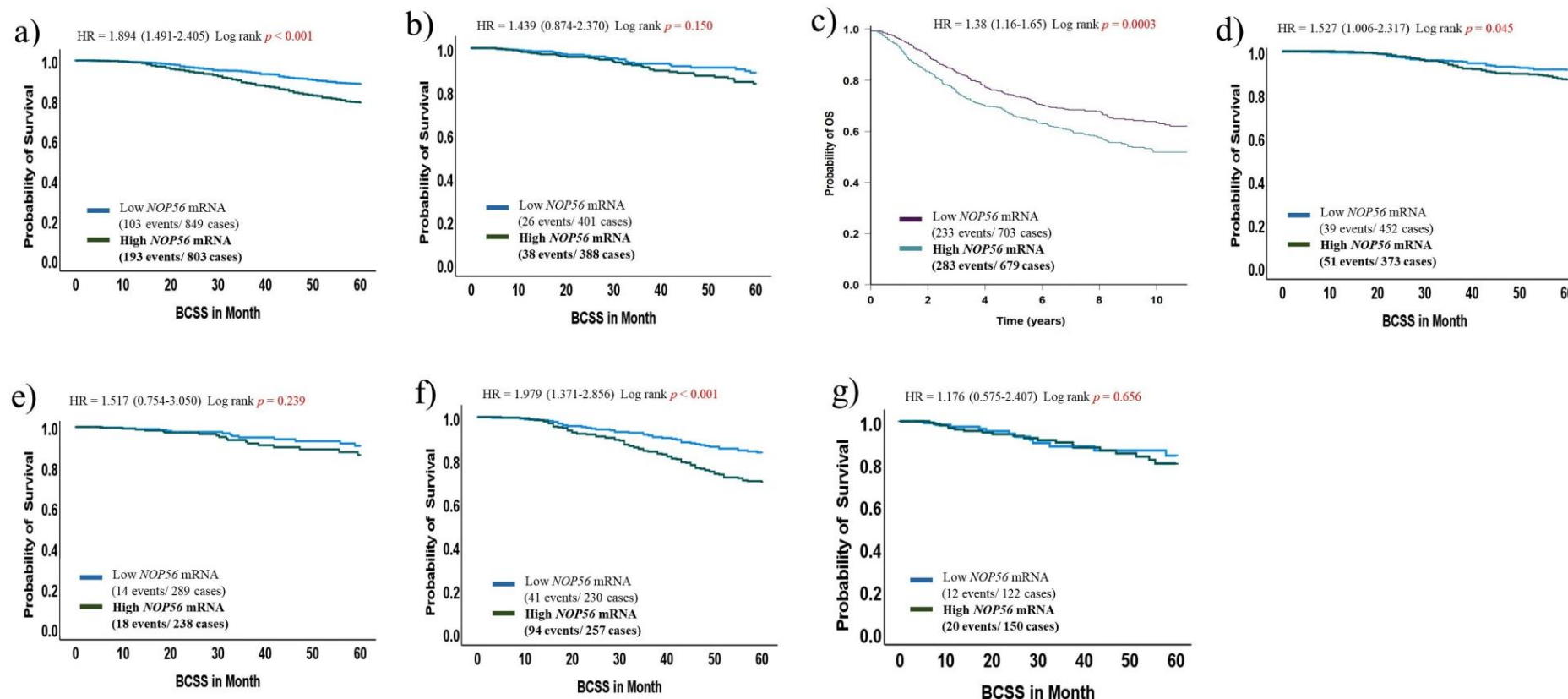


Figure 6.4: The Kaplan-Meier survival plots showing cumulative survival of patients with breast cancer stratified by *NOP56* mRNA expression. METABRIC (a), TCGA-BRCA (b) Breast Cancer Gene-Expression Miner (Affymetrix) (c), LVI- in METABRIC (d), LVI- in TCGA-BRCA (e), LVI+ in METABRIC (f) and LVI+ in TCGA-BRCA (g). OS, overall survival; BCSS, breast cancer-specific survival.

6.4.4 NOP56 Protein Expression

The protein expression of NOP56 in full-face sections showed homogeneous distribution staining in the cytoplasm of invasive tumour cells. A range of NOP56 immunoreactivity from absent to high was observed, and the median H-score of 35 was used to dichotomise the cases into the following two groups: negative/low and high immunoreactivity (Figure 6.5). There was more high than low NOP56 expression among all cases ($n = 803$; 59%,). The correlation between NOP56 transcriptomic and proteomic expression showed a positive uphill association (Spearman's coefficient: 0.21; $p = 0.001$).

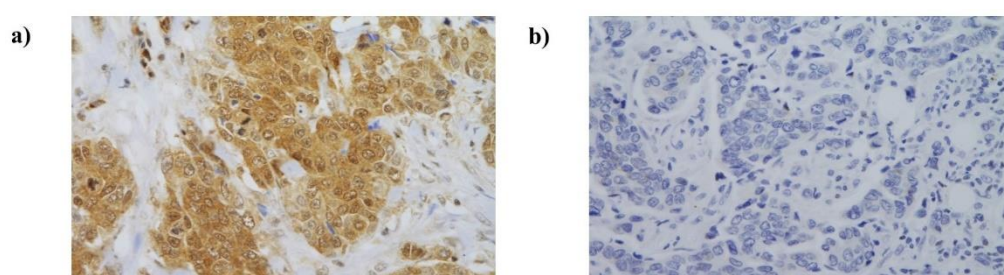


Figure 6.5 Immunohistochemically protein expression for NOP56 in tissue microarray cores of breast cancer tissue. Invasive cancer cells (intense staining in the cytoplasm of cancer cells) (a); and no staining (b) (images are at 20× magnification).

6.4.5 NOP56 Protein Expression and its Association with the Clinicopathological Parameters

A significant association was observed between high NOP56 expression and poor prognostic clinicopathological parameters, including LVI positivity ($p = 0.001$), larger tumour size, high tumour grade, ER negativity, PR negativity, HER2 positivity and the NPI poor prognostic group. High NOP56 protein expression was also associated with high levels of PIK3CA ($p < 0.001$), pAKT ($p = 0.025$), p53 ($p < 0.001$) and EGFR ($p = 0.019$) (Table 6.3). The stratification of NOP56 expression according to intrinsic molecular subtypes showed a significant association of high NOP56 expression with HER2 and TNBC ($p < 0.001$) than with other subtypes (Table 6.3).

Table 6.3: Relationship of NOP56 protein expression with clinicopathological characteristics in the Nottingham cohort.

Factors	Expression of NOP56			Factors	Expression of NOP56		
	Low	High	p-value		Low	High	p-value
	N (%)	N (%)			N (%)	N (%)	
Tumour size				Age			
≤2 cm	288 (45.7)	342 (54.3)	<0.001	≤50 years	181(35.5)	329 (64.5)	0.002
>2 cm	261 (36.6)	452 (63.4)		>50 years	370 (43.8)	474 (56.2)	
Histological grade				Nodal status			
1,2	322 (50.6)	314 (49.4)	<0.001	Negative	342 (42.4)	464 (57.6)	0.148
3	227 (32.2)	477 (67.8)		Positive	207 (38.5)	331 (61.5)	
Lymphovascular invasion				Oestrogen receptor			
Negative	381 (43.4)	497 (56.6)	0.01	Negative	110 (30.6)	249 (69.4)	<0.001
Positive	166 (36.2)	293 (63.8)		Positive	435 (44.4)	544 (55.6)	
Progesterone receptor				Human epidermal growth factor receptor-2			
Negative	192 (35.2)	353 (64.8)	0.001	Negative	469 (42.4)	636 (57.6)	<0.001
Positive	329 (44.2)	461 (55.8)		Positive	50 (27.9)	129 (72.1)	
Ki67				Phosphoinositide-3-kinase, Catalytic, Alpha polypeptide			
Low	138 (40.7)	201 (59.3)	0.459	Negative	135 (56.3)	105 (43.8)	<0.001
High	197(38.2)	319 (61.8)		Positive	274 (36.9)	468 (63.1)	
Phosphorylated AKT				P53			
Negative	97 (44.1)	123 (55.9)	0.025	Negative	399 (44.3)	502 (55.7)	<0.001
Positive	235 (35.6)	425 (64.4)		Positive	125 (32.4)	261 (67.6)	
Epidermal growth factor receptor				Triple-negative			
Negative	425 (42.1)	585 (57.9)	0.019	Non-TNBC	459 (43.1)	606 (56.9)	<0.001
Positive	96 (34.3)	184 (65.7)		TNBC	71 (29.2)	172 (70.8)	
Intrinsic Subtypes				Nottingham prognostic index			
Luminal A	73 (44.2)	92 (55.8)	<0.001	Good prognostic group	195 (52.8)	174 (47.2)	<0.001
Luminal B	114 (44.2)	144 (55.8)		Moderate prognostic group	276 (38.2)	447 (61.8)	
HER2 enriched BC	36 (29.3)	87 (70.7)		Poor prognostic group	76 (31.5)	165 (68.5)	
TNBC	71 (29.2)	172 (70.8)					
Values in bold are statistically significant.							

6.4.6 Outcome Analysis for NOP56 Protein Expression

In the outcome analysis, patients with high NOP56 protein expression showed significantly shorter BCSS (HR = 1.42, 95% CI = 1.1–2.047; $p = 0.003$; Figure 6.6a) than patients with low expression. The categorisation of NOP56 protein expression based on LVI status revealed no significant association with high NOP56 expression (LVI–, HR = 1.351; 95% CI = 0.900–2.028, $p = 0.144$; Figure 6.6b). Interestingly, a significant association was observed between high NOP56 expression and poor patient outcome in (LVI+, HR = 1.593; 95% CI = 1.052–2.412, $p = 0.026$; Figure 6.6c).

The multivariate Cox regression analysis showed that high NOP56 protein expression was an independent predictor for shorter BCSS (HR = 1.434; 95% CI = 1.075–1.934, $p = 0.026$; Table 6.4) when including LVI, LN, tumour size and age.

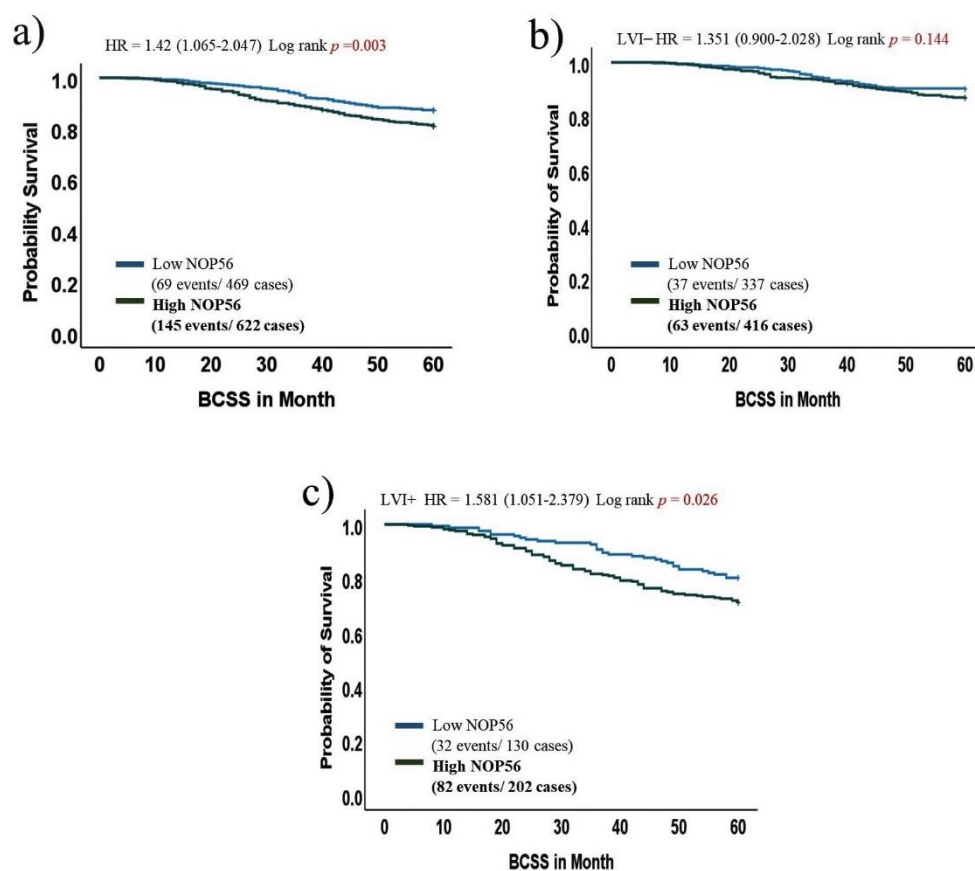


Figure 6.6: Kaplan-Meier survival plots showing breast cancer specific survival of BC patients stratified by NOP56 expression in Nottingham cohort. NOP56 protein expression revealed that high expression was associated with poor patient outcome. Breast cancer specific survival in the Nottingham cohort (a), in the LVI- cases (b) and the LVI+ cases(c).

Table 6.4: Multivariate Cox regression analysis of clinicopathological factors and protein expression of NOP56 to predict breast cancer specific survival.

Factors	Multivariate analysis BCSS		
	Hazard Ratio	95% CI	p-value
NOP56 expression	1.434	1.075-1.914	0.014
Lymphovascular invasion	1.681	1.263-2.238	<0.001
Lymph node	2.183	1.630-2.924	<0.001
Tumour size	1.785	1.314-2.425	<0.001
Age	1.272	0.963-1.680	0.091
Values in bold are statistically significant.			

6.5 Discussion

Globally, BC is recognised as the most life-affecting and life-threatening disease for women (394). Moreover, LVI is one of the reasons for tumour metastasis and cancer progression (121). It is hypothesised that primary tumours can migrate towards the LN, this process includes detaching from the primary site, managing the connection with the tumour microenvironment, enhancing the invasion of LVs and protecting tumour cells during circulation in LVs (395). Even though LVI is a major metastasis event facilitating tumour progression, the exact mechanisms and the critical molecular factors involved in this process are unclear. Thus, the deciphering of novel molecular mechanisms associated with cancer cell migration and invasion to enable the discovery of prognostic predictors in BC is needed.

According to our previous gene signature study (223), *NOP56* mRNA significantly correlates with LVI-positivity. It has been suggested that *NOP56* and the other snoRNPs enhance cell alteration and tumourigenesis (387, 396). *NOP56* is also thought to play an essential role in MYC regulation, which enhances BC growth and progression (387, 397).

This study identified a significant association between high *NOP56* mRNA and/or protein expression and aggressive tumour clinicopathological parameters, including larger tumour size, high tumour grade, LVI-positivity, NPI poor prognostic groups, ER and PR negativity, induced p53 expression, lymph node stage, and the TNBC in addition to poor patient outcomes. Interestingly, *NOP56* showed a similar significant association with LVI at the proteomic level. Furthermore, a positive correlation was also observed between *NOP56* and the high expression of the proliferative marker Ki67 and EGFR. These results link that *NOP56* is associated with poor prognosis and an undefined oncogenic role, as previously stated in several types of research (398, 399), which might highlight the importance of *NOP56* in BC development.

MYC, the cells' metabolic and signalling pathways regulator, help BC cells to survive under treatment stress (400-402). There is also evidence that high-grade BC tumours are more likely to express MYC than low-grade tumours (403, 404). An *in vitro* study showed that Burkitt's lymphoma tumour growth and MYC-induced cell transformation are mediated by NOP56 (387). In this research, *NOP56* and *MYC* showed a significant positive relationship. Thus, the positive correlation of NOP56 and LVI+, nodal status, and advanced tumour stage found in this study might highlight NOP56 functions in cancer invasiveness through manipulating MYC-binding sites. Nonetheless, there is an association between these poor prognostic factors and the expression of c-Myc in muscle-invasive urethral carcinoma of the urinary bladder (405). In addition to the correlation between *MYC* and *NOP56*. Likewise, NOP56 showed a significant association with EGFR, PI3K and Akt, which were previously reported as MYC regulators. It has been hypothesised that EGFR regulates MYC through enhancing specific pathways such as PI3K-Akt and Ras/Raf/MEK/ERK (397). Another study in BC cell models found a correlation between blocking PI3K/AKT and MYC inhibition (406). In addition, EGFR, PI3K, and AKT have a crucial role in BC cell proliferation, migration and metastasis (407, 408). Moreover, the mammalian target of rapamycin (mTOR) signalling pathway that controls cell growth and autophagy is regulated by PI3k and AKT (409). Hence, the current study results showed a significant association between the aforementioned biomarkers and NOP56 expression, which requires further mechanistic examination to understand the molecular profile underlying NOP56 and its association with LVI positivity in BC. The correlation between mTOR and NOP56 has also been shown to play an essential role in Kirsten rat sarcoma viral oncogene homolog (KRAS-mutant) lung cancer treatment (388). It has been reported that KRAS-mutant cancer cells' response to oxidative stress is weakened in the presence of low NOP56 levels, which forces these cells to depend on mTOR signalling to survive. Therefore, knockdown of NOP56 would stimulate KRAS-mutant lung cancer death through apoptosis. This finding supports the correlation between BC progression and invasion cascades and the presence of high NOP56 expression.

Compared to other BC subtypes, TNBC has aggressive clinical behaviour, a worse prognosis and a higher mortality rate (407). Furthermore, patients with TNBC have a higher incidence of LVI than those without TNBC (20.4% vs 13.5%, $p = 0.03$) (408), and LVI in TNBC is associated with early onset of relapse (409). Additionally, high NOP56 expression in this study was associated with TNBC at both the transcriptomic and proteomic levels. Thus, targeting NOP56 may influence the treatment of TNBC and other BC molecular subtypes.

Even though this study revealed promising results highlighting a potential critical role of NOP56 in BC LVI+ tumours, this study carries certain limitations. First, the semi-quantitative H-score method was used to score the stained sections, and the reliability was increased by double-scoring at least 10% of the cores via well-trained observers. Moreover, functional assessment was essential in this study to uncover the molecular mechanisms that underline that marker. However, the COVID-19 pandemic interrupted the work, which made it impossible to proceed with further functional studies. This should be applied to NOP56 in the future to evaluate its prognostic/predictive value in BC LVI.

In conclusion, NOP56 is an independent prognostic biomarker of poor patient survival and is associated with LVI positivity and worse tumour prognostic characteristics. Furthermore, NOP56 overexpression is associated with MYC and may play a critical role in tumour cell survival, proliferation and metastasis. At both the *in vivo* and *in vitro* levels, more functional assessments are needed such as cell cycle, apoptosis, proliferation, migration, transmigration and survival assays to estimate the treatment abilities of NOP56 in BC and its effects on TNBC progression in the presence of LVI.

7 Chapter 7: Mechanistic insights of HMGB3, a biomarker that was chosen from the LVI-related gene signature set

7.1 Overview

In the past few decades, improvements in diagnostic accuracy, early cancer detection and personalised drug treatments have been made, which has improved patient survival for early stage BC (410). Although outcomes have significantly improved, metastasis, which impacts around 20% of patients with BC, is still the primary reason for deaths among patients with BC (411). Tumour size, histologic grade of the tumour and lymph node status are only a few histological traits that are used as prognostic factors in BC, all being strongly associated with disease progression (412, 413). Another important prognostic factor is LVI, which is regarded as an early step towards BC progression and dissemination (65). Although the molecular patterns associated with tumour differentiation, including tumour grade, histological type of the tumour and progression of LN dissemination, have been thoroughly investigated (215, 216), currently, there is a lack of knowledge about the molecular mechanisms, pathways and genes linked to LVI status, which may be utilised as biomarkers or therapeutic targets to enhance patient survival. Thus, targeting of LVI is needed and might be promising for identifying tumour metastasis in patients with early stage BC.

The high mobility group (HMG) proteins, which are essential for the overall assembly of chromatin domains, are considered the second most frequently found proteins in the cell. They can provide more transcriptional fine-tuning in response to sudden changes in the surrounding environment through interactions with nucleosomes, transcription factors, nucleosome remodelling complexes and histone H1 (414). Moreover, chromosomal proteins that take part in DNA synthesis, transcription, breaking and recombination, and repair are members of the HMG box (HMGB) group, a subset of the HMG superfamily (415, 416). The relevance of the HMGB family in cancer development is demonstrated by the correlation between the high expression of HMGB family members and a number of cancer hallmarks, such as uncontrolled replicative potential, resistance to programmed cell death, activation of tissue invasion and metastasis (414, 417). The HMGB family contains several genes, including *HMGB1*, *HMGB2*, *HMGB3* and *HMGB4* (418). Despite these members' 80% similarity in sequences of amino acids, they have different biological functions in cells (414, 417). Various research works have shown that the proteins HMGB1, HMGB2 and HMGB3 play a crucial part in several cancers, such as colon cancer (419), HCC (420) and pancreatic cancer (421). Interestingly, *HMGB3* has been identified as being among the highly DEGs correlated with LVI positivity using the WAD approach, a bioinformatics analysis (223). The *HMGB3* gene is localised in chromosome Xq28 (416), and its protein form is mostly expressed in the cytoplasm and nucleus (420). The

expression of *HMGB3* is typically strong throughout embryogenesis; however, in adults, it is downregulated or absent in their normal tissue (422, 423). The extensive expression of this protein in mammalian cells highlights its wide range of contributions to several physiological and pathological processes, including DNA damage repair and programmed cell death. While *HMGB3* may help maintain homeostasis, it may also have an impact on the development of diseases such as HCC and benign chronic liver diseases (424). Several studies have shown a significant correlation between high levels of *HMGB3* expression and the development of tumours as well as a poor prognosis for late-stage lung tumours (425), prostate cancer (426) and bladder cancer (427). Through the regulation of the cell cycle and stimulation of cancer cell proliferation and invasion via the Wnt/ β -catenin, AKT and several other signalling pathways, *HMGB3* acquires its ability to govern tumour progression. By enhancing the expression of hypoxia-inducible factor-1, *HMGB3* can also boost the production of reactive oxygen species and the development of tumour cells (424). However, the precise involvement of *HMGB3* in LVI in BC has yet to be defined.

7.2 Aim and Hypothesis

The *HMGB3* gene manipulates proliferation, invasion, and several cell processes and pathways in various cancer types. The research hypothesis was that dysregulation of *HMGB3* promotes LVI+ status sequence steps in BC, including tumour cell proliferation, migration, adhesion and transmigration. Thus, the objectives of this research were to (1) assess the mechanistic role of *HMGB3* in BC cell lines *in vitro*, with a focus on its contribution to LVI progress, and (2) evaluate the clinicopathological significance of *HMGB3* at the mRNA and protein levels utilising large, well-characterised and long-term-follow-up BC cohorts. The Nottingham Ps cohort has been stained for HMGB3 by Abrar Aljohani and reported in her dissertation. Therefore this study focuses on functional assessment for HMGB3.

7.3 Materials and Methods

7.3.1 Selecting and Handling Breast Cancer Tumour Cells

Breast cancer cell lines were chosen for this study's *in vitro* experiments based on the *HMGB3* expression levels determined by an *HMGB3* WB analysis. The BC cell lines BT-474 (ER+/PR+/HER2+), a luminal B–HER2 positive molecular class subtype, and MDA-MB-231 (ER–/PR–/HER2–), a triple-negative subtype, both of which have been linked to the highest prevalence of LVI (121), were employed in this study. All BC cell lines underwent WB to determine the relative protein expression, and the LI-COR Odyssey machine was utilised for quantification. Cells that expressed *HMGB3* at the highest levels were chosen for the

knockdown experiment and other *in vitro* experiments. More details on preserving the environment for cell growth, culture and passaging are provided in **Chapter 2 (section 2.2)**.

7.3.2 Endothelial Cells

Primary HUVECs and dermal lymphatic endothelial cells (DLECs) were obtained from Promocell (Heidelberg, Germany). Both cell types were cultured in the medium used for endothelial cell growth (C-22022, Promocell, Germany). The cell culture process was conducted in a class II cabinet considering sterile conditions, with an adjusted temperature of 37°C and 5% CO₂ as the incubation conditions. The primary HUVECs and DLECs were used to examine the relationship between BC and endothelial cells in order to investigate the effects of *HMGB3* on LVI. Further details on the culture conditions and cell growth maintenance of these cells are available in **Chapter 2 (section 2.2)**.

7.3.3 Silencing *HMGB3* using Transient Small Double-Stranded Interfering RNA

The potential functional consequences of *HMGB3* knockdown (KD) and its involvement in BC development and LVI were investigated using a type of RNA interference (RNAi) known as a small double-stranded interfering RNA (siRNA)-based method with BC cell lines, which silences genes by causing degradation of mRNA. This KD technique is based on introducing the RNAi (siRNA) into the cytoplasm, where it is integrated into RNA-induced silencing complex (RISC), a multi-protein complex; this step activates RISC in preparing for a gene's suppression. The RISC activation results in the cleavage of one of the siRNA strands, which is then released from the complex, while the other strand serves as a template on the mRNA that is being targeted. After that, a ribonuclease cleaves and degrades the mRNA (428). In this experiment, the target gene is transiently knocked down by introducing siRNA to cells using lipofection. This forwards transfection approach employs scrambled negative control siRNA (does not target MDA-MB-231 and BT-474) or pre-validated silencer select siRNA constructs (primarily for *HMGB3*). The *HMGB3* siRNA is composed of the following ribonucleotide bases: 5'-GCACCCUGAAACUGUAUCAAtt-3'.

In this study, three silencer select siRNA constructs were used to target *HMGB3* (s6652, 115646 and 145247); all were obtained from Thermo Fisher Scientific (UK, 4392420 and AM16708). Lipofectamine RNAiMAX Transfection Reagent (13778150, Thermo Fisher Scientific, Lithuania) was used for transfection. The siRNAs were optimised alongside the

negative control, scrambled siRNA (AM4611, Thermo Fisher Scientific, UK). To achieve the highest KD efficiency, three different concentrations of IDs of *HMGB3* siRNA and lipofectamine were included in the optimisation process and then determined by applying the WB technique. Moreover, the transfection step was accomplished using a forwards transfection technique, where after seeding the targeted cells for 24 h, the transfection mix (siRNA constructs and lipofectamine) was introduced to the wells.

7.3.4 The Procedure of siRNA Transient Transfection

Transfection was performed on a six-well plate. Cells were seeded in the plate at a density of 3×10^5 per well and were then incubated overnight at 37°C and 5% CO₂ incubation conditions. The next day, when the cells were about 40% confluent, transfection was performed with 10 nm of three different *HMGB3* siRNA IDs. For this purpose, 1.5 µl of siRNA in Opti-MEM Reduced-Serum Medium (31985062, Thermo Fisher Scientific) and 5 µl of Lipofectamine RNAiMAX Transfection Reagent were diluted in the same medium. The diluted siRNA and lipofectamine (reagent 1:1 ratio) were then combined, gently mixed and incubated at room temperature for 15 minutes.

Then, every well that contained cells received an siRNA liquid complex. The cells were incubated for three, five and seven days before being lysed for use in WB analysis or trypsinised for use in functional experiments, the latter of which will be described below.

7.3.5 Cell Lysis for siRNA Samples

Luminal B–HER2 positive (BT-474) and TNBC (MDA-MB-231) cell lines were utilised in this study. More information about the cell passaging procedure is given in **Chapter 2 (section 2.2.3)**.

7.3.6 Assays of in vitro Studies (Preclinical)

The potential role of *HMGB3* in LVI and other biological processes was investigated in preclinical studies utilising the following cells and assays.

7.3.6.1 CellTiter 96® Aqueous One Solution Cell Proliferation Assay (MTS assay)

7.3.6.1.1 Overview Principles

Cell viability is measured using the colorimetric technique known as the MTS assay. The proliferation assay reagent CellTiter 96® Aqueous One Solution Reagent (G3580, Promega, UK, LTD) contains the tetrazolium compound [3-(4,5-dimethylthiazol-2-yl)-5-(carboxymethoxyphenyl)-2-yl]-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H tetrazolium, inner salt; MTS] and an electron coupling agent (phenazine methosulfate). The

previously mentioned compound is reduced by metabolically active (viable) cells with the help of dehydrogenase enzymes. Due to this chemical reaction, coloured formazan soluble in a cell culture medium is generated. Nicotinamide adenine dinucleotide hydrogen and nicotinamide adenine dinucleotide phosphate hydrogen are produced in conjunction with this chemical reaction. There is a clear correlation between the number of living cells and the quantity of formazan produced, as the amount of the latter reflects the number of viable cells (429).

7.3.6.1.2 Assay Procedure

The MTS assay was carried out in accordance with the manufacturer's instructions. A 96-well plate was seeded with cells in triplicate, with a density of 3,000 cells in each well, and the cells were left to attach. The proliferation reagent was added at 20 µl to each well and incubated under standard conditions away from light for an hour. With the aid of a microplate reader (Tecan Infinite F50, Tecan, UK), the absorbance was measured at 490 nm. Every assay condition was done in triplicate to ensure the best results, and the average was calculated. The graphs were plotted using percent absorbance taken by comparing the untreated cells' control population.

7.3.6.2 Clonogenic Survival Assay

7.3.6.2.1 Overview Principles

Clonogenic survival assay is used to evaluate the viability/survival of cells after KD or exposure to cytotoxic agents. This test method counts living cells that have been replicated at least 5–6 times to make a colony and have thus survived (430, 431).

