

University of Nottingham
UK | CHINA | MALAYSIA

Exploring Molecular Portraits Defining Lymphovascular Invasion in Invasive Breast Cancer Patients

Sultan Naeem M Sonbul
B.Sc., M.Sc.

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

December 2018

ABSTRACT

BACKGROUND: Lymphovascular invasion (LVI) is a robust clinicopathological parameter of poor prognosis in invasive breast cancer (BC). The accurate detection of tumour emboli within lymphovascular cavities in invasive BC can predict the unfavourable outcome for BC patients, including distant metastasis, recurrence, and short survival. The molecular drivers of LVI in invasive BC patients remain poorly defined. We sought to determine if specific transcriptomic and proteomic profiles regulate the LVI mechanisms in invasive BC.

METHODOLOGIES: Meticulous detection of LVI occurrence was performed with histological and immunohistochemical staining on invasive BC full-face tissue sections to construct the LVI-supervised (LVI+/-) cohort, (n=170). All LVI- patients in this cohort were lymph node negative (LN-). The extraction of total RNA from the LVI+/- specimens followed by RNA microarray experiment and data analysis, utilising the linear models for microarray data (LIMMA) package for differential expression, yielded a large cluster of differentially expressed genes (DEG) associated with LVI, (n=1154). The last DEGs were validated against an external cohort of invasive BC patients under the same experimental conditions. We aimed to focus and prioritise our selection of LVI candidates through the statistical assessment of copy number aberrations (CNA) focusing on the top DEGs, (n=100), associated significantly with LVI according to the LIMMA analysis. CNA gain ratios were also calculated for each DEG in the LVI+ and LVI- categories. To avoid the dismissing of crucial LVI-drivers, an exhaustive and investigative approach was also applied and included all of the 1154 DEGs to identify the upregulated gene signatures associated with the LVI+ and LVI- cases in the LVI+/- cohort. Clinicopathological relevance and LVI associations were assessed for the identified LVI-DEGs in the METABRIC cohort, (n=1980), a subset of METABRIC cohort, including LVI-/LN+/- cases, (n=296), and the LVI+/- cohort, (n=170). Fewer LVI-candidates were selected for protein studies. Immunohistochemistry (IHC) protocols were performed to assess the altered protein expression for the qualified LVI-DEGs on tissue microarray (TMA) sections derived from the Nottingham primary series (NPS) of invasive BC, (n=2497). The following experiment aimed to identify the LVI differentially expressed proteins (DEP) utilising high-performance liquid chromatography/mass spectrometry (HPLC/MS) as a high-throughput screening technique. An LVI-supervised cohort consisted of invasive BC luminal-A phenotype was constructed, (n=80). The LVI status for these patients was defined histologically and immunohistochemically, and it comprised LVI+, (n=41), and LVI-, (n=39) cases. These specimens were used for

total protein extraction, digestion, purification, and injected in Eksigent ekspert nano LC 425 system, then, analysed on a Sciex TripleTOF 6600 mass spectrometer.

RESULTS: The CNA/CNA gain ratios analyses revealed a promising cluster of LVI-DEGs, (n=12). These LVI-DEGs are: *MCFD2* (p-value $\leq$ 0.02, CNA-gain = 4.56); *CCT8* (p-value $\leq$ 0.03, CNA-gain = 2.73); *CUL4A* (p-value $\leq$ 0.002, CNA-gain = 1.59); *ASCC3* (p-value $\leq$ 0.01, CNA-gain = 1.36); *DDX59* (p-value $\leq$ 0.0001, CNA-gain = 1.11); *ARFGEF1* (p-value $\leq$ 1.0E-9, CNA-gain = 1.07); *CD46* (p-value $\leq$ 0.01, CNA-gain = 1.03); *HRSP12* (p-value $\leq$ 0.00008, CNA-gain = 0.91); *MED23* (p-value $\leq$ 0.008, CNA-gain = 0.91); *EDEM* (p-value $\leq$ 0.01, CNA-gain = 0.86); *FAM103A1* (p-value $\leq$ 0.007, CNA-gain = 0.30); and *PPP2CB* (p-value $\leq$ 0.0007, CNA-gain = 0.18). Exploratory bioinformatics search identified that *SNPA23* is a co-expressed gene with *MCFD2* and *CCT8*, the top LVI-candidates in the CNA analysis. The associations between MCFD2, CCT8, and SNAP23 against LVI were interrogated on the protein level with IHC.

The exhaustive transcriptomic/LVI analyses has also discovered two unique clusters of LVI-DEGs. The first cluster included upregulated transcripts in the LVI^+ category, (UP-13/LVI^+). These transcripts are: *PPM1G* (p-value $\leq$ 0.0001), *SEC14L1* (p-value $\leq$ 0.005), *P4HA2* (p-value $\leq$ 0.007), *CTSD* (p-value $\leq$ 0.009), *ITPRIP* (p-value $\leq$ 0.016), *ZDHHC9* (p-value $\leq$ 0.031), *PILRA* (p-value $\leq$ 0.031), *PDLIM7* (p-value $\leq$ 0.032), *RNASE1* (p-value $\leq$ 0.032), *SLC3A2* (p-value $\leq$ 0.032), *IKZF3* (p-value $\leq$ 0.036), *SIGLEC1* (p-value $\leq$ 0.039), and *SLC6A9* (p-value $\leq$ 0.047). The second cluster included upregulated transcripts in the LVI^- category, (UP-22/LVI^-). These transcripts are: *LPAR1* (p-value $\leq$ 0.0002), *CXCL12* (p-value $\leq$ 0.001), *ELP5* (p-value $\leq$ 0.002), *ECM2* (p-value $\leq$ 0.004), *APCDD1* (p-value $\leq$ 0.005), *FBLN1* (p-value $\leq$ 0.005), *ANXA5* (p-value $\leq$ 0.008), *INKA1* (p-value $\leq$ 0.011), *TMEM98* (p-value $\leq$ 0.012), *FEZ1* (p-value $\leq$ 0.016), *THAP4* (p-value $\leq$ 0.016), *TNFRSF19* (p-value $\leq$ 0.016), *NBR1* (p-value $\leq$ 0.018), *WISP2* (p-value $\leq$ 0.019), *CNRIP* (p-value $\leq$ 0.020), *ITSN1* (p-value $\leq$ 0.021), *C1S* (p-value $\leq$ 0.022), *TNC* (p-value $\leq$ 0.028), *PRRX1* (p-value $\leq$ 0.030), *STAT3* (p-value $\leq$ 0.035), *TSC1* (p-value $\leq$ 0.036), and *TWIST2* (p-value $\leq$ 0.047). The association between SEC14L1 and LVI was investigated on the protein level with IHC.

The cytosolic protein expressions of MCFD2 and SEC14L1, the nuclear protein expression of CCT8, and the membranous protein expression of SNAP23 illustrated strong associations with LVI in invasive BC. Cytosolic MCFD2 and nuclear CCT8 exhibited qualities of LVI inhibitors, (p-value $\leq$ 0.003) and (p-value $\leq$ 0.010)

respectively. On the other hand, cytosolic SEC14L1 and membranous SNAP23 displayed strong associations with actual LVI occurrences, (p-value $\leq$ 0.0003) and (p-value $\leq$ 0.020) correspondingly.

The HPLC/MS differential protein expression identified 370 DEP associated with LVI. Among these DEPs, we identified 56 DEP overlapped with the identifiers of DEGs produced by the LVI/RNA microarray experiment. These DEPs are: ACTB, ACTG1, ANP32A, ARHGDIA, ARPC5, CCT8, COL6A2, COMT, DBI, DDX17, DYNLRB1, EIF4E, ERH, FBLN1, H2AFY, HMGB2, HNRNPA2B1, HNRNPH3, HNRNPK, HSP90AA1, HSPA4, ILF3, KCTD12, LGALS3, LMNA, OLA1, PGAM1, PPIA, PRRC1, PSMA6, PTGES3, PTMA, RAC1, RBBP4, RPL10A, RPL7, RPL9, RPLP1, RPS11, RPS18, RPS2, RPS28, RPS4X, RPS7, RPSA, SAR1A, SEC61B, SNRPE, SRSF3, SUB1, TPI1, TUBB6, UFM1, UPF1, VDAC2, and YWHAE. As the flow of LVI-triggering genetic codes should be transcripted first into mRNA then translated into functional proteins, we utilised the METABRIC transcriptomic data and applied Mann-Whitney test for median comparison to confirm the correlations between LVI and the 56 LVI-DEP on the mRNA level. We identified a cluster of DEGs/DEPs according to the RNA microarray and HPLC/MS experiments. These LVI-DEGs/DEPs are: *HSP90AA1* (p-value $\leq$ 0.0001), *ARHGDIA* (p-value $\leq$ 0.004), *FBLN1* (p-value $\leq$ 0.005), *DYNLRB1* (p-value $\leq$ 0.018), *VDAC2* (p-value $\leq$ 0.030), *RPS18* (p-value $\leq$ 0.037), *RPL10A* (p-value $\leq$ 0.046), and *PPIA* (p-value $\leq$ 0.047).

CONCLUSION: To conclude, the LVI initiation or repression in invasive BC may be attributed to altered expression of a cluster of genes and proteins. On the transcriptomic level, the LVI definite incidences were associated with the UP-13/LVI^{+} cluster of mRNAs that were preferentially upregulated in the luminal-B, HER2-enriched, and basal-like intrinsic subtypes. The proteins encoded by these mRNAs are involved in several pathways including intracellular protein trafficking, adaptive immune system, signalling by nuclear receptors, transmembrane glycoprotein interactions, extracellular matrix organisation and degradation, Na^{+}/Cl^{-} dependent neurotransmitter transporters, SLC-mediated transmembrane transport. In contrast, the absence of LVI was associated with a different cluster of mRNAs which were overexpressed in the vast majority of normal-like and luminal-A tumours. This explorative study of LVI provided novel LVI-signature in invasive BC with relevant prognostic utilities. In the future, elaborated and well-designed functional studies may reveal better insights on the contributions for each identified LVI-candidate in driving or inhibiting LVI mechanisms in invasive BC patients.

PUBLICATIONS ARISING FROM THIS THESIS

SONBUL, S. N., ALESKANDARANY, M. A., KUROZUMI, S., JOSEPH, C., TOSS, M. S., DIEZ-RODRIGUEZ, M., NOLAN, C. C., MUKHERJEE, A., MARTIN, S., CALDAS, C., ELLIS, I. O., GREEN, A. R. & RAKHA, E. A. 2018. Saccharomyces cerevisiae-like 1 (SEC14L1) is a prognostic factor in breast cancer associated with lymphovascular invasion. Mod Pathol.

KUROZUMI, S., JOSEPH, C., **SONBUL, S.**, ALESKANDARANY, M. A., PIGERA, M., ALSALEEM, M., ALSAEED, S., KARIRI, Y., NOLAN, C. C., DIEZ-RODRIGUEZ, M., JOHNSTON, S., MONGAN, N. P., FUJII, T., SHIRABE, K., MARTIN, S. G., ELLIS, I. O., GREEN, A. R. & RAKHA, E. A. 2018. Clinicopathological and prognostic significance of Ras association and pleckstrin homology domains 1 (RAPH1) in breast cancer. Breast Cancer Res Treat.

JOSEPH, C., MACNAMARA, O., CRAZE, M., RUSSELL, R., PROVENZANO, E., NOLAN, C. C., DIEZ-RODRIGUEZ, M., **SONBUL, S. N**., ALESKANDARANY, M. A., GREEN, A. R., RAKHA, E. A., ELLIS, I. O. & MUKHERJEE, A. 2018. Mediator complex (MED) 7: a biomarker associated with good prognosis in invasive breast cancer, especially ER+ luminal subtypes. Br J Cancer, 118, 1142-1151.

KUROZUMI, S., JOSEPH, C., **SONBUL, S.**, GORRINGE, K. L., PIGERA, M., ALESKANDARANY, M. A., DIEZ-RODRIGUEZ, M., NOLAN, C. C., FUJII, T., SHIRABE, K., KUWANO, H., STORR, S., MARTIN, S. G., ELLIS, I. O., GREEN, A. R. & RAKHA, E. A. 2018. Clinical and biological roles of Kelch-like family member 7 in breast cancer: a marker of poor prognosis. Breast Cancer Res Treat, 170, 525-533.

SONBUL, S. N., GORRINGE, K. L., ALESKANDARANY, M. A., MUKHERJEE, A., GREEN, A. R., ELLIS, I. O. & RAKHA, E. A. 2017. Chemokine (C-C motif) receptor 7 (CCR7) associates with the tumour immune microenvironment but not progression in invasive breast carcinoma. J Pathol Clin Res, 3, 105-114.

ALESKANDARANY, M. A., **SONBUL, S.**, SURRIDGE, R., MUKHERJEE, A., CALDAS, C., DIEZ-RODRIGUEZ, M., ASHANKYTY, I., ALBRAHIM, K. I., ELMOUNA, A. M., ANEJA, R., MARTIN, S. G., ELLIS, I. O., GREEN, A. R. & RAKHA, E. A. 2017. Rho-GTPase activating-protein 18: a biomarker associated with good prognosis in invasive breast cancer. Br J Cancer, 117, 1176-1184.

ALESKANDARANY, M. A., **SONBUL, S. N.**, MUKHERJEE, A. & RAKHA, E. A. 2015. Molecular Mechanisms Underlying Lymphovascular Invasion in Invasive Breast Cancer. Pathobiology, 82, 113-23.

KUROZUMI, S., JOSEPH, C., RAAFAT, S., **SONBUL, S.**, KARIRI, Y., ALSAEED, S., PIGERA, M., ALSALEEM, M., NOLAN, C. C., JOHNSTON, S. J., ALESKANDARANY, M. A., OGDEN, A., FUJII, T., SHIRABE, K., MARTIN, S. G., ALSHANKYTY, I., MONGAN, N. P., ELLIS, I. O., GREEN, A. R. & RAKHA, E. A. 2019. Utility of ankyrin 3 as a prognostic marker in androgen-receptor-positive breast cancer. Breast Cancer Research and Treatment. (**A post-viva publication**)

OTHER PUBLICATIONS

MUFTAH, A. A., ALESKANDARANY, M., **SONBUL, S. N.,** NOLAN, C. C., DIEZ RODRIGUEZ, M., CALDAS, C., ELLIS, I. O., GREEN, A. R. & RAKHA, E. A. 2017a. Further evidence to support bimodality of oestrogen receptor expression in breast cancer. Histopathology, 70, 456-465.

MUFTAH, A. A., ALESKANDARANY, M. A., AL-KAABI, M. M., **SONBUL, S. N.,** DIEZ-RODRIGUEZ, M., NOLAN, C. C., CALDAS, C., ELLIS, I. O., RAKHA, E. A. & GREEN, A. R. 2017b. Ki67 expression in invasive breast cancer: the use of tissue microarrays compared with whole tissue sections. Breast Cancer Res Treat, 164, 341-348.

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

CONFERENCE PRESENTATIONS

As the main presenter:

Oral presentation "Exploring Molecular Mechanisms Underlying Lymphovascular Invasion in Breast Cancer." 8th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, June 23-25, 2015, Dublin, Republic of Ireland.

Oral presentation "Vesicular Transporter Proteins in Invasive Breast Cancer: Clinicopathological Relevance of Synaptosomal-Associated Protein-23 (SNAP23)." 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, June 28 – July 1, 2016, Nottingham, United Kingdom.

Oral presentation "Multiple Coagulation Factor Deficiency Protein-2 (MCFD2) as a Potential Prognostic Germ Cell Marker in Invasive Breast Carcinoma: a Transcriptome-Driven Study." 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, June 28 – July 1, 2016, Nottingham, United Kingdom.

Poster presentation "Clinicopathological Impact of C-C Chemokine Receptor Type 7 (CCR7) Protein Expression in Triple Negative Breast Carcinoma (TNBC)." 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, June 28 – July 1, 2016, Nottingham, United Kingdom.

Poster presentation "Lympho-Vascular Invasion (LVI) Supervised Transcriptomic Study Disclosed Clusters of Genes with Significant Prognostic Importance in Invasive Breast Carcinoma." 207th Scientific Meeting of the Pathological Society of Great Britain & Ireland, January 19-20, 2017, London, United Kingdom.

As a co-author:

Oral presentation "Ankyrin 3 Protein may Influence DNA-Damage Response Pathways and the Immune Response in Invasive Breast Carcinoma." 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, June 28 – July 1, 2016, Nottingham, United Kingdom.

Oral presentation "Expression of CUL4A in invasive Breast Cancer: a Biomarker of Good Prognosis Related to Tumour Differentiation." 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, June 28 – July 1, 2016, Nottingham, United Kingdom.

Oral presentation "Further Evidence to Support that Oestrogen Receptor Expression in Breast Cancer is Bimodal." 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, June 28 – July 1, 2016, Nottingham, United Kingdom.

Poster presentation "Rho-GTPase Activating-protein 18 (ARHGAP18): A Biomarker Associated with Lymphovascular Invasion in Invasive Breast Cancer." 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, June 28 – July 1, 2016, Nottingham, United Kingdom.

OTHER PRESENTATIONS

As the main presenter:

Oral & Poster presentations "Loss of MCFD2 expression in breast cancer is associated with lymphovascular invasion and poor outcome." Cancer and Stem Cells (CASC) Divisional Meeting, May 2, 2017, Main University Campus, Nottingham, United Kingdom.

ACKNOWLEDGEMENTS

All praise and gratitude go to the giver of strength, patience, and wisdom, the almighty GOD.

My special gratitude is for my late mother and father. I wished you could see this thesis completed, but I know you are seeing me now.

I want to express my highest appreciation to my supervisors, Prof. Emad Rakha, Prof. Ian Ellis, Prof. Stewart Martin, and Dr Abhik Mukherjee, for all their support and guidance throughout this project and beyond.

I am deeply grateful to Dr Andrew Green, Dr Sasagu Kurouzmi, Dr Chitra Joseph, Mr Christopher Nolan, and Mrs Maria Diez-Rodriguez for their academic, technical, and much needed moral support.

I am also thankful to all members of Cancer and Stem Cells Division and Histopathology Department who have always been caring and supportive.

My deepest and heartfelt appreciation to my wife, Ola, my son, Faris, my daughter, Sulaf, and to the little one who was one month old on my viva day, Omar, for all their love and support.

My sincerest appreciation to my sponsor in the Kingdom of Saudi Arabia, King Abdulaziz University, the Saudi Arabian Cultural Bureau in the United Kingdom, for the financial support for my PhD study.

Thank you.