7.3.6.2.2 Assay Procedure

The cell lines were seeded in triplicate into six-well plates at varying densities. The incubation period was determined to be from 12–14 days at 37°C and 5% CO₂. Then, the medium was aspirated off, and the wells were gently washed off with DPBS. Colonies were manually counted after the plates were stained with 0.5% crystal violet (Cat#: C0775, Sigma Aldrich, UK), which contained a mixture of 500 mg of the stain with 25 ml methanol and 75 ml DW. Visible colonies (containing ≥50 cells) stained with blue dye were manually counted, while the surviving fraction was determined using the formula described below:

$$\text{Surviving Fraction (SF)} = \frac{\text{No. of colonies formed}}{\text{No. of cells seeded} \times \text{Plating efficiency}}$$

Plating efficiency was calculated as follows:

$$\text{Plating efficiency (PE)} = \frac{\text{No. of colonies formed}}{\text{No. of cells seeded}} \times 100$$

7.3.6.3 *In vitro Validation of Invasion Metastatic Cascade (Adhesion, Migration and Transmigration)*

7.3.6.3.1 Overview Principle

Adhesion, migration and transmigration are major sequential processes in the tumour cell metastasis cascade. The most important routes cancer cells utilise to spread are blood capillaries and lymphatic circulation. However, because the LVs lack supporting structures and have basement membrane irregularity and loose junctions caused by wide lumens, the lymphatic vascular system has been found to engage in metastasis more than the BV system. Therefore, the invasion process, including adhesion, migration and transmigration through blood and lymphatic endothelium, is thought to be driving the initial metastatic stages. The adhesive interaction of tumour cells with endothelial cells is required to facilitate this step. Thus, adhesion is an essential step in the intravasation process to promote migration and transmigration. These migratory stages, including the migration of tumour cells through the ECM and the transmigration across the lymphatic endothelial cells, are critical in cancer progression. Due to their ability to migrate and transmigrate, cancer cells can invade secondary tissues, ultimately resulting in metastasis. Thus, cancer research needs to investigate potential novel methods to inhibit cancer cells' adhesion, migration and transmigration.

The purpose of using adhesion, migration and transmigration assays *in vitro* is to evaluate the ability of cancer cells to invade the lymphatic and vascular endothelium. In adhesion and transmigration assays, the interactions between tumour cells and the endothelial layer that coat the LVs are investigated *in vitro* by seeding an endothelial cell monolayer in plates to mimic the *in vivo* interaction. Two types of endothelial cells were used in this study, HUVECs and DLECs. The HUVECs were obtained from umbilical cord vein endothelial cells. They have been used extensively as a theoretical model for research into the control of endothelial activity and the involvement of the endothelium layer in the sensitivity of the BV wall to various physical functions. Meanwhile, DLECs, derived from skin tissue, are the essential assembly of the LV wall and aid in controlling lymphocyte recycling and the body's immunological response, contributing to the maintenance of internal fluid homeostasis. More details on the lymphatic vascular structure are given in **Chapter 1 (section 1.3.4)**.

In addition, a wound healing experiment is a common method for studying cell migration *in vitro* by applying different environmental factors, such as gene silencing utilising siRNA or treating cells with a particular compound. In this assay, a physical gap is made in a cell monolayer. The cell migration process, involving filling a cell-free gap, is then recorded by capturing pictures at various intervals. Automated software assists the measurement and analysis of the gap closure rate (432).

7.3.6.3.2 Wound healing Assay (Migration) Procedure

The Ibidi Culture-Insert 2 Well in μ -Dish 35 mm (81176, Thistle Scientific Ltd) was used to perform a wound healing assay following the supplier's instructions. Initially, the treated and untreated cells were trypsinised and processed to prepare them for counting and then suspended in $(15 \times 10^4)/140 \mu\text{l}$ of cultured medium. Each well received $70 \mu\text{l}$ of cell suspension, which was then incubated at 37°C with 5% CO_2 . The cells needed at least 24 h to attach and form a confluent layer, and once that was achieved, the dish was scratched by gently removing the Culture-Insert 2 Well with sterile tweezers and then gently washing with DPBS. Next, 2 ml of 1% foetal bovine serum (FBS-Heat Inactivated, F9665, Sigma Aldrich, UK) cultured medium was added to the dish. Images of the wound were taken with an inverted microscope (Leica DMI3000B, Leica Microsystems, Germany) on the day the well was removed and over the following days. The ImageJ (version 1.52) analysis program was utilised to quantify migrated cells via transwell inserts.

7.3.6.3.3 Adhesion Assay Procedure

A 24-well plate was pre-marked with a grid line on the back to draw five fields before use. Before the endothelial cells were seeded, the transwell was coated with a base mixture of collagen gel and DPBS in a ratio of 1:9 μl and incubated for 30 minutes before being washed gently with DPBS. This step enhanced the endothelial cell monolayer attachment. Then, the endothelial cells were seeded and incubated for 24 h or until they became confluent. Ensuring the confluency of the endothelial monolayer was a crucial step to avoid any confounding factors that could affect a proper reading of the findings. The tumour cells were then harvested, counted and labelled with $1 \mu\text{M}$ CellTracker Green CMFDA Dye (C2925, Invitrogen, UK) and incubated at 37°C for 30 minutes. Following labelling, the tumour cells were resuspended in a medium containing 10% serum and incubated in the plate that contained endothelial cell monolayers at 37°C for 35 minutes. To ensure better results, the cells were washed using cancer cell media to eliminate all non-adherent cells. Then, a fluorescence microscope (Leica DMI3000B, Leica Microsystems, Germany) was utilised to count the remaining cancer cells at

10× magnification. With the aid of the plate's back grid, five fields of view were counted in the well's centre. To express the results, the percentage of cells adhered to the endothelial layer was calculated by ratioing their total number to the number of control cells.

7.3.6.3.4 Transmigration Assay Procedure

Hanging Transwell FluoroBlok inserts (351152, Corning Aldrich) with an 8-µm pore size were used to grow a monolayer of endothelial cells. This coloured type of transwell has a unique design that contains a black-coloured membrane. This blocks light transmission through the insert to enhance the detection of labelled cells migrating across endothelial cells to under the insert and to save time by avoiding washing or manual cell scraping off the insert. Before the endothelial cells were seeded, the transwell was coated with a base mixture of collagen gel and DPBS in a ratio of 1:9 µl and incubated for 30 minutes before being washed gently with DPBS. After that, the endothelial cells were seeded and incubated for 24 h or until they became confluent. Then, tumour cells were harvested and labelled with 1 µm of CellTracker Green CMFDA Dye (C2925, Invitrogen, UK) and seeded on the endothelial cells. The transwell chamber was filled with regular tumour cell medium, and then, both cell types were incubated for 16 h at 37°C with 5% CO₂. Using a fluorescence microscope (Leica DMI3000B, Leica Microsystems, Germany) at 10×, transmigration was assessed by counting the number of transmigrated tumour cells on the chamber's bottom across the endothelial cells through the filter.

7.3.7 Clinical Methodology

7.3.7.1 Study Cohort

The transcriptome expression of *HMGB3* was evaluated in breast tissue using gene expression TPM values from the TNM plot database. As discovery cohorts, the METABRIC cohort's (n = 1,980) cases of primary operable invasive BC and the TCGA-BCA cohort's (n = 854) cases were utilised to evaluate the clinical significance of *HMGB3*. Additionally, *HMGB3* mRNA patient outcome significance was examined using the bc-GenExMiner v4.9 dataset (Affymetrix; n = 4,904). The clinical history, outcome information and clinicopathological characteristics of the cohorts are described in **Chapter 2 (section 2.1.1)**.

7.3.7.2 Western Blot

The *HMGB3* primary antibody's (rabbit polyclonal HPA062583; Sigma-Aldrich, UK) specificity was confirmed using WB and with the assistance of lysates from the BC cell lines

BT-474 and MD-MB-231. More information on the procedures, concentration of HMGB3 antibody and incubation time are provided in **Chapter 2 (section 2.2 and Table 2.5)**.

7.3.8 Statistical Analysis

The SPSS 28.0 program (IBM, Chicago, IL, USA) was used to analyse clinical data and *HMGB3* mRNA expression based on the median. The determination of the association between *HMGB3* mRNA and clinicopathological factors was performed utilising the chi-squared test. The Kaplan–Meier curve test was utilised for the survival analysis. The GraphPad Prism software version 9.5.0 was utilised to perform the statistical analysis. The *in vitro* data were shown as the mean $\pm$ SEM of three separate experiments done in triplicate. A Student's t-test was performed to determine the significant differences between the control non-treated and knocked-down *HMGB3*. The statistically significant value was measured as a *p*-value of <0.05 . More details regarding the statistical analysis were previously described in **Chapter 2 (section 2.11)**.

7.4 Results

7.4.1 Pre-Clinical Results

To validate the potential role of *HMGB3* reduction and investigate its function in survival efficiency, proliferation, adhesion, migration and transmigration, this research utilised an siRNA-based method in nominated BC cell lines.

7.4.1.1 Antibody Specificity Validation Utilising Western Blot

The WB image of anti-HMGB3 AB showed a single band at MW (~25 kDa; Figure 7.1). The housekeeping mouse β -actin antibody (A5441, clone AC-15, Sigma-Aldrich, Gillingham, UK) was used with MW of 42 KDa (Figure 7.1).

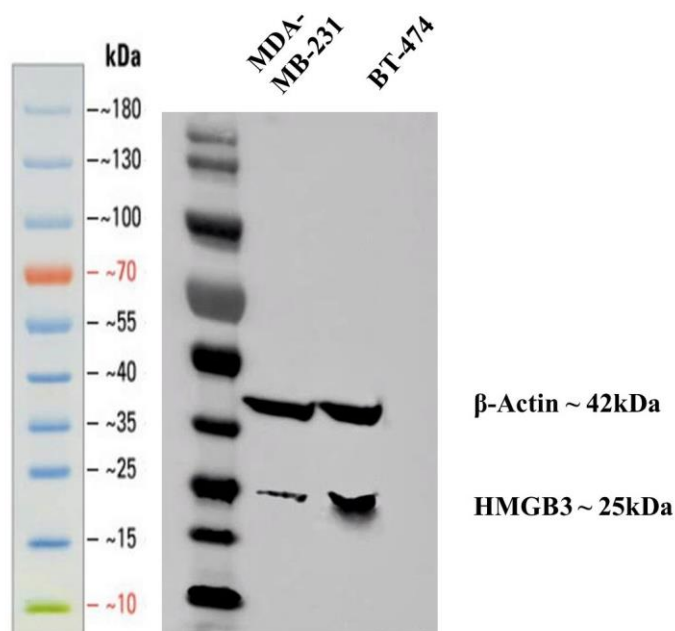


Figure 7.1: Western blotting validation indicated HMGB3 primary antibody specificity based on a single band observed at approximately 25 kDa.

7.4.1.2 HMGB3 Protein Expression Validation

High expression of *HMGB3* mRNA showed a significant association with the luminal B, HER2-enriched and MDA-TNBC subgroup (Figure 7.8 and Table 7.1). Therefore, the WB method was used to validate the HMGB3 protein expression utilising the available luminal B–HER2 positive and MDA-TNBC cell lines. Between these cell lines, the BT-474 (ER+/ PR+/ HER2+), a luminal B–HER2 positive molecular class, and MDA-MB-231 (ER–/PR–/HER2–) TNBC cell lines showed the highest expression of HMGB3 protein. The efficacy of *HMGB3* KD was validated using three siRNA constructs targeting *HMGB3* (IDs: 115646, 145247 and s66520) compared to the non-targeting scramble control siRNA (Cat# AM4611). The KD process shared the same results of HMGB3 protein expression loss in the BT-474 cell lines with all of the siRNAs used to target *HMGB3* (Figure 7.2). Thus, in all subsequent functional studies, only one siRNA (115646) was utilised.

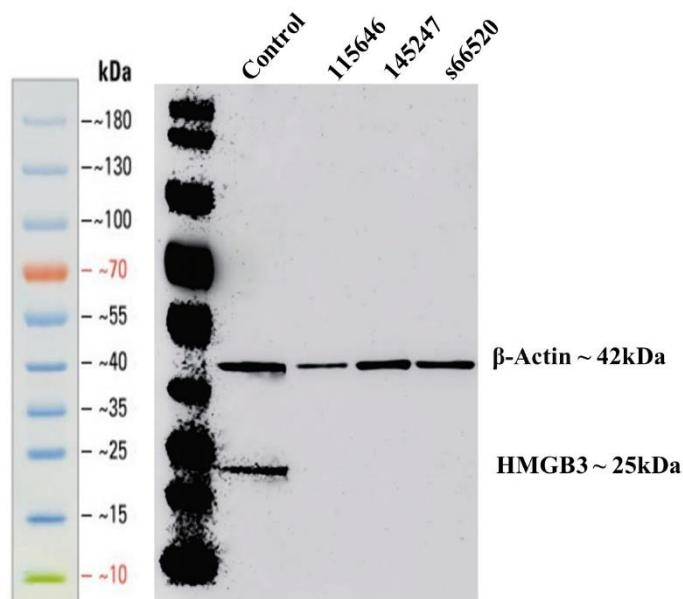


Figure 7.2: The HMGB3 protein knockdown utilising various siRNA IDs in the BT-474 breast cancer cell line. Western blotting showed *HMGB3* knockdown utilising siRNA (IDs: 115646, 145247 and s66520) and negative control scrambled cells.

7.4.1.3 *HMGB3 Enhances Cell Viability, Survival Efficiency and Proliferation*

The *HMGB3* KD showed about 95%–97% reduction in the *HMGB3* mRNA cells in the cell lines MDA-MB-231 and BT-474 ($p < 0.001$ and $p < 0.001$, respectively; Figure 7.4a). In line with the *HMGB3* protein expression, the downregulation of *HMGB3* reduced the capability of the tumour cells to produce a colony and, therefore, survived in MDA-MB-231 ($p < 0.001$; Figure 7.3); however, no colonisation was seen in BT-474. Moreover, *HMGB3* KD showed a significant reduction in the cells' proliferation ability in T24 (MDA-MB-231: $p = 0.0014$; BT-474: $p < 0.001$), T48 (MDA-MB-231: $p = 0.0092$; BT-474: $p < 0.001$) and T72 (MDA-MB-231: $p = 0.0060$; BT-474: $p < 0.001$; Figure 7.4c).

7.4.1.4 *HMGB3 Suppresses the Ability of Cells to Migrate*

A significant result seen in the migratory assay was that KD of *HMGB3* caused a large unhealed wound compared to the control, which showed increased wound closure ability at (T24) with MDA-MB-231: $p = 0.0012$; BT-474: $p = 0.0208$) and (T48) with MDA-MB-231: $p < 0.001$; BT-474: $p = 0.0171$; Figure 7.4d–f).

7.4.1.5 HMGB3 Promotes Breast Cancer Cell Adhesion and Transmigration to Monolayer Endothelial Cells

Compared to the KD cells, a high adherence percentage to the endothelial cells was observed in both the MDA-MB-231 and BT-474 cell lines (MDA-MB-231 with HUVECs and DLECs: $p < 0.001$ and $p < 0.001$, respectively; BT-474 with HUVECs and DLECs: $p < 0.001$ and $p = 0.003$, respectively; Figure 7.5a and Figure 7.5 b, respectively). In the transmigration assay, both BC cell lines showed a significant reduction in number of transmigrated cells across the endothelial monolayer cells in the KD model in comparison to the negative control model (MDA-MB-231 with HUVECs and DLECs: $p < 0.001$ and $p < 0.001$, respectively; BT-474 with HUVECs and DLECs: $p = 0.0194$ and $p = 0.0257$, respectively; Figure 7.6a and b, respectively).

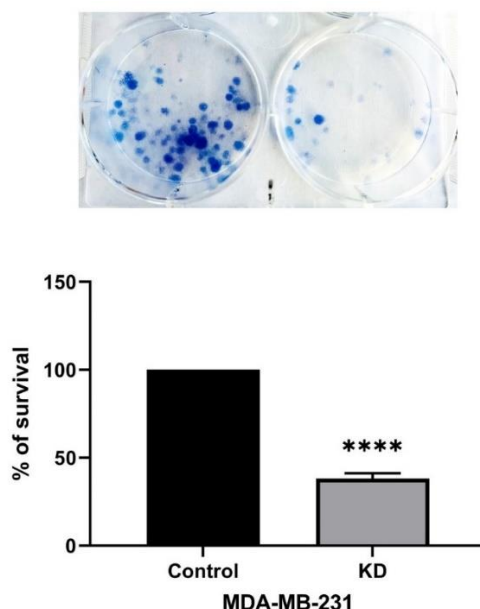


Figure 7.3: The HMGB3 protein expression assessed in the triple negative breast cancer (MDA-MB-231) cell line. The ability of the tumour cells to form colonies was reduced after HMGB3 siRNA transfection. The p -value was interpreted as **** $p < 0.0001$.

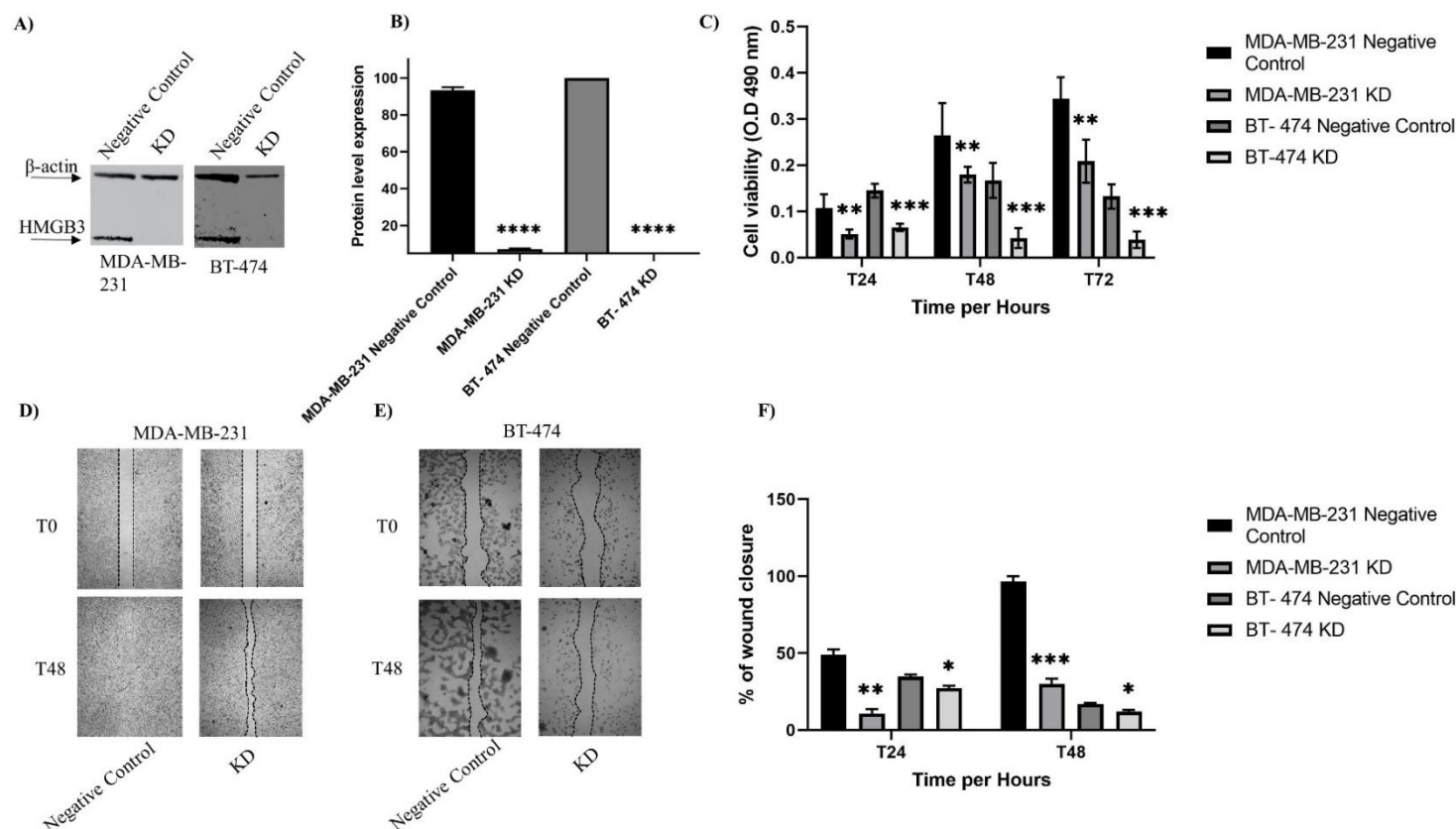


Figure 7.4: The HMGB3 protein expression in the triple negative breast cancer (MDA-MB-231) and luminal B–HER2 positive (BT-474) cell lines. Evaluation of the HMGB3 expression cell lines after transfection with *HMGB3* siRNA and negative control scrambled siRNA (a); the level of protein expression of HMGB3 in the KD cell compared to the negative control quantified by densitometry and normalised to β -actin levels showed almost complete loss of HMGB3 protein expression (b); the effect of KD *HMGB3* using siRNA in the proliferation ability of the cell line compared to the non-transfected negative control (c); the migration (wound closure) assay images showed the effect of *HMGB3* KD in MDA-MB-231 (d) and BT-474 (e) cell line migration; the migration effect in T0 and T48 in both the KD and negative control (f). The standard error of the mean was calculated in all experiments. The *p*-values were interpreted as **p* < 0.05, ***p* < 0.01, ****p* < 0.001 and *****p* < 0.0001.

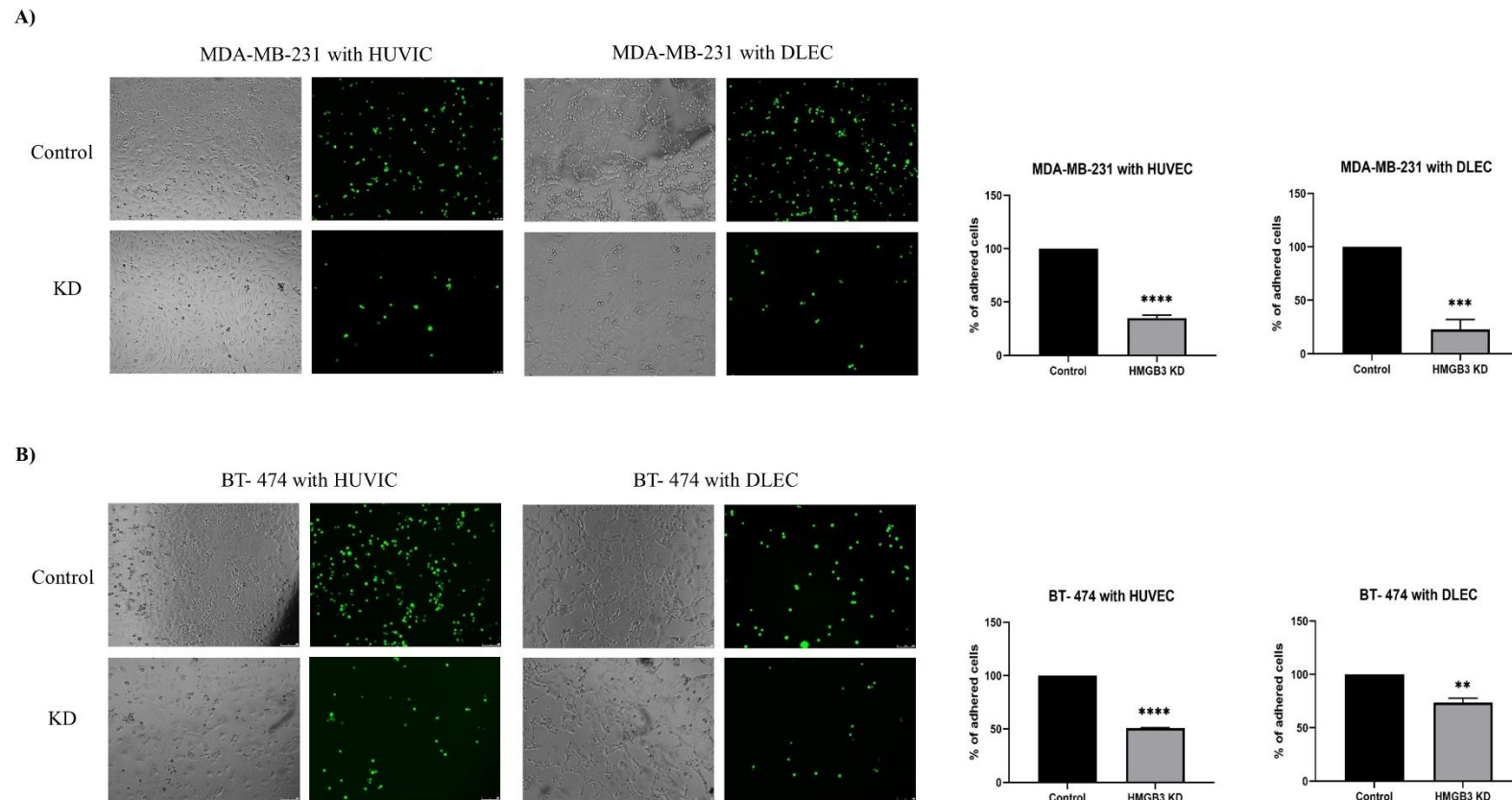


Figure 7.5: The photomicrograph of the triple negative breast cancer (MDA-MB-231) and luminal B–HER2 positive (BT-474) cell line adhesion with vascular and lymphatic endothelial cells (HUVECs and DLECs). The number of adhering cancer cells to the endothelial cells decreased after *HMGB3* knockdown in MDA-MB-231 (a) and BT-474 (b) BC with both the HUVECs and DLECs. The standard error of the mean was implicated in the experiments. The *p*-values were interpreted as **p* < 0.05, ***p* < 0.01, ****p* < 0.001 and *****p* < 0.0001.

Chapter 7

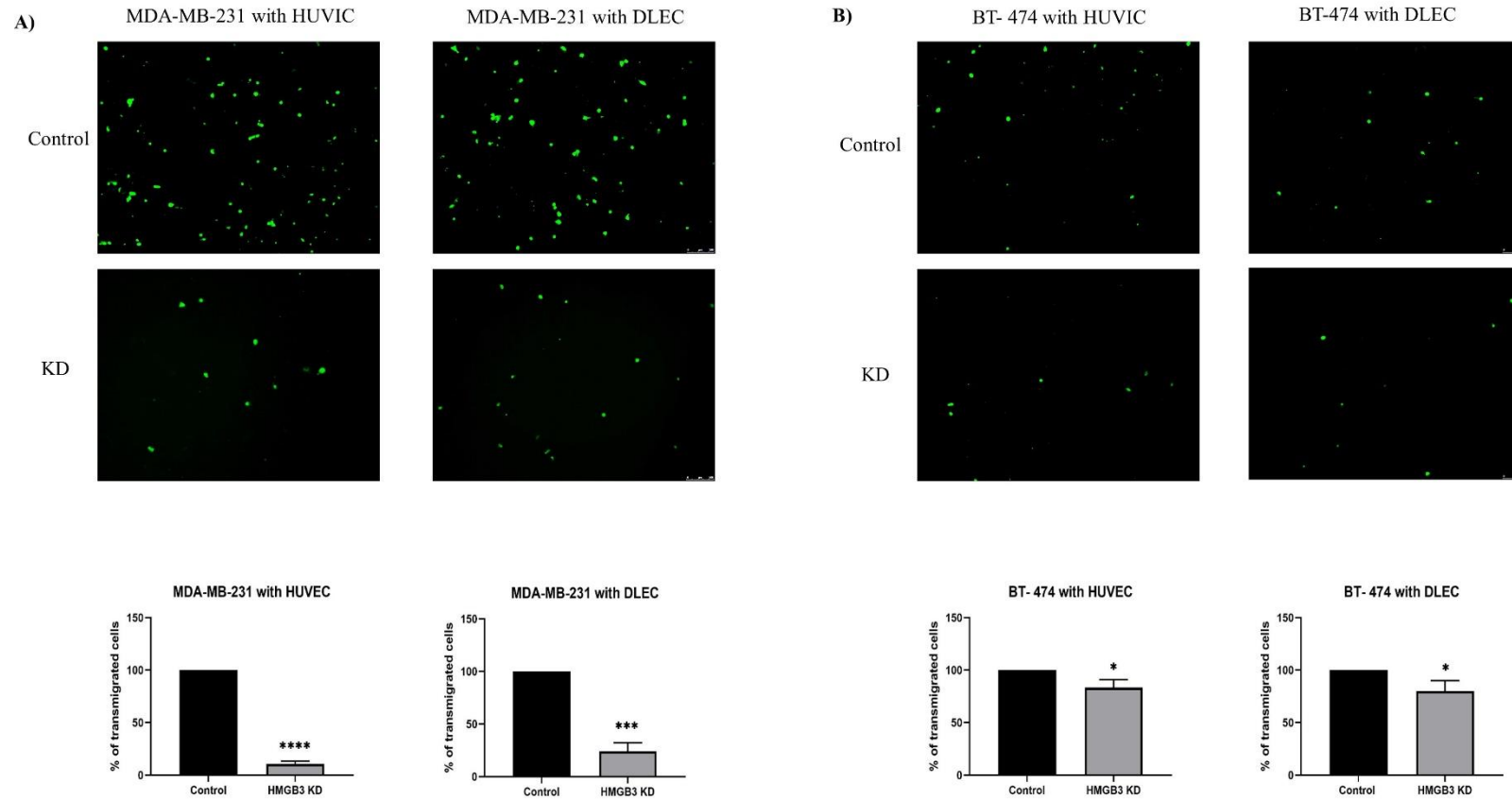


Figure 7.6: The photomicrograph of the triple negative breast cancer (MDA-MB-231) and luminal B–HER2 positive (BT-474) cell lines’ transmigration across vascular and lymphatic endothelial cells (HUVECs and DLECs). The number of transmigrated cancer cells across the endothelial cells decreased after *HMGB3* knockdown in MDA-MB-231 (**a**) and BT-474 (**b**) BC with both the HUVECs and DLECs. The standard error of the mean was implicated in the experiments. The *p*-values were interpreted as **p* < 0.05, ***p* < 0.01, ****p* < 0.001 and *****p* < 0.0001.

7.4.2 Clinical Results

7.4.2.1 Clinicopathological Significance of *HMGB3* mRNA Expression

The TNM plot portal data were utilised to investigate the prognostic value of *HMGB3* mRNA expression. High *HMGB3* mRNA expression was found more in tumoural and metastatic cases than in normal breast tissue samples (Figure 7.7). High *HMGB3* mRNA expression showed a significant association with BC poor prognostic features, including LVI positivity, high histological grade, and hormonal receptor ER and PR negativity (in both the METABRIC and TCGA-BRCA cohorts: $p < 0.001$) (Table 7.1). Considering molecular subtypes, high *HMGB3* mRNA expression was significant in luminal B and HER2 cases in the METABRIC cohort and luminal B and TNBC cases in the TCGA-BRCA cohort (both cohorts: $p < 0.0001$) (Table 7.1). The *HMGB3* mRNA expression level was higher in the luminal B, HER2-enriched and MDA-TNBC BC subgroup compared to other BC subtypes ($p < 0.0001$) using bc-GenExMiner v4.9 (Figure 7.8). Moreover, a high *HMGB3* level was significantly correlated with Ki67 and AKT levels (in both the METABRIC and TCGA-BRCA cohorts: $p < 0.001$) (Table 7.2).

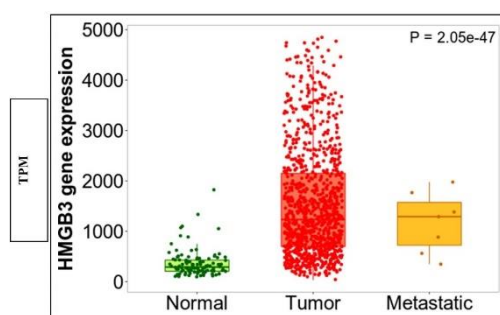


Figure 7.7: The *HMGB3* mRNA gene expression in breast cancer tissue based on sample type in RNA-Seq Data. (p -value is a comparison for multiple independent samples).

Table 7.1: Association of *HMGB3* mRNA expression with clinicopathological characteristics in the METABRIC (n = 1980) and TCGA-BRCA (n = 854) cohorts.

Factors	METABRIC			TCGA-BRCA		
	Low <i>HMGB3</i>	High <i>HMGB3</i>	<i>p</i> -value	Low <i>HMGB3</i>	High <i>HMGB3</i>	<i>p</i> -value
	N (%)	N (%)		N (%)	N (%)	
Lymphovascular invasion						
Negative	521 (56.0)	409 (44.0)	<0.001	306 (54.7)	253 (45.3)	<0.001
Positive	283 (44.6)	352 (55.4)		121 (41.0)	174 (59.0)	
Tumour size						
≤2.0 cm	328 (52.7)	294 (47.3)	0.081	123 (51.5)	116 (48.5)	0.594
>2.0 cm	649 (48.5)	689 (51.5)		304 (49.4)	311 (50.6)	
Age						
≤50	209 (49.3)	215 (50.7)	0.742	128 (49.4)	131 (50.6)	0.823
>50	781 (50.2)	775 (49.8)		299 (50.3)	296 (49.7)	
Nodal status						
Negative	555 (53.1)	485 (46.9)	0.004	219 (51.4)	207 (48.6)	0.319
Positive	437 (46.6)	501 (53.4)		203 (48.0)	220 (52.0)	
Histological grade						
Grades 1 and 2	566 (60.2)	374 (39.8)	<0.001	307 (66.2)	157 (33.8)	<0.001
Grade 3	375 (39.4)	577 (60.6)		96 (27.3)	256 (72.7)	
Oestrogen receptor						
Negative	170 (35.9)	304 (64.1)	<0.001	30 (16.2)	155 (83.8)	<0.001
Positive	820 (54.4)	686 (45.6)		383 (59.9)	256 (40.1)	
Progesterone receptor						
Negative	433 (46.1)	507 (53.9)	<0.001	80 (29.4)	192 (70.6)	<0.001
Positive	557 (53.6)	483 (46.4)		329 (60.3)	217 (39.7)	
Human epidermal growth factor receptor 2						
Negative	935 (54)	798 (46)	0.008	294 (51.9)	273 (48.1)	0.013
Positive	55 (22.3)	192 (77.7)		53 (39.8)	80 (60.2)	
Nottingham prognostic index						
Good prognostic group	421 (61.9)	259 (38.1)	<0.001			
Moderate prognostic group	490 (44.5)	611 (55.5)				
Poor prognostic group	79 (39.7)	120 (60.3)				
Molecular subtypes						
Luminal A	456 (63.5)	262 (36.5)	<0.001	285 (70.4)	120 (29.6)	<0.001
Luminal B	194 (39.8)	294 (60.2)		56 (39.7)	85 (60.3)	
HER2-enriched	55 (22.9)	185 (77.1)		5 (8.9)	51 (91.1)	
Basal-like	140 (42.4)	190 (57.6)		15 (11.3)	118 (88.7)	
Normal-like	144 (72.4)	55 (27.6)		19 (63.3)	11 (36.7)	
Values in bold are statistically significant.						

Table 7.2: Association between *HMGB3* mRNA expression and other genes of interest in the METABRIC and TCGA-BRCA invasive breast cancer datasets.

Gene of Interest	<i>HMGB3</i> mRNA Level (METABRIC)		Mean Rank		Z-score	p-value	<i>HMGB3</i> mRNA Level (TCGA-BRCA)		Mean Rank		Z-score	p-value
	Low (%)	High (%)	Low	High			Low (%)	High (%)	Low	High		
<i>Ki67</i>	990 (50)	990 (50)	834	1146	-12.177	<0.001	427 (50)	427 (50)	322	532	-12.461	<0.001
<i>PIK3CA</i>			989	991	-0.395	0.693			420	434	-0.810	0.418
<i>AKT</i>			891	1089	-7.688	<0.001			406	448	-2.514	0.012

Values in bold are statistically significant.

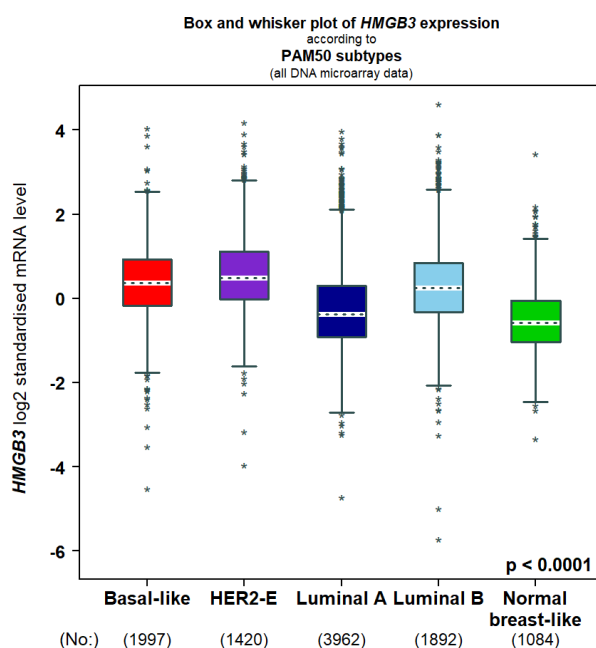


Figure 7.8: The *HMGB3* mRNA gene expression on PAM50 subtypes using bc-GenExMiner v4.9.

7.4.2.2 Patient Outcome Analysis of HMGB3 mRNA Expression

A significant positive correlation was found between high *HMGB* mRNA expression and shorter patient survival in METABRIC (HR = 1.519, 95% CI = 1.206–1.913, $p < 0.001$; Figure 7.9a) and TCGA-BRCA (HR = 2.330, 95% CI = 1.382–3.927, $p = 0.001$; Figure 7.9b). High *HMGB3* mRNA expression was also significantly associated with poor patient outcome in bc-GenExMiner v4.9 (Affymetrix; HR = 1.64, 95% CI = 1.137–1.96, $p < 0.001$; Figure 7.9c).

High *HMGB3* mRNA expression was a significant predictor of shorter survival in BCSS based on LVI status in both the METABRIC and TCGA-BRCA cohorts. However, LVI– cases showed poor prognosis but were not statically significant in METABRIC (HR = 1.427, 95% CI = 0.943–2.158, $p = 0.092$) and TCGA-BRCA (HR = 1.1782, 95% CI = 0.886–3.584, $p = 0.105$) (Figure 7.9d and Figure 7.9e, respectively). In contrast, the LVI+ cases showed a significant association with poor outcomes in the METABRIC dataset (HR = 1.447, 95% CI = 1.023–2.046, $p = 0.034$; Figure 7.9 f) and TCGA-BRCA dataset (HR = 2.826, 95% CI = 1.222–6.539, $p = 0.015$; Figure 7.9 g).

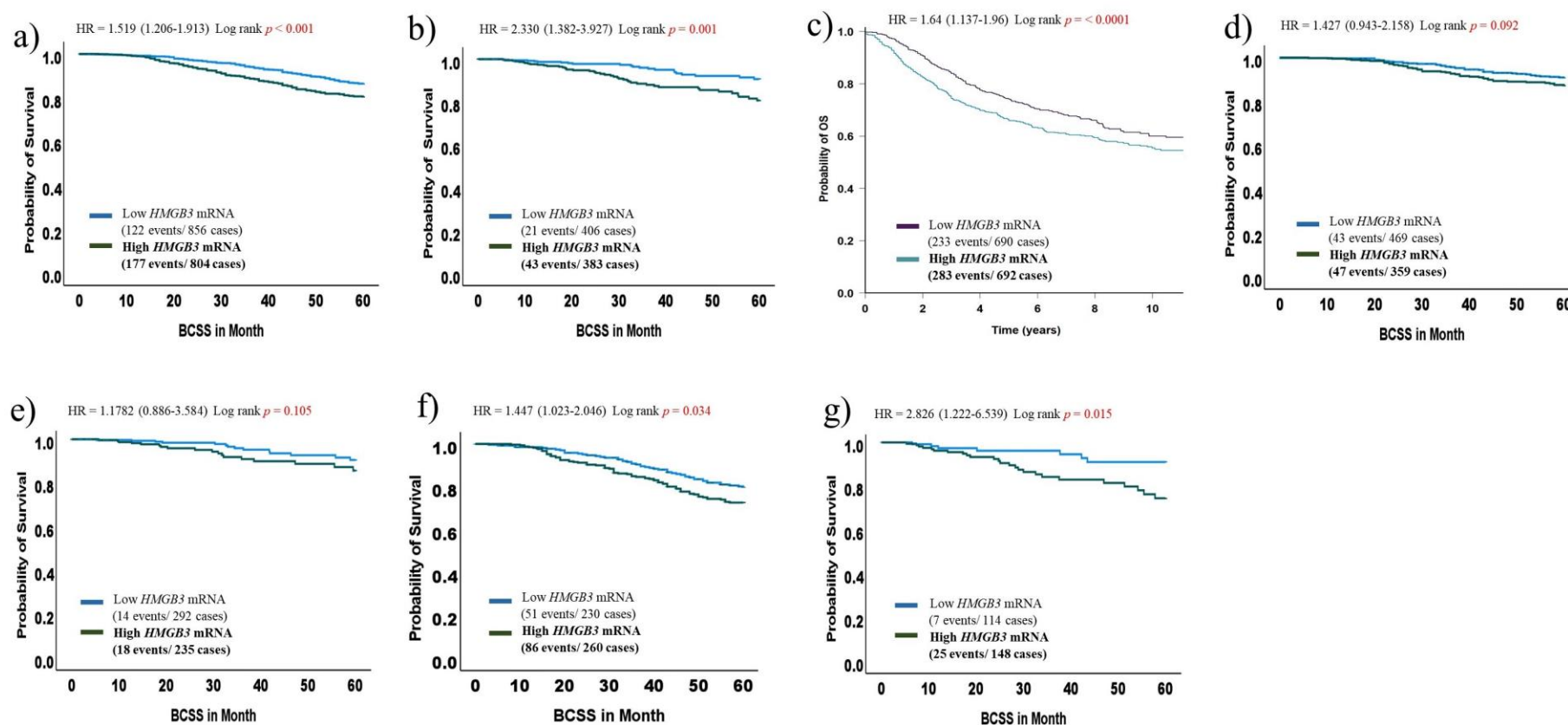


Figure 7.9: The Kaplan–Meier survival plots showing cumulative survival of patients with breast cancer stratified by *HMGB3* mRNA expression: METABRIC (a), TCGA-BRCA (b), Breast Cancer Gene-Expression Miner (Affymetrix) (c), LVI– in METABRIC (d), LVI– in TCGA-BRCA (e), LVI+ in METABRIC (f) and LVI+ in TCGA-BRCA (g). OS, overall survival; BCSS, breast cancer-specific survival.

7.5 Discussion

The engagement of *HMGB3* in various cellular functions is one of the characteristics of this gene which highlights its possible role in regulating cancer progression (433). According to recent studies, the mutant form of *HMGB3* is a procarcinogen that promotes cancer growth, proliferation, invasion and dissemination (418). This has been recorded in various cancer types, including lung cancer (434), urinary bladder carcinoma (427), gastric cancer (GC) (435) and colorectal carcinoma (418). However, research regarding *HMGB3* potential role in BC LVI has yet to be conducted.

In a previous study, *HMGB3* was identified via a rigorous bioinformatics approach as one of the most DEGs linked to LVI+ status (223). Likewise, a study using bioinformatics analysis also found a connection between *HMGB3* and tumour progression in GC (414). The current study aimed to (1) conduct an *in vitro* examination to explore the mechanistic function of *HMGB3* in different BC cell lines and (2) examine the clinicopathological importance of *HMGB3* at the mRNA level utilising large BC cohorts.

Tumour cell proliferation is a critical prognostic indicator throughout the cancer diagnostic process (436). The preclinical investigations that were carried out for this research revealed that inhibiting *HMGB3* leads to a reduction in the proliferation of tumour cells. These results support earlier reports (422, 427, 437). A previous study showed a decrease in the growth of xenograft tumours originating from *HMGB3*-KD HCC cells compared to negative control tumours, with a significant decrease in the Ki67 proliferation marker level. It proposed the contribution of *HMGB3* in cancer development *in vivo* by manipulating the proliferation process in HCC cells (433). The positive correlation between high *HMGB3* mRNA expression and Ki67, a proliferation-related marker (436), in this study supports this. To invade neighbouring tissues and initiate the process of metastasis, cancer cells must first proliferate and then escape apoptosis. As a consequence of this process, BC cells eventually invade the lymphatic channels.

Moreover, silencing the *HMGB3* gene *in vitro* resulted in minimised BC cell migration. This finding is in agreement with an earlier report showing that *HMGB3* causes stimulation of colorectal cancer cell migration through the Wnt/ β -catenin signalling pathway. This effect was observed by assessment expression of β -catenin, a main WNT pathway protein. Moreover, *HMGB3* increases the expression of not only β -catenin but also the downstream genes, including *MMP-7* and *c-Myc* (418).

The current understanding of *HMGB3* suggests it is implicated in tumour progression; however, its carcinogenic effects and mechanism of action in BC LVI remain unclear. Notably, *HMGB3* KD in this study decreased adhesion and transmigration across endothelial cell lines in the MDA-MB-231 and BT-474 cells. However, the MDA-MB-231 cells showed higher affinity for all endothelial cell assays than the BT-474 cells, reflecting its invasiveness and aggressiveness, which is in agreement with other studies (438-440). The positive correlation between PI3K, AKT and *HMGB3* at the mRNA level supports these findings. The PI3K/AKT/mTOR pathway significantly contributes to several intracellular mechanisms, such as cell metabolism, proliferation, angiogenesis and survival (441). Moreover, it has been shown to play an important role in BC progression and metastasis (439) and contribute to tumour resistance to targeted treatments (442, 443). In addition, a previous study suggested that the PI3K/AKT signalling pathway enhances tumour cell progression by contributing to integrin signalling, and *in vivo* inhibition of the PI3K/AKT pathway reduces scirrhous gastric carcinoma cell adhesion, consequently minimising its spread to the ECM (444). Moreover, AKT affects adhesion through regulating focal adhesion kinase (FAK), which is an AKT upstream. Focal adhesion kinase plays a role as a regulator in cancer progression and has been found to be linked to lymph node metastasis and shorter survival in oral squamous cell carcinoma (445). Furthermore, AKT regulates FAK via response to extracellular pressure, which enhances cancer cells' adhesion, and its reduction *in vitro* has been found to minimise tumour cell metastasis in colon cancer by affecting adhesion (446). In BC, PI3K regulates the Rac pathway (447), which is critical for cancer invasion and dissemination through regulation of the cytoskeletal remodelling, cell–cell adhesion, motility and proliferation of the cell (448, 449). Interestingly, upregulation of Rac1 activity is linked to lymph node metastasis and LVI and is an independent prognostic survival factor for patients with an unfavourable prognosis of urothelial carcinoma of the upper urinary tract (450). Another study revealed that inhibition of Rac or PI3K reduces the adhesion and transmigration of MDA-MB-231 in both melanoma and BC cells (438). The significant reduction seen in adhesive and transmigrated tumour cells after *HMGB3* silencing highlights the vital role of *HMGB3* in these steps, which might mean that *HMGB3* stimulates the PIK3/AKT/mTOR and Rac pathways to facilitate tumour cell invasion and metastasis.

Evidence exists supporting the role of *HMGB3* in triggering the LVI process in the BT-474 and MDA-MB-231 cell lines. Such an effect has been observed by overexpression of *HMGB3*, which promotes migration, adhesion and transmigration of cancer cells via endothelial cell lines. However, *in vivo* confirmation of these findings and a mechanistic understanding of *HMGB3*'s role in triggering LVI are needed.

This research investigated the clinicopathological and prognostic significance of *HMGB3* mRNA utilising well-characterised BC datasets. At the transcriptomic level, high *HMGB3* mRNA expression was significantly correlated with the most aggressive BC parameters, including LVI+ status, high histological grade, NPI poor prognostic group, hormonal receptor negativity and HER2 subgroup positivity. The survival plot analysis showed a significant correlation between high *HMGB3* mRNA expression and unfavourable BCSS, which aligns with prior research (419, 434). The results support the notion that *HMGB3* plays an oncogenic role in BC and raise the possibility that it contributes to LVI pathogenesis. The positive correlation of *HMGB3* and several pathways emphasises its potential role in BC cell proliferation and might be involved in BC LVI status. The influence of *HMGB3* silencing on some tumour formation processes, such as adhesion, migration and transmigration, was demonstrated in this study; however, the molecular mechanisms underlying LVI development in BC were not examined. Even though the findings on lymphatic invasion and the precursor contributions to the metastatic cascade of *HMGB3* in this research suggest that it is an important marker associated with LVI, additional studies are needed to confirm these results.

8 Chapter 8: General discussion

8.1 Overview

Globally, BC is considered the leading cause of mortality in women (2). Thus, improving the strategies of prevention, diagnosis and treatment will enhance patient survival and ensure women's good health. However, obstacles are always exhibited in each improvement plan, and continuous efforts to improve patient health are always needed. Breast cancer features heterogeneity, often accompanied by distinct morphological variations and unique molecular characteristics for each patient (451, 452). Therefore, targeting BC cells at an early cancer stage has become the focus of many studies to design new effective treatments that can prevent tumour progression and metastasis. This approach could be enhanced by better understanding of BC's molecular, biological, and inter- and intra-tumoural mechanisms (157-159).

Interestingly, BCSS and LVI are strongly correlated in clinical studies (453). Molecularly, BC cells can react and behave uniquely to enhance the pre-metastasis steps, including by reducing adhesion, migration and invasion, to facilitate metastasis through LVI (161, 454). Currently, there is limited knowledge about LVI-driving genes. Thus, finding tumour-associated pathways and locating clusters of DEGs through investigating transcriptomic- and proteomic-based approaches has grabbed researchers' attention and shown efficacy in enhancing understanding of cancer development (455). The chance to explore new methodologies to understand cancer's molecular mechanisms exists. Thus, determining LVI drivers that enhance cancer cells' aggressiveness is urgent to create novel prognostic biomarkers and treatment possibilities for IBC in poor-outcome patients. This study aimed to create a panel of targeted genes associated with LVI status and uncover the pathways that potentially contribute to BC LVI positivity by understanding the prognostic association between candidate genes and LVI. Due to a lack of well-studied LVI signatures in IBC, this research attempted to apply a high-throughput technique that evaluated the association of DEGs with LVI status in patients with BC. In this technique, both the transcriptomic and proteomic levels were investigated.

Although transcriptomic approaches to LVI candidate identification may provide a huge number of markers from which to choose, understanding LVI pathways requires functional proteins. Biologically, cancer cells require various cellular, biophysical and biochemical machines to accomplish the metastasis journey through LVI. As discussed in **Chapter 1 (section 1.3.1)**, the LVI process in IBC may be controlled and manipulated by certain gene and protein expression patterns. Among several studies that have investigated LVI gene signatures, only a few have been performed on BC (262), HCC (261) and EC (260), and their findings hinted at distinct groups of DEGs linked to LVI processes. However, these results were

accompanied by a few study limitations, such as use of a small cohort or lack of LVI status confirmation despite utilisation of advanced techniques; this has been discussed more in **Chapter 3 (section 3.5)**.

8.2 Hypothesis, Aim and Methodology

It was hypothesised that specific genes and proteins can activate or inhibit LVI progression in patients with BC through different converged or parallel molecular and biological pathways. Therefore, to validate this hypothesis, this study conducted several experiments with different aims: First, exploring the main drivers of LVI in BC was the initial aim through conducting a robust analysis of DGEs. To uncover the underlining LVI mechanisms in IBC, LVI status was investigated in two transcriptomic cohorts by employing high-throughput statistical analysis to investigate DEGs associated with LVI positivity. Second, to validate mass spectrometry analysis work previously carried out, which nominated *HSP90AA1* as a promising gene associated with LVI that was later revealed to be associated with LVI status based on the DGE analysis. Lastly, the Nottingham BC cohort was stained as an external validation for identified prognostic biomarkers to validate their role in LVI positivity and IBC patient outcomes.

8.3 Chapter 3: Bioinformatics approach and differential gene expression

Efforts to investigate the underlying molecular mechanisms of BC LVI have been met with only modest success. This is likely because most of these studies either evaluated the prognostic expression of genes that were already known or used a small number of cases. Furthermore, the lack of IHC evaluation of identified prognostic genes related to LVI in the earlier efforts to discover BC gene signatures might reflect a lack of clinical participation during such investigations. Thus, the focus of this study was to identify a novel set of genes associated with LVI from well-established mRNA cohorts with the help of the DGE technique. As a result, this study identified 42 genes that are positively and significantly associated with LVI; however, 57 genes exhibited a negative correlation with LVI status (Table 3.3) (223). The molecular and biological activities of the genes that were correlated positively with LVI+ were identified using several online biological tools to understand the underlying molecular function of the designated genes (Table 3.4 and Figure 3.2). Moreover, *SLC9A3R1*, *NOP56* and *HMGB3* were among the genes with a positive association that were chosen for further validation and investigation of their role in BC and the linked pathways between these genes and LVI+ status.

8.4 Chapter 4: HSP90 α , a gene associated with LVI status selected from the DGE analysis and previous mass spectrometry analysis

The mechanisms that drive BC progression and poor outcome are not fully understood. However, LVI is known as an essential step in the progression and metastasis of various tumours, which has encouraged many researchers to study the molecular mechanisms behind LVI to enhance patient management and choose the most effective treatment to improve patients' chances of improved outcome (255, 262). In the current study, HSP90 α was chosen to explore its clinical and prognostic importance in BC. Also, HSP90 α has been shown to have a vital role in cellular responses to stress and is implicated in the development and progression of many cancers (289), (292-294).

High *HSP90AA1* transcriptomic and proteomic expression is also associated with poor prognostic BC parameters (Table 4.1 Table 4.3). Moreover, *HSP90AA1* has shown significant association with many oncogenic signalling pathways, such as PIK3CA and p53; (Table 4.2) most of these pathways contribute to the cell cycle, apoptotic inhibition and, consequently, tumour progression and metastasis (309, 310), which may be utilised to enhance the LVI process. This study also showed a correlation between *HSP90AA1* and aggressive BC subtypes, such as luminal B, HER2 and TNBC, which show a significant association with LVI+ status (Table 4.1 Table 4.3). Short patient survival was observed in a high *HSP90AA1* expression group (Figure 4.2). Therefore, the outcome analysis revealed that high HSP90 α protein expression is an independent predictor of shorter BCSS regardless of tumour size, axillary nodal stage and LVI (Table 4.4). Thus, HSP90 α could be used as a potential prognostic marker in BC. Further mechanistic studies are warranted to determine the underlying molecular mechanisms mediated by HSP90 α in BC LVI.

8.5 Chapter 5: SLC9A3R1, a biomarker that was chosen from the LVI-related gene signature set

Lymphovascular invasion is linked with early stage BC invasiveness (121). However, the signalling pathways that contribute to LVI have not yet been fully discovered. High-throughput analysis of multiple BC cohorts has demonstrated a strong association between the *SLC9A3R1* gene and LVI+ status (223). Thus, this study aimed to explore the prognostic significance of SLC9A3R1 in large cohorts of IBC and its relationship with LVI+.

In this study, high *SLC9A3R1* mRNA and/or SLC9A3R1 protein expression was significantly correlated with aggressive tumour behaviour, including LVI positivity (Table 5.3 Table 5.5).

Moreover, SLC9A3R1 showed amplification at both transcriptomic levels, which increased its expression in BC (Table 5.1 Table 5.2). The relationship between LVI and SLC9A3R1 is indirect; therefore, the mechanisms accompanying SLC9A3R1 overexpression, which may lead to LVI, were explored. For instance, SLC9A3R1 shares the same chromosome site with *ErbB2*, elevating the VEGF family when overexpressed (359). When the tumour cell overexpresses VEGFR-3, it is thought to contribute to cancer invasion, migration and metastatic spread through lymph vessels (359). Moreover, *ErbB2* overexpression could manipulate MMP-9 and MMP-2 activities, which would enhance LNM and LVI (360). **Chapter 1 (section1.3.1)** describes more details about the above mechanisms.

Nevertheless, the high expression of HER2 associated with SLC9A3R1 in this study (Table 5.5) and the opposite KD relationship previously found between them (344) highlighted the role of SLC9A3R in therapeutic improvement. This study's results indicated that SLC9A3R1 plays a role in BC progression and may contribute to LVI. Consequently, poor patient survival and high SLC9A3R1 protein expression are significant independent predictors for shorter BCSS (Table 5.6).

8.6 Chapter 6: NOP56, a biomarker that was chosen from the LVI-related gene signature set

The NOP56 protein is part of a nucleolus complex that plays a role in pre-rRNA processing (376, 377), which may influence BC cell survival, proliferation and metastasis. It was selected here according to its strong relationship with LVI+ status (223). Lymphovascular invasion is linked with invasiveness in BC and short survival (65). This study aimed to investigate the clinicopathological significance of NOP56 expression in multiple cohorts of BC and evaluate its relationship with LVI.

High *NOP56* mRNA and/or NOP56 protein expression was significantly correlated with aggressive tumour behaviour, including high histological grade, larger tumour size, positive nodal status, LVI positivity, hormone receptor negativity and NPI poor prognostic group (Table 6.1Table 6.3). High NOP56 protein expression was associated with high Ki67, PI3KCA and EGFR (Table 6.2), which all participate in BC cell proliferation, migration and metastasis (407, 408, 456), the primary mechanism utilised by BC cells to facilitate the LVI process. Moreover, *NOP56* showed a significant correlation with MYC (Table 6.2), a critical signalling pathway that induces BC cell survival during the treatment (387, 397). Also, NOP56 expressed more TNB positivity (Table 6.3), associated with a high incidence of LVI and BC recurrence (407).

Therefore, targeting NOP56 could help in patients' TNBC treatment. High NOP56 protein expression was an independent biomarker presenting a poor BCSS prognosis (Table 6.4). The association of NOP56 with the most aggressive clinicopathological parameters and signalling pathways points to its essential role in BC progression, taking together apoptosis inhibition, proliferation, tumour cell migration, invasion and metastasis. Therefore, NOP56 could be considered an independent prognostic biomarker of LVI+ IBC.

8.7 Chapter 7: Mechanistic insights of HMGB3, a biomarker that was chosen from the LVI-related gene signature set

The contribution of HMGB3 to several intracellular mechanisms explains its role in cancer development and metastasis (433). Its ability as a procarcinogen facilitates tumour cascade sequence steps from tumour initiation to metastasis (418). Moreover, *HMGB3* has been identified as one of the most DEGs linked to LVI+ status (223). In addition, research into *HMGB3* mRNA expression has illustrated the possibility of *HMGB3* engagement in the BC progression process, starting from various physiological and pathological mechanisms to LVI formation via detaching tumour cells from the primary site after losing adhesion, starting migration, communicating with TME and, eventually, invading the lymphovascular system. This has been seen at the transcriptomic level of *HMGB3* utilising large, well-defined and characterised BC cohorts. The current investigation revealed that high *HMGB3* mRNA expression is significantly correlated with poor prognostic clinicopathological characteristics, including LVI positivity, high tumour grade, NPI poor prognostic group, hormonal receptor negativity, high level of HER2 (Table 7.1), high level of proliferative marker Ki67, high level of AKT and positive association with PI3K (Table 7.2).

According to the aforementioned clinical observations and the prognostic significance of HMGB3 relative to BC, this study aimed to evaluate the effect of HMGB3 in BC cell lines utilising *in vitro* assays and connect this to its relevance to LVI. This study showed a significant reduction of HMGB3 protein expression after the KD process in both the MDA-MB-231 and BT-474 cell lines (Figure 7.2), and silencing *HMGB3* enhanced survival efficiency, proliferation, adhesion, migration and transmigration ability (Figure 7.3, Figure 7.4, Figure 7.5 and Figure 7.6). Overall, the mentioned research will help pave the path for additional strategies to target HMGB3 activity utilising aggressive BC cells.