ABBREVIATIONS

Unicode special characters	
× g	relative centrifugal force
µl	microlitres
µm	micrometres
A	
A1AT	α-1-antitrypsin
Ab	antibody
a-CGH	array comparative genomic hybridisation
ADAMTS3	a disintegrin and metalloproteinase with thrombospondin motifs-3
ADH	atypical ductal hyperplasia
Ag	antigen
ALDH	aldehyde dehydrogenase
ALH	atypical lobular hyperplasia
ANN	artificial neural network
ANXA5	Annexin A5
APCDD1	APC down-regulated 1
AR	androgen receptor
ARHGAP18	Rho GTPase-activating protein-18
ARHGDIA	Rho-GDP dissociation inhibitor-1
ARSG	arylsulfatase G
ASCO	American Society of Clinical Oncology
AvE	average expression
B	
BC	breast cancer
bcGenExMiner	breast cancer gene expression miner
BCSS	BC-specific survival
BioGRID	biological general repository for interaction datasets
BP	biological processes
BRCA1	breast cancer type-1 susceptibility protein
BRCA2	breast cancer type-2 susceptibility protein
BSA	bovine serum albumin
BVI	blood vascular invasion
C	
C17orf8/ELP5	Elongator acetyltransferase complex protein 5
C1S	Complement C1s subcomponent
CAFs	cancer-associated fibroblasts
CAP	College of American Pathologists
CC	cellular compartment
CCL21	C-C motif chemokine 21
CCNE1	G1/S-specific cyclin-E1
CCR7	C-C chemokine receptor type-7
CDH1	E-cadherin
CDH2	N-cadherin
CDH3	P-cadherin
cDNA	complementary-DNA
CHEK1	serine/threonine-protein kinase Chk1
CI	confidence intervals
CK5/6	type II cytoskeletal 5 and 6

CLVS1	clavesin-1
CNA	copy number aberrations
CNRIP1	CB1 cannabinoid receptor-interacting protein 1
CNS	central nervous system
CNV	copy number variance
CRC	colorectal carcinoma
CRUK	Cancer Research UK
CSC	cancer stem cells
CSF-1	colony-stimulating factor 1
CTSD	Cathepsin D
CUH	Cambridge University Hospitals
CvC	cervical cancer
CXCL12	Stromal cell-derived factor 1
D	
DCIS	ductal carcinoma in situ
DDX58	retinoic acid-inducible gene 1 protein
DEGs	differentially expressed genes
DEPs	differentially expressed proteins
DFS	disease-free survival
DGE	Differential gene expression
DMFS	distant metastasis-free survival
DNAH9	dynein heavy chain-9
DYNLRB1	dynein light chain roadblock-type-1
E	
EC	endometrial cancer
E-cadherin	epithelial cadherin
ECM	extracellular matrix
ECM2	Extracellular matrix protein 2
EDG2	endothelial differentiation gene-2
EGF	Epidermal growth factor
EGFR	epidermal growth factor receptor
EMT	epithelial-mesenchymal transition
EnR	endoplasmic reticulum
EPG	excellent prognostic group
ER	Oestrogen receptor
ERG	Ets-related gene
ERGIC	endoplasmic reticulum-Golgi intermediate compartment
EtOH	Ethanol
F	
F5F8D	combined factor-V and factor-VIII deficiency
Fab	fragment antigen-binding
FBLN1	Fibulin-1
Fc region	fragment crystallisable region
FDR	false discovery rate
FEZ1	Fasciculation and elongation protein zeta-1
FFPE	formalin-fixed paraffin embedded
FITC-labelled Abs	Fluorescein isothiocyanate-labelled antibodies
FMN2	formin-2
FN1	fibronectin-1

G	
GATA3	T-cell-specific transcription factor
GBM	glioblastoma
GEO	Gene Expression Omnibus
gm	gram
GO	Gene ontology
GPG	good prognostic group
GSEA	gene set enrichment analysis
GSK3	glycogen synthase kinase-3
H	
H&E	haematoxylin and eosin
H2O2	hydrogen peroxide
HCC	hepatocellular carcinoma
HCl	hydrochloric acid
HER2	human epidermal growth factor receptor-2
HIER	heat-induced epitope retrieval
HNSCC	head and neck squamous cell carcinoma
HPLC	High-performance liquid chromatography
HR	hazard ratio
HRP	Horse-radish peroxide
HRT^-	hormonal receptor-negative
HRT^+	hormonal receptor-positive
H-score	Histological score
HSD17B12	very-long-chain 3-oxoacyl-CoA reductase
HSP70	70 kDa heat shock proteins
HSP90	90 kDa heat shock proteins
HSP90B1	endoplasmin
HSP90-α	heat shock protein 90 α
HTA	Health Technology assessment
I	
iBC	Inflammatory BC
IC-2	intermediate chain-2
ICC	intra-class correlation coefficient
IDC	invasive ductal carcinoma
IgG	Immunoglobulin G
IHC	immunohistochemistry
IKZF3	Zinc finger protein Aiolos
ILC	invasive lobular carcinoma
IMPC	Invasive micropapillary carcinoma
IMS	of industrial methylated spirits
INKA1	PAK4-inhibitor INKA1
IntClusts	integrative clusters
IPB	immunising peptide blocking
IRAC	International Agency for Research on Cancer
ITPRIP	Inositol 1,4,5-trisphosphate receptor-interacting protein
ITSN1	Intersectin-1
IκB	nuclear factor-κB
J	
JvGCRC	John van Geest Cancer Research Centre
K	

Kb	kilobase
KB	knowledge-bases
kDa	kilo Daltons
KLHL7	kelch-like protein-7
KM	Kaplan–Meier
KU70	X-ray repair cross-complementing protein-6
KU80	X-ray repair cross-complementing protein-5
L	
LCIS	lobular carcinoma in situ
Leica AXL	Leica Autostainer XL system
limma	Linear Models for Microarray Data
LMAN1	Lectin mannose-binding 1
LN	Lymph node
log2FC	the logarithm to base 2 fold change
LPA	lysophosphatidic acid
LPAR1	Lysophosphatidic acid receptor 1
LVD	lymphatic vessel density
LVI	lymphovascular invasion
$LVI^{+/-}$	LVI-supervised
M	
MAPK	phosphorylation levels of MAP kinase
MB	METABRIC cohort
MCFD2	Multiple coagulation factor deficiency protein-2
MED1	mediator of RNA polymerase II transcription subunit-1
MED7	mediator of RNA polymerase II transcription subunit-7
MeOH	Methanol
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MF	molecular functions
mg/ml	milligram/millilitre
miRNA	microRNA
MMP	matrix metalloproteinase
MMP1	interstitial collagenase
MPA	medroxyprogesterone
MPG-I	moderate prognostic group-I
MPG-II	moderate prognostic group-II
MS	Mass spectroscopy
msec	milliseconds
MSI-H	microsatellite unstable
MT1	membrane type-1
MWW	Mann-Whitney-Wilcoxon
MYCBPAP	MYCBP-associated protein
MYH11	smooth muscle myosin protein
N	
NBR1	Next to BRCA1 gene 1 protein
ng	nanograms
NHSB	Nottingham Health Science Biobank
NIHR	National Institute for Health Research
nm	nanometre

NPC	nasopharyngeal carcinoma
NPI	Nottingham Prognostic Index
NPI+	NPI-plus
NPS	Nottingham Primary Series
NTU	Nottingham Trent University
NUH	Nottingham University Hospitals
O	
OC	ovarian cancer
OeSCC	oesophageal squamous cell carcinoma
OS	overall survival
OSCC	oral squamous cell carcinoma
P	
P4HA2	Prolyl 4-hydroxylase subunit alpha-2
p53	tumour protein 53
PAM50	prediction analysis of microarray of 50 genes
PANTHER	Protein Analysis Through Evolutionary Relationships
PARP1	poly [ADP-ribose] polymerase-1
PBS	phosphate-buffered saline
PC	Protein class
PCL	papillary thyroid carcinoma
PDAC	pancreatic ductal adenocarcinoma
PDLIM7	PDZ and LIM domain protein 7
PDPN	Podoplanin
PEB	protein extraction buffer
pERK1/2	phosphorylated extracellular-regulated kinase 1/2
PIAS1	E3 SUMO-protein ligase PIAS1
PIAS4	E3 SUMO-protein ligase PIAS4
PIK3C2G	phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit gamma
PILRA	Paired immunoglobulin-like type 2 receptor alpha
PLC	pregnancy-lactation cycle
PPG	poor prognostic group
PPI	protein-protein interaction
PPIA	peptidyl-prolyl cis-trans isomerase-A
PPM1F	protein phosphatase 1F
PPM1G	Protein phosphatase 1G
PR	progesterone receptor
PrC	Prostate cancer
PRRX1	Paired mesoderm homeobox protein 1
PTEN	phosphatase and tensin homolog gene
PTM	post-translational modifications
PW	Pathways
Q	
QMC	Queen's Medical Centre
R	
RAPH1	pleckstrin homology domains-containing protein-1
RERGL	Ras-related and oestrogen-regulated growth inhibitor-like protein
RNASE1	Ribonuclease pancreatic
ROS	reactive oxygen species

RPL10A	60S ribosomal protein L10a
rpm	rounds per minute
RPS18	40S ribosomal protein S18
RT-qPCR	quantitative reverse transcription-polymerase chain reaction
S	
S100A8	protein S100-A8
SAM	significance analysis of the microarrays
SCCHN	squamous cell carcinoma of the head and neck
SDF-1	the stromal cell-derived-α
SDNSF	neural stem/progenitor cell supporting factor
SEC14L1	SEC14-like protein 1
SEC63	translocation protein SEC63
shRNA	short hairpin RNA
SIGLEC1	Sialoadhesin
SLC3A2	4F2 cell-surface antigen heavy chain
SLC6A9	Sodium- and chloride-dependent glycine transporter 1
SMA	smooth muscle actin
SNAP23	Synaptosomal-associated protein 23
SNAPs	synaptosomal-associated proteins
SNPs	single nucleotide polymorphism
STAT3	Signal transducer and activator of transcription 3
STMN1	Stathmin
STXs	syntaxins
SUMO	small ubiquitin-like modifier protein
T	
TBS	Tris Buffered Saline
TCAB1	telomerase Cajal body protein-1
TDLU	terminal duct lobular unit
TGF-α	transforming growth factor alpha
THAP4	THAP domain-containing protein 4
TMA	tissue microarray
TMEM98	Transmembrane protein 98
TMPRSS2	transmembrane protease, serine 2
TNC	Tenascin
TNFRSF19	Tumour necrosis factor receptor superfamily member 19
TNF-α	tumour necrosis factor alpha
TP63	tumour protein 63
TSC1	Hamartin
t-SNARE	target SNARE
TWIST2	Twist-related protein 2
U	
UBC9	SUMO-conjugating enzyme UBC9
UMOD	uromodulin
V	
VAMPs	vesicle-associated membrane proteins
VDAC2	voltage-dependent anion-selective channel protein-2
VEGFC	vascular endothelial growth factor-C
VEGFD	vascular endothelial growth factor-D

VEGFR3	vascular endothelial growth factor receptor-3
VHL	von Hippel-Lindau
VM	Vasculogenic mimicry
VPG	very poor prognostic group
v-SNARE	vesicles SNARE
W	
WB	Western blotting
WISP2	WNT1-inducible-signalling pathway protein 2
X	
XIC	extracted ion chromatogram
Z	
ZDHHC9	Palmitoyltransferase ZDHHC9
ZFYVE26	zinc finger FYVE domain-containing protein-26
ZNF217	zinc finger protein 217

Table of Contents

ABSTRACT i
PUBLICATIONS ARISING FROM THIS THESIS iv
OTHER PUBLICATIONS v
CONFERENCE PRESENTATIONS vi
OTHER PRESENTATIONS vii
ACKNOWLEDGEMENTS viii
ABBREVIATIONS ix
Table of Contents xvi
LIST OF TABLES xxi
LIST OF FIGURES xxiii
Chapter 1: Introduction 1
1.1. General introduction to breast and breast cancer (BC) 2
1.1.1. Anatomy and histology of the normal breast tissue 2
1.1.2. Development of invasive BC 2
1.1.3. Breast cancer: Incidence and mortality 3
1.1.4. Risk factors for breast cancer 4
1.1.4.1. Exposure to oestrogen 4
1.1.4.2. Age 4
1.1.4.3. Family history and genetic predisposition 4
1.1.4.4. Alcohol consumption, smoking, and obesity 5
1.1.4.5. Breast density 5
1.2. Prognostic and predictive factors in BC 5
1.2.1. Prognostic factors 6
1.2.1.1. Phenotypes of BC 6
1.2.1.2. Tumour grade 7
1.2.1.3. Tumour size 7
1.2.1.4. Lymph node stage 8
1.2.1.5. Nottingham Prognostic Index 8
1.2.1.6. Lymphovascular invasion 9
1.2.2. Predictive factors 9
1.2.2.1. Hormone receptors, ER and PR 9
1.2.2.2. HER2 10
1.3. Molecular mechanisms underlying LVI in invasive BC 10
1.3.1. Introduction to LVI 10
1.3.2. Morphological assessment of LVI 11
1.3.3. Tumour histological features and LVI 12
1.3.4. Vascular or lymphatic invasion 14
1.3.5. Tumour cell invasiveness, tumour microenvironment, and LVI 15
1.3.6. Gene expression profiling, gene signatures, and LVI 18
1.3.7. Proteomics and LVI 20
1.3.8. Unified Plot of LVI 21
1.3.9. Conclusions 22
1.4. Hypotheses and aims 23
1.4.1. Hypotheses 23
1.4.2. Aims and objectives 23
Chapter 2: Materials and Methodologies 24
2.1. Study cohorts 25
2.1.1. The METABRIC subset cohort 25
2.1.2. The METABRIC cohort 29
2.1.3. NPS of invasive BC 31
2.1.4. Luminal-A subtype cohort of invasive BC 32
2.1.5. Other cohorts collected or used in this study 34
2.2. Tissue handling and processing 35
2.2.1. FFPE blocks retrieval from NHSB 35

2.2.2. Tissue sectioning 35
2.2.2.1. Documentations 35
2.2.2.2. Sectioning in the microtome 35
2.2.3. Preparation of donor blocks and TMA construction 36
2.3. H&E histological staining 37
2.3.1. Overview 37
2.3.2. H&E procedure 37
2.4. Determination of binding specificity for selected antibodies with western blotting.... 38
2.4.1. Overview 38
2.4.2. Cell line lysates 39
2.4.3. Preparation of western blotting reagents and tools 39
2.4.4. Sample preparation 40
2.4.5. Loading samples and separation of proteins mixture 40
2.4.6. Transferring protein bands onto nitrocellulose membrane 41
2.4.7. Visualisation of protein bands 42
2.5. Immunising peptide blocking (IPB) 42
2.6. IHC staining 43
2.6.1. Overview 43
2.6.2. Immunohistochemistry tools, chemicals, and procedures 43
2.6.2.1. Labelling of slides 44
2.6.2.2. Deparaffinisation and rehydration of tissue sections 44
2.6.2.3. Heat-induced epitope retrieval: Citrate buffer, pH 6.00 44
2.6.2.4. Preparation of sodium citrate buffer, pH 6.00 44
2.6.2.5. Slide assembly onto Shandon™ Sequenza coverplates™ 45
2.6.2.6. Blocking of endogenous peroxidases 45
2.6.2.7. Blocking non-specific reacting proteins 46
2.6.2.8. Applying the optimised primary antibody 46
2.6.2.9. Applying post-primary antibody 47
2.6.2.10. Applying Novolink™ polymer 47
2.6.2.11. Blotting with 3,3'-diaminobenzidine (DAB) 47
2.6.2.12. Counterstaining with haematoxylin and tissue dehydration 47
2.7. Imaging of tissue sections and histological scoring (H-score) system 48
2.7.1. Whole slide imaging 48
2.7.2. H-score 48
2.8. Differential gene expression (DGE) and gene selection 49
2.8.1. Overview 49
2.8.2. Nucleic acids extraction and microarray hybridisation 49
2.8.3. Profile analysis of differentially expressed genes 50
2.8.4. Gene selection based on LIMMA and CNA results 50
2.8.5. Gene selection based on accessible online resources 51
2.8.6. Gene selection based on transcriptomic and cliniccal profiles of the $LVI^{+/-}$ cohort 52
2.9. Characterisation of protein expression patterns on full-face FFPE tissue sections.... 52
2.9.1. Overview 52
2.9.2. Sampling and the characterisation methodology 52
2.10. Total protein extraction and quantification from FFPE 53
2.10.1. Overview 53
2.10.2. Tissue harvesting and collection 53
2.10.3. Deparaffinisation of harvested tissues 54
2.10.4. Protein extraction using Qproteome protein extraction buffer 54
2.10.4.1. Bradford protein assay and the determination of total protein concentration 54
2.11. Preparation of protein sample for tryptic digestion 55
2.12. HPLC/MS procedures 56
2.12.1. Tryptic digestion of extracted proteins for mass spectrometry 56
2.12.2. Reverse-phase high-performance liquid chromatography electrospray ionisation tandem mass spectrometry (RP-HPLC-ESI-MS/MS) in SWATH and IDA modes 56
2.12.3. Construction of peptides library, spectral normalisation, and fold change analysis 57
2.13. Statistics 57
2.13.1. Introduction 57
2.13.2. Descriptive statistics 57

2.13.3. Intra-class correlation coefficient (ICC) and Cohen's kappa coefficient 58
2.13.4. Chi-square test (goodness to fit) 58
2.13.5. Mann-Whitney U test (median analysis) 58
2.13.6. Student's t-test (mean analysis) 58
2.13.7. Pearson rank test (model/line of best fit) 59
2.13.8. Kaplan-Meier survival analysis and log-rank test 59
2.13.9. Proportional hazards model (Cox regression model) 59
2.14. Ethical approval 59
Chapter 3: Integrated Differential Gene Expression and Bioinformatics Analyses 60
3.1. Introduction 61
3.2. Materials and methods 62
3.2.1. Study cohorts 62
3.2.1.1. LVI-supervised cohort 62
3.2.1.2. The METABRIC cohort 64
3.2.2. DGE and retrieval of copy number aberration 65
3.2.2.1. Nucleic acids extraction 65
3.2.2.2. Microarray hybridisation and scanning of Illumina HumanHT-12 v3 expression beadchips 67
3.2.2.3. Analysis of CNA and obtaining the complete transcriptomic data of the METABRIC cohort 68
3.2.3. Strategies of gene selection 69
3.2.3.1. Profile analysis of DEGs 69
3.2.3.2. Pathway analysis and clinicopathological interrogations 70
3.2.4. Statistics 71
3.3. Results 72
3.3.1. DGE profile of the LVI-NT1154 72
3.3.2. DGE profile of the LVI-NT100 72
3.3.2.1. CNA data analysis of the LVI-NT100 cluster of DEGs 73
3.3.2.2. Clinicopathological impacts for the DEGs in the LVI-NT100 cluster 75
3.3.3. DGE profile of the LVI-NT283 82
3.3.3.1. Transcriptomic/LVI analysis of the LVI-NT283 cluster of DEGs 82
3.3.3.2. Clinicopathological impact for transcripts in the UP-13/LVI^{+} cluster 84
3.3.3.3. Clinicopathological impact for transcripts in the UP-22/LVI^{-} cluster 123
3.4. Discussion 187
3.4.1. Overview 187
3.4.2. LVI-related transcripts from the LVI-NT100 cluster 188
3.4.2.1. MCFD2 188
3.4.2.2. CCT8 189
3.4.3. LVI-related transcripts from the LVI-NT283 cluster 190
3.4.3.1. LVI-related transcripts from the UP-13/LVI^{+} signature 190
3.4.3.2. LVI-related transcripts from the UP-22/LVI^{-} signature 193
3.5. Conclusion 196
Chapter 4: IHC Assessment of MCFD2, CCT8, SNAP23, and SEC14L1 as Potential Regulators for LVI in Invasive BC 198
4.1. Introduction 199
4.1.1. Multiple coagulation factor deficiency protein 2 200
4.1.2. Chaperonin containing T-complex protein 1 (TCP1) subunit theta 201
4.1.3. Synaptosomal-associated protein 23 203
4.1.4. SEC14-like protein 1 204
4.2. Materials and methods 205
4.2.1. Study cohorts 205
4.2.1.1. Nottingham Primary Series of invasive BC (NPS) 205
4.2.1.2. LVI-supervised cohort (LVI$^{+/-}$) 206
4.2.2. Tissue handling and sectioning 206
4.2.3. Purchased antibodies and their optimised conditions for IHC and WB 206
4.2.4. IHC, digital imaging, and scoring of stained sections 207
4.2.5. Statistical analysis 209
4.3. Results 209
4.3.1. Evaluation of the results of WB and IPB 209