This research demonstrated that HMGB3 is critical for the development of IBC tumours. Furthermore, since a rigid preclinical framework for new therapy may be attainable, HMGB3

may constitute a potential candidate for siRNA-targeted treatment in patients with IBC. There are many benefits of using siRNA nanoparticles as therapeutic systems, including the capability to knock genes in living cells, efficiency in suppressing genes by synthesising and optimising siRNA and the increased accessibility of many siRNA design tools that enable siRNA length modification, specificity and nucleotide composition utilising several online platforms (457, 458). Researchers have found that using siRNA nanoparticles to treat various tumour types might reduce tumorigenesis, angiogenesis and distant metastasis while increasing medication efficiency (459-463). For this to be possible, the current study's findings of *HMGB3* siRNA suppression must corroborate through additional *in vivo* mechanistic research employing mice models. When this is done, patients with IBC who will benefit from *HMGB3* siRNA therapy may be identified, taking into account strict conformity with established norms (464). This includes demonstrating the post-transcriptional activity of the *HMGB3* siRNA that was designed, providing a high degree of specificity in the siRNA's ability to downregulate its target and validating the detection that *HMGB3* mRNA expression is reducible through a variety of siRNA sequencing strategies to ban non-specific impacts (465, 466) and identifying the most effective means of siRNA delivery (465-467). Once these aims are achieved and verified data are acquired, a preclinical proposed project including many substantial phases may be implemented (468). In conclusion, the results showed that *HMGB3* might be a potential treatment marker in IBC, implying a possible new treatment strategy to limit the progression of BC.

8.8 Potential Strengths

1. This study designed a gene predictor panel associated with LVI+ status in BC, revealing a novel gene signature for patient risk stratification that could be utilised in future investigation studies.
2. Using a robust approach, DGEs and clustering analysis to investigate LVI-related genes, this study identified DEGs associated with LVI positivity in two large transcriptomic cohorts (METABRIC and TCGA-BRCA). All available transcriptomic and protein cohorts were used for further validation to confirm these findings.
3. Novel prognostic markers (SLC9A3R, NOP56 and *HMGB3*) were identified, all positively associated with LVI status in patients with IBC.
4. The targeted genes were validated using well-characterised primary IBC series (NPS) with a large amount of characteristic patient data including molecular subtypes, responses to therapy, validated tissue markers and patient outcomes.

5. Western blotting was used to determine the specificity of targeted Abs followed by IHC with the inclusion of positive and negative controls to achieve accurate staining results. Also, the the reporting recommendations for tumor marker prognostic studies guidelines (469) were considered in this study for more robust IHC data interpretation and to perform a valid statistical association.
6. The validation of marker expression in the tissue was performed manually and optically with the help of a qualified pathologist; this determination technique is more robust and discriminative than the automated technique. With this method, an individual can make better judgements than an automated program, which differentiates less among various tissue morphologies.

8.9 Potential Limitations

1. Some TMA cores were missing or had limited tumours in the TMA blocks used in this study, which have been ago.
2. While this study was based on retrospective cohort data with long clinical follow-up time, multiple independent data resources are needed to validate these results. In addition, numerous proteomic expression evaluations were unavailable because one centre participated in providing the IHC cohort.
3. Due to insufficient clinicopathological data, the new Nottingham IBC cohort was not used.
4. It took quite some time to obtain and verify LVI status in the transcriptomic cohorts used in this study (METABRIC and TCGA-BRCA).
5. The homogeneity of protein expression was verified using a large sample size of full-face sections. However, this was accompanied by multiple technical limitations, including core damage. Researchers using TMAs for research purposes have approved the single-core use of TMAs (342, 343). However, it would be better to enhance the evaluation by utilising multiple cores for each case to eliminate any tumour heterogeneity.
6. Even though double-scoring of the cores was performed, this study relied on H-scoring, a semi-quantitative method which depends on the scorer's subjective opinion.
7. The coronavirus 2019 restrictions (2020-2022) limited the time available for the study, so the author could not do more experimental studies to uncover the mechanisms behind identified genes and LVI status in BC.

8. The novel biomarkers identified in the study were selected according to their association with LVI. However, it was challenging to prove their interaction, as each has its own biological and signalling pathways. This expanded the search field and added more complexity to the task.

8.10 Future Work

1. Research focusing on IBC should be performed, especially LVI molecular regulator studies, using multiple available data sources and contributing to several related articles.
2. Various mechanistic experiments should be performed including *in vivo* and *in vitro* work, such as animal models, KD and xenograft studies utilising multiple BC cell lines. These experiments should use the most aggressive cell lines to uncover the role of SLC9A3R, NOP56 and HMGB3 in LVI and explore the therapeutic capacity of the target genes
3. Additional functional investigations are necessary to validate the functions of the NOP56 protein. These could consist of the cell cycle, apoptosis, proliferation, migration, transmigration and survival assays.
4. Multiple sources of patient data are required in future studies to assess the targeted biomarker, which will ensure robust results and, consequently, improve the clinical value and quality of the study.
5. Contribute to research by publishing multiple articles to comprehend the molecular mechanism of LVI by deciphering the transcriptomics and proteomics of online global LVI data.

8.11 Final Conclusions

The current study concluded that some specific molecules are essential in driving LVI status and its underlying molecular mechanisms at the transcriptomic and proteomic levels. With the help of the DGE technique and IBC transcriptomic cohorts, various genes and proteins were found to be associated with LVI. Furthermore, the statistical analysis conducted in this study revealed SLC9A3R1, NOP56 and HMGB3 as being highly associated with LVI in IBC. The diligent LVI analysis performed to discover new genes showed that the primary BC phenotype plays a fundamental role during the differential control of LVI mechanisms in IBC and, consequently, is involved in regulating several factors that activate the cancer cells' aggressiveness. Additionally, LVI mechanisms in IBC may also be influenced by cross

pathways depending on various ECM stromal component interactions and TME conditions. Further validation of the current findings should be carried out using functional studies.

9 References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-74.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018;68(6):394-424.
3. Beckmann MW, Niederacher D, Schnürch HG, Gusterson BA, Bender HG. Multistep carcinogenesis of breast cancer and tumour heterogeneity. *Journal of molecular medicine (Berlin, Germany)*. 1997;75(6):429-39.
4. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70.
5. Sugimura T, Terada M, Yokota J, Hirohashi S, Wakabayashi K. Multiple genetic alterations in human carcinogenesis. *Environmental health perspectives*. 1992;98:5-12.
6. Zhu K, Liu Q, Zhou Y, Tao C, Zhao Z, Sun J, et al. Oncogenes and tumor suppressor genes: comparative genomics and network perspectives. *BMC genomics*. 2015;16 Suppl 7(Suppl 7):S8.
7. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery*. 2022;12(1):31-46.
8. Stingl J, Caldas C. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nature Reviews Cancer*. 2007;7(10):791-9.
9. Joseph C, Papadaki A, Althobiti M, Alsaleem M, Aleskandarany MA, Rakha EA. Breast cancer intratumour heterogeneity: current status and clinical implications. *Histopathology*. 2018;73(5):717-31.
10. Weigelt B, Geyer FC, Reis-Filho JS. Histological types of breast cancer: how special are they? *Molecular oncology*. 2010;4(3):192-208.
11. Cancer Research UK LSoHaTMor. Breast Cancer Survival Statistics 2014, November 27 [Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/survival#heading-Two>.
12. Youlten DR, Cramb SM, Dunn NAM, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: An international comparison of screening, incidence, survival and mortality. *Cancer Epidemiology*. 2012;36(3):237-48.
13. Latest global cancer data: Cancer burden rises to 19.3 million new cases and 10.0 million cancer deaths in 2020 [Internet]. 2020, December 15. Available from: <https://www.iarc.who.int/news-events/latest-global-cancer-data-cancer-burden-rises-to-19-3-million-new-cases-and-10-0-million-cancer-deaths-in-2020/>
https://www.iarc.who.int/wp-content/uploads/2020/12/pr292_E.pdf.
14. (IARC) IAFRoc. POPULATION FACT SHEETS. World Health Organization. 2020.
15. Gelband H JP, Sankaranarayanan R, et al.,. Cancer: Disease Control Priorities, Third Edition: Chapter 2 The Changing Global Burden of Cancer: Transitions in Human Development and Implications for Cancer Prevention and Control. 2015, Nov 1;3.
16. Momenimovahed Z, Salehiniya H. Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast cancer (Dove Medical Press)*. 2019;11:151-64.
17. Edwards BK, Howe HL, Ries LA, Thun MJ, Rosenberg HM, Yancik R, et al. Annual report to the nation on the status of cancer, 1973-1999, featuring implications of age and aging on U.S. cancer burden. *Cancer*. 2002;94(10):2766-92.

18. Smigal C, Jemal A, Ward E, Cokkinides V, Smith R, Howe HL, et al. Trends in breast cancer by race and ethnicity: update 2006. *CA: a cancer journal for clinicians*. 2006;56(3):168-83.
19. Sandip K Mishra* DS, Pranati Sar and Dharmendra K Bhargava,. Molecular Basis of Aging and Breast Cancer. *J Cancer Sci Ther*. 2013;5.2 069-074.
20. Cancer Research UK. Breast Cancer Incidence Statistic: Breast Cancer Incidence by Age. 2021, October 4.
21. Slaoui M, Mouh FZ, Ghanname I, Razine R, El Mzibri M, Amrani M. Outcome of Breast Cancer in Moroccan Young Women Correlated to Clinic-Pathological Features, Risk Factors and Treatment: A Comparative Study of 716 Cases in a Single Institution. *PLoS One*. 2016;11(10):e0164841.
22. 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;5 Suppl 1(Suppl 1):S2-S8.
23. Honrado E, Benítez J, Palacios J. The molecular pathology of hereditary breast cancer: genetic testing and therapeutic implications. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2005;18(10):1305-20.
24. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ (Clinical research ed)*. 2000;321(7261):624-8.
25. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *The Lancet Oncology*. 2001;2(3):133-40.
26. Hulka BS, Moorman PG. Breast cancer: hormones and other risk factors. *Maturitas*. 2008;61(1-2):203-13; discussion 13.
27. Larsen MJ, Thomassen M, Gerdes A-M, Kruse TA. Hereditary breast cancer: clinical, pathological and molecular characteristics. *Breast Cancer (Auckl)*. 2014;8:145-55.
28. Shroff RT, Hendifar A, McWilliams RR, Geva R, Epelbaum R, Rolfe L, et al. Rucaparib Monotherapy in Patients With Pancreatic Cancer and a Known Deleterious BRCA Mutation. *JCO precision oncology*. 2018;2018.
29. Wang GH, Zhao CM, Huang Y, Wang W, Zhang S, Wang X. BRCA1 and BRCA2 expression patterns and prognostic significance in digestive system cancers. *Human pathology*. 2018;71:135-44.
30. Ramus SJ, Gayther SA. The contribution of BRCA1 and BRCA2 to ovarian cancer. *Molecular oncology*. 2009;3(2):138-50.
31. Mars N, Widén E, Kerminen S, Meretoja T, Pirinen M, Della Briotta Parolo P, et al. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat Commun*. 2020;11(1):6383.
32. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017;551(7678):92-4.
33. Yanes T, Young M-A, Meiser B, James PA. Clinical applications of polygenic breast cancer risk: a critical review and perspectives of an emerging field. *Breast Cancer Research*. 2020;22(1):21.
34. Waters R. Maintaining genome integrity. *EMBO reports*. 2006;7(4):377-81.
35. Antoniou A, Anton-Culver H, Borowsky A, Broeders M, Brooks J, Chiarelli A, et al. A response to "Personalised medicine and population health: breast and ovarian cancer". *Human genetics*. 2019;138(3):287-9.

36. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*. 2018;50(9):1219-24.
37. Dossus L, Boutron-Ruault MC, Kaaks R, Gram IT, Vilier A, Fervers B, et al. Active and passive cigarette smoking and breast cancer risk: results from the EPIC cohort. *Int J Cancer*. 2014;134(8):1871-88.
38. Flom JD, Ferris JS, Tehranifar P, Terry MB. Alcohol intake over the life course and mammographic density. *Breast Cancer Res Treat*. 2009;117(3):643-51.
39. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *The New England journal of medicine*. 2003;348(17):1625-38.
40. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *Bmj*. 2007;335(7630):1134.
41. Picon-Ruiz M, Morata-Tarifa C, Valle-Goffin JJ, Friedman ER, Slingerland JM. Obesity and adverse breast cancer risk and outcome: Mechanistic insights and strategies for intervention. *CA: a cancer journal for clinicians*. 2017;67(5):378-97.
42. Begum P, Richardson CE, Carmichael AR. Obesity in post menopausal women with a family history of breast cancer: prevalence and risk awareness. *International seminars in surgical oncology : ISSO*. 2009;6:1.
43. Neuhaus ML, Aragaki AK, Prentice RL, Manson JE, Chlebowski R, Carty CL, et al. Overweight, Obesity, and Postmenopausal Invasive Breast Cancer Risk: A Secondary Analysis of the Women's Health Initiative Randomized Clinical Trials. *JAMA oncology*. 2015;1(5):611-21.
44. McDonnell DP, Park S, Goulet MT, Jasper J, Wardell SE, Chang CY, et al. Obesity, cholesterol metabolism, and breast cancer pathogenesis. *Cancer Res*. 2014;74(18):4976-82.
45. Jung AY, Hüsing A, Behrens S, Krzykalla J, Obi N, Becher H, et al. Postdiagnosis weight change is associated with poorer survival in breast cancer survivors: A prospective population-based patient cohort study. *Int J Cancer*. 2021;148(1):18-27.
46. Baglietto L, Krishnan K, Stone J, Apicella C, Southey MC, English DR, et al. Associations of mammographic dense and nondense areas and body mass index with risk of breast cancer. *American journal of epidemiology*. 2014;179(4):475-83.
47. Gierisch JM, Coeytaux RR, Urrutia RP, Havrilesky LJ, Moorman PG, Lowery WJ, et al. Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: a systematic review. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2013;22(11):1931-43.
48. Russo J, Moral R, Balogh GA, Mailo D, Russo IH. The protective role of pregnancy in breast cancer. *Breast Cancer Res*. 2005;7(3):131-42.
49. Zhou Y, Chen J, Li Q, Huang W, Lan H, Jiang H. Association between breastfeeding and breast cancer risk: evidence from a meta-analysis. *Breastfeeding medicine : the official journal of the Academy of Breastfeeding Medicine*. 2015;10(3):175-82.
50. Crandall CJ, Aragaki AK, Cauley JA, McTiernan A, Manson JE, Anderson G, et al. 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;132(1):275-85.
51. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet (London, England)*. 2003;362(9382):419-27.

52. Payne SJ, Bowen RL, Jones JL, Wells CA. Predictive markers in breast cancer--the present. *Histopathology*. 2008;52(1):82-90.
53. Kim JY, Lee YS, Yu J, Park Y, Lee SK, Lee M, et al. Deep Learning-Based Prediction Model for Breast Cancer Recurrence Using Adjuvant Breast Cancer Cohort in Tertiary Cancer Center Registry. *Frontiers in oncology*. 2021;11:596364.
54. Rakha EA, El-Sayed ME, Lee AH, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008;26(19):3153-8.
55. Lee AH, Ellis IO. The Nottingham prognostic index for invasive carcinoma of the breast. *Pathology oncology research : POR*. 2008;14(2):113-5.
56. Joseph C, Macnamara O, Craze M, Russell R, Provenzano E, Nolan CC, et al. Mediator complex (MED) 7: a biomarker associated with good prognosis in invasive breast cancer, especially ER+ luminal subtypes. *British Journal of Cancer*. 2018;118(8):1142-51.
57. Rakha EA, Tan PH, Varga Z, Tse GM, Shaaban AM, Climent F, et al. Prognostic factors in metaplastic carcinoma of the breast: a multi-institutional study. *British journal of cancer*. 2015;112(2):283-9.
58. Mojarad S, Venturini B, Fulgenzi P, Papaleo R, Brisigotti M, Monti F, et al. 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 research*. 2013;33(9):3925-33.
59. Cserni G, Chmielik E, Cserni B, Tot T. The new TNM-based staging of breast cancer. *Virchows Archiv : an international journal of pathology*. 2018;472(5):697-703.
60. Saadatmand S, Bretveld R, Siesling S, Tilanus-Linthorst MMA. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173 797 patients. *BMJ*. 2015;351:h4901.
61. Cianfrocca M, Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. *The oncologist*. 2004;9(6):606-16.
62. 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 Journal of Surgical Oncology*. 2014;12(1):243.
63. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham prognostic index in primary breast cancer. *Breast Cancer Research and Treatment*. 1992;22(3):207-19.
64. Blamey RW, Ellis IO, Pinder SE, Lee AH, Macmillan RD, Morgan DA, et al. Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999. *European journal of cancer (Oxford, England : 1990)*. 2007;43(10):1548-55.
65. Rakha EA, Martin S, Lee AHS, Morgan D, Pharoah PDP, Hodi Z, et al. The prognostic significance of lymphovascular invasion in invasive breast carcinoma. *Cancer*. 2012;118(15):3670-80.
66. Al-Rohil RN, Milton DR, Nagarajan P, Curry JL, Feldmeyer L, Torres-Cabala CA, et al. Intratumoral and peritumoral lymphovascular invasion detected by D2-40 immunohistochemistry correlates with metastasis in primary cutaneous Merkel cell carcinoma. *Human pathology*. 2018;77:98-107.
67. Ellis IO, Galea M, Broughton N, Locker A, Blamey RW, Elston CW. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. *Histopathology*. 1992;20(6):479-89.

68. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-52.
69. Geyer FC, Lopez-Garcia MA, Lambros MB, Reis-Filho JS. Genetic characterization of breast cancer and implications for clinical management. *Journal of cellular and molecular medicine*. 2009;13(10):4090-103.
70. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869-74.
71. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, et al. Triple-Negative Breast Cancer: Distinguishing between Basal and Nonbasal Subtypes. *Clinical Cancer Research*. 2009;15(7):2302.
72. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-52.
73. Dawson SJ, Rueda OM, Aparicio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J*. 2013;32(5):617-28.
74. Luck AA, Evans AJ, Green AR, Rakha EA, Paish C, Ellis IO. The influence of basal phenotype on the metastatic pattern of breast cancer. *Clinical oncology (Royal College of Radiologists (Great Britain))*. 2008;20(1):40-5.
75. Inic Z, Zegarac M, Inic M, Markovic I, Kozomara Z, Djuriscic I, et al. Difference between Luminal A and Luminal B Subtypes According to Ki-67, Tumor Size, and Progesterone Receptor Negativity Providing Prognostic Information. *Clinical Medicine Insights Oncology*. 2014;8:107-11.
76. Dolle JM, Daling JR, White E, Brinton LA, Doody DR, Porter PL, et al. Risk factors for triple-negative breast cancer in women under the age of 45 years. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2009;18(4):1157-66.
77. Mohammed RA, Ellis IO, Lee AH, Martin SG. Vascular invasion in breast cancer; an overview of recent prognostic developments and molecular pathophysiological mechanisms. *Histopathology*. 2009;55(1):1-9.
78. Pereira B, Chin S-F, Rueda OM, Vollan H-KM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nature Communications*. 2016;7(1):11479.
79. Sabatier R, Finetti P, Guille A, Adelaide J, Chaffanet M, Viens P, et al. Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. *Molecular cancer*. 2014;13:228.
80. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res*. 2010;12(5):R68.
81. Sun WY, Lee J, Kim BK, Kim JO, Park J. Distinct Somatic Alteration Features Identified by Gene Panel Sequencing in Korean Triple-Negative Breast Cancer with High Ki67 Expression. *Diagnostics (Basel)*. 2021;11(3).
82. Oh DY, Bang YJ. HER2-targeted therapies - a role beyond breast cancer. *Nature reviews Clinical oncology*. 2020;17(1):33-48.
83. Effi AB, Aman NA, Kouli BS, Koffi KD, Traore ZC, Kouyate M. Breast Cancer Molecular Subtypes Defined by ER/PR and HER2 Status: Association with Clinicopathologic Parameters in Ivorian Patients. *Asian Pacific journal of cancer prevention : APJCP*. 2016;17(4):1973-8.

84. Eroles P, Bosch A, Pérez-Fidalgo JA, Lluch A. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer treatment reviews*. 2012;38(6):698-707.
85. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010;28(16):2784-95.
86. Lee E-S, Han W, Kim MK, Kim J, Yoo T-k, Lee MH, et al. Factors associated with late recurrence after completion of 5-year adjuvant tamoxifen in estrogen receptor positive breast cancer. *BMC Cancer*. 2016;16(1):430.
87. Rakha EA, Reis-Filho JS, Ellis IO. Combinatorial biomarker expression in breast cancer. *Breast Cancer Res Treat*. 2010;120(2):293-308.
88. Cancellio G, Montagna E. Treatment of breast cancer in young women: do we need more aggressive therapies? *J Thorac Dis*. 2013;5 Suppl 1(Suppl 1):S47-S54.
89. Zhou X-L, Fan W, Yang G, Yu M-X. The clinical significance of PR, ER, NF- κ B, and TNF- α in breast cancer. *Dis Markers*. 2014;2014:494581-.
90. Ingthorsson S, Andersen K, Hilmarsson 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;35(32):4244-55.
91. Lai X, Li Q, Wu F, Lin J, Chen J, Zheng H, et al. Epithelial-Mesenchymal Transition and Metabolic Switching in Cancer: Lessons From Somatic Cell Reprogramming. *Frontiers in cell and developmental biology*. 2020;8:760.
92. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation*. 2009;119(6):1420-8.
93. Feigin ME, Muthuswamy SK. Polarity proteins regulate mammalian cell-cell junctions and cancer pathogenesis. *Current opinion in cell biology*. 2009;21(5):694-700.
94. Mohammed ZM, McMillan DC, Edwards J, Mallon E, Doughty JC, Orange C, et al. The relationship between lymphovascular invasion and angiogenesis, hormone receptors, cell proliferation and survival in patients with primary operable invasive ductal breast cancer. *BMC clinical pathology*. 2013;13(1):31.
95. Hamy AS, Lam GT, Laas E, Darrigues L, Balezeau T, Guerin J, et al. Lymphovascular invasion after neoadjuvant chemotherapy is strongly associated with poor prognosis in breast carcinoma. *Breast Cancer Res Treat*. 2018;169(2):295-304.
96. Ahmed S, Sami A, Xiang J. HER2-directed therapy: current treatment options for HER2-positive breast cancer. *Breast cancer (Tokyo, Japan)*. 2015;22(2):101-16.
97. Hunter KW, Crawford NP, Alsarraj J. Mechanisms of metastasis. *Breast Cancer Res*. 2008;10 Suppl 1(Suppl 1):S2.
98. Talmadge JE, Fidler IJ. AACR Centennial Series: The Biology of Cancer Metastasis: Historical Perspective. *Cancer Research*. 2010;70(14):5649-69.
99. Scully OJ, Bay BH, Yip G, Yu Y. Breast cancer metastasis. *Cancer genomics & proteomics*. 2012;9(5):311-20.
100. Fidler IJ, Gersten DM, Hart IR. The biology of cancer invasion and metastasis. *Advances in cancer research*. 1978;28:149-250.
101. Talmadge JE. Clonal Selection of Metastasis within the Life History of a Tumor. *Cancer Research*. 2007;67(24):11471-5.
102. Weigelt B, Peterse JL, van 't Veer LJ. Breast cancer metastasis: markers and models. *Nature reviews Cancer*. 2005;5(8):591-602.

103. Chen LL, Blumm N, Christakis NA, Barabási AL, Deisboeck TS. Cancer metastasis networks and the prediction of progression patterns. *British journal of cancer*. 2009;101(5):749-58.
104. Paget S. THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST. *The Lancet*. 1889;133(3421):571-3.
105. Faraji F, Eissenberg JC. Seed and soil: A conceptual framework of metastasis for clinicians. *Mo Med*. 2013;110(4):302-8.
106. Kuwashima N. Organ-specific adhesion of neuroblastoma cells in vitro: Correlation with their hepatic metastasis potential. *Journal of Pediatric Surgery*. 1997;32(4):546-51.
107. Paschos KA, Canovas D, Bird NC. The role of cell adhesion molecules in the progression of colorectal cancer and the development of liver metastasis. *Cellular signalling*. 2009;21(5):665-74.
108. Mochizuki H, Matsubara A, Teishima J, Mutaguchi K, Yasumoto H, Dahiya R, et al. Interaction of ligand-receptor system between stromal-cell-derived factor-1 and CXCR4 chemokine receptor 4 in human prostate cancer: a possible predictor of metastasis. *Biochem Biophys Res Commun*. 2004;320(3):656-63.
109. Ribatti D, Mangialardi G, Vacca A. Stephen Paget and the 'seed and soil' theory of metastatic dissemination. *Clinical and experimental medicine*. 2006;6(4):145-9.
110. Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. *Nature*. 2016;529(7586):298-306.
111. Kiricuta I, Willner J, Kolbl O. The clinical relevance of lymphatic and blood-vessel invasion in breast-cancer patients. *Oncology reports*. 1995;2(1):137-41.
112. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging Biological Principles of Metastasis. *Cell*. 2017;168(4):670-91.
113. Nathanson SD, Krag D, Kuerer HM, Newman LA, Brown M, Kerjaschki D, et al. Breast cancer metastasis through the lympho-vascular system. *Clinical & experimental metastasis*. 2018;35(5-6):443-54.
114. Paduch R. The role of lymphangiogenesis and angiogenesis in tumor metastasis. *Cell Oncol (Dordr)*. 2016;39(5):397-410.
115. Mohammed RAA, Martin SG, Mahmmoud AM, Macmillan RD, Green AR, Paish EC, et al. 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. *The Journal of Pathology*. 2011;223(3):358-65.
116. Barbieri E, Anghelone CAP, Gentile D, La Raja C, Bottini A, Tinterri C. Metastases from Occult Breast Cancer: A Case Report of Carcinoma of Unknown Primary Syndrome. *Case reports in oncology*. 2020;13(3):1158-63.
117. Pereira ER, Kedrin D, Seano G, Gautier O, Meijer EFJ, Jones D, et al. Lymph node metastases can invade local blood vessels, exit the node, and colonize distant organs in mice. *Science (New York, NY)*. 2018;359(6382):1403-7.
118. Brown M, Assen FP, Leithner A, Abe J, Schachner H, Asfour G, et al. Lymph node blood vessels provide exit routes for metastatic tumor cell dissemination in mice. *Science (New York, NY)*. 2018;359(6382):1408-11.
119. Wu JK, Kitajewski C, Reiley M, Keung CH, Monteagudo J, Andrews JP, et al. Aberrant lymphatic endothelial progenitors in lymphatic malformation development. *PLoS One*. 2015;10(2):e0117352.
120. Steinskog ES, Sagstad SJ, Wagner M, Karlsen TV, Yang N, Markhus CE, et al. Impaired lymphatic function accelerates cancer growth. *Oncotarget*. 2016;7(29):45789-802.

121. Aleskandarany MA, Sonbul SN, Mukherjee A, Rakha EA. Molecular Mechanisms Underlying Lymphovascular Invasion in Invasive Breast Cancer. *Pathobiology*. 2015;82(3-4):113-23.
122. 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. *Journal of breast cancer*. 2011;14(3):198-203.
123. Gupta GP, Massagué J. Cancer metastasis: building a framework. *Cell*. 2006;127(4):679-95.
124. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. *Nature Reviews Cancer*. 2009;9(4):239-52.
125. Vestweber D. VE-cadherin: the major endothelial adhesion molecule controlling cellular junctions and blood vessel formation. *Arteriosclerosis, thrombosis, and vascular biology*. 2008;28(2):223-32.
126. Li L, Liu C, Amato RJ, Chang JT, Du G, Li W. CDKL2 promotes epithelial-mesenchymal transition and breast cancer progression. *Oncotarget*. 2014;5(21):10840-53.
127. Li DM, Feng YM. Signaling mechanism of cell adhesion molecules in breast cancer metastasis: potential therapeutic targets. *Breast Cancer Res Treat*. 2011;128(1):7-21.
128. Bruner HC, Derksen PWB. Loss of E-Cadherin-Dependent Cell-Cell Adhesion and the Development and Progression of Cancer. *Cold Spring Harbor perspectives in biology*. 2018;10(3).
129. Gould Rothberg BE, Bracken MB. E-cadherin immunohistochemical expression as a prognostic factor in infiltrating ductal carcinoma of the breast: a systematic review and meta-analysis. *Breast Cancer Res Treat*. 2006;100(2):139-48.
130. Kotb AM, Hierholzer A, Kemler R. Replacement of E-cadherin by N-cadherin in the mammary gland leads to fibrocystic changes and tumor formation. *Breast Cancer Res*. 2011;13(5):R104-R.
131. Cao ZQ, Wang Z, Leng P. Aberrant N-cadherin expression in cancer. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2019;118:109320.
132. Bonnomet A, Brysse A, Tachsidis A, Waltham M, Thompson EW, Polette M, et al. Epithelial-to-mesenchymal transitions and circulating tumor cells. *Journal of mammary gland biology and neoplasia*. 2010;15(2):261-73.
133. Scott LE, Weinberg SH, Lemmon CA. Mechanochemical Signaling of the Extracellular Matrix in Epithelial-Mesenchymal Transition. *Frontiers in cell and developmental biology*. 2019;7:135.
134. Yilmaz M, Christofori G. Mechanisms of Motility in Metastasizing Cells. *Molecular Cancer Research*. 2010;8(5):629-42.
135. van Roy F. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nature reviews Cancer*. 2014;14(2):121-34.
136. 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;6(13):11652-63.
137. Bhat-Nakshatri P, Appaiah H, Ballas C, Pick-Franke P, Goulet R, Jr., Badve S, et al. SLUG/SNAI2 and tumor necrosis factor generate breast cells with CD44+/CD24-phenotype. *BMC Cancer*. 2010;10:411.
138. Otake S, Itoh Y, Omata C, Saitoh M, Miyazawa K. ZEB1 and oncogenic Ras constitute a regulatory switch for stimulus-dependent E-cadherin downregulation. *Cancer Sci*. 2021;112(1):205-16.
139. Kechagia JZ, Ivaska J, Roca-Cusachs P. Integrins as biomechanical sensors of the microenvironment. *Nature reviews Molecular cell biology*. 2019;20(8):457-73.

140. Seguin L, Desgrosellier JS, Weis SM, Cheres DA. Integrins and cancer: regulators of cancer stemness, metastasis, and drug resistance. *Trends in cell biology*. 2015;25(4):234-40.
141. Bergmeier W, Hynes RO. Extracellular matrix proteins in hemostasis and thrombosis. *Cold Spring Harbor perspectives in biology*. 2012;4(2):a005132.
142. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. *Nature reviews Molecular cell biology*. 2014;15(12):786-801.
143. Mitchell K, Svenson KB, Longmate WM, Gkirtzimanaki K, Sadej R, Wang X, et al. Suppression of integrin $\alpha 3 \beta 1$ in breast cancer cells reduces cyclooxygenase-2 gene expression and inhibits tumorigenesis, invasion, and cross-talk to endothelial cells. *Cancer research*. 2010;70(15):6359-67.
144. Tang J, Wang J, Fan L, Li X, Liu N, Luo W, et al. cRGD inhibits vasculogenic mimicry formation by down-regulating uPA expression and reducing EMT in ovarian cancer. *Oncotarget*. 2016;7(17):24050-62.
145. Harbeck N, Kates RE, Schmitt M, Gauger K, Kiechle M, Janicke F, et al. Urokinase-type plasminogen activator and its inhibitor type 1 predict disease outcome and therapy response in primary breast cancer. *Clinical breast cancer*. 2004;5(5):348-52.
146. Völker H-U, Weigel M, Strehl A, Frey L. Levels of uPA and PAI-1 in breast cancer and its correlation to Ki67-index and results of a 21-multigene-array. *Diagn Pathol*. 2018;13(1):67-.
147. Song T, Meng S, Xu S-T, Jin S-J, Zeng Q-Z, Gu G-J. The overexpression of uPA promotes the proliferation and fibrinolytic activity of human umbilical vein endothelial cells. *Int J Clin Exp Pathol*. 2019;12(8):2959-66.
148. Fidalgo F, Rodrigues TC, Pinilla M, Silva AG, Maciel Mdo S, Rosenberg C, et al. Lymphovascular invasion and histologic grade are associated with specific genomic profiles in invasive carcinomas of the breast. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2015;36(3):1835-48.
149. Yousef EM, Tahir MR, St-Pierre Y, Gaboury LA. MMP-9 expression varies according to molecular subtypes of breast cancer. *BMC cancer*. 2014;14:609-.
150. Wu Q-W, Yang Q-M, Huang Y-F, She H-Q, Liang J, Yang Q-L, et al. Expression and Clinical Significance of Matrix Metalloproteinase-9 in Lymphatic Invasiveness and Metastasis of Breast Cancer. *PLOS ONE*. 2014;9(5):e97804.
151. Stacker SA, Williams SP, Karnezis T, Shayan R, Fox SB, Achen MG. Lymphangiogenesis and lymphatic vessel remodelling in cancer. *Nature reviews Cancer*. 2014;14(3):159-72.
152. Guan X. Cancer metastases: challenges and opportunities. *Acta pharmaceutica Sinica B*. 2015;5(5):402-18.
153. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nature reviews Molecular cell biology*. 2014;15(3):178-96.
154. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet (London, England)*. 2005;365(9472):1687-717.
155. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010;28(20):3271-7.
156. 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;150(3):547-57.

157. Sun Y, Wang X, Zhou Q, Lu Y, Zhang H, Chen Q, et al. Inhibitory effect of emodin on migration, invasion and metastasis of human breast cancer MDA-MB-231 cells in vitro and in vivo. *Oncology reports*. 2015;33(1):338-46.
158. Fazilaty H, Gardaneh M, Bahrami T, Salmaninejad A, Behnam B. Crosstalk between breast cancer stem cells and metastatic niche: emerging molecular metastasis pathway? *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2013;34(4):2019-30.
159. González de Vega R, Clases D, Fernández-Sánchez ML, Eiró N, González LO, Vizoso FJ, et al. MMP-11 as a biomarker for metastatic breast cancer by immunohistochemical-assisted imaging mass spectrometry. *Analytical and bioanalytical chemistry*. 2019;411(3):639-46.
160. Lee CC, Lin JC, Hwang WL, Kuo YJ, Chen HK, Tai SK, et al. Macrophage-secreted interleukin-35 regulates cancer cell plasticity to facilitate metastatic colonization. *Nat Commun*. 2018;9(1):3763.
161. Shinde A, Wilmanski T, Chen H, Teegarden D, Wendt MK. Pyruvate carboxylase supports the pulmonary tropism of metastatic breast cancer. *Breast Cancer Res*. 2018;20(1):76-.
162. 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? *Pathology oncology research : POR*. 2012;18(2):253-62.
163. Hishida T, Yoshida J, Maeda R, Ishii G, Aokage K, Nishimura M, et al. Prognostic impact of intratumoural microvascular invasion and microlymphatic permeation on node-negative non-small-cell lung cancer: which indicator is the stronger prognostic factor? *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery*. 2013;43(4):772-7.
164. Agarwal D, Pardhe N, Bajpai M, Gupta S, Mathur N, Vanaki SS, et al. 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;8(10):ZC86-ZC9.
165. Rack B, Schindlbeck C, Jückstock J, Andergassen U, Hepp P, Zwingers T, et al. Circulating tumor cells predict survival in early average-to-high risk breast cancer patients. *Journal of the National Cancer Institute*. 2014;106(5):dju066.
166. Kuo MC, Kothari AN, Kuo PC, Mi Z. Cancer stemness in bone marrow micrometastases of human breast cancer. *Surgery*. 2018;163(2):330-5.
167. Cornwell LB, McMasters KM, Chagpar AB. The impact of lymphovascular invasion on lymph node status in patients with breast cancer. *The American surgeon*. 2011;77(7):874-7.
168. Dadiani M, Kalchenko V, Yosepovich A, Margalit R, Hassid Y, Degani H, et al. Real-time imaging of lymphogenic metastasis in orthotopic human breast cancer. *Cancer Res*. 2006;66(16):8037-41.
169. Karpanen T, Wirzenius M, Mäkinen T, Veikkola T, Haisma HJ, Achen MG, et al. Lymphangiogenic growth factor responsiveness is modulated by postnatal lymphatic vessel maturation. *Am J Pathol*. 2006;169(2):708-18.
170. 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;11(6):526-38.
171. Achen MG, Stacker SA. Molecular control of lymphatic metastasis. *Annals of the New York Academy of Sciences*. 2008;1131:225-34.
172. Nathanson SD. Insights into the mechanisms of lymph node metastasis. *Cancer*. 2003;98(2):413-23.

173. Pepper MS. Lymphangiogenesis and tumor metastasis: myth or reality? *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2001;7(3):462-8.
174. Wong SY, Hynes RO. Lymphatic or hematogenous dissemination: how does a metastatic tumor cell decide? *Cell Cycle*. 2006;5(8):812-7.
175. Kahn HJ, Marks A. A new monoclonal antibody, D2-40, for detection of lymphatic invasion in primary tumors. *Laboratory investigation; a journal of technical methods and pathology*. 2002;82(9):1255-7.
176. Breiteneder-Geleff S, Soleiman A, Kowalski H, Horvat R, Amann G, Kriehuber E, et al. Angiosarcomas express mixed endothelial phenotypes of blood and lymphatic capillaries: podoplanin as a specific marker for lymphatic endothelium. *Am J Pathol*. 1999;154(2):385-94.
177. Joukov V, Pajusola K, Kaipainen A, Chilov D, Lahtinen I, Kukk E, et al. A novel vascular endothelial growth factor, VEGF-C, is a ligand for the Flt4 (VEGFR-3) and KDR (VEGFR-2) receptor tyrosine kinases. *EMBO J*. 1996;15(2):290-98.
178. Righi L, Deaglio S, Pecchioni C, Gregorini A, Horenstein AL, Bussolati G, et al. Role of CD31/platelet endothelial cell adhesion molecule-1 expression in in vitro and in vivo growth and differentiation of human breast cancer cells. *Am J Pathol*. 2003;162(4):1163-74.
179. Schacht V, Dadras SS, Johnson LA, Jackson DG, Hong YK, Detmar M. Up-regulation of the lymphatic marker podoplanin, a mucin-type transmembrane glycoprotein, in human squamous cell carcinomas and germ cell tumors. *Am J Pathol*. 2005;166(3):913-21.
180. Van der Auwera I, Van den Eynden GG, Colpaert CG, Van Laere SJ, van Dam P, Van Marck EA, et al. Tumor lymphangiogenesis in inflammatory breast carcinoma: a histomorphometric study. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2005;11(21):7637-42.
181. Evangelou E, Kyzas PA, Trikalinos TA. Comparison of the diagnostic accuracy of lymphatic endothelium markers: Bayesian approach. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2005;18(11):1490-7.
182. Gujam FJ, Going JJ, Mohammed ZM, Orange C, Edwards J, McMillan DC. Immunohistochemical detection improves the prognostic value of lymphatic and blood vessel invasion in primary ductal breast cancer. *BMC Cancer*. 2014;14:676.
183. Vosough Z, Golbini S, Sharbatdaran M, Hosseini A. D2-40 A Helpful Marker in Assessment of Lymphatic Vessel Invasion in Carcinoma of Breast. *Iran J Pathol*. 2021;16(2):96-102.
184. He K-W, Sun J-J, Liu Z-B, Zhuo P-Y, Ma Q-H, Liu Z-Y, et al. Prognostic significance of lymphatic vessel invasion diagnosed by D2-40 in Chinese invasive breast cancers. *Medicine (Baltimore)*. 2017;96(44):e8490-e.
185. Zhang S, Zhang D, Yi S, Gong M, Lu C, Cai Y, et al. The relationship of lymphatic vessel density, lymphovascular invasion, and lymph node metastasis in breast cancer: a systematic review and meta-analysis. *Oncotarget*. 2017;8(2):2863-73.
186. Chen W, Chen M, Liao Z, Wang Y, Zhan Q, Cai G. Lymphatic vessel density as predictive marker for the local recurrence of rectal cancer. *Diseases of the colon and rectum*. 2009;52(3):513-9.
187. Martin SG, Rakha E, Storr SJ. Quantifying Lymphatic Vessel Density in Human Tissue Samples. *Methods in molecular biology (Clifton, NJ)*. 2022;2441:183-9.

188. Chen J-M, Luo B, Ma R, Luo X-X, Chen Y-S, Li Y. Lymphatic Endothelial Markers and Tumor Lymphangiogenesis Assessment in Human Breast Cancer. *Diagnostics (Basel)*. 2021;12(1):4.
189. Britto AV, Schenka AA, Moraes-Schenka NG, Alvarenga M, Shinzato JY, Vassallo J, et al. Immunostaining with D2-40 improves evaluation of lymphovascular invasion, but may not predict sentinel lymph node status in early breast cancer. *BMC Cancer*. 2009;9:109.
190. Schwartz AM, Henson DE, Chen D, Rajamarthandan S. Histologic grade remains a prognostic factor for breast cancer regardless of the number of positive lymph nodes and tumor size: a study of 161 708 cases of breast cancer from the SEER Program. *Archives of pathology & laboratory medicine*. 2014;138(8):1048-52.
191. Rudno-Rudzińska J, Kielan W, Grzebieniak Z, Dziegiel P, Donizy P, Mazur G, et al. High density of peritumoral lymphatic vessels measured by D2-40/podoplanin and LYVE-1 expression in gastric cancer patients: an excellent prognostic indicator or a false friend? *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association*. 2013;16(4):513-20.
192. Suchanski J, Tejchman A, Zacharski M, Piotrowska A, Grzegorzolka J, Chodaczek G, et al. Podoplanin increases the migration of human fibroblasts and affects the endothelial cell network formation: A possible role for cancer-associated fibroblasts in breast cancer progression. *PloS one*. 2017;12(9):e0184970-e.
193. Martín-Villar E, Megías D, Castel S, Yurrita MM, Vilaró S, Quintanilla M. Podoplanin binds ERM proteins to activate RhoA and promote epithelial-mesenchymal transition. *Journal of cell science*. 2006;119(Pt 21):4541-53.
194. Sikorska J, Gawel D, Domek H, Rudzińska M, Czarnocka B. Podoplanin (PDPN) affects the invasiveness of thyroid carcinoma cells by inducing ezrin, radixin and moesin (E/R/M) phosphorylation in association with matrix metalloproteinases. *BMC cancer*. 2019;19(1):85-.
195. Neri S, Ishii G, Hashimoto H, Kuwata T, Nagai K, Date H, et al. Podoplanin-expressing cancer-associated fibroblasts lead and enhance the local invasion of cancer cells in lung adenocarcinoma. *Int J Cancer*. 2015;137(4):784-96.
196. Ochoa-Alvarez JA, Krishnan H, Pastorino JG, Nevel E, Kephart D, Lee JJ, et al. Antibody and lectin target podoplanin to inhibit oral squamous carcinoma cell migration and viability by distinct mechanisms. *Oncotarget*. 2015;6(11):9045-60.
197. Pula B, Wojnar A, Weryńska B, Ambicka A, Kruczek A, Witkiewicz W, et al. Impact of different tumour stroma assessment methods regarding podoplanin expression on clinical outcome in patients with invasive ductal breast carcinoma. *Anticancer research*. 2013;33(4):1447-55.
198. 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.
199. Salven P, Mustjoki S, Alitalo R, Alitalo K, Rafii S. VEGFR-3 and CD133 identify a population of CD34+ lymphatic/vascular endothelial precursor cells. *Blood*. 2003;101(1):168-72.
200. Kaipainen A, Korhonen J, Mustonen T, van Hinsbergh VW, Fang GH, Dumont D, et al. Expression of the fms-like tyrosine kinase 4 gene becomes restricted to lymphatic endothelium during development. *Proc Natl Acad Sci U S A*. 1995;92(8):3566-70.
201. Bieniasz-Krzywiec P, Martín-Pérez R, Ehling M, García-Caballero M, Pinioti S, Pretto S, et al. Podoplanin-Expressing Macrophages Promote Lymphangiogenesis and Lymphoinvasion in Breast Cancer. *Cell metabolism*. 2019;30(5):917-36.e10.

202. Plate K. From angiogenesis to lymphangiogenesis. *Nature medicine*. 2001;7(2):151-2.
203. Skobe M, Hawighorst T, Jackson DG, Prevo R, Janes L, Velasco P, et al. Induction of tumor lymphangiogenesis by VEGF-C promotes breast cancer metastasis. *Nature medicine*. 2001;7(2):192-8.
204. Stacker SA, Baldwin ME, Achen MG. The role of tumor lymphangiogenesis in metastatic spread. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2002;16(9):922-34.
205. Ran S, Volk L, Hall K, Flister MJ. Lymphangiogenesis and lymphatic metastasis in breast cancer. *Pathophysiology : the official journal of the International Society for Pathophysiology*. 2010;17(4):229-51.
206. Albelda SM, Muller WA, Buck CA, Newman PJ. Molecular and cellular properties of PECAM-1 (endoCAM/CD31): a novel vascular cell-cell adhesion molecule. *The Journal of cell biology*. 1991;114(5):1059-68.
207. Dileep A, Prasad P. Use of immunomarkers D2-40 and CD31 in detection of lymphovascular invasion in breast carcinoma. *J Med Sci Clin Res*. 2018;6:90-6.
208. Hämmerling B, Grund C, Boda-Heggemann J, Moll R, Franke WW. The complexus adhaerens of mammalian lymphatic endothelia revisited: a junction even more complex than hitherto thought. *Cell and tissue research*. 2006;324(1):55-67.
209. Dejana E, Tournier-Lasserre E, Weinstein BM. The control of vascular integrity by endothelial cell junctions: molecular basis and pathological implications. *Developmental cell*. 2009;16(2):209-21.
210. Breslin JW, Yang Y, Scallan JP, Sweat RS, Adderley SP, Murfee WL. Lymphatic Vessel Network Structure and Physiology. *Comprehensive Physiology*. 2018;9(1):207-99.
211. Baluk P, Fuxe J, Hashizume H, Romano T, Lashnits E, Butz S, et al. Functionally specialized junctions between endothelial cells of lymphatic vessels. *J Exp Med*. 2007;204(10):2349-62.
212. Johnson LA, Clasper S, Holt AP, Lalor PF, Baban D, Jackson DG. An inflammation-induced mechanism for leukocyte transmigration across lymphatic vessel endothelium. *J Exp Med*. 2006;203(12):2763-77.
213. Wick N, Saharinen P, Saharinen J, Gurnhofer E, Steiner CW, Raab I, et al. Transcriptomal comparison of human dermal lymphatic endothelial cells ex vivo and in vitro. *Physiological genomics*. 2007;28(2):179-92.
214. Yates LR, Desmedt C. Translational Genomics: Practical Applications of the Genomic Revolution in Breast Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2017;23(11):2630-9.
215. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *Journal of the National Cancer Institute*. 2006;98(4):262-72.
216. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A*. 2018;115(13):E2970-e9.
217. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70.
218. Pu M, Messer K, Davies SR, Vickery TL, Pittman E, Parker BA, et al. Research-based PAM50 signature and long-term breast cancer survival. *Breast Cancer Res Treat*. 2020;179(1):197-206.
219. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*. 2015;163(2):506-19.

220. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401-4.
221. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling.* 2013;6(269):pl1.
222. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn).* 2015;19(1a):A68-77.
223. Kurozumi S, Joseph C, Sonbul S, Alsaeed S, Kariri Y, Aljohani A, et al. A key genomic subtype associated with lymphovascular invasion in invasive breast cancer. *Br J Cancer.* 2019;120(12):1129-36.
224. Jézéquel P, Campone M, Gouraud W, Guérin-Charbonnel C, Leux C, Ricolleau G, et al. bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast Cancer Research and Treatment.* 2012;131(3):765-75.
225. Yap G. Affymetrix, Inc. Pharmacogenomics. 2002;3(5):709-11.
226. Bartha Á, Györffy B. TNMplot.com: A Web Tool for the Comparison of Gene Expression in Normal, Tumor and Metastatic Tissues. *International Journal of Molecular Sciences.* 2021;22(5):2622.
227. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia.* 2017;19(8):649-58.
228. Chandrashekar DS, Karthikeyan SK, Korla PK, Patel H, Shovon AR, Athar M, et al. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia.* 2022;25:18-27.
229. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, et al. 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;116(3):340-50.
230. Aleskandarany MA, Rakha EA, Macmillan RD, Powe DG, Ellis IO, Green AR. MIB1/Ki-67 labelling index can classify grade 2 breast cancer into two clinically distinct subgroups. *Breast Cancer Research and Treatment.* 2011;127(3):591-9.
231. Cailleau R, Olivé M, Cruciger QV. Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In vitro.* 1978;14(11):911-5.
232. Landry JJ, Pyl PT, Rausch T, Zichner T, Tekkedil MM, Stütz AM, et al. The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda, Md).* 2013;3(8):1213-24.
233. Vasiliou SK, Filippou PS, Clotet-Freixas S, Soosaipillai A, Batruch I, Viktor Tsianos F, et al. Transcriptome profiling and proteomic validation reveals targets of the androgen receptor signaling in the BT-474 breast cancer cell line. *Clinical Proteomics.* 2022;19(1):14.
234. 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;76(9):4350-4.
235. Coons AH, Creech HJ, Jones RN. Immunological Properties of an Antibody Containing a Fluorescent Group. *Proceedings of the Society for Experimental Biology and Medicine.* 1941;47(2):200-2.
236. Ltd LBN. Novolink™ Polymer Detection Systems. 2015.
237. Masuda H, Zhang D, Bartholomeusz C, Doihara H, Hortobagyi GN, Ueno NT. Role of epidermal growth factor receptor in breast cancer. *Breast Cancer Res Treat.* 2012;136(2):331-45.

238. Li H, Prever L, Hirsch E, Gulluni F. Targeting PI3K/AKT/mTOR Signaling Pathway in Breast Cancer. *Cancers*. 2021;13(14).
239. Shimoi T, Hamada A, Yamagishi M, Hirai M, Yoshida M, Nishikawa T, et al. PIK3CA mutation profiling in patients with breast cancer, using a highly sensitive detection system. *Cancer Sci*. 2018;109(8):2558-66.
240. Martorana F, Motta G, Pavone G, Motta L, Stella S, Vitale SR, et al. AKT Inhibitors: New Weapons in the Fight Against Breast Cancer? *Frontiers in pharmacology*. 2021;12:662232.
241. Kreipe H, Harbeck N, Christgen M. Clinical validity and clinical utility of Ki67 in early breast cancer. *Therapeutic advances in medical oncology*. 2022;14:17588359221122725.
242. Gasco M, Shami S, Crook T. The p53 pathway in breast cancer. *Breast Cancer Res*. 2002;4(2):70-6.
243. Berke TP, Slight SH, Hyder SM. Role of Reactivating Mutant p53 Protein in Suppressing Growth and Metastasis of Triple-Negative Breast Cancer. *Onco Targets Ther*. 2022;15:23-30.
244. Abd El-Rehim DM, Pinder SE, Paish CE, Bell JA, Rampaul RS, Blamey RW, et al. Expression and co-expression of the members of the epidermal growth factor receptor (EGFR) family in invasive breast carcinoma. *Br J Cancer*. 2004;91(8):1532-42.
245. Aleskandarany MA, Rakha EA, Ahmed MAH, Powe DG, Paish EC, Macmillan RD, et al. PIK3CA expression in invasive breast cancer: a biomarker of poor prognosis. *Breast Cancer Research and Treatment*. 2010;122(1):45-53.
246. Aleskandarany MA, Rakha EA, Ahmed MA, Powe DG, Ellis IO, Green AR. Clinicopathologic and molecular significance of phospho-Akt expression in early invasive breast cancer. *Breast Cancer Research and Treatment*. 2011;127(2):407-16.
247. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *Journal of the National Cancer Institute*. 2011;103(22):1656-64.
248. Kikuchi S, Nishimura R, Osako T, Okumura Y, Nishiyama Y, Toyozumi Y, et al. Definition of p53 overexpression and its association with the clinicopathological features in luminal/HER2-negative breast cancer. *Anticancer research*. 2013;33(9):3891-7.
249. Kadota K, Nakai Y, Shimizu K. A weighted average difference method for detecting differentially expressed genes from microarray data. *Algorithms for molecular biology : AMB*. 2008;3:8.
250. Alexander-Dann B, Pruteanu LL, Oerton E, Sharma N, Berindan-Neagoe I, Módos D, et al. Developments in toxicogenomics: understanding and predicting compound-induced toxicity from gene expression data. *Molecular omics*. 2018;14(4):218-36.
251. Turashvili G, Brogi E. Tumor Heterogeneity in Breast Cancer. *Frontiers in medicine*. 2017;4:227.
252. Wang Y, Zhang Y, Huang Q, Li C. Integrated bioinformatics analysis reveals key candidate genes and pathways in breast cancer. *Molecular medicine reports*. 2018;17(6):8091-100.
253. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*. 2002;30(1):207-10.
254. Gao H, Yang M, Zhang X. Investigating a multigene prognostic assay based on significant pathways for Luminal A breast cancer through gene expression profile analysis. *Oncology letters*. 2018;15(4):5027-33.

255. Klahan S, Wong HS, Tu SH, Chou WH, Zhang YF, Ho TF, et al. Identification of genes and pathways related to lymphovascular invasion in breast cancer patients: A bioinformatics analysis of gene expression profiles. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2017;39(6):1010428317705573.
256. Diboun I, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC genomics*. 2006;7:252.
257. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics (Oxford, England)*. 2004;20(9):1453-4.
258. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic acids research*. 2005;33(Web Server issue):W741-8.
259. 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 research*. 2017;45(W1):W130-w7.
260. Mannelqvist M, Stefansson IM, Bredholt G, Hellem Bø T, Oyan AM, Jonassen I, et al. Gene expression patterns related to vascular invasion and aggressive features in endometrial cancer. *Am J Pathol*. 2011;178(2):861-71.
261. Mínguez B, Hoshida Y, Villanueva A, Toffanin S, Cabellos L, Thung S, et al. Gene-expression signature of vascular invasion in hepatocellular carcinoma. *Journal of hepatology*. 2011;55(6):1325-31.
262. 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;9(6):e98787.
263. Zhao M, Liu Y, Qu H. Expression of epithelial-mesenchymal transition-related genes increases with copy number in multiple cancer types. *Oncotarget*. 2016;7(17):24688-99.
264. Jena MK, Janjanam J. Role of extracellular matrix in breast cancer development: a brief update. *F1000Research*. 2018;7:274.
265. Musiani D, Konda JD, Pavan S, Torchiano E, Sassi F, Noghero A, et al. Heat-shock protein 27 (HSP27, HSPB1) is up-regulated by MET kinase inhibitors and confers resistance to MET-targeted therapy. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2014;28(9):4055-67.
266. Wei L, Liu TT, Wang HH, Hong HM, Yu AL, Feng HP, et al. Hsp27 participates in the maintenance of breast cancer stem cells through regulation of epithelial-mesenchymal transition and nuclear factor- κ B. *Breast Cancer Res*. 2011;13(5):R101.
267. Song D, Yue L, Zhang J, Ma S, Zhao W, Guo F, et al. Diagnostic and prognostic significance of serum apolipoprotein C-I in triple-negative breast cancer based on mass spectrometry. *Cancer biology & therapy*. 2016;17(6):635-47.
268. Ko HL, Wang YS, Fong WL, Chi MS, Chi KH, Kao SJ. Apolipoprotein C1 (APOC1) as a novel diagnostic and prognostic biomarker for lung cancer: A marker phase I trial. *Thoracic cancer*. 2014;5(6):500-8.
269. Kurozumi S, Fujii T, Matsumoto H, Inoue K, Kurosumi M, Horiguchi J, et al. Significance of evaluating tumor-infiltrating lymphocytes (TILs) and programmed cell death-ligand 1 (PD-L1) expression in breast cancer. *Medical molecular morphology*. 2017;50(4):185-94.
270. Merdad A, Karim S, Schulten HJ, Dallol A, Buhmeida A, Al-Thubaity F, et al. Expression of matrix metalloproteinases (MMPs) in primary human breast cancer:

MMP-9 as a potential biomarker for cancer invasion and metastasis. *Anticancer research*. 2014;34(3):1355-66.

271. Zhang X, Huang S, Guo J, Zhou L, You L, Zhang T, et al. Insights into the distinct roles of MMP-11 in tumor biology and future therapeutics (Review). *International journal of oncology*. 2016;48(5):1783-93.

272. Kim HS, Kim MG, Min KW, Jung US, Kim DH. High MMP-11 expression associated with low CD8⁺ T cells decreases the survival rate in patients with breast cancer. *PLoS One*. 2021;16(5):e0252052.

273. Rakha EA, Abbas A, Pinto Ahumada P, ElSayed ME, Colman D, Pinder SE, et al. Diagnostic concordance of reporting lymphovascular invasion in breast cancer. *Journal of clinical pathology*. 2018;71(9):802-5.

274. Alsaeed SA, Toss M, Alsaleem M, Aleskandarany M, Joseph C, Kurozumi S, et al. Prognostic significance of heat shock protein 90AA1 (HSP90 α) in invasive breast cancer. *Journal of clinical pathology*. 2022;75(4):263-9.

275. Merino Bonilla JA, Torres Tabanera M, Ros Mendoza LH. Breast cancer in the 21st century: from early detection to new therapies. *Radiologia*. 2017;59(5):368-79.

276. Sjöström M, Hartman L, Honeth G, Grabau D, Malmström P, Hegardt C, et al. Stem cell biomarker ALDH1A1 in breast cancer shows an association with prognosis and clinicopathological variables that is highly cut-off dependent. *Journal of clinical pathology*. 2015;68(12):1012-9.

277. Wu H, Li R, Hang X, Yan M, Niu F, Liu L, et al. Can CD44⁺/CD24⁻ Tumor Cells Be Used to Determine the Extent of Breast Cancer Invasion Following Neoadjuvant Chemotherapy? *Journal of breast cancer*. 2011;14(3):175-80.

278. Braunstein LZ, Taghian AG. Molecular Phenotype, Multigene Assays, and the Locoregional Management of Breast Cancer. *Seminars in radiation oncology*. 2016;26(1):9-16.

279. Sabiani L, Houvenaeghel G, Heinemann M, Reyat F, Classe JM, Cohen M, et al. Breast cancer in young women: Pathologic features and molecular phenotype. *Breast (Edinburgh, Scotland)*. 2016;29:109-16.

280. Goksu SS, Tastekin D, Arslan D, Gunduz S, Tatli AM, Unal D, et al. Clinicopathologic features and molecular subtypes of breast cancer in young women (age ≤ 35). *Asian Pacific journal of cancer prevention : APJCP*. 2014;15(16):6665-8.

281. Peart O. Metastatic Breast Cancer. *Radiologic technology*. 2017;88(5):519m-39m.

282. Chircop M, Speidel D. Cellular stress responses in cancer and cancer therapy. *Frontiers in oncology*. 2014;4:304-.

283. Cawthorn TR, Moreno JC, Dharsee M, Tran-Thanh D, Ackloo S, Zhu PH, et al. Proteomic analyses reveal high expression of decorin and endoplasmic reticulum chaperone (HSP90B1) are associated with breast cancer metastasis and decreased survival. *PLoS One*. 2012;7(2):e30992.

284. Ciocca DR, Calderwood SK. Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications. *Cell Stress Chaperones*. 2005;10(2):86-103.

285. Klimczak M, Biecek P, Zylicz A, Zylicz M. Heat shock proteins create a signature to predict the clinical outcome in breast cancer. *Scientific Reports*. 2019;9(1):7507.