4.3.2. Characterisation of protein expression patterns on full-face tissue sections…. 213
4.3.3. Clinicopathological relevance and protein interactions for the selected LVI candidates.....217
4.3.3.1. Clinicopathological impact for MCFD2 protein 224
4.3.3.2. Clinicopathological impact for CCT8 protein 224
4.3.3.3. Synaptosomal-associated protein 23....225
4.3.3.4. SEC14-like protein 1....226
4.4. Discussion....245
4.4.1. Multiple coagulation factor deficiency protein 2 245
4.4.2. T-complex protein 1 subunit theta....249
4.4.3. Synaptosomal-associated protein 23 252
4.4.4. SEC14-like protein 1 255
4.5. Conclusion 257
Chapter 5: LVI Molecular Signatures in Invasive BC: RNA microarray and HPLC/MS Study....259
5.1. Introduction 260
5.2. Materials and methods 262
5.2.1. Study cohorts 262
5.2.1.1. METABRIC cohort....262
5.2.1.2. LVI-supervised cohort 262
5.2.1.3. The luminal-A subtype cohort 262
5.2.2. DGE analyses....262
5.2.3. Total protein extraction and quantification 262
5.2.3.1. Tissue harvesting off FFPE tissue sections 262
5.2.3.2. Dewaxing of the harvested tissues 262
5.2.3.3. Total protein extraction using Qproteome FFPE Tissue Kit 263
5.2.3.4. Bradford protein assay 263
5.2.4. Preparation of protein sample for tryptic digestion 263
5.2.4.1. Tryptic digestion of protein extracts for mass spectroscopy 263
5.2.5. Reverse-phase high-performance liquid chromatography electrospray ionisation tandem mass spectrometry (RP-HPLC-ESI-MS/MS) in SWATH and IDA modes 263
5.2.5.1. Construction of peptides library, spectral normalisation, and fold change analysis 263
5.2.6. Statistical analyses 263
5.2.7. Ethical approval and authentication of material transferring....264
5.3. Results....265
5.3.1. Features of the studied cohorts....265
5.3.1.1. METABRIC cohort....265
5.3.1.2. LVI-supervised sub-cohort of the METABRIC....265
5.3.1.3. LVI-supervised and luminal-A cohort for mass spectroscopy 265
5.3.2. Protein quantitation and differentially expressed proteome 265
5.3.2.1. Biological drivers for LVI and their distribution of expression among the LVI-supervised patients....269
5.3.3. Prognostic importance and molecular interactions of T8 in invasive BC patients 271
5.3.3.1. HSP90AA1 271
5.3.3.2. VDAC2....274
5.3.3.3. ARHGDIA....276
5.3.3.4. DYNLRB1 279
5.3.3.5. FBLN1....281
5.4. Discussion....305
5.4.1. *HSP90AA1*: A molecular chaperone associated with LVI and poor prognosis in invasive BC....305
5.4.1.1. Overview....305
5.4.1.2. Clinicopathological relevance of the HSP90AA1 altered gene expression....306
5.4.1.3. Pivotal gene expression associations between HSP90AA1 expression and matrix metalloproteinases....308
5.4.1.4. Key gene expression associations between HSP90AA1 and immune cells 311
5.4.1.5. HSP90AA1 gene expression and tumour microenvironment and progression biomarker....312
5.4.1.6. Inferences of HSP90AA1 314
5.4.2. VDAC2: A mitochondrial transmembrane protein that may predict LVI, LN metastasis, and poor survival for invasive BC patients....315
5.4.2.1. Overview....315
5.4.2.2. Clinicopathological relevance of the VDAC2 altered gene expression 316

5.4.2.3. Pivotal gene expression associations between VDAC2 expression and matrix metalloproteinases..... 318
5.4.2.4. Key gene expression associations between VDAC2 and immune cells..... 318
5.4.2.5. VDAC2 gene expression and tumour microenvironment and progression biomarker ... 319
5.4.2.6. Co-expression contexts of phosphoglycerate mutase-1 (PGAM1) and triosephosphate isomerase-1 (TPI1) with VDAC2 ... 319
5.4.2.7. Inferences of VDAC2 ... 321
5.4.3. *ARHGDIA*: A signal transducer that regulates the Rho-GTPase cycle and ER-alpha transactivation, and stimulates axonal growth ... 322
5.4.3.1. Overview ... 322
5.4.3.2. Clinicopathological relevance of the ARHGDIA altered gene expression ... 323
5.4.3.3. Pivotal gene expression associations between ARHGDIA expression and matrix metalloproteinases..... 324
5.4.3.4. Key gene expression associations between ARHGDIA and immune cells ... 325
5.4.3.5. ARHGDIA gene expression and tumour microenvironment and progression biomarker 326
5.4.3.6. ARHGDIA: Is it a dependent signalling molecule for invasion?..... ... 327
5.4.3.7. ARHGDIA: A cross-linker in cancer-nerve invasion? ... 328
5.4.3.8. Inferences of ARHGDIA ... 328
5.4.4. DYNLRB1: A member of the dynein superfamily of proteins that is crucial for epithelial cells polarisation..... ... 329
5.4.4.1. Overview ... 329
5.4.4.2. Clinicopathological relevance of the DYNLRB1 altered gene expression ... 330
5.4.4.3. Pivotal gene expression associations between DYNLRB1 expression and matrix metalloproteinases..... ... 331
5.4.4.4. Key gene expression associations between DYNLRB1 and immune cells ... 332
5.4.4.5. DYNLRB1 gene expression and tumour microenvironment and progression biomarker 332
5.4.4.6. Km23-1: A mediator of cellular motility, migration, and invasion ... 333
5.4.4.7. Inferences of DYNLRB1 ... 334
5.4.5. FBLN1: A family member of the basement membrane-associated glycoproteins that suppress BC cellular motility ... 334
5.4.5.1. Overview ... 334
5.4.5.2. Clinicopathological relevance of the FBLN1 altered gene expression ... 335
5.4.5.3. Pivotal gene expression associations between FBLN1 expression and matrix metalloproteinases..... ... 336
5.4.5.4. Key gene expression associations between FBLN1 and immune cells ... 338
5.4.5.5. FBLN1 gene expression and tumour microenvironment and progression biomarker ... 338
5.4.5.6. Inferences of FBLN1 ... 346
5.5. Molecular architecture of LVI in invasive BC patients: A chapter conclusion ... 347
Chapter 6: General Discussion & Conclusion ... 351
6.1. General discussion ... 352
6.2. Preeminent findings of the study ... 354
6.3. Strengths and limitations of the study ... 357
6.3.1. Potential strengths ... 357
6.3.2. Potential limitations ... 358
6.3.3. Potential hypotheses for future testing ... 359
6.4. Future work ... 360
6.5. Final conclusions ... 360
REFERENCES ... 362
Appendices ... 389
APPENDIX A: SUPPLEMENTARY INFORMATION FOR CHAPTER 2 ... 390
APPENDIX B: SUPPLEMENTARY INFORMATION FOR CHAPTER 3 ... 391
APPENDIX C: SUPPLEMENTARY INFORMATION FOR CHAPTER 4 ... 413
APPENDIX D: SUPPLEMENTARY INFORMATION FOR CHAPTER 5 ... 417

LIST OF TABLES

Table 2. 1: clinical features of the METABRIC sub-cohort 26
Table 2. 2: clinical features of the LVI-supervised cohort 28
Table 2. 3: clinical features of the METABRIC cohort 30
Table 2. 4: Clinicopathological characteristics of Nottingham primary series of invasive BC 32
Table 2. 5: Clinicopathological characteristics of the Luminal-A molecular subtype 34

Table 3. 1: Clinical relevance for MCFD2 and CCT8 in the LVI-supervised patients 77
Table 3. 2: Clinical relevance for MCFD2 and CCT8 in the METABRIC patients 78
Table 3. 3: MCFD2 and CCT8: Multivariate Cox regression analysis 79
Table 3. 4: Differential mean rank scores for upregulated mRNAs in the LVI-definite patients 83
Table 3. 5: Differential mean rank scores for upregulated mRNAs in the LVI-absent patients 84
Table 3. 6: Clinical significance of positively correlated genes with LVI in the LVI-supervised cohort 94
Table 3. 7: Clinical significance of positively correlated genes with LVI in the METABRIC cohort 100
Table 3. 8: Multivariate Cox regression analysis for prediction of breast cancer specific survival, UP-13/LVI^{+} 106
Table 3. 9: Clinical significance of upregulated genes with LVI^{-} patients in the LVI-supervised cohort 135
Table 3. 10: Clinical significance of upregulated genes with LVI^{-} patients in the METABRIC cohort 147
Table 3. 11: Multivariate Cox regression analysis for prediction of breast cancer specific survival, UP-22/LVI^{-} 159

Table 4. 1: Clinical significance of the LVI-correlated proteins in the Nottingham primary series of invasive breast carcinoma 227
Table 4. 2: Protein-protein interactions for the LVI-correlated proteins in the Nottingham primary series of invasive breast carcinoma 229
Table 4. 3: Multivariate Cox regression analyses for altered protein expression of MCFD2. 231
Table 4. 4: Multivariate Cox regression analyses for altered protein expression of cytoplasmic CCT8. 232
Table 4. 5: Multivariate Cox regression analyses for altered protein expression of nuclear CCT8. 233
Table 4. 6: Multivariate Cox regression analyses for p altered protein expression of cytoplasmic SNAP23. 234
Table 4. 7: Multivariate Cox regression analyses for altered protein expression of membranous SNAP23. 235
Table 4. 8: Clinical significance of SEC14L1 protein in the LVI-supervised cohort. 241
Table 4. 9: Protein-protein interactions of SEC14L1 protein in the LVI-supervised cohort 242
Table 4. 10: Multivariate Cox regression analyses for altered protein expression of cytoplasmic SEC14L1. 243

Table 5. 1: The cluster of LVI-DEM produced by RNA microarray and HPLC-MS protocols 267

Table 5. 2: LVI statistical associations for the T8 list of LVI-DEM in the LVI-supervised cohort 270
Table 5. 3: Clinical significance of the T8 upregulated genes associated LVI positivity in the LVI-supervised cohort 285
Table 5. 4: Clinical significance of the T8 upregulated genes associated LVI positivity in the METABRIC cohort . 286
Table 5. 5: Distribution of the LVI-supervised and METABRIC patients on the PAM50 and IntClusts classification systems against gene expression profiles of the T8 upregulated genes associated LVI positivity 287
Table 5. 6: Clinicopathological significance of the T8 upregulated genes associated LVI negativity in the LVI-supervised cohort 289
Table 5. 7: Clinicopathological significance of the T8 upregulated genes associated LVI negativity in the METABRIC cohort 290
Table 5. 8: Distribution of the LVI-supervised and METABRIC patients on the PAM50 and IntClusts classification systems against gene expression profiles of the T8 upregulated genes associated LVI negativity 291
Table 5. 9: Gene expression correlations between the T8 genes and a cluster of MMP-encoding genes in the LVI-supervised patients..... 293
Table 5. 10: Gene expression correlations between the T8 genes and a cluster of MMP-encoding genes in the METABRIC patients 294
Table 5. 11: Gene expression correlations between the T8 genes and a cluster of CD-encoding genes in the LVI-supervised patients..... 295
Table 5. 12: Gene expression correlations between the T8 genes and a cluster of CD-encoding genes in the METABRIC patients. 296
Table 5. 13: Gene expression correlations between the T8 genes and a cluster of tumour progression genes in the LVI-supervised patients..... 297
Table 5. 14: Gene expression correlations between the T8 genes and a cluster of tumour progression genes in the METABRIC patients 298
Table 5. 15: Multivariate Cox regression analyses for prediction of breast cancer-specific survival, altered gene expression of *HSP90AA1*, *VDAC2*, *ARHGDIA*, *DYNLRB1*, and *FBLN1*. 304

LIST OF FIGURES

Figure 1.1: A case of invasive BC with extensive LVI, H&E 12

Figure 3. 1: The IHC detection of LVI 63
Figure 3. 2: Differentially expressed genes of the LVI-NT100 cluster 74
Figure 3. 3: KM estimates of the BCSS for *MCFD2* mRNA in the LVI-supervised and METABRIC cohorts 80
Figure 3. 4: KM estimates of the BCSS for *CCT8* mRNA in the LVI-supervised and METABRIC cohorts 81
Figure 3. 5: KM estimates of the BCSS for *CTSD* mRNA in the LVI-supervised and METABRIC cohorts 110
Figure 3. 6: KM estimates of the BCSS for *IKZF3* mRNA in the LVI-supervised and METABRIC cohorts 111
Figure 3. 7: KM estimates of the BCSS for *ITPRIP* mRNA in the LVI-supervised and METABRIC cohorts 112
Figure 3. 8: KM estimates of the BCSS for *P4HA2* mRNA in the LVI-supervised and METABRIC cohorts 113
Figure 3. 9: KM estimates of the BCSS for *PDLIM7* mRNA in the LVI-supervised and METABRIC cohorts 114
Figure 3. 10: KM estimates of the BCSS for *PILRA* mRNA in the LVI-supervised and METABRIC cohorts 115
Figure 3. 11: KM estimates of the BCSS for *PPM1G* mRNA in the LVI-supervised and METABRIC cohorts 116
Figure 3. 12: KM estimates of the BCSS for *RNASE1* mRNA in the LVI-supervised and METABRIC cohorts 117
Figure 3. 13: KM estimates of the BCSS for *SEC14L1* mRNA in the LVI-supervised and METABRIC cohorts 118
Figure 3. 14: KM estimates of the BCSS for *SIGLEC1* mRNA in the LVI-supervised and METABRIC cohorts 119
Figure 3. 15: KM estimates of the BCSS for *SLC3A2* mRNA in the LVI-supervised and METABRIC cohorts 120
Figure 3. 16: KM estimates of the BCSS for *SLC6A9* mRNA in the LVI-supervised and METABRIC cohorts 121
Figure 3. 17: KM estimates of the BCSS for *ZDHHC9* mRNA in the LVI-supervised and METABRIC cohorts 122
Figure 3. 18: KM estimates of the BCSS for *ANXA5* mRNA in the LVI-supervised and METABRIC cohorts 165

Figure 3. 19: KM estimates of the BCSS for *APCDD1* mRNA in the LVI-supervised and METABRIC cohorts 166
Figure 3. 20: KM estimates of the BCSS for *ELP5* mRNA in the LVI-supervised and METABRIC cohorts 167
Figure 3. 21: KM estimates of the BCSS for *C1S* mRNA in the LVI-supervised and METABRIC cohorts 168
Figure 3. 22: KM estimates of the BCSS for *CNRIP1* mRNA in the LVI-supervised and METABRIC cohorts 169
Figure 3. 23: KM estimates of the BCSS for *CXCL12* mRNA in the LVI-supervised and METABRIC cohorts 170
Figure 3. 24: KM estimates of the BCSS for *ECM2* mRNA in the LVI-supervised and METABRIC cohorts 171
Figure 3. 25: KM estimates of the BCSS for *FBLN1* mRNA in the LVI-supervised and METABRIC cohorts 172
Figure 3. 26: KM estimates of the BCSS for *FEZ1* mRNA in the LVI-supervised and METABRIC cohorts 173
Figure 3. 27: KM estimates of the BCSS for *INKA1* mRNA in the LVI-supervised and METABRIC cohorts 174
Figure 3. 28: KM estimates of the BCSS for *ITSN1* mRNA in the LVI-supervised and METABRIC cohorts 175
Figure 3. 29: KM estimates of the BCSS for *LPAR1* mRNA in the LVI-supervised and METABRIC cohorts 176
Figure 3. 30: KM estimates of the BCSS for *NBR1* mRNA in the LVI-supervised and METABRIC cohorts 177
Figure 3. 31: KM estimates of the BCSS for *PRRX1* mRNA in the LVI-supervised and METABRIC cohorts 178
Figure 3. 32: KM estimates of the BCSS for *STAT3* mRNA in the LVI-supervised and METABRIC cohorts 179
Figure 3. 33: KM estimates of the BCSS for *THAP4* mRNA in the LVI-supervised and METABRIC cohorts 180
Figure 3. 34: KM estimates of the BCSS for *TMEM98* mRNA in the LVI-supervised and METABRIC cohorts 181
Figure 3. 35: KM estimates of the BCSS for *TNC* mRNA in the LVI-supervised and METABRIC cohorts 182
Figure 3. 36: KM estimates of the BCSS for *TNFRSF19* mRNA in the LVI-supervised and METABRIC cohorts 183
Figure 3. 37: KM estimates of the BCSS for *TSC1* mRNA in the LVI-supervised and METABRIC cohorts 184

Figure 3. 38: KM estimates of the BCSS for *TWIST2* mRNA in the LVI-supervised and METABRIC cohorts 185
Figure 3. 39: KM estimates of the BCSS for WISP2 mRNA in the LVI-supervised and METABRIC cohorts 186

Figure 4. 1: Binding specificity of the primary anti-MCFD2 antibody, western blotting..... 210
Figure 4. 2: Confirmation of anti-MCFD2 antibody specificity with immunising peptide blocking method 210
Figure 4. 3: Binding specificity of the anti-CCT8 antibody, western blotting..... 211
Figure 4. 4: Binding specificity of the anti-SNAP23 antibody, western blotting 211
Figure 4. 5: Binding specificity of the anti-SEC14L1 antibody, western blotting 212
Figure 4. 6: Protein expression characterisation of MCFD2 in invasive BC, full-face tissue sections 213
Figure 4. 7: Protein expression characterisation of CCT8 in invasive BC, full-face tissue sections 214
Figure 4. 8: Protein expression characterisation of SNAP23 in invasive BC, full-face tissue sections 215
Figure 4. 9: Protein expression characterisation of SEC14L1 in invasive BC, full-face tissue sections 216
Figure 4. 10: Immunohistochemistry expression of MCFD2, tissue microarray images. 218
Figure 4. 11: Immunohistochemistry expression of SEC14L1, tissue microarray images. 219
Figure 4. 12: Immunohistochemistry expression of cytoplasmic CCT8, tissue microarray images. .. 220
Figure 4. 13: Immunohistochemistry expression of nuclear CCT8, tissue microarray images. 221
Figure 4. 14: Immunohistochemistry expression of cytoplasmic SNAP23, tissue microarray images. 222
Figure 4. 15: Immunohistochemistry expression of membranous SNAP23, tissue microarray images. 223
Figure 4. 16: KM estimates of the BCSS, DMFS, and DFS for cytoplasmic MCFD2 protein in the Nottingham primary series of invasive breast carcinoma..... 236
Figure 4. 17: KM estimates of the BCSS, DMFS, and DFS for cytoplasmic CCT8 protein in the Nottingham primary series of invasive breast carcinoma..... 237
Figure 4. 18: KM estimates of the BCSS, DMFS, and DFS for nuclear CCT8 protein in the Nottingham primary series of invasive breast carcinoma..... 238
Figure 4. 19: KM estimates of the BCSS, DMFS, and DFS for cytoplasmic SNAP23 protein in the Nottingham primary series of invasive breast carcinoma..... 239
Figure 4. 20: KM estimates of the BCSS, DMFS, and DFS for membranous SNAP23 protein in the Nottingham primary series of invasive breast carcinoma..... 240
Figure 4. 21: KM estimates of the BCSS, DMFS, and DFS for cytoplasmic SEC14L1 protein in the Nottingham primary series of invasive breast carcinoma..... 244

Figure 5. 1: Bovine serum albumin standard curve for the calculation of total protein concentrations 266
Figure 5. 2: KM estimates of the BCSS for *HSP90AA1* mRNA in the LVI-supervised and METABRIC cohorts 299
Figure 5. 3: KM estimates of the BCSS for VDAC2 mRNA in the LVI-supervised and METABRIC cohorts 300
Figure 5. 4: KM estimates of the BCSS for *ARHGDIA* mRNA in the LVI-supervised and METABRIC cohorts 301
Figure 5. 5: KM estimates of the BCSS for *DYNLRB1* mRNA in the LVI-supervised and METABRIC cohorts 302
Figure 5. 6: KM estimates of the BCSS for *FBLN1* mRNA in the LVI-supervised and METABRIC cohorts 303

Chapter 1: Introduction

1.1. General introduction to breast and breast cancer (BC)

1.1.1. Anatomy and histology of the normal breast tissue

The female breast is an organ that lies over the pectoral muscle in front of the rib cage. In females, the breast glandular tissue is a complex structure that produces and secretes milk. Morphologically, the female breast is round shaped. In this structure, the milk-producing mammary gland is surrounded by fatty and connective tissues that help to maintain its arrangement integrity. The mammary gland of the breast is comprised of 12 to 20 lobes, each of which is subdivided into several thousands of smaller lobules. Each lobule can contain up to 20–200 pouch-shaped structures called acini, which secrete milk. Ducts from these acini converge to form the terminal duct and, with the associated lobule, the basic structure of the breast, the terminal duct lobular unit (TDLU), is formed. The TDLU drains into the subsegmental and segmental ducts, which drain into the lactiferous duct and open into the lactiferous sinus that leads to the nipple [1]. Normal breast development starts early during the embryonic phase with the formation of the mammary crest and subsequent primitive mammary buds. As a unique organ, the functional matured status of the breast is activated during the pregnancy-lactation cycle (PLC) [2], which consumes a high level of energy due to the increased metabolic needs. The glandular tissue of the breast is exposed to hormonal and structural changes during PLC and also after puberty and in older ages.

1.1.2. Development of invasive BC

Uncontrolled cellular division, or hyperplasia, and evading apoptosis pathways by breast epithelial cells are one of initial steps to establish tumorous breast tissues [3]. Some benign breast lesions could be influenced to form invasive BC. However, highly proliferative atypical ductal hyperplasia (ADH) and atypical lobular hyperplasia (ALH) are considered to be risk factors for some individuals [4]. The epithelial hyperplasia could develop into cellular atypia that obstructs the breast duct and subsequently progress into ductal carcinoma in situ (DCIS). Principally, mammographic screening is an administrated method applied on patients for the detection of asymptomatic DCIS. The detection of the "pre-invasive" lesions of the breast may suggest a sequential progression from normal breast tissue to invasive BC; especially, in the high grade and young DCIS patients [5]. Further development to the DCIS lesion would beget a locally invasive ductal carcinoma (IDC). The different types of BC could originate from different subtypes of epithelial cells; hence, specific BCs tend to show aggressive biological behaviour and activate an invasion-metastasis cascade of events while some other BC types are slow growing and sensitive to different treatments [6]. The claimed cellular

and molecular distinctive features that allow non-invasive DCIS to breach the myoepithelium lining and basement membrane depend on genomic, transcriptomic, and proteomic aberrations between DCIS and IDC.

Malignant cells of the breast are governed by highly sophisticated mechanisms that may assure their propagation and survival. Evading growth suppressors and immune surveillance, activating inflammation, invasion, and metastasis pathways, and genomic instability are a few examples of how BC is a multi-factorial disease [7]. Some reports state that invasive BC with a high burden of cancer stem cells (CSC) may exhibit high invasion potentials and poor survival outcome [8, 9]. Other studies attribute BC invasiveness to specific molecular intrinsic subtypes that dictate hormonal signalling pathways in invasive BC [10]. Recently, it was stated that younger BC patients are susceptible to aggressive phenotypes of invasive BC such as luminal-B [11]. Another study states that young luminal-B invasive BC patients show strong correlations to lymphovascular invasion (LVI) definite events [12]. The certain presence of LVI is an unfavourable sign for invasive BC patients, and it adversely impacts their survival dramatically. However, the true nature of LVI is not fully determined as BC is a heterogeneous disease with distinct molecular features. Later in this chapter, the molecular mechanisms underlying LVI in invasive BC patients will be introduced.

1.1.3. Breast cancer: Incidence and mortality

Invasive BC is a leading cause of death among women worldwide. According to the World Cancer Report 2014, BC had a remarkably high incidence rate of 43.3 per 100,000 females worldwide. The next highest incidence of cancer was colorectal carcinoma (CRC) with an incidence rate of 14.3 per 100,000. The global incidence burden of BC in 2012 was 1.67 million cases. In 2008, the number of women affected by BC was around 1.4 million [13]. Therefore, the total number of BC patients is increasing. In developed and less-developed countries, BC was the most common cancer with 794,000 and 883,000 cases, respectively, according to the World Cancer Report 2014. Globally, BC as a death-causing disease was ranked fifth behind other cancers, with 522,000 deaths. However, BC is the leading cause of death in less-developed countries, with 324,000 deaths, and the second cause of death in more developed countries, with 198,000 deaths [14]. The estimated BC incidence proportions in European females in 2012 was 464,202 (28.8%), whereas the estimated BC mortality was 131,257 (16.8%). The highest incidence rates for all cancers, except non-melanoma skin cancer, were found in high-income countries, including United States, Canada, Belgium, France, Ireland, Luxembourg, Monaco, Netherlands, United Kingdom, Japan, Australia,

New Zealand, and South Korea. However, 70% of all cancer-related deaths were associated with developing countries in Africa, Asia, Central and South America. By 2025, it is predicted that 20 million new cancer cases could occur, specifically in low-income countries [15]. In Saudi Arabia, there has been a significant increase in the number of female BC patients from 1,152 in 2008 to 1,473 in 2010 [16]. BC screening in Saudi Arabia is found to be very low [17]. Early detection of BC may enhance treatment options. Avoiding risk factors associated with BC and increasing the awareness of BC through educational campaigns could halt the increasing number of new BC cases.

1.1.4. **Risk factors for breast cancer**

The exact mechanisms of BC are not fully comprehended. Generally, BC-related risk factors could be either hereditary, environmental, or related to lifestyle. In the following sub-sections, we will introduce some of the risk factors associated with BC incidence or progression.

1.1.4.1. *Exposure to oestrogen*

Hormonal factors are known to be the main cause of BC initiation and progression. However, BC also occurs in menopausal women. Oestrogen is the primary female sex hormone that moderates a myriad of biological functions. Exposure to oestrogen from oral contraceptive medications or menopausal hormone therapies could stimulate breast growth, hence increasing BC risk [18]. It is stated that the combination of oestrogen and medroxyprogesterone (MPA), a form of progesterone that regulates ovulation, increases the risk of invasive BC in menopausal women [19]. Pregnancy and longer duration of breastfeeding can break the exposure to oestrogen and decrease the incidence of invasive BC via the protective effect of lobule differentiation [20]. Occurrences of BC in full-term pregnancy and lactating women are significantly less than in women who have never been pregnant, nulliparous [21].

1.1.4.2. *Age*

Age is another crucial factor that is correlated with an increased risk of BC in women. The risk of developing BC is statistically greater in women aged over 50 in comparison to younger patients; albeit, younger patients suffer from an aggressive phenotype of advanced-stage BC such as human epidermal growth factor receptor-2 (HER2)-enriched and triple-negative, which are strongly related to poor survival [22, 23].

1.1.4.3. *Family history and genetic predisposition*

Hereditary BC roughly denotes 10% of all BC incidences [24]. For some women, BC is linked to their family history. Women with a first- or a second-degree relative, mother, daughter, or sister who has had BC, or with multiple relatives who have suffered from ovarian cancer (OC) could be at a higher risk of developing BC. Germline mutations in the genes breast cancer type-1 susceptibility protein (*BRCA1*) and breast cancer type-2 susceptibility protein (*BRCA2*) were previously reported to correlate with a large number of hereditary BCs. A recent study has confirmed that BRCA1/BRCA2 mutation accounts for 25% of family risk [25]. Mutated BRCA1/BRCA2 BCs tend to be bilateral diseases with early onset [26]. Individuals carrying mutations in BRCA1/BRCA2, besides being a risk factor for BC, are susceptible to cancers of the digestive system such as colon, pancreatic, and gastric carcinomas [27, 28].

1.1.4.4. *Alcohol consumption, smoking, and obesity*

Alcohol intake is also related to an increased risk of BC in women. According to the International Agency for Research on Cancer (IARC), daily consumption of 10 grams (gm) of alcohol in a drink can increase the risk of developing BC in adult women by up to 10% [29]. A relationship was identified between high and regular alcohol intake and increased breast mammographic density [30]. Longer smoking duration is also a risk factor associated with BC. Active smokers show high susceptibility to invasive BC compared to passive smokers. Passive smokers are also at risk of developing BC compared with women who have never smoked or been exposed to passive smoking [31]. Being overweight or obese is also associated with an increased risk of BC in postmenopausal women [32]. The association between obesity and developing BC could be attributed to the aromatase-driven oestrogen production in adipose tissue. In addition, obesity is associated with increased secretion of inflammatory cytokines, insulin-like growth factor I, and circulating insulin, all of which are associated with BC progression [33].

1.1.4.5. *Breast density*

Dense breast tissues signify an abundance of glandular and supportive fibrous tissues compared to fatty tissues. Dense and non-dense breast tissues can be distinguished by mammogram. Mammographic breast density is a robust predictor of BC risk that labels women with dense breast tissues at higher risk of developing BC compared to individuals with lower breast densities [34].

1.2. Prognostic and predictive factors in BC

The definition of prognostic factors in BC refers to any tumour characteristic that is capable of predicting the patient's outcome measured in terms of BC-specific survival (BCSS), distant metastasis-free survival (DMFS), and disease-free survival (DFS). On the other hand, predictive factors are any clinicopathological-based features of BC that can estimate the probability of a response to a specific treatment. Specific clinicopathological features may illustrate prognostic and predictive qualities. Identifying accurate prognostic and predictive parameters could aid in the selection of therapeutic options and improve outcomes. In the following sub-sections, we will present the most commonly used prognostic and predictive factors of BC.

1.2.1. **Prognostic factors**

1.2.1.1. *Phenotypes of BC*

As a heterogeneous disease, BC manifests as different phenotypes with distinguishable clinical relevance. The gene expression patterns of BC were described early in 2001 [35]. By performing DNA microarray experiments and hierarchical clustering analysis, two distinct sets of hormonal receptor tumours were identified: 1) hormonal receptor-positive (HRT^+), and 2) hormonal receptor-negative (HRT^-). The considered hormonal receptors were oestrogen receptor (ER) and progesterone receptor (PR). Further clustering of the HRT^+ set identified two other subgroups based on their gene expression profiles that resemble luminal epithelial mammary cells. On the other hand, the clustering analysis of HRT^- produced three distinct tumour classes: 1) HER2-amplified, 2), basal-like, and 3) normal breast-like. These phenotypes were denoted according to their ER, PR, and HER2 expression profiles as luminal-A (ER^+, $PR^{+/-}$, $HER2^-$), luminal-B (ER^+, $PR^{+/-}$, $HER2^+$), $HER2^+$ (ER^-, PR^-, $HER2^+$), and basal-like (ER^-, PR^-, $HER2^-$) [35, 36]. Normal-like BC was not depicted as a true subtype. Later, the basal-like BCs were distinguished from the non-basal triple-negative by the positive immunohistochemistry (IHC) expression of type II cytoskeletal 5 and 6 ($CK5/6^+$) and/or epidermal growth factor receptor ($EGFR^+$) [37]. Basal-like BC exhibits distinctive biological features to those of non-basal triple-negative BC.

In a different approach, a comprehensive genomic and transcriptomic profiling of 2,000 BC samples stratified the studied cohort into 10 integrative clusters (IntClusts). Each IntClust illustrated different clinical and molecular features, genomic landscapes, and unique ER, PR, and HER2 expression profiles [38]. BCSS, up to ten years after initial treatment, was associated with good outcomes in IntClust-3, IntClust-4, IntClust-7, and IntClust-8. Moderate BCSS was related to patients in IntClust-1, IntClust-6, and IntClust-9. Poor BCSS was

correlated with IntClust-2, IntClust-5 and IntClust-10 [39]. The molecular features for the IntClusts are: 1) IntClust-1, 17q23 amplification; 2) IntClust-2, 11q13/14 amplification; 3) IntClust-3, paucity of copy number changes; 4) IntClust-4, copy number devoid; 5) IntClust-5, HER2 amplification; 6) IntClust-6, 8p12 amplification; 7) IntClust-7, 16p gain, 17q loss, 8q amplification; 8) IntClust-8, 1q gain, 16q loss; 9) IntClust-9, 8q gain, 20q amplification; 10) IntClust-10, 5q loss, 8q gain, 10p gain, 12p gain [39].

1.2.1.2. *Tumour grade*

Tumour grade is a prevailing prognostic indicator of a three-tier grading scheme that assesses the degree of resemblance between BC cells and normal breast cells, and how BC cells are organised in a malignant lesion. By definition, a tumour grading system has three distinctive categories: 1) tumour grade-I – BC cells appear very similar to normal breast cells, are well differentiated and slow growing; 2) tumour grade-II – BC cells do not seem like normal breast cells, are differentiated and grow rapidly; and 3) tumour grade-III – BC cells look extremely different to normal breast cells, are poorly differentiated and grow rapidly [40]. If the grade cannot be determined, some pathologists may denote that corresponding patient with grade-X [40]. Other reports add a fourth grading category, grade-IV, when the cancer is high grade but undifferentiated. The calculation methods for tumour grading depend on three components. The first component is tubule or acinus formation: score=1 if majority of tumour (>75%), score=2 if moderate degree (10–75%), and score=3 if little or none (<10%). The second component is the nuclear pleomorphism: score=1 if small or regular uniform cells, score=2 if a moderate increase in size and variability, and score=3 if marked variation. The third component is mitotic count: score=1 if mitotic count $\leq$9, score=2 if 10$\leq$ mitotic count $\leq$19, and score=3 if mitotic count $\geq$ 20. The counting of mitosis is performed in ten high-power field using light microscopy with a field diameter of 0.58mm. If the examined tumour section is digitally scanned, the image area should be 512×512μm^2. By summing each component for a case, a tumour grade is assigned. Tumour grade-I has a total score of 3, 4, or 5 and a good prognostic indication. Tumour grade-II has a total score of 6 or 7 and a moderate prognostic indication. Tumour grade-III has a total score of 8 or 9 and a poor prognostic indication [40, 41]. This tumour grade system was integrated with the Nottingham Prognostic Index (NPI) [42].

1.2.1.3. *Tumour size*

Tumour size is considered to be one of the most significant factors in predicting BC behaviour and patient survival. Tumour size shows important

prognostic implications in invasive BC [43-45]. Tumour size is stated as a strong predictor of nodal metastasis in invasive BC [46]. In this study, tumour size is binarised into two groups of small and large tumours. If the tumour size ≤20mm, a small tumour is considered, and if the tumour size >20mm, a large tumour is considered. Tumour size is integrated with NPI [42].

1.2.1.4. *Lymph node stage*

The lymph node (LN) stage is also regarded as one of the robust prognostic factors in invasive BC, especially for BC at an early stage, as there is a strong correlation between LN stage and DFS and BCSS [47, 48]. The applied pathological system for LN staging in this study divided the studied cases into three distinctive categories: 1) LN stage-I, if no LN is involved; 2) LN stage-II, if one, two, or three LNs are involved; and 3) LN stage-III, if more than three LNs are involved [49, 50]. There is a direct proportion between the increased mortality among invasive BC patients and their LN stage status as better survival is associated with a lower LN stage and vice versa. The percentage of BC patients with LN stage-I who survived for a five-year period is around 83% [48]. This LN staging system is integrated with NPI [42].

1.2.1.5. *Nottingham Prognostic Index*

The NPI is viewed as a survival-setting and prognostic stratification tool for invasive BC patients [49]. As explained above, NPI comprises tumour grade, tumour size, and LN stage as three robust prognostic parameters. Therefore, NPI provides an enhanced prognostic power for invasive BC patients. NPI has three distinctive groups: 1) good-NPI, 2) moderate-NPI, and 3) poor-NPI. The calculation of NPI grouping for each patient is based on a formula that includes their clinicopathological scores/data of tumour grade, tumour size, and LN stage. The NPI is calculated for each patient by applying this formula:

NPI = [0.2 × tumour size (in centimetres)] + LN stage (1 to 3) + tumour grade (1 to 3)

The original NPI grouping assigns patients with a calculated NPI of ≤3.4 in the good prognosis-NPI group. Patients with NPI of 3.4–5.4 are grouped into the moderate prognosis-NPI group. Finally, patients with NPI >5.4 are grouped into the poor prognosis-NPI group [42]. An improved NPI grouping system is also suggested by sub-dividing the aforementioned three NPI prognostic groups into six groups. The improved NPI groups are: excellent prognostic group (EPG) if NPI is 2.08–2.4; good prognostic group (GPG) if NPI is 2.41–3.40; moderate prognostic group-I (MPG-I) if NPI is 3.41–4.40; moderate prognostic group-II (MPG-II) if NPI is 4.41–5.40; poor prognostic group (PPG) if NPI is 5.41–6.40;

and very poor prognostic group (VPG) if NPI is >6.40 [51]. We apply the original NPI grouping system to all cases in this study.

1.2.1.6. *Lymphovascular invasion*

LVI is an independent prognostic factor that is strongly correlated with BCSS and DMFS in invasive BC patients [52]. LVI denotes the presence of tumour cells within lymphatic spaces, blood vessels, or both at the peritumoural area and is identified morphologically by microscopic examination of the primary tumour with or without endothelial-specific markers. Detection of LVI in the primary tumour is a marker of metastatic potential. The prognostic value of LVI has been demonstrated by several independent studies [52-54]. In this study, we focus on identifying the molecular drivers for LVI in a large cohort of invasive BC patients on the mRNA and protein levels of expression. The successful identification of LVI-related genes and proteins, which could provide independent prognostic information and pinpoint high-risk subgroups of invasive BC patients, may enhance therapeutic strategies in the future. The underlying molecular mechanisms of LVI are discussed in depth in section 1.3.

1.2.2. **Predictive factors**

The predictive factors or biomarkers help to predict prognosis and response to treatment. The predictive biomarkers are crucial for the management and stratification of BC patients, especially when they are combined with the clinicopathological data of the prognostic parameters discussed above [55].

1.2.2.1. *Hormone receptors, ER and PR*

According to the guidelines from the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP), the recommended thresholds of positive detection for ER and PR in invasive BC patients was set at 1% of tumour nuclei [56]. ER provides a powerful predictive indication for extended adjuvant endocrine treatment; however, the extended duration of a hormonal therapy regimen for ER^+ patients is still debated [57]. Practically, considering the prognostic relevance of ER positivity in post-menopausal BC patients could be more beneficial. Younger patients are usually affected by aggressive BC phenotypes; however, considering ER status would marginalise the effect of age as an independent prognostic factor and allow the patient to benefit from intensive chemotherapy treatment [58, 59]. The co-expression of PR in ER^+ patients is also considered in routine clinical practice. The co-expression status of ER and PR has been used to predict patients' responses to adjuvant hormonal therapy. When ER and PR are co-expressed, the prediction of hormonal therapy

response reaches 80%, while it is only 50% in ER^+/PR^- patients [58]. Together, ER and PR could be associated with the invasion and metastasis of BC as they correlate with the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) and the tumour necrosis factor alpha (TNF-α) pathways [60].

1.2.2.2. *HER2*

The HER2 protein is a receptor tyrosine-protein kinase erbB-2 encoded by the *ERBB2* gene. HER2 is overexpressed in approximately 15% of BC patients, and it is strongly correlated with aggressive phenotypes of invasive BC and poor survival [61]. HER2 positivity marks discohesive BC cells undergoing epithelial-mesenchymal transition (EMT) and illustrates oncogenic potentials [62]. The strong association between HER2 positivity and LVI in ER^+ invasive BC patients was identified [63]. HER2 positivity in invasive BC patients is associated with aberrant cellular polarity, malignant growth, invasion, and unfavourable outcomes [64]. Targeting $HER2^+$ tumours with Trastuzumab, a monoclonal antibody that is used to treat $HER2^+$ BCs, was previously established; however, the best effect was relevantly restricted to early-stage BC patients [65]. A recent report states that LVI was associated with 30% of a large cohort of invasive BC patients who received neoadjuvant chemotherapy. In these patients, LVI was associated with poor DFS, BCSS, and DMFS in all invasive BC subtypes; however, remarkably, with $HER2^+$, the hazard ratio (HR) = 6.21 [66].