286. Kurozumi S, Yamaguchi Y, Kurosumi M, Ohira M, Matsumoto H, Horiguchi J. Recent trends in microRNA research into breast cancer with particular focus on the associations between microRNAs and intrinsic subtypes. *Journal of Human Genetics*. 2017;62(1):15-24.

287. Garrido C, Brunet M, Didelot C, Zermati Y, Schmitt E, Kroemer G. Heat Shock Proteins 27 and 70: Anti-Apoptotic Proteins with Tumorigenic Properties. *Cell Cycle*. 2006;5(22):2592-601.

288. Zuehlke AD, Beebe K, Neckers L, Prince T. Regulation and function of the human HSP90AA1 gene. *Gene*. 2015;570(1):8-16.
289. Pick E, Kluger Y, Giltneane JM, Moeder C, Camp RL, Rimm DL, et al. High HSP90 Expression Is Associated with Decreased Survival in Breast Cancer. *Cancer Research*. 2007;67(7):2932.
290. Li P, Wang J, Zou Y, Sun Z, Zhang M, Geng Z, et al. Interaction of Hsp90AA1 with phospholipids stabilizes membranes under stress conditions. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2019;1861(2):457-65.
291. Pandey P, Saleh A, Nakazawa A, Kumar S, Srinivasula SM, Kumar V, et al. Negative regulation of cytochrome c-mediated oligomerization of Apaf-1 and activation of procaspase-9 by heat shock protein 90. *EMBO J*. 2000;19(16):4310-22.
292. Nimmanapalli R, O'Bryan E, Bhalla K. Geldanamycin and Its Analogue 17-Allylamino-17-demethoxygeldanamycin Lowers Bcr-Abl Levels and Induces Apoptosis and Differentiation of Bcr-Abl-positive Human Leukemic Blasts. *Cancer Research*. 2001;61(5):1799.
293. Xiang X, You X-M, Li L-Q. Expression of HSP90AA1/HSPA8 in hepatocellular carcinoma patients with depression. *Onco Targets Ther*. 2018;11:3013-23.
294. Dong Z, Yang P, Qiu X, Liang S, Guan B, Yang H, et al. KCNQ1OT1 facilitates progression of non-small-cell lung carcinoma via modulating miRNA-27b-3p/HSP90AA1 axis. *Journal of Cellular Physiology*. 2019;234(7):11304-14.
295. Pawłowska E, Szczepanska J, Blasiak J. The Long Noncoding RNA HOTAIR in Breast Cancer: Does Autophagy Play a Role? *International Journal of Molecular Sciences*. 2017;18(11):2317.
296. Nazio F, Bordi M, Cianfanelli V, Locatelli F, Cecconi F. Autophagy and cancer stem cells: molecular mechanisms and therapeutic applications. *Cell Death & Differentiation*. 2019;26(4):690-702.
297. Xiao X, Wang W, Li Y, Yang D, Li X, Shen C, et al. HSP90AA1-mediated autophagy promotes drug resistance in osteosarcoma. *Journal of Experimental & Clinical Cancer Research*. 2018;37(1):201.
298. Birbo B, Madu EE, Madu CO, Jain A, Lu Y. Role of HSP90 in Cancer. *Int J Mol Sci*. 2021;22(19).
299. Whitley D, Goldberg SP, Jordan WD. Heat shock proteins: A review of the molecular chaperones. *Journal of Vascular Surgery*. 1999;29(4):748-51.
300. Bhatia A, O'Brien K, Chen M, Woodley DT, Li W. Keratinocyte-Secreted Heat Shock Protein-90alpha: Leading Wound Reepithelialization and Closure. *Advances in Wound Care*. 2015;5(4):176-84.
301. Calderwood SK. Heat shock proteins in breast cancer progression—A suitable case for treatment? *International Journal of Hyperthermia*. 2010;26(7):681-5.
302. Eustace BK, Sakurai T, Stewart JK, Yimlamai D, Unger C, Zehetmeier C, et al. Functional proteomic screens reveal an essential extracellular role for hsp90α in cancer cell invasiveness. *Nature Cell Biology*. 2004;6(6):507-14.
303. Whitesell L, Lindquist SL. HSP90 and the chaperoning of cancer. *Nature reviews Cancer*. 2005;5(10):761-72.
304. Neckers L, Kern A, Tsutsumi S. Hsp90 inhibitors disrupt mitochondrial homeostasis in cancer cells. *Chemistry & biology*. 2007;14(11):1204-6.
305. Taipale M, Jarosz DF, Lindquist S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature reviews Molecular cell biology*. 2010;11(7):515-28.
306. Xiao Y, Liu Y. Recent Advances in the Discovery of Novel HSP90 Inhibitors: An Update from 2014. *Current drug targets*. 2020;21(3):302-17.

307. Park H-K, Yoon NG, Lee J-E, Hu S, Yoon S, Kim SY, et al. Unleashing the full potential of Hsp90 inhibitors as cancer therapeutics through simultaneous inactivation of Hsp90, Grp94, and TRAP1. *Experimental & Molecular Medicine*. 2020;52(1):79-91.
308. Liu K, Kang M, Li J, Qin W, Wang R. Prognostic value of the mRNA expression of members of the HSP90 family in non-small cell lung cancer. *Experimental and therapeutic medicine*. 2019;17(4):2657-65.
309. Giulino-Roth L, van Besien HJ, Dalton T, Totonchy JE, Rodina A, Taldone T, et al. Inhibition of Hsp90 Suppresses PI3K/AKT/mTOR Signaling and Has Antitumor Activity in Burkitt Lymphoma. *Molecular cancer therapeutics*. 2017;16(9):1779-90.
310. Peng Y, Chen L, Li C, Lu W, Chen J. Inhibition of MDM2 by hsp90 contributes to mutant p53 stabilization. *The Journal of biological chemistry*. 2001;276(44):40583-90.
311. Rumman M, Jung KH, Fang Z, Yan HH, Son MK, Kim SJ, et al. HS-173, a novel PI3K inhibitor suppresses EMT and metastasis in pancreatic cancer. *Oncotarget*. 2016;7(47):78029-47.
312. Franke TF, Kaplan DR, Cantley LC. PI3K: downstream AKTion blocks apoptosis. *Cell*. 1997;88(4):435-7.
313. Sato S, Fujita N, Tsuruo T. Modulation of Akt kinase activity by binding to Hsp90. *Proc Natl Acad Sci U S A*. 2000;97(20):10832-7.
314. Fujiwara H, Yamakuni T, Ueno M, Ishizuka M, Shinkawa T, Isobe T, et al. IC101 induces apoptosis by Akt dephosphorylation via an inhibition of heat shock protein 90-ATP binding activity accompanied by preventing the interaction with Akt in L1210 cells. *The Journal of pharmacology and experimental therapeutics*. 2004;310(3):1288-95.
315. Paplomata E, O'Regan R. The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. *Therapeutic advances in medical oncology*. 2014;6(4):154-66.
316. Sabapathy K. The Contrived Mutant p53 Oncogene - Beyond Loss of Functions. *Frontiers in oncology*. 2015;5:276.
317. Soussi T, Wiman KG. TP53: an oncogene in disguise. *Cell death and differentiation*. 2015;22(8):1239-49.
318. Park SJ, Kostic M, Dyson HJ. Dynamic Interaction of Hsp90 with Its Client Protein p53. *Journal of molecular biology*. 2011;411(1):158-73.
319. Pinder SE, Ellis IO, Galea M, O'Rourke S, Blamey RW, Elston CW. Pathological prognostic factors in breast cancer. III. Vascular invasion: relationship with recurrence and survival in a large study with long-term follow-up. *Histopathology*. 1994;24(1):41-7.
320. Liotta LA, Tryggvason K, Garbisa S, Hart I, Foltz CM, Shafie S. Metastatic potential correlates with enzymatic degradation of basement membrane collagen. *Nature*. 1980;284(5751):67-8.
321. Itoh Y, Nagase H. Matrix metalloproteinases in cancer. *Essays in biochemistry*. 2002;38:21-36.
322. Johansson N, Ahonen M, Kähäri VM. Matrix metalloproteinases in tumor invasion. *Cellular and molecular life sciences : CMLS*. 2000;57(1):5-15.
323. Niland S, Riscanevo AX, Eble JA. Matrix Metalloproteinases Shape the Tumor Microenvironment in Cancer Progression. *Int J Mol Sci*. 2021;23(1).
324. Song X, Wang X, Zhuo W, Shi H, Feng D, Sun Y, et al. The regulatory mechanism of extracellular Hsp90{alpha} on matrix metalloproteinase-2 processing and tumor angiogenesis. *The Journal of biological chemistry*. 2010;285(51):40039-49.
325. Stellas D, El Hamidieh A, Patsavoudi E. Monoclonal antibody 4C5 prevents activation of MMP2 and MMP9 by disrupting their interaction with extracellular HSP90 and inhibits formation of metastatic breast cancer cell deposits. *BMC cell biology*. 2010;11:51.

326. Gao W, Wu J, Chen X, Lin L, Fei X, Shen K, et al. Clinical validation of Ki67 by quantitative reverse transcription-polymerase chain reaction (RT-PCR) in HR+/HER2-early breast cancer. *J Cancer*. 2019;10(5):1110-6.
327. Creighton CJ. The molecular profile of luminal B breast cancer. *Biologics*. 2012;6:289-97.
328. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol*. 2014;5(3):412-24.
329. Banin-Hirata BK, de Oliveira CEC, Losi-Guembarovski R, Ozawa PMM, Vitiello GAF, de Almeida FC, et al. The prognostic value of regulatory T cells infiltration in HER2-enriched breast cancer microenvironment. *International Reviews of Immunology*. 2018;37(3):144-50.
330. Wang H, Lu M, Yao M, Zhu W. Effects of treatment with an Hsp90 inhibitor in tumors based on 15 phase II clinical trials. *Molecular and clinical oncology*. 2016;5(3):326-34.
331. Kurozumi S, Joseph C, Sonbul S, Gorringer KL, Pigera M, Aleskandarany MA, et al. Clinical and biological roles of Kelch-like family member 7 in breast cancer: a marker of poor prognosis. *Breast Cancer Research and Treatment*. 2018;170(3):525-33.
332. Kurozumi S, Joseph C, Sonbul S, Aleskandarany MA, Pigera M, Alsaleem M, et al. Clinicopathological and prognostic significance of Ras association and pleckstrin homology domains 1 (RAPH1) in breast cancer. *Breast Cancer Research and Treatment*. 2018;172(1):61-8.
333. Ejlerlsen B, Jensen M-B, Rank F, Rasmussen BB, Christiansen P, Kroman N, et al. Population-Based Study of Peritumoral Lymphovascular Invasion and Outcome Among Patients With Operable Breast Cancer. *JNCI: Journal of the National Cancer Institute*. 2009;101(10):729-35.
334. Weinman EJ, Steplock D, Shenolikar S. NHERF-1 uniquely transduces the cAMP signals that inhibit sodium-hydrogen exchange in mouse renal apical membranes. *FEBS letters*. 2003;536(1-3):141-4.
335. Staal FJ, Clevers HC. WNT signalling and haematopoiesis: a WNT-WNT situation. *Nature reviews Immunology*. 2005;5(1):21-30.
336. Prasad CP, Chaurasiya SK, Axelsson L, Andersson T. WNT-5A triggers Cdc42 activation leading to an ERK1/2 dependent decrease in MMP9 activity and invasive migration of breast cancer cells. *Molecular oncology*. 2013;7(5):870-83.
337. Xu X, Zhang M, Xu F, Jiang S. Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. *Molecular cancer*. 2020;19(1):165.
338. Söderholm S, Cantù C. The WNT/ β -catenin dependent transcription: A tissue-specific business. *WIREs mechanisms of disease*. 2021;13(3):e1511.
339. Speirs V. Beta-catenin in breast cancer. *Breast Cancer Research*. 2000;2(1):66674.
340. Lin SY, Xia W, Wang JC, Kwong KY, Spohn B, Wen Y, et al. Beta-catenin, a novel prognostic marker for breast cancer: its roles in cyclin D1 expression and cancer progression. *Proc Natl Acad Sci U S A*. 2000;97(8):4262-6.
341. Cantù C, Pagella P, Shajiei TD, Zimmerli D, Valenta T, Hausmann G, et al. A cytoplasmic role of Wnt/ β -catenin transcriptional cofactors Bcl9, Bcl9l, and Pygopus in tooth enamel formation. *Science signaling*. 2017;10(465).
342. Liu C, Li Y, Semenov M, Han C, Baeg GH, Tan Y, et al. Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell*. 2002;108(6):837-47.
343. Shibata T, Chuma M, Kokubu A, Sakamoto M, Hirohashi S. EBP50, a beta-catenin-associating protein, enhances Wnt signaling and is over-expressed in hepatocellular carcinoma. *Hepatology (Baltimore, Md)*. 2003;38(1):178-86.

344. Jeong J, Kim W, Hens J, Dann P, Schedin P, Friedman PA, et al. NHERF1 Is Required for Localization of PMCA2 and Suppression of Early Involution in the Female Lactating Mammary Gland. *Endocrinology*. 2019;160(8):1797-810.
345. Liu H, Ma Y, He H-W, Wang J-P, Jiang J-D, Shao R-G. SLC9A3R1 stimulates autophagy via BECN1 stabilization in breast cancer cells. *Autophagy*. 2015;11(12):2323-34.
346. Takahashi Y, Morales FC, Kreimann EL, Georgescu MM. PTEN tumor suppressor associates with NHERF proteins to attenuate PDGF receptor signaling. *EMBO J*. 2006;25(4):910-20.
347. Lazar CS, Cresson CM, Lauffenburger DA, Gill GN. The Na⁺/H⁺ exchanger regulatory factor stabilizes epidermal growth factor receptors at the cell surface. *Molecular biology of the cell*. 2004;15(12):5470-80.
348. Mangia A, Saponaro C, Malfettone A, Bisceglie D, Bellizzi A, Asselti M, et al. Involvement of nuclear NHERF1 in colorectal cancer progression. *Oncology reports*. 2012;28(3):889-94.
349. Mangia A, Partipilo G, Schirosi L, Saponaro C, Galetta D, Catino A, et al. Fine Needle Aspiration Cytology: A Tool to Study NHERF1 Expression as a Potential Marker of Aggressiveness in Lung Cancer. *Molecular biotechnology*. 2015;57(6):549-57.
350. Bellizzi A, Malfettone A, Cardone RA, Mangia A. NHERF1/EBP50 in Breast Cancer: Clinical Perspectives. *Breast care (Basel, Switzerland)*. 2010;5(2):86-90.
351. Hayashi Y, Molina JR, Hamilton SR, Georgescu MM. NHERF1/EBP50 is a new marker in colorectal cancer. *Neoplasia*. 2010;12(12):1013-22.
352. Ma Q, Jiao Y, Hao Y, Yan S, Lyu N, Gao H, et al. Targeting of NHERF1 through RNA interference inhibits the proliferation and migration of metastatic prostate cancer cells. *Oncology letters*. 2016;11(2):1149-54.
353. Dai JL, Wang L, Sahin AA, Broemeling LD, Schutte M, Pan Y. NHERF (Na⁺/H⁺ exchanger regulatory factor) gene mutations in human breast cancer. *Oncogene*. 2004;23(53):8681-7.
354. Stemmer-Rachamimov AO, Wiederhold T, Nielsen GP, James M, Pinney-Michalowski D, Roy JE, et al. NHE-RF, a merlin-interacting protein, is primarily expressed in luminal epithelia, proliferative endometrium, and estrogen receptor-positive breast carcinomas. *Am J Pathol*. 2001;158(1):57-62.
355. Cardone RA, Bellizzi A, Busco G, Weinman EJ, Dell'Aquila ME, Casavola V, et al. The NHERF1 PDZ2 domain regulates PKA-RhoA-p38-mediated NHE1 activation and invasion in breast tumor cells. *Molecular biology of the cell*. 2007;18(5):1768-80.
356. Sun J, Li P, Yang J. Repressing of NHERF1 inhibits liver cancer progression by promoting the production of ROS. *Biochem Biophys Res Commun*. 2019;509(1):8-15.
357. Criscitiello C, Disalvatore D, De Laurentiis M, Gelao L, Fumagalli L, Locatelli M, et al. High Ki-67 score is indicative of a greater benefit from adjuvant chemotherapy when added to endocrine therapy in luminal B HER2 negative and node-positive breast cancer. *Breast (Edinburgh, Scotland)*. 2014;23(1):69-75.
358. Karn T, Ruckhäberle E, Hanker L, Müller V, Schmidt M, Solbach C, et al. Gene expression profiling of luminal B breast cancers reveals NHERF1 as a new marker of endocrine resistance. *Breast Cancer Research and Treatment*. 2011;130(2):409-20.
359. Yu D, Hung M-C. Overexpression of ErbB2 in cancer and ErbB2-targeting strategies. *Oncogene*. 2000;19(53):6115-21.
360. Petit AM, Rak J, Hung MC, Rockwell P, Goldstein N, Fendly B, et al. Neutralizing antibodies against epidermal growth factor and ErbB-2/neu receptor tyrosine kinases down-regulate vascular endothelial growth factor production by tumor cells in vitro and

in vivo: angiogenic implications for signal transduction therapy of solid tumors. *Am J Pathol.* 1997;151(6):1523-30.

361. Ahmed ARH. HER2 expression is a strong independent predictor of nodal metastasis in breast cancer. *Journal of the Egyptian National Cancer Institute.* 2016;28(4):219-27.

362. Yao Z, Yuan T, Wang H, Yao S, Zhao Y, Liu Y, et al. MMP-2 together with MMP-9 overexpression correlated with lymph node metastasis and poor prognosis in early gastric carcinoma. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine.* 2017;39(6):1010428317700411.

363. Song Z, Wang J, Su Q, Luan M, Chen X, Xu X. The role of MMP-2 and MMP-9 in the metastasis and development of hypopharyngeal carcinoma. *Brazilian journal of otorhinolaryngology.* 2021;87(5):521-8.

364. Li H, Qiu Z, Li F, Wang C. The relationship between MMP-2 and MMP-9 expression levels with breast cancer incidence and prognosis. *Oncology letters.* 2017;14(5):5865-70.

365. Gooch JC, Schnabel F, Chun J, Pirraglia E, Troxel AB, Guth A, et al. A Nomogram to Predict Factors Associated with Lymph Node Metastasis in Ductal Carcinoma In Situ with Microinvasion. *Annals of surgical oncology.* 2019;26(13):4302-9.

366. Mangia A, Chiriatti A, Bellizzi A, Malfettone A, Stea B, Zito FA, et al. Biological role of NHERF1 protein expression in breast cancer. *Histopathology.* 2009;55(5):600-8.

367. Li BT, Michelini F, Misale S, Cocco E, Baldino L, Cai Y, et al. HER2-Mediated Internalization of Cytotoxic Agents in ERBB2 Amplified or Mutant Lung Cancers. *Cancer Discovery.* 2020;10(5):674.

368. Jeong J, VanHouten JN, Kim W, Dann P, Sullivan C, Choi J, et al. The scaffolding protein NHERF1 regulates the stability and activity of the tyrosine kinase HER2. *The Journal of biological chemistry.* 2017;292(16):6555-68.

369. Timmermans-Sprang EPM, Gracanin A, Mol JA. High basal Wnt signaling is further induced by PI3K/mTor inhibition but sensitive to cSRC inhibition in mammary carcinoma cell lines with HER2/3 overexpression. *BMC cancer.* 2015;15:545-.

370. Castaneda CA, Cortes-Funes H, Gomez HL, Ciruelos EM. The phosphatidyl inositol 3-kinase/AKT signaling pathway in breast cancer. *Cancer and Metastasis Reviews.* 2010;29(4):751-9.

371. Dillon RL, White DE, Muller WJ. The phosphatidyl inositol 3-kinase signaling network: implications for human breast cancer. *Oncogene.* 2007;26(9):1338-45.

372. Yang SX, Polley E, Lipkowitz S. New insights on PI3K/AKT pathway alterations and clinical outcomes in breast cancer. *Cancer treatment reviews.* 2016;45:87-96.

373. Tserga A, Chatziandreou I, Michalopoulos NV, Patsouris E, Saetta AA. Mutation of genes of the PI3K/AKT pathway in breast cancer supports their potential importance as biomarker for breast cancer aggressiveness. *Virchows Archiv.* 2016;469(1):35-43.

374. Meric-Bernstam F, Gonzalez-Angulo AM. Targeting the mTOR Signaling Network for Cancer Therapy. *Journal of Clinical Oncology.* 2009;27(13):2278-87.

375. Pan Y, Weinman EJ, Dai JL. Na⁺/H⁺ exchanger regulatory factor 1 inhibits platelet-derived growth factor signaling in breast cancer cells. *Breast Cancer Res.* 2008;10(1):R5.

376. Gautier T, Bergès T, Tollervey D, Hurt E. Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol Cell Biol.* 1997;17(12):7088-98.

377. Newman DR, Kuhn JF, Shanab GM, Maxwell ES. Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA (New York, NY).* 2000;6(6):861-79.

378. Lin LM, Pan Q, Sun YM, Wang WT. Small nucleolar RNA is potential as a novel player in leukemogenesis and clinical application. *Blood science* (Baltimore, Md). 2021;3(4):122-31.
379. Bergès T, Petfalski E, Tollervey D, Hurt EC. Synthetic lethality with fibrillarin identifies NOP77p, a nucleolar protein required for pre-rRNA processing and modification. *EMBO J.* 1994;13(13):3136-48.
380. Weaver PL, Sun C, Chang TH. Dbp3p, a putative RNA helicase in *Saccharomyces cerevisiae*, is required for efficient pre-rRNA processing predominantly at site A3. *Mol Cell Biol.* 1997;17(3):1354-65.
381. Brown SJ, Cole MD, Erives AJ. Evolution of the holozoan ribosome biogenesis regulon. *BMC genomics.* 2008;9:442.
382. Ryu H, Sun XX, Chen Y, Li Y, Wang X, Dai RS, et al. The deubiquitinase USP36 promotes snoRNP group SUMOylation and is essential for ribosome biogenesis. *EMBO reports.* 2021;22(6):e50684.
383. Zhao S, Zhang D, Liu S, Huang J. The roles of NOP56 in cancer and SCA36. *Pathology oncology research : POR.* 2023;29:1610884.
384. Perna D, Fagà G, Verrecchia A, Gorski MM, Barozzi I, Narang V, et al. Genome-wide mapping of Myc binding and gene regulation in serum-stimulated fibroblasts. *Oncogene.* 2012;31(13):1695-709.
385. Schlosser I, Hölzel M, Mürnseer M, Burtscher H, Weidle UH, Eick D. A role for c-Myc in the regulation of ribosomal RNA processing. *Nucleic acids research.* 2003;31(21):6148-56.
386. Ruggero D. The role of Myc-induced protein synthesis in cancer. *Cancer Res.* 2009;69(23):8839-43.
387. Cowling VH, Turner SA, Cole MD. Burkitt's lymphoma-associated c-Myc mutations converge on a dramatically altered target gene response and implicate Nop56 in oncogenesis. *Oncogene.* 2014;33(27):3519-27.
388. Yang Z, Liang SQ, Zhao L, Yang H, Marti TM, Hegedüs B, et al. Metabolic synthetic lethality by targeting NOP56 and mTOR in KRAS-mutant lung cancer. *Journal of experimental & clinical cancer research : CR.* 2022;41(1):25.
389. Zhang Y, Lin Z, Lin X, Zhang X, Zhao Q, Sun Y. A gene module identification algorithm and its applications to identify gene modules and key genes of hepatocellular carcinoma. *Sci Rep.* 2021;11(1):5517.
390. Ojha S, Malla S, Lyons SM. snoRNPs: Functions in Ribosome Biogenesis. *Biomolecules.* 2020;10(5).
391. Pelletier J, Thomas G, Volarević S. Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nature reviews Cancer.* 2018;18(1):51-63.
392. Liang J, Wen J, Huang Z, Chen XP, Zhang BX, Chu L. Small Nucleolar RNAs: Insight Into Their Function in Cancer. *Frontiers in oncology.* 2019;9:587.
393. Bohnsack KE, Bohnsack MT. Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J.* 2019;38(13):e100278-e.
394. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: a cancer journal for clinicians.* 2017;67(6):439-48.
395. Kariri YA, Aleskandarany MA, Joseph C, Kurozumi S, Mohammed OJ, Toss MS, et al. Molecular Complexity of Lymphovascular Invasion: The Role of Cell Migration in Breast Cancer as a Prototype. *Pathobiology.* 2020;87(4):218-31.
396. Su H, Xu T, Ganapathy S, Shadfan M, Long M, Huang TH, et al. Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene.* 2014;33(11):1348-58.

397. Xu J, Chen Y, Olopade OI. MYC and Breast Cancer. *Genes Cancer*. 2010;1(6):629-40.
398. Gong J, Li Y, Liu CJ, Xiang Y, Li C, Ye Y, et al. A Pan-cancer Analysis of the Expression and Clinical Relevance of Small Nucleolar RNAs in Human Cancer. *Cell reports*. 2017;21(7):1968-81.
399. Qingshan QJLPLXWLL. Expression of NOP56 in breast cancer tissues and its clinical significance. *Chinese Journal of Bioinformatics*. 2019;17(2):9.
400. Ohshima K, Morii E. Metabolic Reprogramming of Cancer Cells during Tumor Progression and Metastasis. *Metabolites*. 2021;11(1).
401. Gao F-y, Li X-t, Xu K, Wang R-t, Guan X-x. c-MYC mediates the crosstalk between breast cancer cells and tumor microenvironment. *Cell Communication and Signaling*. 2023;21(1):28.
402. Massó-Vallés D, Beaulieu M-E, Jauset T, Giuntini F, Zacarías-Fluck MF, Foradada L, et al. MYC Inhibition Halts Metastatic Breast Cancer Progression by Blocking Growth, Invasion, and Seeding. *Cancer Research Communications*. 2022;2(2):110-30.
403. Deming SL, Nass SJ, Dickson RB, Trock BJ. C-myc amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. *British Journal of Cancer*. 2000;83(12):1688-95.
404. Blancato J, Singh B, Liu A, Liao DJ, Dickson RB. Correlation of amplification and overexpression of the c-myc oncogene in high-grade breast cancer: FISH, in situ hybridisation and immunohistochemical analyses. *Br J Cancer*. 2004;90(8):1612-9.
405. Elwy AE, Elsaba TM, Abd Elzaher AR, Nassar MI. Prognostic Value of c-Myc Immunohistochemical Expression in Muscle Invasive Urothelial Carcinoma of the Urinary Bladder: A Retrospective Study. *Asian Pacific journal of cancer prevention : APJCP*. 2019;20(12):3735-46.
406. Venturutti L, Cordo Russo RI, Rivas MA, Mercogliano MF, Izzo F, Oakley RH, et al. MiR-16 mediates trastuzumab and lapatinib response in ErbB-2-positive breast and gastric cancer via its novel targets CCNJ and FUBP1. *Oncogene*. 2016;35(48):6189-202.
407. Cadoo KA, Fornier MN, Morris PG. Biological subtypes of breast cancer: current concepts and implications for recurrence patterns. *The quarterly journal of nuclear medicine and molecular imaging : official publication of the Italian Association of Nuclear Medicine (AIMN) [and] the International Association of Radiopharmacology (IAR), [and] Section of the So*. 2013;57(4):312-21.
408. Agarwal G, Nanda G, Lal P, Mishra A, Agarwal A, Agrawal V, et al. Outcomes of Triple-Negative Breast Cancers (TNBC) Compared with Non-TNBC: Does the Survival Vary for All Stages? *World journal of surgery*. 2016;40(6):1362-72.
409. Na YM, Ryu YJ, Cho JS, Park MH, Yoon JH. Lymphovascular Invasion as a Predictive Factor for Recurrence in Triple-Negative Breast Cancer. *Indian Journal of Surgery*. 2021;83(2):475-83.
410. Marshall DC, Webb TE, Hall RA, Saliccioli JD, Ali R, Maruthappu M. Trends in UK regional cancer mortality 1991-2007. *Br J Cancer*. 2016;114(3):340-7.
411. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008;26(8):1275-81.
412. 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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2011;29(19):2619-27.

413. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res.* 2010;12(4):207.
414. Fang J, Ge X, Xu W, Xie J, Qin Z, Shi L, et al. Bioinformatics analysis of the prognosis and biological significance of HMGB1, HMGB2, and HMGB3 in gastric cancer. *J Cell Physiol.* 2020;235(4):3438-46.
415. Agresti A, Bianchi ME. HMGB proteins and gene expression. *Current opinion in genetics & development.* 2003;13(2):170-8.
416. Stros M. HMGB proteins: interactions with DNA and chromatin. *Biochimica et biophysica acta.* 2010;1799(1-2):101-13.
417. Guo ZS, Liu Z, Bartlett DL, Tang D, Lotze MT. Life after death: targeting high mobility group box 1 in emergent cancer therapies. *American journal of cancer research.* 2013;3(1):1-20.
418. Zhang Z, Chang Y, Zhang J, Lu Y, Zheng L, Hu Y, et al. HMGB3 promotes growth and migration in colorectal cancer by regulating WNT/ β -catenin pathway. *PLOS ONE.* 2017;12(7):e0179741.
419. Ueda M, Takahashi Y, Shinden Y, Sakimura S, Hirata H, Uchi R, et al. Prognostic significance of high mobility group box 1 (HMGB1) expression in patients with colorectal cancer. *Anticancer research.* 2014;34(10):5357-62.
420. Zheng W, Yang J, Dong Z, Wang L, Fang M, Wu W, et al. High mobility group box 3 as an emerging biomarker in diagnosis and prognosis of hepatocellular carcinoma. *Cancer management and research.* 2018;10:5979-89.
421. Takeda T, Izumi H, Kitada S, Uramoto H, Tasaki T, Zhi L, et al. 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 biology : the journal of the International Society for Oncodevelopmental Biology and Medicine.* 2014;35(10):10555-69.
422. Guo S, Wang Y, Gao Y, Zhang Y, Chen M, Xu M, et al. Knockdown of High Mobility Group-Box 3 (HMGB3) Expression Inhibits Proliferation, Reduces Migration, and Affects Chemosensitivity in Gastric Cancer Cells. *Medical science monitor : international medical journal of experimental and clinical research.* 2016;22:3951-60.
423. Wang K, Yin Y-H, Yang Z-Q, Yu H-F, Wang Y-S, Guo B, et al. Hmgb3 Induces the Differentiation of Uterine Stromal Cells Through Targeting Ptn. *Reproductive Sciences.* 2018;26(7):891-9.
424. Wen B, Wei YT, Zhao K. The role of high mobility group protein B3 (HMGB3) in tumor proliferation and drug resistance. *Molecular and cellular biochemistry.* 2021;476(4):1729-39.
425. Xie Y, Wang L, Yang D. CircEPSTI1 Promotes the Progression of Non-Small Cell Lung Cancer Through miR-145/HMGB3 Axis. *Cancer management and research.* 2020;12:6827-36.
426. Yamada Y, Nishikawa R, Kato M, Okato A, Arai T, Kojima S, et al. Regulation of HMGB3 by antitumor miR-205-5p inhibits cancer cell aggressiveness and is involved in prostate cancer pathogenesis. *J Hum Genet.* 2018;63(2):195-205.
427. Li M, Cai Y, Zhao H, Xu Z, Sun Q, Luo M, et al. Overexpression of HMGB3 protein promotes cell proliferation, migration and is associated with poor prognosis in urinary bladder cancer patients. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine.* 2015;36(6):4785-92.
428. Petrova NS, Zenkova MA, Chernolovskaya EL, editors. *Structure - Functions Relations in Small Interfering RNAs* 2013.

429. Riss TL, Moravec RA, Niles AL. Cytotoxicity testing: measuring viable cells, dead cells, and detecting mechanism of cell death. *Methods in molecular biology* (Clifton, NJ). 2011;740:103-14.
430. Munshi A, Hobbs M, Meyn RE. Clonogenic cell survival assay. *Methods in molecular medicine*. 2005;110:21-8.
431. Rafehi H, Orlowski C, Georgiadis GT, Ververis K, El-Osta A, Karagiannis TC. Clonogenic assay: adherent cells. *Journal of visualized experiments : JoVE*. 2011(49).
432. Jonkman JE, Cathcart JA, Xu F, Bartolini ME, Amon JE, Stevens KM, et al. An introduction to the wound healing assay using live-cell microscopy. *Cell adhesion & migration*. 2014;8(5):440-51.
433. 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 journal of gastroenterology*. 2018;24(32):3650-62.
434. Song N, Liu B, Wu JL, Zhang RF, Duan L, He WS, et al. Prognostic value of HMGB3 expression in patients with non-small cell lung cancer. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2013;34(5):2599-603.
435. Gong Y, Cao Y, Song L, Zhou J, Wang C, Wu B. HMGB3 characterization in gastric cancer. *Genetics and molecular research : GMR*. 2013;12(4):6032-9.
436. Juríková M, Danihel L, Polák Š, Varga I. Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer. *Acta histochemica*. 2016;118(5):544-52.
437. 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 biochemistry and function*. 2018;36(7):357-65.
438. Safuan S, Storr SJ, Patel PM, Martin SG. A comparative study of adhesion of melanoma and breast cancer cells to blood and lymphatic endothelium. *Lymphatic research and biology*. 2012;10(4):173-81.
439. Walter-Yohrling J, Cao X, Callahan M, Weber W, Morgenbesser S, Madden SL, et al. Identification of genes expressed in malignant cells that promote invasion. *Cancer Res*. 2003;63(24):8939-47.
440. Kim HG, Kim JY, Han EH, Hwang YP, Choi JH, Park BH, et al. Metallothionein-2A overexpression increases the expression of matrix metalloproteinase-9 and invasion of breast cancer cells. *FEBS letters*. 2011;585(2):421-8.
441. Yang J, Nie J, Ma X, Wei Y, Peng Y, Wei X. Targeting PI3K in cancer: mechanisms and advances in clinical trials. *Molecular cancer*. 2019;18(1):26.
442. Martini M, De Santis MC, Braccini L, Gulluni F, Hirsch E. PI3K/AKT signaling pathway and cancer: an updated review. *Annals of medicine*. 2014;46(6):372-83.
443. Akinleye A, Avvaru P, Furqan M, Song Y, Liu D. Phosphatidylinositol 3-kinase (PI3K) inhibitors as cancer therapeutics. *Journal of hematology & oncology*. 2013;6(1):88.
444. Matsuoka T, Yashiro M, Nishioka N, Hirakawa K, Olden K, Roberts JD. PI3K/Akt signalling is required for the attachment and spreading, and growth in vivo of metastatic scirrhous gastric carcinoma. *Br J Cancer*. 2012;106(9):1535-42.
445. de Vicente JC, Rosado P, Lequerica-Fernández P, Allonca E, Villallaín L, Hernández-Vallejo G. Focal adhesion kinase overexpression: correlation with lymph node metastasis and shorter survival in oral squamous cell carcinoma. *Head & neck*. 2013;35(6):826-30.

446. Wang S, Basson MD. Akt directly regulates focal adhesion kinase through association and serine phosphorylation: implication for pressure-induced colon cancer metastasis. *American journal of physiology Cell physiology*. 2011;300(3):C657-70.
447. Ebi H, Costa C, Faber AC, Nishtala M, Kotani H, Juric D, et al. PI3K regulates MEK/ERK signaling in breast cancer via the Rac-GEF, P-Rex1. *Proc Natl Acad Sci U S A*. 2013;110(52):21124-9.
448. Sahai E, Marshall CJ. RHO-GTPases and cancer. *Nature Reviews Cancer*. 2002;2(2):133-42.
449. Kumar R, Gururaj AE, Barnes CJ. p21-activated kinases in cancer. *Nature Reviews Cancer*. 2006;6(6):459-71.
450. Kamai T, Shirataki H, Nakanishi K, Furuya N, Kambara T, Abe H, et al. Increased Rac1 activity and Pak1 overexpression are associated with lymphovascular invasion and lymph node metastasis of upper urinary tract cancer. *BMC Cancer*. 2010;10(1):164.
451. Lindström LS, Yau C, Czene K, Thompson CK, Hoadley KA, Van't Veer LJ, et al. Intratumor Heterogeneity of the Estrogen Receptor and the Long-term Risk of Fatal Breast Cancer. *Journal of the National Cancer Institute*. 2018;110(7):726-33.
452. Hong D, Fritz AJ, Zaidi SK, van Wijnen AJ, Nickerson JA, Imbalzano AN, et al. Epithelial-to-mesenchymal transition and cancer stem cells contribute to breast cancer heterogeneity. *J Cell Physiol*. 2018;233(12):9136-44.
453. García-Fernández A, Barco I, Fraile M, Lain JM, Carmona A, Gonzalez S, et al. Factors predictive of mortality in a cohort of women surgically treated for breast cancer from 1997 to 2014. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*. 2016;134(2):212-6.
454. Insua-Rodríguez J, Pein M, Hongu T, Meier J, Descot A, Lowy CM, et al. Stress signaling in breast cancer cells induces matrix components that promote chemoresistant metastasis. *EMBO molecular medicine*. 2018;10(10).
455. Abdel-Fatah TMA, Agarwal D, Liu DX, Russell R, Rueda OM, Liu K, et al. SPAG5 as a prognostic biomarker and chemotherapy sensitivity predictor in breast cancer: a retrospective, integrated genomic, transcriptomic, and protein analysis. *The Lancet Oncology*. 2016;17(7):1004-18.
456. Soliman NA, Yussif SM. Ki-67 as a prognostic marker according to breast cancer molecular subtype. *Cancer biology & medicine*. 2016;13(4):496-504.
457. Hoelder S, Clarke PA, Workman P. Discovery of small molecule cancer drugs: successes, challenges and opportunities. *Molecular oncology*. 2012;6(2):155-76.
458. Fakhr E, Zare F, Teimoori-Toolabi L. Precise and efficient siRNA design: a key point in competent gene silencing. *Cancer gene therapy*. 2016;23(4):73-82.
459. Zhao X, Li F, Li Y, Wang H, Ren H, Chen J, et al. Co-delivery of HIF1 α siRNA and gemcitabine via biocompatible lipid-polymer hybrid nanoparticles for effective treatment of pancreatic cancer. *Biomaterials*. 2015;46:13-25.
460. Parvani JG, Gujrati MD, Mack MA, Schiemann WP, Lu ZR. Silencing β 3 Integrin by Targeted ECO/siRNA Nanoparticles Inhibits EMT and Metastasis of Triple-Negative Breast Cancer. *Cancer Res*. 2015;75(11):2316-25.
461. Su S, Tian Y, Li Y, Ding Y, Ji T, Wu M, et al. "Triple-punch" strategy for triple negative breast cancer therapy with minimized drug dosage and improved antitumor efficacy. *ACS nano*. 2015;9(2):1367-78.
462. Zhang X, Wang Q, Qin L, Fu H, Fang Y, Han B, et al. EGF-modified mPEG-PLGA-PLL nanoparticle for delivering doxorubicin combined with Bcl-2 siRNA as a potential treatment strategy for lung cancer. *Drug delivery*. 2016;23(8):2936-45.

463. Liu JY, Chiang T, Liu CH, Chern GG, Lin TT, Gao DY, et al. Delivery of siRNA Using CXCR4-targeted Nanoparticles Modulates Tumor Microenvironment and Achieves a Potent Antitumor Response in Liver Cancer. *Molecular therapy : the journal of the American Society of Gene Therapy*. 2015;23(11):1772-82.
464. Singh A, Trivedi P, Jain NK. Advances in siRNA delivery in cancer therapy. *Artificial cells, nanomedicine, and biotechnology*. 2018;46(2):274-83.
465. Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, et al. Expression profiling reveals off-target gene regulation by RNAi. *Nature biotechnology*. 2003;21(6):635-7.
466. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for RNA interference. *Nature biotechnology*. 2004;22(3):326-30.
467. Ozcan G, Ozpolat B, Coleman RL, Sood AK, Lopez-Berestein G. Preclinical and clinical development of siRNA-based therapeutics. *Advanced drug delivery reviews*. 2015;87:108-19.
468. Steinmetz KL, Spack EG. The basics of preclinical drug development for neurodegenerative disease indications. *BMC neurology*. 2009;9 Suppl 1(Suppl 1):S2.
469. Sauerbrei W, Taube SE, McShane LM, Cavenagh MM, Altman DG. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An Abridged Explanation and Elaboration. *Journal of the National Cancer Institute*. 2018;110(8):803-11.

10 Appendices

10.1 Copyright approval for re-use Figure 1.1

Thank you for your order!

Dear Dr. Sami Alsaeed,

Thank you for placing your order through **Copyright** Clearance Center's RightsLink® service.

Order Summary

Licensee:	University Of Nottingham
Order Date:	Jan 18, 2023
Order Number:	5471981506210
Publication:	Cancer Discovery
Title:	Hallmarks of Cancer: New Dimensions
Type of Use:	Thesis/Dissertation
Order Ref:	1
Order Total:	0.00 GBP

View or print complete [details](#) of your order and the publisher's terms and conditions.

10.2 Copyright approval for re-use “A key genomic subtype with lymphovascular invasion in invasive breast cancer” (223).

SPRINGER NATURE

A key genomic subtype associated with lymphovascular invasion in invasive breast cancer

Author: Sasagu Kurozumi et al

Publication: British Journal of Cancer

Publisher: Springer Nature

Date: May 22, 2019

Copyright © 2019, Cancer Research UK

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](#)

10.3 Copyright approval for the re-use “Prognostic significance of heat shock protein 90AA1 (HSP90 α in invasive breast cancer” (274).

ET editorial.thebmj <editorial.thebmj+canned.response@bmj.com>     
To: SAMI ALSAEED Mon 16/01/2023 13:29

Thank you for your email to bmj.permissions@bmj.com.

If you are an author and wish to use your own article in a thesis please note that this is permitted by your author licence and you do not require permission. Please see your author rights here - <https://www.bmj.com/company/products-services/rights-and-licensing/author-self-archiving-and-permissions/>.

Students wishing to reuse figures and tables in their theses/dissertations may also do so without permission; or if your institution requires proof you can acquire a free licence through RightsLink as below

B bdoro@bmj.com on behalf of bmj.permissions <bmj.permissions@bmj.com>  
To: SAMI ALSAEED
Cc: bmj.permissions@bmj.com Mon 16/01/2023 14:25

Dear Sami,

Thank you for contacting BMJ.

I am happy to confirm that as the author of the article you are allowed to reuse it in your thesis.

If you wish to, you can read more about authors' rights here:
<https://www.bmj.com/company/products-services/rights-and-licensing/author-self-archiving-and-permissions/>

I hope this helps.

Best wishes,
Benedetta

10.4 List of top 350 genes significantly associated with lymphovascular invasion in the Nottingham cohort.

WAD ranking	WAD value	Gene symbols	WAD ranking	WAD value	Gene symbols	WAD ranking	WAD value	Gene symbols	WAD ranking	WAD value	Gene symbols
1	-0.339	<i>C10orf116</i>	89	0.114	<i>TFAP2A</i>	177	-0.092	<i>WBP5</i>	265	-0.078	<i>XBP1</i>
2	-0.291	<i>MGP</i>	90	0.114	<i>KRT8</i>	178	-0.091	<i>ALDH2</i>	266	0.078	<i>EPN1</i>
3	0.256	<i>EEF1A2</i>	91	-0.114	<i>PRRX1</i>	179	0.091	<i>IDH2</i>	267	0.078	<i>FOXA1</i>
4	-0.236	<i>CFB</i>	92	-0.113	<i>PPP1R1B</i>	180	-0.091	<i>ACOX2</i>	268	-0.078	<i>CCND2</i>
5	0.217	<i>S100P</i>	93	-0.113	<i>RPS20</i>	181	-0.091	<i>SERPINF1</i>	269	-0.078	<i>LSM1</i>
6	-0.217	<i>STC2</i>	94	-0.113	<i>BEX1</i>	182	-0.091	<i>CIDEA</i>	270	0.078	<i>SNAR-A3</i>
7	-0.215	<i>SUSD3</i>	95	-0.112	<i>ANXA1</i>	183	-0.091	<i>RPL27A</i>	271	0.077	<i>ZNF217</i>
8	0.212	<i>CDH1</i>	96	0.112	<i>RUSC1</i>	184	0.090	<i>MMP11</i>	272	-0.077	<i>RPSA</i>
9	0.209	<i>MX1</i>	97	-0.111	<i>IFITM1</i>	185	-0.090	<i>EFEMP1</i>	273	-0.077	<i>CD36</i>
10	-0.201	<i>CFD</i>	98	-0.111	<i>CAP2</i>	186	-0.090	<i>ANG</i>	274	0.077	<i>ELF3</i>
11	0.199	<i>PITX1</i>	99	-0.111	<i>PLAC9</i>	187	-0.090	<i>CCL15</i>	275	-0.077	<i>TPM2</i>
12	-0.194	<i>COMP</i>	100	-0.111	<i>ACTG2</i>	188	-0.090	<i>HLA-DQA1</i>	276	-0.077	<i>SAPS2</i>
13	-0.190	<i>FABP4</i>	101	0.111	<i>GNAS</i>	189	0.090	<i>UCP2</i>	277	-0.077	<i>NFIB</i>
14	-0.187	<i>C1orf64</i>	102	-0.111	<i>RARRES2</i>	190	-0.089	<i>RPL36</i>	278	0.077	<i>MBOAT7</i>
15	-0.174	<i>MT1E</i>	103	-0.110	<i>HLA-DRA</i>	191	-0.089	<i>ECM2</i>	279	0.076	<i>ATP5B</i>
16	-0.173	<i>SERPINE2</i>	104	-0.110	<i>RPL10A</i>	192	0.089	<i>S100A9</i>	280	-0.076	<i>C7orf41</i>
17	-0.170	<i>FBLN1</i>	105	-0.110	<i>RPS18</i>	193	-0.088	<i>BTG1</i>	281	0.076	<i>ABCB9</i>
18	-0.169	<i>PLIN4</i>	106	0.110	<i>UBE2C</i>	194	-0.088	<i>C13orf15</i>	282	0.076	<i>CDR2L</i>
19	-0.167	<i>FCGBP</i>	107	-0.109	<i>GLTSCR2</i>	195	-0.088	<i>CITED2</i>	283	-0.076	<i>RPS28</i>
20	-0.165	<i>CIDEA</i>	108	0.109	<i>SLC9A3R1</i>	196	0.088	<i>HOXB2</i>	284	0.076	<i>LMTK3</i>
21	-0.164	<i>FGD3</i>	109	-0.109	<i>GFRA1</i>	197	-0.088	<i>CDC42EP4</i>	285	0.075	<i>P4HB</i>
22	0.164	<i>FASN</i>	110	0.109	<i>CTSD</i>	198	-0.088	<i>CAV1</i>	286	0.075	<i>ATP5C1</i>
23	-0.162	<i>ESR1</i>	111	-0.108	<i>AKR1C2</i>	199	0.088	<i>PGAP3</i>	287	-0.075	<i>F13A1</i>
24	0.161	<i>IFI27</i>	112	-0.108	<i>RPL3</i>	200	0.087	<i>SCD</i>	288	0.075	<i>ULK1</i>
25	0.157	<i>APOC1</i>	113	0.107	<i>HSP90AB1</i>	201	-0.087	<i>FAU</i>	289	-0.075	<i>KLHDC9</i>
26	-0.153	<i>SCUBE2</i>	114	0.107	<i>HSPA1A</i>	202	-0.087	<i>LRRC17</i>	290	0.074	<i>ZG16B</i>
27	0.151	<i>SLC7A5</i>	115	-0.107	<i>ARHGEF6</i>	203	0.087	<i>PROM2</i>	291	0.074	<i>TMED9</i>
28	-0.150	<i>TFP1</i>	116	-0.107	<i>AKR1C3</i>	204	-0.087	<i>CCL5</i>	292	0.074	<i>ZMIZ1</i>
29	-0.150	<i>LGALS1</i>	117	-0.106	<i>RPL13AP6</i>	205	0.087	<i>DNAJA4</i>	293	0.074	<i>ATP2A2</i>
30	0.149	<i>ALDOA</i>	118	-0.106	<i>LPAR1</i>	206	-0.087	<i>IFITM2</i>	294	-0.074	<i>RPL27</i>
31	0.147	<i>APOE</i>	119	-0.106	<i>SOCS2</i>	207	-0.087	<i>ARHGEF3</i>	295	0.074	<i>GPI</i>
32	0.146	<i>HSPB1</i>	120	0.105	<i>FLJ40504</i>	208	-0.087	<i>HCST</i>	296	-0.074	<i>WNK4</i>
33	-0.145	<i>GSTP1</i>	121	-0.105	<i>SAI1</i>	209	-0.087	<i>S100A4</i>	297	-0.074	<i>RPL35</i>
34	-0.145	<i>ADH1A</i>	122	-0.105	<i>PYCARD</i>	210	0.086	<i>HIST1H4H</i>	298	-0.074	<i>RS124D1</i>
35	0.144	<i>LGALS3BP</i>	123	-0.105	<i>HLA-DMA</i>	211	-0.086	<i>ALDH3A2</i>	299	0.074	<i>CYB561</i>
36	0.143	<i>KRT18</i>	124	-0.105	<i>COL8A1</i>	212	-0.086	<i>RFTN1</i>	300	-0.073	<i>AK001020</i>
37	-0.143	<i>TOMM7</i>	125	0.105	<i>LRRC26</i>	213	0.086	<i>YWHAQ</i>	301	-0.073	<i>LUM</i>
38	0.142	<i>IFI6</i>	126	-0.105	<i>MYB</i>	214	-0.086	<i>DPYSL2</i>	302	0.073	<i>ACP5</i>
39	-0.142	<i>GPX3</i>	127	0.105	<i>IRX3</i>	215	-0.086	<i>RPL22</i>	303	0.073	<i>HMGA1</i>
40	-0.142	<i>UBD</i>	128	-0.104	<i>RPS17</i>	216	0.085	<i>PFKP</i>	304	-0.073	<i>FBP1</i>
41	-0.142	<i>FST</i>	129	-0.104	<i>TNS3</i>	217	0.085	<i>NME1</i>	305	0.073	<i>FTH1</i>
42	-0.141	<i>SFRP4</i>	130	-0.104	<i>MFAP5</i>	218	-0.085	<i>COMMD6</i>	306	0.073	<i>MELK</i>
43	-0.141	<i>SHISA2</i>	131	-0.103	<i>RPL24</i>	219	-0.085	<i>EEF1B2</i>	307	-0.072	<i>FMOD</i>
44	-0.141	<i>RPS13</i>	132	0.103	<i>TM7SF2</i>	220	0.085	<i>NFKBIZ</i>	308	-0.072	<i>HBA2</i>
45	-0.140	<i>CXCL12</i>	133	0.102	<i>TMSB10</i>	221	-0.084	<i>VCAM1</i>	309	0.072	<i>GPR172A</i>
46	-0.139	<i>SLC44A1</i>	134	-0.102	<i>VTCN1</i>	222	0.084	<i>CALM1</i>	310	-0.072	<i>TIGA1</i>
47	-0.138	<i>AGR3</i>	135	0.102	<i>HIST1H2BK</i>	223	0.084	<i>KRT7</i>	311	0.072	<i>GSTM2</i>
48	0.137	<i>abParts</i>	136	-0.102	<i>AX746718</i>	224	0.083	<i>SLC25A5</i>	312	-0.072	<i>TSPAN9</i>
49	-0.136	<i>CXCL14</i>	137	0.102	<i>MGST1</i>	225	-0.083	<i>MGC87042</i>	313	-0.072	<i>POLB</i>
50	-0.135	<i>CLIC6</i>	138	0.101	<i>ADAM15</i>	226	0.083	<i>BCAP31</i>	314	-0.072	<i>NINJ1</i>
51	0.135	<i>PIP</i>	139	-0.100	<i>RPS14</i>	227	-0.083	<i>GAS1</i>	315	-0.071	<i>RPL5</i>
52	-0.135	<i>TGFBR3</i>	140	0.100	<i>ISG15</i>	228	0.083	<i>MMP9</i>	316	-0.071	<i>CEBPD</i>
53	-0.134	<i>C1S</i>	141	-0.099	<i>MFAP4</i>	229	0.083	<i>FTL</i>	317	0.071	<i>ASNS</i>
54	-0.134	<i>DDK3</i>	142	0.099	<i>APOD</i>	230	0.082	<i>MDH2</i>	318	-0.071	<i>RBP1</i>
55	-0.133	<i>SELM</i>	143	0.099	<i>AZGP1</i>	231	-0.082	<i>C8orf40</i>	319	0.071	<i>UBA1</i>
56	-0.132	<i>DPYSL3</i>	144	-0.099	<i>FCER1A</i>	232	0.082	<i>CDC45</i>	320	0.071	<i>AKT1</i>
57	-0.131	<i>VIM</i>	145	-0.099	<i>MT2A</i>	233	0.082	<i>GLYATL2</i>	321	-0.071	<i>DARC</i>
58	0.131	<i>DBNDD1</i>	146	0.099	<i>CPB1</i>	234	-0.082	<i>TPSAB1</i>	322	-0.071	<i>RERG</i>
59	-0.131	<i>SLC40A1</i>	147	0.098	<i>ATP6V1B1</i>	235	0.082	<i>RNASE1</i>	323	-0.071	<i>PALLD</i>
60	-0.130	<i>MT1A</i>	148	0.098	<i>S100A8</i>	236	-0.081	<i>HLA-DRB6</i>	324	-0.071	<i>OMD</i>
61	0.129	<i>CALML5</i>	149	-0.098	<i>RPL21</i>	237	-0.081	<i>HLA-DQB1</i>	325	-0.071	<i>DCN</i>
62	0.128	<i>ACTG1</i>	150	-0.098	<i>RPS26P11</i>	238	-0.081	<i>CRYAB</i>	326	0.071	<i>CCNB2</i>
63	-0.127	<i>SERPINA3</i>	151	-0.097	<i>DPT</i>	239	-0.081	<i>CPA3</i>	327	-0.071	<i>COL9A2</i>
64	-0.127	<i>CYP4X1</i>	152	-0.097	<i>NDP</i>	240	-0.081	<i>C10orf10</i>	328	-0.071	<i>SRPX</i>
65	0.125	<i>ERBB2</i>	153	-0.097	<i>IL6ST</i>	241	0.081	<i>TUBB</i>	329	-0.071	<i>CTDSPL</i>

66	0.124	MDK	154	-0.097	SMARCA1	242	0.081	NOP56	330	-0.070	ARHGEF2
67	-0.123	PDGFRL	155	-0.096	PK3	243	-0.081	FERMT2	331	-0.070	RPS27
68	-0.123	RPL26	156	-0.096	TPST2	244	-0.080	PRKCDBP	332	-0.070	THBS2
69	0.122	TACSTD2	157	-0.096	GAS6	245	0.080	CD24	333	0.070	HSPD1
70	0.122	UBE2S	158	-0.096	SLC38A1	246	0.080	GRN	334	0.070	ARHGDI
71	-0.119	KIAA0531	159	-0.095	RPL35A	247	-0.080	MXRA5	335	-0.070	ANKRD30A
72	0.118	LY6E	160	-0.095	C9orf46	248	0.080	LASP1	336	-0.070	PTRF
73	-0.118	HBB	161	-0.095	PPP1R3C	249	-0.080	WISP2	337	-0.070	FOXO3
74	-0.118	RPS9	162	-0.095	CYBRD1	250	0.080	POLD2	338	0.070	GIPC1
75	-0.118	RPS3	163	-0.095	CNN3	251	0.080	POTEKP	339	-0.070	CHPT1
76	-0.118	S100A6	164	-0.095	ITPRIPL2	252	-0.080	ARL6IP5	340	-0.070	IL17RB
77	-0.117	RPS6	165	0.094	TUBA1C	253	-0.080	GBP2	341	-0.070	SYBU
78	-0.116	EIF3E	166	0.094	HMGB3	254	0.080	TSPYL5	342	0.070	TUFT1
79	0.116	YWHAZ	167	0.094	ATP6AP1	255	-0.079	FLNB	343	-0.070	CCDC25
80	-0.116	CA12	168	-0.093	HIST1H4C	256	0.079	H2AFY2	344	-0.070	ARID5B
81	-0.116	HLA-DPA1	169	0.093	CSTB	257	0.079	PTTG1	345	-0.070	IGFBP4
82	0.115	PHGDH	170	-0.093	RAI14	258	0.079	COX5A	346	0.070	SPDEF
83	0.115	HSP90AA1	171	-0.093	TCEAL4	259	-0.078	TXNIP	347	0.070	VDAC2
84	-0.115	MAOA	172	0.093	CLDN7	260	-0.078	EIF3L	348	-0.070	DUSP1
85	0.115	LAPTM4B	173	-0.092	PLS3	261	-0.078	ATHL1	349	-0.069	FOS
86	0.115	ACTB	174	0.092	CR610863	262	0.078	CHCHD2	350	0.069	COPG
87	0.115	SEZ6L2	175	-0.092	BTG2	263	0.078	COX6C			
88	0.115	KRT19	176	-0.092	SGCE	264	-0.078	BCL2			