1.3. Molecular mechanisms underlying LVI in invasive BC

1.3.1. **Introduction to LVI**

BC is recognised as a heterogeneous disease with varied presentation, morphology, behaviour, and response to therapy. Although a minority of BC patients (<10%) present with metastatic disease, 20–30% of patients presenting with early-stage disease will develop distant metastases over time [67-69]. Distant metastases, which are considered to be poor prognostic events in cancer, develop through coordinated and highly selective biological processes and interacting pathways involving several related mechanisms, including cell proliferation, cell-matrix interactions, stromal invasion, cellular migration, LVI, and evasion of host immune responses [70-72].

In a previous study of 3,812 cases of BC [52], it was demonstrated that LVI is not only an independent prognostic variable in the whole series but also in the various prognostic subgroups, including the lymph-node-negative cohort. In this subgroup, LVI could be used as a high-risk criterion conferring survival disadvantage equivalent to that provided by the involvement of one or two lymph

nodes and to that provided by one higher size category (pT1 to pT2) [52]. At the molecular level, there are several biological interactions that cancer cells utilise to facilitate their metastatic progression through vascular and lymphatic channels [53, 73-75]. Exploration of these signalling pathways is crucial to finding the novel molecules controlling the critical steps of LVI. The hallmarks of cancer have been elaborated extensively and they all point to the dominant intrinsic and extrinsic qualities of the tumour cells that govern their genetic regulation [7]. Elevated activity of protein expression, resistance to apoptotic signals, increased proliferation rate, adhesion, invasion, and cellular motility leading ultimately to cellular migration are some characteristics of malignant cells [76]. Although these criteria are not directly related to the occurrence of LVI and can be identified in LVI^- and LVI^+ cancers, they are critically required traits for its development, and the molecular mechanisms controlling these processes must remain active, at least in some cellular clones, throughout the carcinogenic pathway from initiation to dissemination through vascular channels. In this chapter, the clinicopathological features of LVI in BC and the molecular alteration characteristics of invasion and LVI are reviewed.

1.3.2. Morphological assessment of LVI

The morphological assessment of LVI has been performed traditionally in the peritumoural areas. This is based predominantly on the morphological ease of detecting LVI outside the tumorous tissue, avoiding the controversy regarding the existence of intratumoural lymphatics, the presence of tumour retraction artefacts, the prognostic significance of peritumoural rather than intratumoural LVI, and the rarity of pure intratumoural LVI [77-80]. In our own laboratory, we have also observed LVI in both peritumoural and intratumoural vessels [78]. In LVI-positive cases, three quarters had peritumoural LVI and almost one quarter had intratumoural LVI. Of the intratumoural LVI-positive cases, only a handful (~4%) had LVI in the intratumoural compartment alone. Such small numbers can only lead to tenuous conclusions regarding the significance of intratumoural LVI. This leaves us with the conundrum of whether or not intratumoural LVI is important. In cases in which there is intratumoural LVI alone but with positive lymph node status, the intratumoural LVI may explain the metastatic cascade. However, when such node involvement is not apparent, care should be taken before disregarding the value of intratumoural LVI. The pragmatic approach would be to acknowledge that the tumour has found its mechanisms of spreading into the vasculature and is, therefore, worth clinical vigil. There are several lines of evidence to indicate the occurrence of LVI in breast tumours lacking morphological

evidence of LVI such as the demonstration of circulating tumour cells [81], bone marrow micro-metastases [82], and lymph node positivity [83]. This may indirectly influence our ability to understand the molecular mechanisms driving LVI in BC since identification of metastatic potential in cases lacking LVI will be challenging.

In addition, there may be a possibility that molecular alterations identified in LVI-positive tumours compared to LVI-negative cases reflect the quantity and not the quality of the biological processes. In other words, LVI in BC may comprise a biological continuum akin to other morphological surrogates of tumour biology, such as tumour grade; at one end of the spectrum are cases lacking LVI and at the other end are cases with extensive LVI (Figure 1.1). In addition to the complexity of the molecular mechanisms underlying LVI and their overlap with the metastatic cascade of invasion and the lack of suitable models to investigate BC cell intravasation, this further complicates the identification of driver molecules, pathways, and potential therapeutic targets associated with LVI in BC.

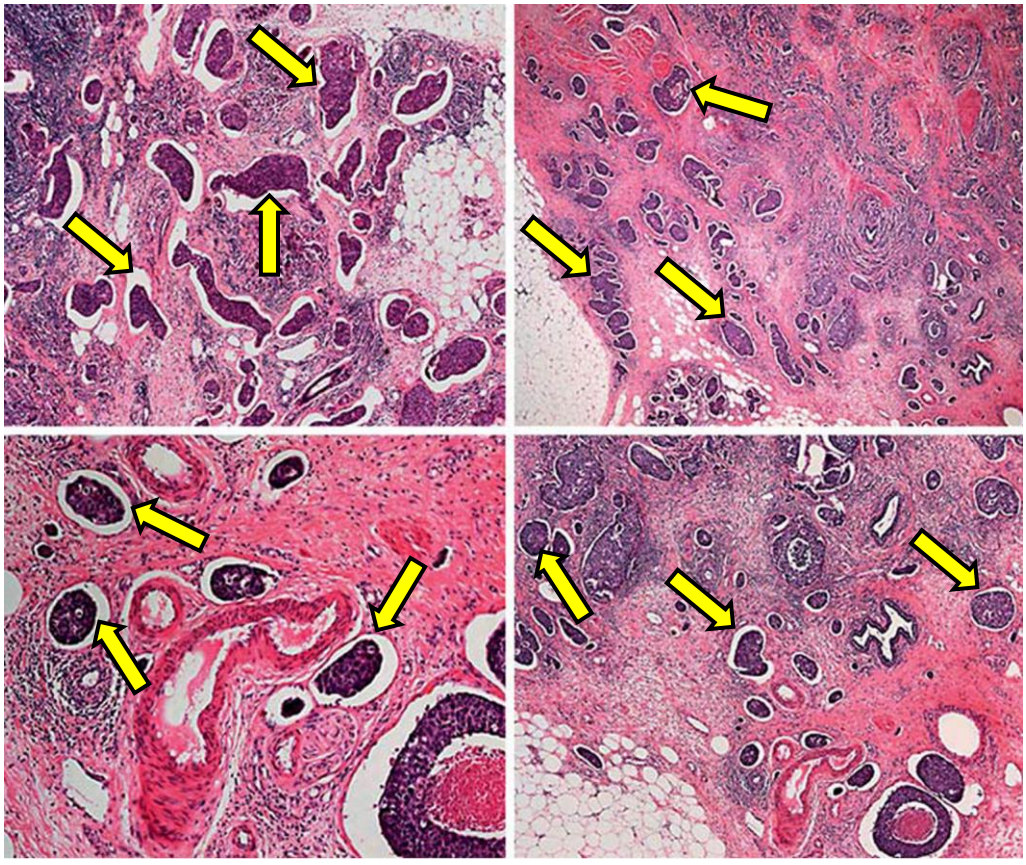


Figure 1.1: case of invasive BC of no special type with extensive LVI, yellow arrows, showing tumour cell emboli within vascular and lymphatic spaces in the peritumoural area. H&E staining.

1.3.3. **Tumour histological features and LVI**

Certain tumour morphological features are associated with LVI, including tumour type, grade, and size. One of the characteristic features of inflammatory BC (IBC) is extensive LVI, particularly the involvement of dermal lymphatics [84]. This extensive lymphatic spread of IBC is associated with high lymphatic endothelial cell proliferation [85]. It was reported that the lymphovascular tumour emboli in an IBC human xenograft mouse model formed cohesive spheroids that resembled embryonal blastocyst [86]. Cells in these spheroids co-expressed E-cadherin along with the stem cell phenotype $CD44^{+}$ /$CD24^{-/low}$ and were able to generate tumour nodules exhibiting florid LVI in vivo with subsequent development of metastases irrespective of their site of injection and molecular subtype [87]. Interestingly, different authorities reported the cooperative role of E-cadherin overexpression and marked a reduction in sialyl-Lewis X/A antigen by the intravascular emboli, rendering these emboli both coherent to each other yet poorly adherent to the E-selectin of the vascular endothelium. Accordingly, the tumour cell-endothelial cell aversion contributes towards the passive propagation of IBC spheroids in metastatic dissemination [88, 89]. Invasive micropapillary carcinoma (IMPC) is another rare subtype of invasive BC characterised by the frequent development of LVI [90]. Through exome and RNA sequencing of IMPC, and compared to IDC, genomic aberrations have been identified for several protein-encoding genes that could be responsible for the abnormal cellular polarity, cytoskeletal organisation, and primary cilium biogenesis including formin-2 (*FMN2*), translocation protein SEC63 (*SEC63*), dynein heavy chain-9 (*DNAH9*), and zinc finger FYVE domain-containing protein-26 (*ZFYVE26*) [91]. It is also important to note that IMPCs show a high frequency of LVI regardless of the proportion of micropapillary components in an invasive tumour [92], suggesting an intrinsic biological feature of this tumour type that promotes the development of LVI.

In contrast, invasive lobular carcinoma (ILC) is known to show a low frequency of LVI, even when LVI was detected using lymphatic endothelial-specific biomarkers [93], despite having an equal frequency of nodal involvement and a tendency to metastasise to more nodes than grade-matched IDC [94]. The disproportionate balance between the metastatic rate and the frequency of morphologically detected LVI in ILC can be explained by the discohesive single-cell infiltration pattern of ILC with the inherited loss of E-cadherin resulting in intratumoural or peritumoural intravasation of single small uniform cells rather than the easily identifiable cohesive clusters of IDC cells. In the rare pleomorphic variant of ILC, LVI is more detectable, although it shares the same molecular alterations with the classical ILC [95-97]. Benign papillary lesions of the breast

are a unique type of benign breast tumour that may show LVI. This phenomenon is likely to be related to their inherent friability resulting in epithelial intravascular displacement following needle instrumentation procedures rather than biological features associated with metastatic potential [98].

There is a direct correlation between LVI and tumour grade and primary tumour size. For instance, LVI is rarely seen in grade-I tumours while up to one-half of grade-3 tumours are associated with LVI [52, 99, 100]. Similar associations are reported with tumour size [101]. Although infrequent, detection of LVI in small BC is an independent predictor of node metastases and a poor outcome [101, 102], which may reflect the aggressive nature of these tumours. There is also a correlation between LVI and BC molecular subtype. Triple-negative or basal-like molecular classes show less association with LVI despite their poor prognosis and a higher tendency of local recurrence [103]. In contrast, luminal and $HER2^+$ classes show the highest incidence of LVI [104]. One explanation attributes this difference to the lower expression of the tight junction proteins, claudins, and the higher expression of proteins that are crucial to transforming the malignant epithelial cell to the mesenchymal form in triple-negative and basal-like tumours compared with luminal and $HER2^+$ BC classes [105, 106]. LVI is more frequent in symptomatic compared to screen-detected BC and in young premenopausal BC patients [52, 100]. Therefore, studies on the molecular determinants of LVI should endorse the appropriate choice of cohorts as well as the control samples for more reliable identification of LVI-controlling genes.

1.3.4. Vascular or lymphatic invasion

Invasive BC can disseminate through blood vessels or lymph vessels; however, current evidence indicates that lymph vessels rather than blood vessels play the major role in the initial dissemination of BC, which may be related to the intrinsic tropism of the malignant cells towards lymphatic channels [78]. A recent report states that blood vascular invasion (BVI), not LVI, is strongly associated with the recurrence of endometrial cancer (EC) [107]. Structural differences between lymphatics and blood vessels expectedly determine the initial route of spread and, in this regard, lymphatic capillaries lack the tight junctions typically present between endothelial cells lining blood capillaries [108]. Accordingly, this renders lymphatics ectatic and leakier than blood vessels, therefore easing tumour cell intravasation. In other words, peritumoural lymphatic vessels may represent the primary route for the intravasation pathway for invasive BC cells incapable of overcoming the physical restrictions of vascular endothelial cells [109].

It has also been suggested that the interstitial fluid pressure is responsible for the directional motility of the metastatic malignant cells toward the lymphatics at the tumour periphery, thus passively increasing the chance of their intravasation into lymphatic channels [110]. Nevertheless, active molecular mechanisms attracting malignant cells more towards either type of vessel is another likely explanation [111]. Expression of lymphangiogenic factors may influence the route of metastatic spread through the lymphatic vessels due to their specific structure, and this may support the tumour growth [108, 112]. For instance, vascular endothelial growth factor receptor-3 (*VEGFR3*) is exclusively expressed on lymphatic endothelial cells, and its ligand vascular endothelial growth factor-C (VEGFC), which is overexpressed by malignant cells, influences the process of lymphangiogenesis. However, some studies did not find significant associations between VEGFC and LVI or node metastasis, but this may be attributed to the probable role of multiple factors besides VEGFC and vascular endothelial growth factor-D (VEGFD) and the heterogeneity of the study samples [113]. Regardless of the nature of the involved vessels, whether blood or lymphatic, native or newly formed, it is imperative to be clear that neither angiogenesis nor lymphangiogenesis is equivalent to LVI despite their strong associations. Angiogenesis is the development of neovasculature to sustain tumour growth and promote invasion, whereas LVI is the process whereby tumour emboli lodge into vascular/lymphatic channels early in the metastatic cascade. Although the presence of the vasculature is necessary to provide the fertile ground for LVI, it does not equate to LVI itself.

1.3.5. Tumour cell invasiveness, tumour microenvironment, and LVI

Before BC cells access the vascular lumina, they have to detach themselves from the growing tumour cell mass, followed by invasion and migration through the extracellular matrix (ECM) towards the vascular walls. These steps are, therefore, considered prerequisites for the LVI process. Malignant cells infiltrate the surrounding frontline of ECM as single cells or as clusters of malignant cells. BC cells on the tumour boundaries upregulate genes such as *SNAIL*, *SLUG*, and *ZEB1*, and other transcriptional factors that accelerate the potential transformation of the malignant epithelial cells to the mesenchymal-like phase through repression of E-cadherin and other cell-cell adhesion molecules, which acquire more motility [114, 115]. The tumour microenvironment controls the mechanobiological interactions between malignant cells and the components of the stroma [116]. The interaction, through genetic and epigenetic regulations between malignant cells and the stromal components of the tumour

microenvironment (TME), has been reported to alter adhesion and promote invasion and migration of malignant cells through the adjacent ECM with a subsequent increase in their penetration into vascular channels [7].

Stromal cells assist in proteolytic degradation via matrix metalloproteinase (MMP) dictated by the epigenetic regulations from malignant cells. Strong correlations have been reported between LVI and MMP9 [117, 118], and MMP1 expression [119]. Overexpression of cancer cell-associated membrane type-1 (MT1)-MMP or MMP14 by BC cells has been reported to induce degradation of the vascular basement membrane and cancer cell intravasation [120]. It has been demonstrated that MMP14 expression in cancer cells induces its expression in vascular endothelial cells and leads to vascular basement membrane remodelling and blood vessel invasion but not lymphatic vessel invasion. It was suggested that MMP14 should be considered as a predictor and therapeutic target of haematogenous metastasis in triple-negative BC patients [121]. Podoplanin is a glycoprotein that encoded by the *PDPN* gene and expressed in the lymphatic endothelium. Together with other markers such as LYVE-1 and Prox-1, podoplanin is considered to be a specific lymphatic endothelial marker. Podoplanin is also expressed by cancer-associated fibroblasts (CAFs) and is known to boost cellular motility and reduce cellular adhesion. Some studies have reported an association between podoplanin expression by CAFs in BC and aggressive features, including LVI, node metastasis, and poor outcome [122]. Podoplanin expression in CAFs is also correlated with BC cell VEGFC expression and intratumoural microvessel count [123]. Interestingly, high podoplanin expression at the invasive front of BC plays an alternative role in invasion and metastasis by forming invasive protrusions and filopodia without activation of endothelial-mesenchymal transition [122]. This phenomenon could be considered a potent synergism between metastasis-related paracrine signalling from CAFs and epigenetics from malignant cells to facilitate lymphatic vessel invasion.

Possibly one of the best examples illustrating the concert played by malignant cells, vascular endothelium and dendritic cells in LVI is the C-C chemokine receptor type-7 (CCR7), a lymphatic-homing chemokine receptor. It has been reported that CCR7 and VEGFC expression by tumour cells acts synergistically to promote LVI [124]. The transcellular lymphatic drainage that is caused by the elevated interstitial flow of lymphatic fluids may also polarise malignant cells toward the nearest lymphatic vessel through the autologous secretion of CCR7 [110]. The ligand of CCR7 (CCL21), C-C motif chemokine 21, is located on the vascular endothelial surface. Under the influence of VEGFC secreted

by malignant cells and the upstream signalling of its receptor VEGFR3, the expression of CCL21 by vascular endothelium is significantly elevated. Therefore, $CCR7^+$ malignant cells will actively migrate toward lymphatic vessels and, thus, the chances of LVI increase significantly. In other words, this illustrates how the malignant cells can intensify and utilise the chemokinetics of lymphangiogenesis to direct themselves to the preferential invasion target.

Although the presence of macrophages in primary tumours is associated with increased metastatic potential and development of LVI, the mechanistic basis for this observation remains unknown. It is reported that the interaction between tumour cells and macrophages facilitates the migration of carcinoma cells [125]. EGF and/or colony-stimulating factor 1 (CSF-1) expression can stimulate a collection of tumour cells and macrophages into microneedles while inhibition of either EGF or CSF-1-stimulated signalling reduces the migration of both cell types [125]. Overexpression of the EGF receptor by mammary tumour cells renders cells chemotactic to EGF released by macrophages lining blood vessels [125]. This work provides evidence of a synergistic interaction between macrophages and tumour cells during cell migration in vivo and could explain, at least in part, the increased tendency of basal-like BCs, which overexpress the EGF receptor, to be more node negative and show haematogenous rather than lymphatic dissemination to viscera and bone [126, 127]. Macrophages are well known to produce interleukin-10 and angiogenic factors [128]. Within the hypoxic environment, growth factors that promote angiogenesis and lymphangiogenesis are upregulated and secreted by malignant cells [129] and serve as chemo-attractants for macrophages. In a different study, an association between perivascular macrophages and malignant cell intravasation was demonstrated [130]. Recruited macrophages promote invasion and metastasis through a dual-direction reaction loop of paracrine signalling with malignant epithelial cells. Therefore, intravasation into vascular channels and migration through the lumen of vessels is facilitated [131]. Accordingly, the use of immunomodulatory drug therapy may interfere with tumour-induced immunosuppression mechanisms and may provide useful adjuncts to non-specific cytotoxic drugs [132].

Vasculogenic mimicry or vascular mimicry (VM) was so termed to emphasise the de novo generation of microvascular channels without participation by endothelial cells independent of angiogenesis. VM was initially introduced and demonstrated in vitro in highly aggressive and primary and metastatic melanoma cells [133]. The generation of these highly patterned microvascular channels by genetically deregulated aggressive tumour cells occurs in the absence of

endothelial cells and fibroblasts [134]. It was subsequently identified in a xenograft IBC model and cell lines [135, 136]. The presence of VM in tumour tissues has been reported to be associated with a poor clinical outcome and node metastases in patients [137]. This poor prognostic impact of VM was attributed to the possible advantage of the survival of aggressive tumour cells and the potential facilitation for metastatic tumour cells to access the bloodstream and, subsequently, distant organs [138]. Diverse molecular biomarkers have been reported to be overexpressed in VM BC cells, including VEGF, COX2, TWIST2, and molecules conferring the cancer stem cell phenotype and those undergoing endothelial-mesenchymal transition [139].

1.3.6. Gene expression profiling, gene signatures, and LVI

Identification of genes differentially expressed between LVI-positive and LVI-negative BC may help identify key genes and pathways driving LVI. With the advancement of high-throughput molecular and bioinformatics techniques, it has been possible to investigate genes associated with LVI on a large scale [140]. One of the main problems in the identification of LVI-associated genes in genome-wide gene profiling studies is the nature of the cohorts investigated (as described above) and the use of archival frozen tissue with inaccurate or incomplete LVI data. In a previous LVI gene-signature study of endometrial carcinoma, 18 genes were identified to be associated with LVI [141] and subsequently applied to BC utilising 11 open-access gene expression datasets comprising data on more than 2,400 BC patients [142]. This 18-gene signature was found to be significantly associated with morphologically detected LVI and other poor prognostic features in BC, including high grade, advanced stage, hormone receptor negativity, HER2 positivity, and basal-like phenotype [142]. Candidate genes included seven upregulated and 11 downregulated genes, both spanning a diverse array of genes controlling ECM collagens, ECM proteolysis, interleukins, angiopoietin, tumour necrosis factor, and fibroblast growth factor receptors. Although this gene signature is associated with LVI in BC, it appears to reflect features characteristic of tumour cell aggressiveness and interactions with the surrounding microenvironment in general rather than specific molecular mechanisms driving the process of LVI. Our efforts in studying LVI mechanisms in invasive BC has identified SEC14L1, ARHGAP18, KLHL7, and MED7 as novel LVI-associated proteins.

Another study investigated 214 hepatocellular carcinomas (HCC) to develop the gene signature that predicts the presence of LVI in patients with surgically treated HCC [143]. Interestingly, they identified 14 genes upregulated and 21

genes down-regulated with LVI, but none overlap with the 18-gene signature described above [142], though this could be due to different cancer types. Some of the genes, e.g. *CD24*, *CDKN3*, *UBE2C*, and *PIK3R1*, play a role in BC development and progression [143]. In a recent study, array comparative genomic hybridisation (a-CGH), quantitative reverse transcription-polymerase chain reaction (RT-qPCR), and complementary DNA (cDNA) microarray were applied to identify DNA copy number aberrations (CNA) and the gene expression profile linked to LVI in a cohort of invasive BC [119]. Although they did not find a significant association between LVI and increased genomic instability considering the total number of chromosomal alterations, they identified an increased frequency of specific alterations associated with LVI, including gains at 5p, 16p, 17q12, and 19, and losses at 8p, 11q, 18q, and 21. They also detected three novel small-scale rearrangements in LVI-positive grade-III tumours that harbour putative BC genes, including amplicons at 11p11.2, harbouring very-long-chain 3-oxoacyl-CoA reductase (*HSD17B12*); 20q13.2, harbouring zinc finger protein 217 (*ZNF217*); 19q12, harbouring G1/S-specific cyclin-E1 (*CCNE1*); and 4q13.3, harbouring a disintegrin and metalloproteinase with thrombospondin motifs-3 (*ADAMTS3*). Furthermore, they detected a deletion at 12p12.3, harbouring Ras-related and oestrogen-regulated growth inhibitor-like protein (*RERGL*) and phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit gamma (*PIK3C2G*). Gene expression analysis revealed networks highlighting interstitial collagenase (*MMP1*), mediator of RNA polymerase II transcription subunit-1 (*MED1*), and protein S100-A8 (*S100A8*) as potential candidate genes involved in LVI-positive grade-3 tumours [119]. Genes that showed a concordant pattern of upregulated gene expression and copy number gains in LVI-positive samples were uromodulin (*UMOD*), arylsulfatase G (*ARSG*), MYCBP-associated protein (*MYCBPAP*), and *MED1* [119]. Despite the promising results of this study, it is important to note that the BC cohort was small (n=57), and frozen tissue was utilised with no further validation of LVI status. The cohort included a high proportion of LVI-positive cases (44%) with more LN^+ (65%) and ER^- (42%) cases, which may influence the interpretation of the final results.

Although data on the role of microRNA (miRNA) in BC is conflicting, some authors have demonstrated an association between LVI and levels of some miRNAs such as miR-9-3 [144], miRNA-205 [145], and miR-10b [146]. Applying analytical techniques of bioinformatics to identify target genes to a well-selected and empowered BC cohort stratified based on strict histopathological criteria for LVI may be a promising approach. Only cases with definite LVI should be considered and, as a control, the LVI-negative group should include cases with no

morphological or biological evidence of LVI such as circulating tumour cells or positive LN ideally in the discovery and validation sets. Candidate genes should be analysed based not only on their differential expression but also on their biological function, their functional interactions, and specific molecular pathway involvements. Correlation with gene CNA may help to identify driver genes. This approach is likely to narrow down the number of potential genes that can be investigated in detail using in vivo models and in vitro functional assays.

1.3.7. **Proteomics and LVI**

Correlations of proteins to LVI have been mostly indirect or incidental to date. In an IHC-based analysis of 177 paraffin-embedded BC sections stained with VEGFA, VEGFC, VEGFD, podoplanin, and CD34, increased expression of VEGFA and VEGFC, but not VEGFD, was associated with significantly higher lymph vessel and microvessel density, nodal and distant metastases, and a shorter survival [147]. This provides indirect evidence that high VEGFA and VEGFC augur a poor prognosis by angiogenesis and lymphangiogenesis and are related to LVI. In another IHC-based study on 77 patients, VEGFC and VEGFD expression in BC was significantly correlated with peritumoural lymph vessel density, which was significantly correlated with LVI and nodal metastases [148]. In another study, VEGFC and COX2 IHC stains on 70 BCs indicated an elevated VEGFC and COX2 expression were associated not only with higher lymph vessel density but also D2-40-positive LVI [149]. Hence, despite other contrasting reports [113], there is enough indirect and direct evidence of the importance of the VEGF pathways in the early metastatic cascade and LVI. Autocrine functions of VEGF on tumour cells include adhesion, survival, migration, and invasion [150]. Moreover, deconstruction of the pathway and its interacting members will further aid in understanding the LVI biology. However, some studies have attempted to discover novel proteins associated with LVI taking leads from gene analysis [151].

The artificial neural network (ANN) approach to supervising a gene expression array of 91 BC patients identified 89 transcripts associated with LVI. Low expression of one transcript, calpastatin, an inhibitor of the calpain proteases, was correlated with intratumoural LVI on PCR and validated on IHC platforms. On the other hand, calpain-1 and calpain-2, which are involved in tumorigenesis and the proteolysis of substrates such as inhibitors of nuclear factor-κB (IκB) and focal adhesion proteins, were not correlated with LVI [152]. Thus, the inhibitor of calpains, calpastatin, is the key regulator of LVI within this pathway. Other protein correlates of LVI have been incidental findings while investigating the roles of novel proteins in BC in general. For example, in a study investigating members of

the mitogen-activated protein kinases in BC, the phosphorylated extracellular-regulated kinase 1/2 (pERK1/2) was associated with a negative LVI status [153]. This correlation was also true for the ER^+ subgroup in this series. Traditionally, LVI has been shown to have an inverse relationship with ER and progesterone receptor expression [154], and molecules such as pERK1/2 may provide the clues to the underlying biology. The investigation into DNA repair proteins in BC has also revealed correlations with LVI. For example, a high expression of X-ray repair cross-complementing protein-6/X-ray repair cross-complementing protein-5 (KU70/KU80) is correlated with a positive LVI status [155]. Members of the small ubiquitin-like modifier protein (SUMO) family have shown correlations with LVI. Although positive LVI was associated with a loss of SUMO nuclear expression, it was associated with a positive cytoplasmic expression of E3 SUMO-protein ligase PIAS1 (*PIAS1*), SUMO-conjugating enzyme UBC9 (*UBC9*), and E3 SUMO-protein ligase PIAS4 (*PIAS4*) [156]. This exemplifies the importance of the SUMO pathways in LVI. On the other hand, poly [ADP-ribose] polymerase-1 (*PARP1*), a key facilitator of DNA repair, is not related to LVI in either cleaved or non-cleaved forms [157], nor is the serine/threonine-protein kinase Chk1 (*CHEK1*), a key member of the DNA damage-activated checkpoint signalling responses [158]. The aforementioned findings, though incidental, provide a picture of the DNA response pathways in relation to the LVI-positive phenotype. As increasing numbers of proteins are investigated in BC in relation to invasion, migration, apoptosis, and proliferation, more relationships with LVI will be unearthed.

1.3.8. **Unified Plot of LVI**

The different key mechanisms underlying LVI collaborate progressively under the dictation of genetic, transcriptomic, and proteomic tumour profiles, and the resulting biological interactions are the guarantor of cancer cells to invade vascular spaces and metastasise. Initially, ECM is the field of the vibrant interactions between malignant cells and stromal non-tumoural cells in the event of LVI, and when malignant cells induce significant molecular modifications in the ECM, they will secure their progressive pathways of invasion and, hence, metastasis. These remodelling actions are the synthesis, alignment, crosslinking, and proteolysis of the ECM, and malignant cells will select a combination of remodelling actions to set up the ECM microenvironment to their optimum conditions to invade the targeted vessels and disseminate to distal sites of the body [159]. Fibroblasts, macrophages and other immune cells, and endothelial cells are perceived as LVI mediators due to their different contributions of LVI mechanisms. Angiogenesis and lymphangiogenesis are crucial biological processes

that are governed by several chemokines and cytokines in the tumour microenvironment. Hypoxia, the growing tumour mass, and high interstitial fluid pressure are reasons why the formation of new lymphatic/blood vessels is not only needed to maintain accessibility to nourishment supplies and to replace damaged intratumoural vessels but also to increase the contact surface areas and, subsequently, the chances of LVI.

1.3.9. **Conclusions**

Several molecular factors governing the occurrence of LVI have been unravelled, although other major contributors/players still await discovery. Questions around molecular mechanisms driving BC tumour cells to invade vascular spaces and to disseminate remain largely ambiguous. Therefore, large-scale genomic and transcriptomic profiling of histologically validated LVI in BC may potentially yield a high throughput of genes and pathways within which LVI master regulators could exist. The candidate genes composing these signatures and pathways could be further scrutinised using different research methodologies to identify novel therapeutic targets.

1.4. Hypotheses and aims

1.4.1. Hypotheses

1. The manifestation of LVI in invasive BC patients is modulated by specific genes and proteins that can activate or inhibit LVI progression through several converged or parallel pathways.

2. The progression of LVI in invasive BC patients can be limited to specific molecular phenotypes, i.e. luminal-B BC.

3. LVI and LVI-related genes/proteins can identify subsets of high-risk invasive BC patients that require better attention.

1.4.2. Aims and objectives

1. Identify differentially expressed genes that associate LVI in invasive BC patients by methodically analysing high-throughput transcriptomic data produced by RNA microarray.

2. Assess the clinicopathological implications for a selected cluster of LVI-related transcripts in a large cohort of invasive BC patients.

3. Examine a cluster of novel proteins correlated with LVI and evaluate their prognostic features utilising IHC methodologies.

4. Explore the LVI differentially expressed proteins by applying high-performance liquid chromatography/mass spectroscopy (HPLC/MS) procedures, and assess the consistency of results between RNA microarray and HPLC/MS methods.

5. Propose LVI-gene and LVI-protein key signatures that can be used as a source for functional studies in the future.

Chapter 2: Materials and Methodologies

2.1. Study cohorts

2.1.1. The METABRIC subset cohort

A subset of the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort of invasive BC was used to investigate genome-wide expression profiles and identify reliable differential gene expression biomarkers that trigger LVI mechanisms. The METABRIC subset cohort was originally derived from the Nottingham Primary Series (NPS) of invasive BC. Previously, 344 primary invasive BC fresh frozen biopsy specimens were sent from Nottingham City Hospital (in the UK), part of the Nottingham University Hospitals (NUH), and received by Addenbrooke's Hospital, which is affiliated to Cambridge University Hospitals (CUH). This contribution was required for the construction of the complete METABRIC cohort comprising 1,980 patients and for the integrated analyses of gene expression and copy number to reveal new BC subgroups based on their genomic and transcriptomic architecture [38]. The total number of Nottingham's specimens that were included in the METABRIC study with genomic and transcriptomic data was 296. The clinicopathological data for this cohort are presented in Table 2.1.

In order to obtain the DGE profiles of the samples from the METABRIC subset and to assess the molecular regulation of LVI accurately, strict criteria for sample selection were applied to this cohort. First, the LVI status of all patients was evaluated morphologically by staining BC full-face sections with haematoxylin and eosin (H&E) (detailed methods in section 2.3) molecularly by targeting the sialoglycoprotein found on the lymphatic endothelium surface, PDPN, and by confirming the expression of the hematopoietic progenitor cell antigen CD34 (detailed methods in section 2.4). Secondly, the distinction between LVI^+ and LVI^- samples was made on the basis of the observation of malignant cells within the lymphatic lumen; the LN status was considered when the sample was categorised as LVI^-. If the sample showed LN positivity in the latter category, it indicated a preparative artefact or an actual invasion but via another route, e.g. through the vascular endothelium. Therefore, all LVI^-/LN^+ samples were considered as a unique LVI subset, beside LVI^+ and LVI^-/LN^-. In Chapter 4, the IHC evaluation of the selected LVI candidates considered LVI^-/LN^-, LVI^-/LN^+, and $LVI^+/LN^{+/-}$ as three LVI independent categories. For all LVI^+ samples, the LN status was not considered as the LVI was definitely observed. This step was essential to ensure the credibility of sample selection and to nominate molecules associated with LVI^+ primarily.

Clinicopathological characteristics	n (%)
Age	
≤ 50 years	94 (31.8)
> 50 years	202 (68.2)
Menopausal status	
Pre-	102 (34.5)
Peri-	0 (0.0)
Post-	194 (65.5)
Tumour size	
≤ 2cm	170 (57.4)
> 2cm	126 (42.6)
Grade	
Grade-I	37 (12.5)
Grade-II	107 (36.3)
Grade-III	151 (51.2)
Tumour type	
Invasive ductal/NST	212 (71.6)
Lobular	19 (6.4)
Medullary-like	7 (2.4)
Special types, Tubular/Mucinous	53 (17.9)
Miscellaneous	5 (1.7)
Nottingham prognostic index (NPI)	
Good prognostic group (GPG)	106 (35.8)
Moderate prognostic group (MPG)	171 (57.8)
Poor prognostic group (PPG)	19 (6.4)
Lymph Node Stage	
Stage-I	170 (57.4)
Stage-II	98 (33.1)
Stage-III	28 (9.5)
Lympho-vascular Invasion, (H&E/IHC:Podoplanin/CD34)	
Negative	164 (55.4)
Definite	132 (44.6)
Follow-up Status	
Alive	173 (58.4)
Died from breast cancer	74 (25.0)
Died from other causes/unknown	31 (10.5)
Lost to follow-up	18 (6.1)
PAM50	
Normal-like	19 (6.4)
Luminal-A	110 (37.2)
Luminal-B	82 (27.7)
HER2-enriched	37 (12.5)
Basal-like	48 (16.2)
Estrogen receptor status (IHC)	
Negative	70 (23.6)
Positive	226 (76.4)
Progesterone receptor status (IHC)	
Negative	137 (46.3)
Positive	159 (53.7)
HER2 (IHC/CISH)	
Negative	254 (85.8)
Positive	42 (14.2)
Triple-negative (IHC)	
No	248 (83.8)
Yes	48 (16.2)

Table 2. 1: clinical features of the METABRIC sub-cohort utilised in the construction of the LVI-supervised cohort, n =296

The valid number of LVI-supervised ($LVI^{+/-}$) samples in the METABRIC subset that were considered when retrieving CNA data decreased to 170 (LVI^{+}, 90, and

LVI^-/LN^-, 80). The reduction in sample size was due to strict qualitative and quantitative assessments applied to the extracted nucleic acids upon processing and retrieving CNA analyses and the reserved histological criteria for sample selection. A disadvantage for this careful selection of the $LVI^{+/-}$ samples in this study is that it favours the accumulation of the aggressive molecular subtype of $ER^+/PR^+/HER2^+$, and it dismisses the good prognosis subtype of $ER^+/PR^+/HER2^-$ in the Nottingham cohort of patients.

DGE analyses, including CNA, was suggested as a robust selection tool for candidate biomarkers. The requested DGE and CNA files were received as text files. Both of these files were forwarded to Dr Abhik Mukherjee, Clinical Associate Professor of Histopathology, Faculty of Medicine and Health Sciences, University of Nottingham, who received the files initially from Dr Roslin Russell, Computational Biologist, Cancer Research UK, Cambridge Institute, University of Cambridge. Two separate files were received, on 6 September 2014, including DGE data for a discovery and an external validation set. Parts of CNA data were received in the following months and the database was assembled completely in the middle of 2015. Clinicopathological data for the $LVI^{+/-}$ cohort is detailed in Table 2.2.

Clinicopathological characteristics	n (%)
Age	
≤ 50 years	66 (38.8)
> 50 years	104 (61.2)
Menopausal status	
Pre	66 (38.8)
Peri	0 (0.0)
Post	104 (61.2)
Tumour size	
≤ 2cm	100 (58.8)
> 2cm	70 (41.2)
Grade	
Grade-I	26 (15.4)
Grade-II	55 (32.5)
Grade-III	88 (52.1)
Tumour type	
Invasive ductal/NST	121 (71.2)
Lobular	11 (6.5)
Medullary-like	5 (2.9)
Special types, Tubular/Mucinous	28 (16.5)
Miscellaneous	5 (2.9)
Nottingham prognostic index (NPI)	
Good prognostic group (GPG)	69 (40.6)
Moderate prognostic group (MPG)	88 (51.8)
Poor prognostic group (PPG)	13 (7.6)
Lymph Node Stage	
Stage-I	105 (61.8)
Stage-II	49 (28.8)
Stage-III	16 (9.4)
Lympho-vascular Invasion, (H&E/IHC:Podoplanin/CD34)	
Negative	80 (47.1)
Definite	90 (52.9)
Follow-up Status	
Alive	98 (57.6)
Died from breast cancer	49 (28.8)
Died from other causes/unknown	14 (8.2)
Lost to follow-up	9 (5.3)
PAM50	
Normal-like	12 (7.1)
Luminal-A	64 (37.6)
Luminal-B	45 (26.5)
HER2-enriched	14 (8.2)
Basal-like	35 (20.6)
Estrogen receptor status (IHC)	
Negative	42 (24.7)
Positive	128 (75.3)
Progesterone receptor status (IHC)	
Negative	148 (87.1)
Positive	22 (12.1)
HER2 (IHC/CISH)	
Negative	74 (43.5)
Positive	96 (56.5)
Triple-negative (IHC)	
No	140 (82.4)
Yes	30 (17.6)

Table 2. 2: clinical features of the LVI-supervised cohort, n =170

2.1.2. **The METABRIC cohort**

A large list of 1,980 patients was provided by the METABRIC cohort (MB) for this study, with validated clinicopathological and transcriptomic data. The decoded and revised data file was received on 5 September 2016, courtesy of Dr Oscar M. Rueda, Senior Research Associate, Cancer Research UK, Cambridge Institute, University of Cambridge. The determination of prognostic indications for differentially expressed genes (DEGs) related to the molecular regulation of LVI for a large number of patients and, at the transcriptomic level, worked as independent and reciprocal confirmatory approaches.