10.5 List of top 350 genes significantly associated with lymphovascular invasion in the remaining METABRIC cases.

WAD ranking	WAD value	Gene symbols	WAD ranking	WAD value	Gene symbols	WAD ranking	WAD value	Gene symbols	WAD ranking	WAD value	Gene symbols
1	0.191	ERBB2	89	-0.068	FST	177	0.054	ORMDL3	265	-0.047	CH13L2
2	-0.170	SERPINA3	90	-0.068	DKK3	178	-0.054	MAOA	266	-0.047	GATA3
3	0.145	S100P	91	-0.068	PAM	179	-0.053	ANG	267	0.047	AURKA
4	0.141	TFF3	92	0.068	NME4	180	0.053	SERHL2	268	-0.047	SCPEP1
5	-0.123	PIP	93	-0.068	ZAK	181	-0.053	FBLN2	269	-0.047	TXNIP
6	-0.120	MGP	94	0.067	UBE2S	182	-0.053	C9orf46	270	0.047	ZNF148
7	-0.120	ACTG2	95	-0.066	MFGF8	183	-0.053	MMP7	271	-0.047	QPCT
8	-0.120	GSTP1	96	0.066	PUF60	184	0.053	TMEM106C	272	-0.047	CD248
9	-0.119	HBB	97	-0.066	MT1X	185	0.053	ALCAM	273	-0.047	PRDX2
10	0.116	UBE2C	98	-0.066	EGR1	186	0.053	CCNB2	274	0.047	BOLA2
11	0.115	MMP11	99	0.065	TUBA1B	187	0.053	KRT7	275	0.047	GRINA
12	-0.112	ACTG1	100	0.065	CDCA5	188	-0.053	SGK223	276	0.047	hNp95
13	-0.107	FOS	101	-0.064	NAT1	189	0.053	MFSD3	277	0.047	COX6C
14	0.107	LY6E	102	-0.064	SRP9	190	-0.053	ALOX5	278	0.047	RPL30
15	0.106	CNTNAP2	103	-0.064	VIM	191	-0.053	ALOX5AP	279	-0.047	IGJ
16	-0.105	HBA2	104	-0.064	PDLIM1	192	-0.053	CA2	280	-0.047	TGFBR2
17	0.104	YWHAZ	105	0.064	EEF1A2	193	0.052	ATP6V0B	281	0.047	STIP1
18	0.104	FLJ40504	106	0.064	SQLE	194	0.052	FGFR3	282	0.047	TDG
19	-0.102	CLIC6	107	-0.064	DPYSL2	195	-0.052	APOD	283	-0.047	KRT6B
20	0.101	PGAP3	108	-0.064	COL16A1	196	-0.052	TGOLN2	284	0.047	CLN3
21	-0.099	SFRP1	109	0.063	SLC9A3R1	197	-0.052	ZDHHC8	285	-0.046	PTGDS
22	-0.097	HLA-A	110	0.063	NAPRT1	198	-0.052	ZFP36L2	286	0.046	NOP56
23	0.096	SPDEF	111	0.063	RRM1	199	-0.052	MYC	287	0.046	ELF3
24	-0.094	CYP4X1	112	0.063	HIST1H4C	200	0.052	NCRNA00152	288	0.046	ASAP1
25	0.093	STARD10	113	-0.063	SLC40A1	201	-0.052	PCOLCE	289	-0.046	C8orf84
26	0.093	SORD	114	-0.063	PPAP2B	202	0.052	RPL19	290	0.046	SLC1A5
27	-0.092	LTF	115	0.062	EZR	203	0.052	MCM4	291	0.046	MLPH
28	-0.092	STC2	116	0.062	CYC1	204	-0.052	NTN4	292	0.046	KIAA0182
29	0.092	C19orf33	117	0.062	BST2	205	-0.052	FOSB	293	-0.046	FCER1A
30	0.091	SCD	118	0.062	WWP1	206	0.052	LASS6	294	0.046	BZW2
31	0.089	ATP5E	119	-0.062	STC1	207	-0.052	EIF3G	295	-0.046	MAPT
32	-0.089	abParts	120	-0.062	JUN	208	0.052	CKMT1B	296	-0.046	GSN
33	-0.089	CFD	121	-0.062	SRPX	209	-0.051	COL6A1	297	0.046	TMED3
34	-0.087	SELM	122	0.062	RPS26	210	-0.051	TMEM14C	298	0.046	HMGA1
35	0.087	X64709	123	-0.061	SERPINE2	211	-0.051	CST3	299	0.046	ATP5H
36	0.087	TOP2A	124	0.061	TMEM97	212	-0.051	RPL3	300	-0.046	CSNK1E
37	0.086	MX1	125	0.061	PRC1	213	0.051	SLC38A1	301	0.046	CAMK2N1
38	-0.086	FCGBP	126	-0.061	TNC	214	-0.051	FRMD6	302	0.046	ERGIC1
39	-0.086	VTCN1	127	-0.061	CMTM7	215	0.051	SLC5A6	303	0.045	CR613620
40	-0.086	KRT17	128	-0.061	CITED4	216	-0.051	TNS3	304	-0.045	ENPP5
41	-0.086	GSTM2	129	-0.061	SEZ6L2	217	-0.051	C1S	305	0.045	GGCT
42	0.085	CALML5	130	-0.061	TSC22D1	218	-0.051	PLSCR3	306	-0.045	C17orf97
43	0.084	KRT8	131	0.061	IDH2	219	0.050	CANT1	307	0.045	CAPS

44	0.084	<i>NQO1</i>	132	-0.060	<i>HNRNPAIL2</i>	220	0.050	<i>PTPN1</i>	308	0.045	<i>KIAA1598</i>
45	-0.084	<i>SUSD3</i>	133	0.060	<i>38777</i>	221	-0.050	<i>SC5DL</i>	309	-0.045	<i>SERPINA1</i>
46	0.083	<i>TMBIM6</i>	134	-0.060	<i>S100A4</i>	222	-0.050	<i>ITM2B</i>	310	0.045	<i>RPS19</i>
47	-0.083	<i>EEF1G</i>	135	0.060	<i>VPS28</i>	223	0.050	<i>MYL6</i>	311	0.045	<i>SLC39A11</i>
48	-0.082	<i>DUSP1</i>	136	-0.060	<i>ZFP36</i>	224	-0.050	<i>IL17RB</i>	312	-0.045	<i>SAT1</i>
49	-0.082	<i>CXCL12</i>	137	-0.059	<i>CCDC130</i>	225	-0.050	<i>MT1E</i>	313	-0.045	<i>ACTB</i>
50	-0.082	<i>ANXA1</i>	138	-0.059	<i>SLC44A1</i>	226	-0.050	<i>CSDA</i>	314	-0.045	<i>NUCB1</i>
51	-0.081	<i>NFIX</i>	139	-0.058	<i>CD24</i>	227	0.050	<i>DNAJA4</i>	315	-0.045	<i>SEMA6A</i>
52	0.080	<i>BOLA2B</i>	140	0.058	<i>EIF3E</i>	228	-0.050	<i>TNFSF10</i>	316	-0.045	<i>CRISPLD2</i>
53	0.080	<i>HSPB1</i>	141	-0.058	<i>PRNP</i>	229	-0.049	<i>NDP</i>	317	0.045	<i>TMEM62</i>
54	-0.080	<i>FOXC1</i>	142	-0.058	<i>PLAT</i>	230	0.049	<i>C12orf44</i>	318	-0.045	<i>CYBRD1</i>
55	0.079	<i>FAM83H</i>	143	0.058	<i>MAL2</i>	231	0.049	<i>SERF2</i>	319	-0.045	<i>MT1G</i>
56	0.078	<i>C10orf116</i>	144	0.058	<i>DDIT4</i>	232	-0.049	<i>CDC42EP4</i>	320	0.045	<i>PTTG3</i>
57	0.078	<i>STAT1</i>	145	-0.058	<i>CGNL1</i>	233	-0.049	<i>CYP4Z1</i>	321	-0.045	<i>GADD45A</i>
58	0.078	<i>NUSAP1</i>	146	-0.058	<i>PDGFRL</i>	234	-0.049	<i>LOC389493</i>	322	-0.045	<i>RNASE1</i>
59	-0.078	<i>MYH11</i>	147	-0.058	<i>ITM2A</i>	235	-0.049	<i>ADM</i>	323	-0.045	<i>PYCARD</i>
60	0.077	<i>S100A16</i>	148	0.058	<i>ARL6IP1</i>	236	-0.049	<i>TMEM101</i>	324	-0.045	<i>LPIN1</i>
61	0.077	<i>PSMB3</i>	149	0.057	<i>KRT19</i>	237	-0.049	<i>HERPUD1</i>	325	-0.045	<i>PPIC</i>
62	0.076	<i>GINS2</i>	150	0.057	<i>CRIP1</i>	238	0.049	<i>DENND1B</i>	326	-0.045	<i>DQ893812</i>
63	-0.076	<i>COL4A5</i>	151	-0.057	<i>NINJ1</i>	239	0.049	<i>IFI44L</i>	327	0.045	<i>HMGB3</i>
64	0.076	<i>APOC1</i>	152	0.057	<i>TSPAN13</i>	240	0.049	<i>MRPL27</i>	328	0.044	<i>ZHX1</i>
65	-0.076	<i>CXCL14</i>	153	0.057	<i>CPNE3</i>	241	-0.049	<i>ALPL</i>	329	0.044	<i>NUDT1</i>
66	0.076	<i>KIAA0101</i>	154	-0.057	<i>ECHDC2</i>	242	-0.049	<i>WLS</i>	330	-0.044	<i>POLR3GL</i>
67	0.075	<i>NME1</i>	155	0.057	<i>CSE1L</i>	243	0.049	<i>CXCL10</i>	331	0.044	<i>TP53INP1</i>
68	0.075	<i>GRB7</i>	156	-0.057	<i>GLA</i>	244	-0.049	<i>ARMCX1</i>	332	0.044	<i>TUFM</i>
69	0.075	<i>ISG15</i>	157	-0.057	<i>SLC7A2</i>	245	-0.049	<i>KRT15</i>	333	-0.044	<i>CEBPD</i>
70	0.074	<i>AGR2</i>	158	0.056	<i>CUEDC1</i>	246	0.049	<i>LAPTM4B</i>	334	0.044	<i>IFI27</i>
71	-0.073	<i>HIST1H2AC</i>	159	-0.056	<i>EEF1B2</i>	247	0.049	<i>CLDN3</i>	335	-0.044	<i>SOX18</i>
72	-0.073	<i>CFB</i>	160	0.056	<i>PRDX1</i>	248	-0.049	<i>ZBTB20</i>	336	0.044	<i>RACGAP1</i>
73	0.073	<i>PTTG1</i>	161	0.056	<i>BCAS4</i>	249	0.048	<i>COPS5</i>	337	0.044	<i>ST3GAL1</i>
74	-0.073	<i>FAM129A</i>	162	-0.056	<i>TUBB2B</i>	250	-0.048	<i>DNAJC12</i>	338	0.044	<i>H2AFX</i>
75	-0.072	<i>DCN</i>	163	-0.056	<i>CSTB</i>	251	-0.048	<i>ID3</i>	339	0.044	<i>PTK2</i>
76	0.072	<i>GPR172A</i>	164	0.055	<i>ATP5EP2</i>	252	-0.048	<i>UBE2E3</i>	340	-0.044	<i>SNX3</i>
77	0.072	<i>ATP9A</i>	165	-0.055	<i>PGM1</i>	253	0.048	<i>PITX1</i>	341	-0.044	<i>CCDC92</i>
78	0.072	<i>CLEC3A</i>	166	0.055	<i>GSDMB</i>	254	0.048	<i>GAPDH</i>	342	-0.044	<i>AK001020</i>
79	0.071	<i>KRT18</i>	167	-0.055	<i>FSCN1</i>	255	-0.048	<i>HLA-B</i>	343	0.044	<i>FAM110A</i>
80	-0.071	<i>GAS1</i>	168	-0.055	<i>SGCE</i>	256	-0.048	<i>SDCBP</i>	344	-0.044	<i>SCGB1D2</i>
81	0.071	<i>APOE</i>	169	0.055	<i>STK3</i>	257	0.048	<i>LRRC26</i>	345	0.044	<i>IFIT1</i>
82	-0.071	<i>TPM1</i>	170	0.055	<i>TM7SF2</i>	258	-0.048	<i>TNFRSF14</i>	346	-0.044	<i>FKBP9L</i>
83	-0.070	<i>MFAP4</i>	171	0.055	<i>EIF2C2</i>	259	-0.048	<i>CRTAP</i>	347	-0.044	<i>HLA-DQA1</i>
84	0.070	<i>UCP2</i>	172	0.055	<i>PABPC1</i>	260	-0.048	<i>C8orf4</i>	348	-0.043	<i>TPM2</i>
85	-0.070	<i>SPP1</i>	173	-0.054	<i>FGD3</i>	261	-0.048	<i>NOSTRIN</i>	349	0.043	<i>CDS1</i>
86	-0.070	<i>FBLN1</i>	174	-0.054	<i>RBBP8</i>	262	0.047	<i>GNAS</i>	350	-0.043	<i>CLIP3</i>
87	0.069	<i>CDC20</i>	175	0.054	<i>NCOA3</i>	263	-0.047	<i>UBD</i>			
88	0.069	<i>C8orf55</i>	176	-0.054	<i>PBX3</i>	264	-0.047	<i>FAM127A</i>			

10.6 Genes full names list of the 99 genes significantly associated with lymphovascular invasion.

Gene symbol	Gene name	Gene symbol	Gene name
ACTG2	actin gamma 2	KRT7	keratin 7
ANG	angiogenin	KRT8	keratin 8
ANXA1	annexin A1	LAPTM4B	lysosomal protein transmembrane 4 beta
APOC1	apolipoprotein C1	LRRC26	leucine rich repeat containing 26
APOE	apolipoprotein E	LY6E	lymphocyte antigen 6 family member E
C1S	complement C1s	MAOA	monoamine oxidase A
CALML5	calmodulin-like 5	MFAP4	microfibrillar-associated protein 4
CCNB2	cyclin B2	MGP	matrix Gla protein
CDC42EP4	CDC42 effector protein 4	MMP11	matrix metalloproteinase 11
CDCA5	cell division cycle associated 5	MT1E	metallothionein 1E
CEBPD	CCAAT/enhancer-binding protein delta	MX1	MX dynamin like GTPase 1
CFB	complement factor B	NDP	NDP, norrin cystine knot growth factor
CFD	complement factor D	NINJ1	ninjurin 1
CLIC6	chloride intracellular channel 6	NME1	NME/NM23 nucleoside diphosphate kinase 1
COX6C	cytochrome c oxidase subunit 6C	NOP56	NOP56 ribonucleoprotein
CXCL12	C-X-C motif chemokine ligand 12	PDGFRL	platelet derived growth factor receptor like
CXCL14	C-X-C motif chemokine ligand 14	PGAP3	post-GPI attachment to proteins 3
CYBRD1	cytochrome b reductase 1	PITX1	paired like homeodomain 1
CYP4X1	cytochrome P450 family 4 subfamily X member 1	PLGRKT	plasminogen receptor with a C-terminal lysine
DCN	decorin	PTTG1	pituitary tumor-transforming 1
DKK3	dickkopf WNT signaling pathway inhibitor 3	PYCARD	PYD and CARD domain containing
DNAJA4	DnaJ heat shock protein family (Hsp40) member A4	RPL3	ribosomal protein L3
DPYSL2	dihydropyrimidinase-like 2	S100A4	S100 calcium binding protein A4
DUSP1	dual specificity phosphatase 1	S100P	S100 calcium binding protein P
EEF1A2	eukaryotic translation elongation factor 1 alpha 2	SCD	stearoyl-CoA desaturase
EEF1B2	eukaryotic translation elongation factor 1 beta 2	SELENOM	selenoprotein M
ELF3	E74 like ETS transcription factor 3	SERPINA3	serpin family A member 3
ERBB2	erb-b2 receptor tyrosine kinase 2	SERPINE2	serpin family E member 2
FBLN1	fibulin-1	SGCE	sarcoglycan epsilon
FCER1A	Fc fragment of IgE receptor 1a	SLC40A1	solute carrier family 40 member 1
FCGBP	Fc fragment of IgG binding protein	SLC44A1	solute carrier family 44 member 1
FGD3	FYVE, RhoGEF and PH domain containing 3	SLC52A2	solute carrier family 52 member 2
FOS	Fos proto-oncogene, AP-1 transcription factor subunit	SLC9A3R1	SLC9A3 regulator 1
FST	follistatin	SPDEF	SAM pointed domain containing ETS transcription factor
GAS1	growth arrest specific 1	SRPX	sushi repeat containing protein, X-linked
GNAS	GNAS complex locus	STC2	stanniocalcin 2
GSTP1	glutathione S-transferase pi 1	SUSD3	sushi domain containing 3
HBA2	hemoglobin subunit alpha 2	TM7SF2	transmembrane 7 superfamily member 2
HBB	hemoglobin subunit beta	TNS3	tensin 3
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	TPM2	tropomyosin 2 (beta)
HMGA1	high mobility group AT-hook 1	TXNIP	thioredoxin interacting protein
HMGB3	high mobility group box 3	UBD	ubiquitin D
HSPB1	heat shock protein family B (small) member 1	UBE2C	ubiquitin conjugating enzyme E2 C
IDH2	isocitrate dehydrogenase (NADP(+)) 2, mitochondrial	UBE2S	ubiquitin conjugating enzyme E2 S
IFI27	interferon alpha inducible protein 27	UCP2	uncoupling protein 2

<i>IL17RB</i>	interleukin 17 receptor B	<i>VIM</i>	vimentin
<i>ISG15</i>	ISG15 ubiquitin-like modifier	<i>VTCN1</i>	V-set domain containing T-cell activation inhibitor 1
<i>KRT18</i>	keratin 18	<i>YWHAZ</i>	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta
<i>KRT18P55</i>	keratin 18 pseudogene 55	<i>ZBTB20</i>	zinc finger and BTB domain containing 20
<i>KRT19</i>	keratin 19		

10.7 Mean value, SEM, subtraction and WAD ranking in the 99 genes significantly associated with lymphovascular invasion.

Upregulated genes												
	Nottingham cases						Remaining METABRIC cases					
LVI	Positive		Negative		Subtraction	WAD ranking	Positive		Negative		Subtraction	WAD ranking
Genes	Mean	SEM	Mean	SEM			Mean	SEM	Mean	SEM		
<i>APOC1</i>	10.28	0.91	10.00	0.98	0.28	25	9.98	0.99	9.84	1.04	0.14	64
<i>APOE</i>	11.73	0.67	11.53	0.74	0.20	31	11.64	0.82	11.54	0.82	0.10	81
<i>CALML5</i>	7.56	2.05	7.04	1.88	0.52	61	7.50	2.16	7.16	1.93	0.34	42
<i>CCNB2</i>	8.25	0.92	8.04	0.93	0.21	326	8.07	0.90	7.90	1.01	0.16	186
<i>CDCA5</i>	8.50	0.96	8.28	0.98	0.22	232	8.48	0.95	8.31	1.04	0.17	100
<i>COX6C</i>	12.98	0.63	12.89	0.65	0.09	263	12.89	0.68	12.83	0.67	0.05	277
<i>DNAJA4</i>	9.17	0.75	8.98	0.81	0.20	205	9.20	0.82	9.09	0.81	0.11	227
<i>EEF1A2</i>	9.08	2.00	8.46	1.92	0.63	3	9.05	2.09	8.90	2.04	0.15	105
<i>ELF3</i>	8.69	0.73	8.49	0.86	0.20	274	8.92	0.79	8.81	0.82	0.11	287
<i>ERBB2</i>	10.83	1.59	10.63	1.33	0.20	65	10.92	1.46	10.62	1.22	0.30	1
<i>GNAS</i>	12.75	0.45	12.62	0.41	0.13	101	12.93	0.53	12.87	0.48	0.05	262
<i>HMGA1</i>	8.48	0.64	8.28	0.77	0.20	303	8.50	0.75	8.38	0.77	0.12	298
<i>HMGB3</i>	7.72	0.89	7.38	0.83	0.34	166	7.64	0.88	7.48	0.91	0.16	327
<i>HSPB1</i>	12.26	0.73	12.07	0.74	0.19	32	12.21	0.79	12.11	0.84	0.10	53
<i>IDH2</i>	9.63	0.88	9.45	0.72	0.18	179	9.63	0.83	9.51	0.86	0.12	131
<i>IFI27</i>	11.95	1.30	11.73	1.15	0.21	24	11.63	1.40	11.57	1.35	0.06	334
<i>ISG15</i>	9.74	1.32	9.54	1.35	0.20	140	9.75	1.35	9.61	1.36	0.14	69
<i>KRT18</i>	11.70	0.96	11.51	1.07	0.20	36	11.84	1.05	11.75	1.07	0.09	79
<i>KRT18P55</i>	10.35	0.97	10.16	1.02	0.18	120	10.17	1.08	9.99	1.10	0.18	18
<i>KRT19</i>	12.51	1.14	12.37	1.33	0.14	88	12.58	1.22	12.51	1.26	0.07	149
<i>KRT7</i>	9.45	1.44	9.27	1.35	0.18	223	9.36	1.53	9.25	1.46	0.11	187
<i>KRT8</i>	10.16	0.92	9.96	0.99	0.21	90	10.46	0.99	10.32	1.01	0.14	43
<i>LAPTM4B</i>	10.43	1.06	10.23	0.92	0.20	85	10.18	1.21	10.09	1.13	0.09	246
<i>LRRC26</i>	9.92	1.63	9.72	1.51	0.20	125	9.92	1.55	9.83	1.53	0.09	257
<i>LY6E</i>	10.45	1.00	10.25	0.95	0.20	72	10.62	1.03	10.45	1.00	0.17	14
<i>MMP11</i>	10.53	1.39	10.38	1.50	0.15	184	10.47	1.36	10.27	1.52	0.19	11
<i>MX1</i>	11.07	1.30	10.75	1.30	0.32	9	11.17	1.32	11.05	1.29	0.13	37
<i>NME1</i>	11.55	0.73	11.43	0.71	0.12	217	11.31	0.72	11.20	0.67	0.11	67
<i>NOP56</i>	9.75	0.47	9.59	0.48	0.16	242	9.95	0.56	9.87	0.58	0.09	286
<i>PGAP3</i>	8.91	1.39	8.69	1.15	0.21	199	8.81	1.26	8.56	0.95	0.25	20
<i>PITX1</i>	9.29	1.55	8.84	1.60	0.45	11	9.34	1.59	9.23	1.61	0.10	253
<i>PTTG1</i>	9.29	0.87	9.12	0.94	0.17	257	9.10	0.91	8.93	1.01	0.16	73
<i>S100P</i>	9.70	2.32	9.26	2.31	0.45	5	9.52	2.34	9.22	2.24	0.30	3
<i>SCD</i>	10.88	0.97	10.75	0.92	0.14	200	10.92	1.11	10.78	1.03	0.14	30
<i>SLC52A2</i>	9.18	0.61	9.02	0.63	0.16	309	9.29	0.72	9.13	0.67	0.16	76
<i>SLC9A3R1</i>	10.77	1.04	10.59	1.01	0.18	108	10.97	0.98	10.87	1.02	0.10	109

<i>SPDEF</i>	9.48	1.39	9.34	1.46	0.14	346	9.74	1.34	9.55	1.45	0.19	23
<i>TM7SF2</i>	8.70	0.93	8.44	0.88	0.26	132	8.70	0.95	8.56	0.91	0.14	170
<i>UBE2C</i>	9.27	1.09	9.03	1.16	0.24	106	9.25	1.17	8.99	1.31	0.25	10
<i>UBE2S</i>	9.29	0.71	9.02	0.73	0.27	70	9.35	0.84	9.21	0.85	0.14	94
<i>UCP2</i>	8.93	0.91	8.71	0.90	0.22	189	9.10	0.94	8.94	0.92	0.16	84
<i>YWHAZ</i>	12.00	0.59	11.84	0.58	0.15	79	12.08	0.63	11.95	0.62	0.13	17
Downregulated genes												
	Nottingham cases						Remaining METABRIC cases					
LVI	Positive		Negative		Subtraction	WAD ranking	Positive		Negative		Subtraction	WAD ranking
Genes	Mean	SEM	Mean	SEM			Mean	SEM	Mean	SEM		
<i>ACTG2</i>	8.75	2.50	9.01	1.81	-0.26	100	8.48	1.62	8.78	1.63	-0.30	7
<i>ANG</i>	8.19	0.94	8.44	1.01	-0.25	186	8.14	1.07	8.29	1.14	-0.15	179
<i>ANXA1</i>	10.91	0.68	11.08	0.71	-0.17	95	10.45	0.91	10.58	0.99	-0.13	50
<i>C1S</i>	10.11	0.93	10.34	0.88	-0.24	53	9.67	1.02	9.77	1.10	-0.10	217
<i>CDC42EP4</i>	10.22	0.38	10.37	0.62	-0.15	197	10.40	0.66	10.48	0.65	-0.08	232
<i>CEBPD</i>	10.09	0.53	10.21	0.70	-0.13	316	10.12	0.84	10.20	0.81	-0.08	333
<i>CFB</i>	10.10	2.45	10.51	1.48	-0.41	4	10.42	1.70	10.54	1.64	-0.12	72
<i>CFD</i>	9.48	1.39	9.87	1.38	-0.40	10	9.24	1.29	9.42	1.36	-0.19	33
<i>CLIC6</i>	8.17	4.43	8.54	2.17	-0.37	50	8.15	2.21	8.43	2.24	-0.28	19
<i>CXCL12</i>	9.44	1.16	9.72	1.00	-0.28	45	9.05	1.10	9.23	1.20	-0.18	49
<i>CXCL14</i>	8.31	2.30	8.67	1.52	-0.36	49	8.18	1.57	8.39	1.61	-0.21	65
<i>CYBRD1</i>	9.73	1.03	9.91	1.03	-0.18	162	9.68	1.15	9.76	1.19	-0.09	318
<i>CYP4X1</i>	8.44	3.82	8.77	1.89	-0.32	64	8.65	1.90	8.88	1.94	-0.23	24
<i>DCN</i>	9.07	1.34	9.23	1.24	-0.16	325	8.46	1.33	8.64	1.43	-0.19	75
<i>DDK3</i>	9.45	0.94	9.72	0.88	-0.27	54	9.07	0.91	9.22	0.93	-0.15	90
<i>DPYSL2</i>	9.82	0.43	9.98	0.60	-0.16	214	9.73	0.68	9.85	0.76	-0.12	107
<i>DUSP1</i>	10.32	0.90	10.44	0.96	-0.12	348	9.89	1.40	10.04	1.45	-0.15	48
<i>EEF1B2</i>	11.20	0.34	11.33	0.52	-0.12	219	10.93	0.78	11.01	0.80	-0.08	159
<i>FBLN1</i>	10.59	1.04	10.86	0.93	-0.27	17	10.51	1.04	10.63	1.12	-0.11	86
<i>FCER1A</i>	7.41	1.20	7.76	1.27	-0.36	144	6.95	1.07	7.15	1.22	-0.21	293
<i>FCGBP</i>	8.72	2.50	9.11	1.64	-0.39	19	8.76	1.61	8.96	1.61	-0.20	38
<i>FGD3</i>	8.81	1.30	9.19	1.11	-0.38	21	9.19	1.20	9.30	1.20	-0.12	173
<i>FOS</i>	10.12	1.85	10.24	1.37	-0.12	349	9.53	1.66	9.74	1.69	-0.21	13
<i>FST</i>	8.22	1.20	8.60	1.10	-0.38	41	8.04	1.02	8.24	1.03	-0.20	89
<i>GAS1</i>	8.92	1.05	9.11	0.92	-0.19	227	8.45	1.06	8.63	1.11	-0.18	80
<i>GSTP1</i>	10.76	1.21	10.99	0.93	-0.23	33	10.61	1.20	10.80	1.10	-0.19	8
<i>HBA2</i>	9.40	2.26	9.55	1.48	-0.15	308	9.03	1.52	9.26	1.57	-0.23	16
<i>HBB</i>	9.34	2.23	9.59	1.47	-0.24	73	8.62	1.62	8.91	1.69	-0.29	9
<i>HLA-DQA1</i>	10.37	1.03	10.52	0.99	-0.15	188	10.04	1.30	10.12	1.31	-0.08	347
<i>IL17RB</i>	7.50	1.18	7.75	1.17	-0.25	340	7.56	1.06	7.73	1.04	-0.17	224
<i>MAOA</i>	7.42	1.62	7.83	1.25	-0.41	84	7.48	1.33	7.67	1.37	-0.19	178
<i>MFAP4</i>	8.48	1.52	8.73	1.26	-0.25	141	8.31	1.17	8.50	1.31	-0.19	83
<i>MGP</i>	12.95	1.44	13.28	1.10	-0.33	2	12.73	1.34	12.87	1.43	-0.14	6

<i>MT1E</i>	9.78	1.26	10.10	1.10	-0.32	15	9.75	1.23	9.84	1.19	-0.09	225
<i>NDP</i>	6.81	2.19	7.26	1.53	-0.45	152	6.91	1.65	7.14	1.68	-0.23	229
<i>NINJ1</i>	10.21	0.29	10.33	0.53	-0.12	314	10.40	0.55	10.49	0.53	-0.09	151
<i>PDGFRL</i>	8.95	1.04	9.23	0.95	-0.28	67	8.51	1.01	8.65	1.09	-0.15	146
<i>PLGRKT</i>	9.88	0.38	10.05	0.63	-0.18	160	9.62	0.78	9.72	0.78	-0.10	182
<i>PYCARD</i>	9.90	0.83	10.10	0.88	-0.19	122	10.05	0.94	10.13	0.94	-0.08	323
<i>RPL3</i>	12.76	0.29	12.89	0.46	-0.13	112	12.70	0.53	12.76	0.53	-0.06	212
<i>S100A4</i>	10.87	0.71	11.00	0.76	-0.13	209	10.46	0.86	10.55	0.90	-0.10	134
<i>SELENOM</i>	10.09	0.49	10.33	0.69	-0.23	55	10.20	0.70	10.35	0.66	-0.15	34
<i>SERPINA3</i>	12.09	3.18	12.25	1.79	-0.16	63	12.05	1.79	12.27	1.68	-0.21	2
<i>SERPINE2</i>	9.97	0.89	10.28	0.96	-0.31	16	9.82	1.03	9.93	1.08	-0.11	123
<i>SGCE</i>	8.87	0.97	9.09	0.89	-0.21	176	8.49	1.10	8.63	1.14	-0.14	168
<i>SLC40A1</i>	9.83	1.18	10.07	1.29	-0.24	59	9.71	1.32	9.83	1.40	-0.12	113
<i>SLC44A1</i>	11.03	0.25	11.24	0.50	-0.21	46	10.93	0.57	11.01	0.55	-0.09	138
<i>SRPX</i>	8.24	1.02	8.43	0.93	-0.20	328	7.84	1.00	8.04	1.14	-0.19	121
<i>STC2</i>	9.26	3.41	9.70	1.94	-0.44	6	9.73	1.96	9.90	1.93	-0.17	28
<i>SUSD3</i>	8.46	2.30	8.99	1.55	-0.53	7	8.67	1.57	8.87	1.57	-0.20	45
<i>TNS3</i>	9.84	0.32	10.04	0.50	-0.19	129	9.98	0.62	10.07	0.58	-0.09	216
<i>TPM2</i>	10.48	0.74	10.61	0.75	-0.13	275	10.32	0.78	10.40	0.82	-0.07	348
<i>TXNIP</i>	10.16	0.35	10.29	0.63	-0.14	259	9.92	0.72	10.00	0.76	-0.09	269
<i>UBD</i>	8.16	2.50	8.55	1.66	-0.39	40	7.98	1.56	8.12	1.61	-0.14	263
<i>VIM</i>	12.25	0.39	12.41	0.58	-0.16	57	12.05	0.77	12.13	0.83	-0.08	103
<i>VTCN1</i>	9.12	3.46	9.34	2.00	-0.22	134	9.08	2.01	9.26	1.96	-0.19	39
<i>ZBTB20</i>	8.79	0.51	8.96	0.63	-0.17	300	8.92	0.68	9.02	0.72	-0.10	342