Clinicopathological profiles for the MB cohort covered 1,980 unselected primary operable invasive BC female patients. The median of the patients' age at diagnosis was 61.8 years, with a range of 74.36 years. Patients with grade-I tumours represented 9.0% of the whole cohort, whereas patients with grade-II and grade-III tumours accounted for 40.7% and 50.3%, respectively. Regional LN metastasis positivity was observed in 47.5%, where the LN involvement at stage-II and stage-III was 31.5% and 16.0%, respectively. The latter LN staging was based on the NPI staging system: stage-I = no lymph node involvement, stage-II = $\leq$3 positive lymph nodes, stage-III = > positive lymph nodes. The BCSS for each patient was calculated in months, up to 300 months, and it started from the date of the primary surgical treatment to the time of death from BC. Last follow-up status revealed that 505 (25.5%) patients died of BC while 1,071 (54.1%) survived the disease. According to histological examination of the surgical specimens of the MB cohort, the histological phenotypes of these patients were as follows: 83.8% as IDC, 7.5% as ILC, and the rest had special types of invasive BC, including mucinous, medullary, and tubular carcinomas.

The overall distribution of intrinsic subtypes of BC in the MB cohort was assessed with prediction analysis of microarray of 50 genes (PAM50), a technique based on RT-qPCR. Normal subtype accounted for 10.4% of the sample, with luminal-A, 36.3%, luminal-B, 24.6%, HER2-enriched, 12.1%, and basal-like, 16.6%. Integrating germline variants, somatic aberrations, and aberration with gene expression data augmented BC classification in this cohort by producing 10 IntClusts (1 to 10). These clusters revealed novel heterogeneous gene expression landscapes within molecular subtypes classified according to their expression of hormonal receptors. Clinical characteristics for the MB cohort are available in Table 2.3.

Clinicopathological characteristics	n (%)
Age	
≤ 50 years	429 (21.7)
> 50 years	1551 (78.3)
Menopausal status	
Pre-menopausal	436 (22.1)
Post-menopausal	1533 (77.9)
Tumour size	
≤ 2cm	858 (43.8)
> 2cm	1102 (56.2)
Grade	
Grade-I	171 (9.0)
Grade-II	770 (40.7)
Grade-III	952 (50.3)
Tumour type	
Invasive ductal/NST	1634 (82.6)
Lobular	147 (7.4)
Medullary-like	32 (1.6)
Special types, Tubular/Mucinous	113 (5.7)
Miscellaneous	52 (2.7)
Nottingham prognostic index (NPI)	
Good prognostic group (GPG)	680 (34.3)
Moderate prognostic group (MPG)	1101 (55.6)
Poor prognostic group (PPG)	199 (10.1)
Lymph Node Stage, NPI staging system	
Stage-I	1035 (52.5)
Stage-II	622 (31.5)
Stage-III	316 (16.0)
Lympho-vascular Invasion, (H&E/IHC:Podoplanin/CD34) *	
Negative	164 (55.4)
Definite	132 (44.6)
Follow-up Status	
Alive	1071 (54.1)
Died from breast cancer	505 (25.5)
Died from other causes/unknown	384 (19.4)
Lost to follow-up	20 (1.0)
PAM50	
Normal-like	205 (10.4)
Luminal-A	718 (36.3)
Luminal-B	488 (24.6)
HER2-enriched	240 (12.1)
Basal-like	329 (16.6)
Estrogen receptor status (IHC)	
Negative	474 (23.9)
Positive	1506 (76.1)
Progesterone receptor status (IHC)	
Negative	940 (47.5)
Positive	1040 (52.5)
HER2 (IHC/CISH)	
Negative	1733 (87.5)
Positive	247 (12.5)
Triple-negative (IHC)	
No	1660 (83.8)
Yes	320 (16.2)

Table 2. 3: clinical features of the METABRIC cohort, n =1980,

* The METABRIC sub-cohort was constructed from the 296 LVI-informative patients, the clinical features of these patients are illustrated in table 2.1

2.1.3. **NPS of invasive BC**

The specimens in this cohort were derived from the well-characterised NPS and consisted of 2,497 unselected primary operable invasive BC female patients. The NPS patients underwent surgery at Nottingham City Hospital between 1987 and 1998. Previously, patients' biopsies from this cohort were used in the construction of tissue microarray (TMA) formalin-fixed paraffin-embedded (FFPE) blocks. Also, few FFPE full-face tissue sections were available and they were used for the general characterisation of protein expression patterns on different invasive BC tissues. Robust clinicopathological profiles for each patient included histological phenotype, molecular subtypes, primary tumour size, histological grade, tumour stage, nodal status, distant metastasis, LVI, and NPI. The DFS and BCSS were recorded for up to 25 years (median, 175 months; range, 1–305 months) and included information on local and regional recurrences and DMFS. Details of the clinical factors are shown in Table 2.4. A broad panel of IHC data of biomarkers is available for this cohort: ER; PR; androgen receptor (AR), HER2, proliferation marker Ki67; DNA-damage repairing and sensing proteins BARD1, RAD51, PARP1, and ATR; cytokeratins such as KRT5, KRT6, KRT7, and KRT8; CSC biomarkers CD24 and CD44; and immune-related markers including CD3, CD8, CD20, CD68, and FOXP3, as previously published [160-163]. The number of patients in statistics, for some variables, could be different from the number mentioned above owing to the absence of valid data and the lack of tumour specimens in the sectioned tissues.

Clinicopathological characteristics	n (%)
Age	
≤ 50 years	669 (34.2)
> 50 years	1285 (65.8)
Menopausal status	
Pre-	906 (36.8)
Peri-	10 (0.4)
Post-	1545 (62.8)
Tumour size	
≤ 2cm	1010 (52.0)
> 2cm	931 (48.0)
Grade	
Grade-I	440 (19.9)
Grade-II	757 (34.2)
Grade-III	1017 (45.9)
Tumour type	
Invasive ductal/NST	1559 (80.3)
Lobular	215 (11.1)
Medullary-like	44 (2.3)
Special types, Tubular/Mucinous	113 (5.8)
Miscellaneous	10 (0.5)
Nottingham prognostic index (NPI)	
Good prognostic group (GPG)	645 (33.2)
Moderate prognostic group (MPG)	997 (51.4)
Poor prognostic group (PPG)	299 (15.4)
Lymph Node Stage, NPI staging system	
Stage-I	1502 (66.1)
Stage-II	595 (26.2)
Stage-III	177 (7.8)
Lympho-vascular Invasion, (H&E/IHC:Podoplanin/CD34)	
Negative	1212 (62.8)
Definite	717 (37.2)
Follow-up Status	
Alive	1356 (56.4)
Died from breast cancer	698 (29.0)
Died from other causes/unknown	314 (13.1)
Lost to follow-up	35 (1.5)
Distant metastasis	
No	1682 (68.6)
Yes	770 (31.4)
Recurrence	
No	1483 (61.6)
Yes	926 (38.4)
Estrogen receptor status (IHC)	
Negative	549 (22.0)
Positive	1601 (74.5)
Progesterone receptor status (IHC)	
Negative	761 (40.7)
Positive	1111 (59.3)
HER2 status (IHC/CISH)	
Negative	1653 (87.7)
Positive	231 (12.3)
Triple-negative (IHC)	
No	1719 (84.3)
Yes	320 (15.7)

Table 2. 4: Clinicopathological characteristics of Nottingham primary series of invasive BC, n= 2497

2.1.4. **Luminal-A subtype cohort of invasive BC**

Luminal-A subtype was a relatively prevalent molecular class of invasive BC. All tumours in this subtype are $ER^{+}/PR^{+/-}/HER2^{-}$. Avoiding the selection of luminal tumour specimens that belong to high-proliferative luminal-B subtype $ER^{+}/PR^{+/-}/HER2^{+}$ was intentional. The HPLC/MS procedures for studying LVI mechanisms in luminal-A subtype were designed with the research group of Prof. Graham

Pockley, under the supervision of Dr Amanda Miles, John van Geest Cancer Research Centre (JvGCRC), School of Science and Technology, Nottingham Trent University (NTU).

The luminal-A cohort was initially constructed by retrieving FFPE blocks from the Nottingham Health Science Biobank (NHSB). High standards of ethical guidelines were followed and the appropriate consent forms for tumour tissue handling and processing were followed strictly. Also, material transferring agreements were signed between NHSB, based in the NUH trust, as the donor institution, and NTU/JvGCRC as the recipient institutions. The authenticated biological materials were divided into two categories based on LVI and followed the sample selection criteria for $LVI^{+/-}$, except they were obtained from only luminal-A patients. Morphological assessment of the tumour burden for each FFPE block was carried out with H&E staining on full-face sections. The total protein concentration for each specimen was quantified spectrophotometrically by the Coomassie (Bradford) protein assay. The minimum protein concentration for each specimen was >1.0mg/ ml. The time-wise and cost-effective maximum required number of specimens for this luminal-A/$LVI^{+/-}$ cohort was set at 80 FFPE blocks, splitting the cohort equally. Robust clinicopathological data are also available as these samples were obtained from the NPS cohort mentioned above. Features of the luminal-A/$LVI^{+/-}$ cohort are available in Table 2.5.

Clinicopathological characteristics	n (%)
Age	
≤ 50 years	27 (33.8)
> 50 years	53 (66.2)
Menopausal status	
Pre-	31 (38.8)
Peri-	1 (1.3)
Post-	48 (60.0)
Tumour size	
≤ 2cm	29 (36.3)
> 2cm	51 (63.7)
Grade	
Grade-I	15 (18.8)
Grade-II	34 (42.5)
Grade-III	31 (38.8)
Tumour type	
Invasive ductal/NST	65 (77.4)
Lobular	13 (15.5)
Medullary-like	0 (0.0)
Special types, Tubular/Mucinous	2 (2.5)
Miscellaneous	0 (0.0)
Nottingham prognostic index (NPI)	
Good prognostic group (GPG)	29 (36.3)
Moderate prognostic group (MPG)	41 (51.2)
Poor prognostic group (PPG)	10 (12.5)
Lymph Node Stage	
Stage-I	64 (76.2)
Stage-II	12 (14.3)
Stage-III	8 (9.5)
Lympho-vascular Invasion, (H&E/IHC:Podoplanin/CD34)	
Negative	39 (48.8)
Definite	41 (51.2)
Follow-up Status	
Alive	35 (43.75)
Died from breast cancer	29 (36.25)
Died from other causes/unknown	16 (20.0)
Lost to follow-up	0 (0.0)
Distant metastasis	
No	47 (58.8)
Yes	33 (41.3)
Distant metastasis site: Bone	
No	62 (77.5)
Yes	18 (22.5)
Distant metastasis site: Brain	
No	78 (97.5)
Yes	2 (2.5)
Distant metastasis site: Liver	
No	69 (86.3)
Yes	11 (13.8)
Distant metastasis site: Lung	
No	72 (90.0)
Yes	8 (10.0)

Table 2. 5: Clinicopathological characteristics of the Luminal-A molecular subtype, n= 80

2.1.5. Other cohorts collected or used in this study

Various attempts to recruit FFPE blocks from several cohorts were carried out in this study. In addition, a cohort of triple-negative subtype with LVI data was collected and tested with H&E staining for HPLC/MS study. However, owing to the lack of funds and time, this experiment was postponed, and some of these blocks are now being used for ongoing research studies related to triple-negative BC heterogeneity. Other efforts were focused on gathering missing LVI and LN

data for FFPE blocks that belong to a specific group of patients. These FFPE blocks were used in the construction of a prospective series, which was designed to serve the NPI-plus (NPI+) project [164]. The H&E results for tumour burden were negative in this series owing to the high consumption of tumour tissues in the NPI+ study. This series served the project for the LVI molecular mechanism with only four FFPE blocks. Finally, full-face sections were prepared from 30 randomly selected FFPE blocks. These FFPE blocks were IHC-stained to characterise protein expression patterns on multiple histological components, including normal epithelium and various stromal cells. A summary diagram for the cohort selection strategies is illustrated on page 390, Appendix A.

2.2. Tissue handling and processing

2.2.1. FFPE blocks retrieval from NHSB

All of the FFPE blocks in this study were accessible through specific requisition applications approved and signed by the NHSB. The "Appendix 1" form in this application includes the research and development approval number that allowed us to use NUH patients' specimens in this research. The retrieved blocks must be archived and their related records kept strictly confidential.

2.2.2. Tissue sectioning

2.2.2.1. *Documentations*

Once the requested FFPE blocks were obtained and indexed in special labelled boxes, tissue sectioning requests were completed. These forms are for human tissue section requests and release records. The form includes two sections, A and B. The 'A' section is to update the Breast Consortium records in the NHS, while the 'B' section contains checklists for the release and dispatch of biological materials by the NHSB. This form must include the Research Ethics Committee approval number, the Health Technology Assessment (HTA) study number, issued by the National Institute for Health Research (NIHR), and the NHSB ACP number. This form applies to full-face sectioning and TMA sectioning as well. This request is completed when H&E and IHC staining is required.

2.2.2.2. *Sectioning in the microtome*

Before sections were cut from the FFPE blocks, safety procedures and risk assessment guidelines were followed. Preparing sections from the FFPE tissues served both of the routine histological examinations with H&E and IHC. Instruments and tools used for this method included: a rotary microtome (Leica RM2125, Leica microsystems, United Kingdom); a paraffin section mounting bath (Thermo Scientific, MH8516, United Kingdom), filled with deionised water to

eliminate distortions of sectioned tissue; and glass slides, with Leica's X-tra® or Leica's Apex adhesive slides preferred for use in H&E for IHC applications. Both types of glass slide exhibit positively charged surfaces, allowing maximum adhesion to occur under hydrophilic conditions. A plastic box containing crushed ice, disposable microtome blades (S35 Feather, CellPath, Newtown, Powys, Wales), curved-tip forceps, and xylene-resistant marker pen, or pencil, were also available.

FFPE blocks were chilled by placing them, face up, on ice. Then, the chilled blocks were inverted, face down, on ice for five minutes maximum. A block was carefully positioned in a microtome chuck; water droplets were removed prior to this step. The movement of the block's face to the blade edge in both axes needed to be parallel; if required, the microtome chuck was unclamped and the block orientation was adjusted. Sectioning thickness was set routinely at 4μm. The microtome handle was turned clockwise until the sections began to be cut; the angle of the block face was adjusted if necessary. The block face was trimmed until a 'full-face' section was produced appropriately. For the TMA blocks, extra sections were cut for optimisation purposes. When the ribbon of full-face sections was produced, forceps were used to grasp its leading edge. Then, it was laid down carefully on the water surface of the water bath, with both ends of the ribbon supported at all times. As the sections expanded, the individual sections were separated with forceps. Single sections were mounted onto microscopic slides by submerging the slides into the water, beneath the tissue, at an angle of approximately 45°. Sections were realigned over the slide when necessary. Mounted slides were placed vertically in a drying rack, then all of the prepared slides were collected and dried overnight in an oven at 37°C. All slides were transferred to a labelled storage box and stored in a cold room at 4°C until needed.

2.2.3. Preparation of donor blocks and TMA construction

Archival FFPE blocks of invasive BC tissue were retrieved from the NHSB. H&E stained sections from these blocks were taken and assessed for sufficient invasive malignant cells for sampling. Representative tumour areas on each slide were marked alongside their corresponding spots on the FFPE blocks; only well-trained pathologists performed the latter step. Marked blocks were loaded onto the TMA Grand Master® (3D HISTECH®, Budapest, Hungary). To make a recipient block for the biological samples, molten paraffin wax (55–58°C) was poured into a mould, onto which a histopathological tissue cassette was mounted. Ice was used to solidify the paraffin and the mould was removed. For TMA construction, the blocks were trimmed. The TMA blocks were designed to hold a maximum of

150 cases. Marked areas of the tumours from the donor blocks were harvested using a 0.6mm needle. Maps were prepared for the TMA blocks, showing medical identifiers for each sample and on each well, as shown in the figure below.

2.3. H&E histological staining

2.3.1. Overview

The H&E staining protocol is applied widely to human biopsies for diagnostic purposes. Distinctive shades of violet and blue colours can be seen microscopically on the H&E tissue sections. Therefore, pathologists distinguish histological components and detect any morphologic alterations that could indicate malignant features occurring on an observed tissue.

Haematoxylin is a positively charged dye that stains the basophilic components of a tissue's genetic material. This gives haematoxylin extreme importance in medical diagnosis, specifically in histological characterisation and diagnosis of malignant neoplasms. The large nuclei of cancer cells are stained a deep purplish-blue colour as the oxidised haematoxylin reacts with the dense nucleic acids, or heterochromatin. Cytoplasmic and extracellular matrix, mainly collagens, are counterstained with eosin. Eosin, non-specifically, stains eosinophilic proteins that involve positively charged amino acids such as arginine, histidine, and lysine. Eosin stains red blood cells with an intense red colour, collagen with pastel pink, and smooth muscle tissue with bright pink. It should be noted that nucleoli are also stained with eosin due to the existence of cationic nuclear proteins.

2.3.2. H&E procedure

Leica Autostainer XL system (Leica AXL) (product code: 14045635136, Leica Biosystems, UK) was used for processing routine IHC and H&E staining of all tissue sections in this study. The use of Leica AXL was semi-automated as the initialisation of the machine, programming, and loading/removing tissue slides were performed by the user, whereas the staining procedure itself was sequentially executed by the machine.

First, tissue sectioning of the FFPE blocks was completed. Second, the deparaffinisation step was made by placing tissue sections face up on the slide on a hot plate at 60°C for ten minutes. This step was essential to melt the hard wax and to prepare it for total dissolving by xylene solution (product code: X/0250/PB17, Fisher Scientific, UK) in the next step. A reagent vessel, specifically made for Leica AXL, was filled with fresh xylene. All tissue sections were sorted into a slide rack fitting for the xylene vessel. The vessel was then placed into a

specific station inside the Leica AXL instrument. The H&E staining programme was selected and activated. The first step was a five-minute incubation in xylene, and this was repeated once for another five minutes in another reagent vessel of fresh xylene. Three consecutive washes in 100% of industrial methylated spirits (IMS) (product code: M/4450/25, Fisher Scientific, UK) were performed for two minutes each. The tissue sections were washed in water twice for two minutes.

Harris's haematoxylin was purchased (product code: 12658626, Fisher Scientific, UK), filtered, and then poured into a clean Leica AXL reagent vessel. The washed tissue sections were then immersed in Harris's haematoxylin vessel for five minutes. Then, the sections were washed for 10 seconds in acid alcohol, and 1% hydrochloric acid (HCl) in IMS was applied, followed by another wash with water for 20 seconds. Tissue sections were dipped into a reagent vessel filled with 1% of borax (product code: B9876, Sigma, UK) for 40 seconds, followed by a wash in water for one minute.

Contrast counterstaining of Harris's haematoxylin staining was achieved by immersing the tissue section in another reagent vessel filled with 1% eosin (product code: 12668616, Fisher Scientific, UK) for 30 seconds, then the sections were washed in water for 15 seconds. The remaining water droplets were totally removed with three dehydration baths in 100% isopropyl alcohol (product code: P/7507/PB17, Fisher Scientific, UK) for 40 seconds each. Alcohol was then displaced from the tissue sections with two xylene baths, as a clearant, for five minutes each. The tissue sections were then mounted with DPX mounting medium. Each slide was cover-slipped with a suitable and clear glass cover.

2.4. Determination of binding specificity for selected antibodies with western blotting

2.4.1. Overview

Western blotting (WB) was performed to assess the binding specificity of each purchased antibody to its corresponding antigenic sequence. The qualitative detection of a single protein, or an epitope, in various BC cell lysates, or lysates from other cell lines, was our tool to describe the binding specificity for each used antibody in this study. Occasionally, the detection of multiple bands in WB infers the presence of non-specific binding or a specific binding with a modified protein with two or more recognisable epitopes by the same primary antibody. The term 'blotting' denotes the transferring of the isolated protein from a separation gel medium onto a surface of non-reactive membrane. This technique was first

introduced in transferring proteins from polyacrylamide gels to nitrocellulose membranes, as detailed in [165].

2.4.2. Cell line lysates

Different available human BC cell lysates were used: MCF7 expressing ER; MDA- MB- 231 and MDA- MB- 468 both expressing EGF and transforming growth factor alpha (TGF-α); SKBR3 derived from a patient treated with radiation, steroids, cytoxan, and 5-fluorouraciland; MCF10A, a non-tumourigenic mammary epithelial cell line with characteristics of luminal ductal cells, with no myoepithelium features; HeLA epithelial cells derived from a human cervix carcinoma patient; HEK293 embryonic kidney cell line, Jurkat a T lymphocyte of an acute T cell leukaemia patient expressing interleukin-2 and CD3; and HT29, an epithelial cell line obtained from a patient diagnosed with primary colorectal adenocarcinoma.

2.4.3. Preparation of western blotting reagents and tools

Prior to sample preparation, all required reagents and experimental tools were prepared and ready to use. The XCell SureLock™ Mini-Cell and XCell II™ Blot Module (catalogue code: EI0002, Invitrogen, Thermo Fisher Scientific, UK) were assembled to use in protein isolation and protein transferring. A NuPAGE™ 4–12% Bis-Tris Protein Gels, ten well (catalogue code: NP0321BOX, Invitrogen, Thermo Fisher Scientific, UK) was purchased and used as pre-casted polyacrylamide gels for denatured and reduced protein separations.

The working solution of the running buffer was freshly prepared by adding 50ml of NuPAGETM MOPS SDS Running Buffer 20X (catalogue code: NP000102, Invitrogen, Thermo Fisher Scientific, UK) to 950ml of distilled water. A standard preparation of the transfer buffer for one gel was then prepared by mixing 50ml of NuPAGETM Transfer Buffer 20X (catalogue code: NP0006, Invitrogen, Thermo Fisher Scientific, UK), 100ml of 10% methanol, 849ml of distilled water, and the addition of 1ml from NuPAGETM Antioxidant (catalogue code: NP0005, Invitrogen, Thermo Fisher Scientific, UK). If two gels were used for protein separation and blotting, methanol concentration was adjusted up to 20% and the distilled water volume was reduced to 749ml. The remaining volumes of the other solution stocks were the same. The wash solution was prepared by dissolving 1ml of Tween-20, non-ionic detergent, in 1,000ml of phosphate-buffered saline (PBS). The blocking solution was prepared by dissolving 2.5gm of Marvel's dried and fat-free milk at 5%, or in 50mL of washing solution. Alternatively, the blocking solution was a solution of 1% of bovine serum albumin (BSA) for some antibodies according to

optimisation conditions. All primary antibodies were prepared upon use and according to their previously optimised conditions.

IRDye® 680RD Donkey (polyclonal) anti-Rabbit IgG (product number: 925-68073, LI-COR Biotechnology, St. John's Innovation Centre, Cowley Road, Cambridge, UK) and IRDye® 800CW Donkey (polyclonal) anti-Mouse IgG (product number: 925-32212, LI-COR Biotechnology, St. John's Innovation Centre, Cowley Road, Cambridge, UK) were used, as appropriate, to detect the primary antibodies of the studied proteins, whether the raising host was mouse or rabbit. The ratio of 1:15000 was applied to dilute fluorescent secondary antibodies in the blocking solution. Primary anti-β-actin antibody produced in rabbit and primary anti-β-actin antibody produced in mouse were also used, as β-actin protein is endogenously expressed in all lysates, and considered as the sample's normal control. The fluorescent secondary antibodies were prepared at the dilution ratio of 1:5000 in the blocking solution, as explained above, while the primary anti-β-actin antibodies were diluted at 1:2000 in Bond™ primary antibody diluent (catalogue number: AR9352, Leica Biosystems, UK).

2.4.4. Sample preparation

Each cell line lysates, the commercially purchased protein standards, MagicMark™ XP western protein standard (catalogue code: LC5602, Invitrogen, Thermo Fisher Scientific, UK) and Novex™ sharp pre-stained protein standard (catalogue code: LC5800, Invitrogen, Thermo Fisher Scientific, UK), were thawed in a bucket of crushed ice. Then, 13µl of each lysate was mixed with 5µl of NuPAGE™ LDS sample buffer 4x (catalogue code: NP0008, Invitrogen, Thermo Fisher Scientific, UK) and 2µl of NuPAGE™ sample reducing agent 10x (catalogue code: NP0004, Invitrogen, Thermo Fisher Scientific, UK). β-mercaptoethanol was also added, 2µl, to cleave the protein disulphide bonds and as an additional protectant. All solution mixtures were mixed in chemical-resistant polypropylene micro-centrifuge tubes. All tubes were spun gently using a laboratory vortex mixer and the protein mixture was digested at 100°C on a heating dry block. All samples were prepared in duplicate, and each sample was sufficient to optimally fill 10µl in each well.

2.4.5. Loading samples and separation of proteins mixture

A ten-well precast NuPAGE Bis-Tris 4-12% protein gel cassette (catalogue code: NP0321BOX, Invitrogen, Thermo Fisher Scientific, UK) was inserted inside a running tank, the XCell SureLock™ Mini-Cell (catalogue code: EI0002, Invitrogen, Thermo Fisher Scientific,UK). The previously prepared running buffer

was poured slowly into the running tank until the maximum fill line was reached by the liquid. No leakage was confirmed. NuPAGE antioxidant reagent (catalogue code: NP0005, Invitrogen, Thermo Fisher Scientific, UK) was added to the running buffer at 1ml to protect the protein from possible oxidation.

Starting from the first well on either left- or right-hand sides, a mix of protein standards, MagicMark™ XP (catalogue code: LC5602, Invitrogen, Thermo Fisher Scientific, UK) at 5µl and Novex™ sharp pre-stained (catalogue code: LC5800, Invitrogen, Thermo Fisher Scientific, UK) at 3µl was loaded onto the first well. Then, each sample mix was loaded at a fixed volume of 10µl into wells consecutively. If the number of samples was less than the number of wells, the empty wells were filled with the loading buffer alone at the same volume. The running tank lid was positioned and the electrodes' cables and plugs were fastened securely. Electrical power supply (Invitrogen, Life Technologies, UK) was connected and set to run for 90 minutes at 150 volts.

2.4.6. Transferring protein bands onto nitrocellulose membrane

The blotting module was assembled in the last 10 minutes of the running time of the protein separation step. The precast gel cassette was opened using a specific metal opener and the gel was removed with extreme caution and merged into a weighting boat filled with running buffer. Sponges and filter papers were cut to the exact size of the protein separation gel and were soaked thoroughly in the transfer buffer, as explained previously. The blotting module was opened and, from the cathode chamber, the gel was layered with blotting membrane, soaked sponges, and the cut filter papers in the following order: 2 x sponges, 1 x filter paper, gel, 1 x nitrocellulose membrane, 1 x filter paper, 2 x sponges. Each layer was flattened with a cylindrical roller to remove any air bubbles that may distort the transfer of protein bands. The blotting module was merged firmly into the previously prepared transfer buffer contained in the running tank, cables and plugs were connected, and the power supply was turned on for 60 minutes at 30 volts.

The blotting module was dismantled and the membrane was laid carefully, facing upward, in a weighting boat filled with proper blocking solution, 5% of Marvel Milk or 1% BSA, and left for 60 minutes at room temperature on a rocker. Then, the membrane was rolled carefully, facing inward, and inserted into a skirted 50ml Falcon tube filled with 5ml of a primary antibody solution for the selected biomarkers at the previously optimised dilution. The membrane was incubated overnight in the cold room. After it had equilibrated to room temperature, the excess blocking solution was poured out carefully and the

membrane was unrolled in a weighting boat filled with washing solution without touching the blotting face. The membrane was washed three time for five minutes with rocking. Then, the secondary antibodies were prepared at 1:15000 by diluting 1µl in 15ml of blocking solution. The membrane was incubated for 60 minutes with continuous rocking. The membrane was kept covered to avoid direct exposure to light at this stage. Finally, the membrane was washed once for 15 minutes and three times for five minutes in the washing buffer.

2.4.7. Visualisation of protein bands

Image capturing of the immunodetected protein bands was performed at two, near-infrared, different emission wavelengths. The infrared dye that was used to label the IRDye® 680RD secondary antibody has an emission wavelength at 694nm whereas the IRDye® 800CW secondary antibody has an emission wavelength at 795nm. The Odyssey Fc Imager (model number: 2800, LI-COR Biotechnology, St. John's Innovation Centre, Cowley Road, Cambridge, UK) was used to capture digital images of the membranes showing red and green fluorescent bands. If the molecular weight of the β-actin overlapped with the targeted antigen of the studied biomarker, a multiplexed image of both anti-rabbit and anti-mouse secondary antibodies was generated digitally. Yellow bands indicated the overlapping proteins with molecular weight around 42 kilo Daltons (kDa). All images were labelled and stored on a personal hard drive.

2.5. Immunising peptide blocking (IPB)

Alternatively, an IPB protocol was performed to identify binding specificity for any primary antibody that showed no positive immune interactions against their specific antigens in the WB gel, i.e. visual bands. The detection of β-actin bands on the same WB gel was necessary to confirm whether there were any faults within the WB procedure itself. First, the primary antibody was neutralised by incubating it with an excessive concentration of its recombinant antigen solution. This reaction blocked the antigen-binding site of this antibody and its fragment antigen-binding (Fab) region was no longer capable of binding to the corresponding epitope. The blocked primary antibody was incubated with the peptide blocking solution (0.25µg/ml) in a labelled 2ml Eppendorf tube for 30 minutes on ice. Similarly, another tube contained the non-blocked primary antibody with exact conditions, except for the addition of the recombinant antigen solution. During the incubation period, two consecutive tissue sections from the same histological sample were prepared according to the routine histological procedures. Then, the IHC protocol was followed and two stained tissue sections were obtained. An absolute absence of positive staining on the blocked primary

antibody sample, in comparison to its negative counterpart, assigned this primary antibody with specific binding affinity to its antigen, and there were no other peptide sequences that the primary antibody attached to it as its antigen by similarities.

2.6. IHC staining

2.6.1. Overview

IHC is an indispensable practice in the clinical management of BC patients. From molecular subtyping to the detection of validated prognostic biomarkers, IHC plays a critical role in enriching the reliability of pathology reports in many neoplastic diseases, including invasive BC. The first IHC experiment was reported in the 1940s [166]. Fluorescein isothiocyanate-labelled antibodies (FITC-labelled Abs) was the first material that allowed researchers to visualise pneumococcal antigens with a fluorescent dye. Continuous efforts were dedicated to developing IHC methods over the next few years, leading to the substitution of fluorophores with enzymes in labelling Abs, for example, peroxidase and alkaline phosphatase.

The IHC procedure is based on a reversible reaction between an antibody (Ab) and an antigen (Ag). Utilising this biochemical bonding between Ab and Ag in the clinical practice of IHC improves patients' clinical profiles and prognosis accuracy by combining the observed histological features of a tumour with its protein expression patterns. In BC pathology, the immunostaining of selected myoepithelial proteins such as smooth muscle actin (*SMA*), tumour protein 63 (*TP63*), and smooth muscle myosin protein (*MYH11*) is to differentiate between invasive and benign, or normal, proliferative compartments of the breast [167], and to identify adenoid cystic carcinoma if the anti-calponin antibody exhibits negative staining [168]. Also, IHC staining of adhesion molecules such as epithelial cadherin (E-cadherin) added high relativity to this technique in revealing discohesive and highly metastatic phenotypes of ILC [169]. Finally, ER, PR, HER2, Ki67, tumour protein 53 (p53), and the apoptosis regulator Bcl-2 were all detected by IHC [170] and helped in the individualisation of therapeutic regimens for each BC patient, especially before and after neoadjuvant chemotherapy.

2.6.2. Immunohistochemistry tools, chemicals, and procedures

In this study, IHC was used to semi-quantify the levels of protein expression on invasive BC TMA sections, 4μm in thickness. In addition, full-face sections, a random number of samples, from 10 to 15, were IHC stained to investigate the heterogeneity of the expressed target; therefore, the use of TMA sections needed to be validated subsequently. All cut tissues were obtained from invasive or

invasive-mixed FFPE specimens. The IHC procedure was performed by using Novolink™ Max Polymer Detection System (product code: RE7280-K, Leica, Biosystems, UK), which is suitable for antigen detection by using mouse immunoglobulin G (IgG) and rabbit IgG.

2.6.2.1. *Labelling of slides*

All tissue sections were loaded onto Leica's X-tra® slides. Then, the slides were labelled specifically with a marker pen for IHC applications. Date, symbol of targeted protein, user name, dilution factor of applied antibody, and specimen's identification number were the basic labelling data for each slide.

2.6.2.2. *Deparaffinisation and rehydration of tissue sections*

Starting with dewaxing the tissues, all TMA sections were placed on a hot plate (Leica HI1220 flattening table, product code: 14042321474, Leica Biosystems, UK), face up, at 60°C for 10 minutes. Deparaffinisation continued with sorting the slides into the Leica's slide rack (product code: 14047533750, Leica Biosystems, UK), then the rack holding the slides was placed into a reagent vessel filled with fresh xylene (product code: X/0250/PB17, Fisher Scientific, UK) and contained inside the chamber of the Leica AXL. The dewaxing programme, previously introduced into the machine, was activated and this programme kept the slides soaked in xylene for five minutes, followed by automated transfer of the holding rack into another fresh xylene vessel, where it was incubated for another five minutes. Three baths of IMS were applied to all slides, as the first two baths were in 100% IMS for two minutes and the third bath was in 70% IMS. The rack was then rehydrated in a vessel filled with flowing water for five minutes.

2.6.2.3. *Heat-induced epitope retrieval: Citrate buffer, pH 6.00*

Formalin or aldehyde fixation and paraffin-embedding procedures may lead to immense physical and chemical modifications in any surgical biopsies. Transposition or alterations of the antigenic site are a few effects of this method on FFPE tissues. Poor exposure to epitope may deteriorate the positivity of IHC staining. The optimum condition to re-expose the antigenic sites of the FFPE tissues was selected in this study. Various heat-induced epitope retrieval (HIER) methods were tested under different conditions. The commonly used protocol of the citrate buffer antigen retrieval method was designated as the optimum antigen retrieval method.

2.6.2.4. *Preparation of sodium citrate buffer, pH 6.00*

The sodium citrate buffer was prepared by the complete dissolving of 2.94gm of Tris-sodium citrate dihydrate in 1,000ml of distilled water. A few drops

of 1 Normal of HCl were added to adjust the pH to exactly 6.00 using a daily calibrated laboratory pH meter. The solution was stored in a fridge at 4°C. A plastic Leica AXL vessel was filled with 500ml of citrate buffer, pH 6.00 solution; the rack from step 2.6.2.2. was immersed in it, and a plastic lid with no metal handles was placed on the top of the vessel. The slides, settled in the citrate (pH 6.00) vessel, were placed inside a laboratory microwave; the microwave's door was locked and it was operated at maximum power for 20 minutes. Protective glasses were worn, the vessel was removed from the turned-off microwave with a heat-resistant pair of gloves, the lid was removed in the sink, not on the bench, and flowing water was applied with a rubber hose into the vessel, displacing the citrate buffer and cooling down the slides for five minutes.

2.6.2.5. *Slide assembly onto Shandon™ Sequenza coverplates™*

In this study, proper application of the IHC reagents onto the tissue sections was managed by reproducibility and cost-effectiveness. Shandon coverplates™ (catalogue number: 72110017, Fisher Scientific, part of Thermo Fisher Scientific, UK) were purchased and used to minimise the consumption of any purchased antibody and the reagents of the Novolink Polymer Detection System kit. The specific matching Shandon Sequenza immunostaining slide rack was also purchased and used (Product number: 36105, Ted Pella, USA).

The working space on the bench was booked and was cleaned in advance. A plastic tank was filled with cold tap water and the Leica vessel from the previous step (2.6.2.4.) was placed beside it. Each tissue section slide was attached to a Shandon coverplate by placing the bottom short edge of the glass slide on the two plastic tips at the rear end of the coverplate at an angle of 45°. This process was performed entirely in the water tank to avoid the presence of air bubbles. The tissues, positioned between the glass slide and the plastic coverplate, were collectively removed from the water tank, held tight, and loaded into their slots in the Shandon Sequenza immunostaining slide rack. The Sequenza reservoir was filled half full with Tris Buffered Saline (TBS) pH 7.6 to rinse the slides. The flow speed and time of the applied TBS were observed accurately for each slide, making sure there were no air bubbles between the slide coverplate complex. Slides were reloaded if total emptiness of the Sequenza reservoirs occurred in less than five minutes. The sections were kept hydrated at all times with TBS and they never dried out.

2.6.2.6. *Blocking of endogenous peroxidases*

Endogenous peroxidase activity is a biochemical feature tissues, which is still chemically active even if the tissue is fixed in formalin and embedded in paraffin wax. Applying a horseradish peroxide (HRP) conjugated antibody may interact with peroxidases presented in the tissues, leading to high and non-specific staining noise. The need for the endogenous peroxidases blocking step could be tested by incubating the studied tissues with DAB substrate; specific details are given in step 2.6.2.11. If the tissues turned a brown colour, the blocking of endogenous peroxidases was obligatory. Calibrated micropipettes, with a volume range of 1-1,000µl, were available. 100µl of Novolink's peroxidase blocking reagent was applied to each slide, containing 3–4% (v/v) hydrogen peroxide (H_2O_2) (product number: RE7101). All slides were left for five minutes to react, then they were washed twice with TBS at intervals of five minutes.

2.6.2.7. *Blocking non-specific reacting proteins*

Non-specific staining could also occur due to the presence of binding sites, peptide sequences that mimic antigenic sites recognisable by the applied antibody. The non-specific protein–protein bindings between the applied antibody and the tissue surface may take different forms such as charge-based and hydrophilic/hydrophobic attractions. Using a new micropipette plastic tip, 100µl of Novolink's protein block reagent, 0.4% casein in PBS, with stabilisers, surfactant, and 0.2% Bronidox L as a broad spectrum anti-bacterial and anti-fungal preservative (product number: RE7102) were applied and the tissue slides were incubated for five minutes, then washed with TBS as in step 2.6.2.6.

2.6.2.8. *Applying the optimised primary antibody*

Several primary Abs were purchased and optimised for this study. In the forthcoming chapters, the optimisation conditions, positive and negative controls, binding specificity, and dilution factor for each primary Ab will be discussed in detail. However, Bond™ primary Ab diluent (catalogue number: AR9352, Leica Biosystems, UK) was used in this study as our universal diluent for all primary Abs. Using a fixed volume micropipette at 100µl, one load of 100µl of the optimally diluted primary antibody was applied to the corresponding tissues only. A few slides were nominated for this study as blank slides to test the user's investigational quality; negative control slides were used to test the binding specificity of the post primary reagent, or secondary antibody, with the fragment crystallisable region (Fc region) of the primary Ab; the application of primary Abs for all experiments was skipped intentionally on those slides; finally, other different tissues, but not BC tissues, were used as positive controls. The incubation time for the primary antibodies was either 60 minutes on the bench or left

overnight in a cold room at 4°C. Washing cycles with TBS were applied as in previous steps after incubation periods.

2.6.2.9. *Applying post-primary antibody*

Post-primary antibody was applied at 100µl to all TBS-washed slides from the previous step. This contains rabbit anti-mouse IgG <10µg/ml in 10% (v/v) animal serum in TBS (0.09%), with ProClin™ 950 as the preservative (product number: RE7111). The slides were left for a further incubation time of 30 minutes, then washed twice in TBS for five minutes.

2.6.2.10. *Applying Novolink™ polymer*

The Novolink™ polymer binds to rabbit IgGs, and it also detects the post-primary and any rabbit-hosted primary antibodies. The chemical composition of this reagent is anti-rabbit Poly-HRP-IgG <25µg/ml containing 10% (v/v) animal serum in TBS (0.09 %), with ProClin™ 950 as a preservative (product number: RE7112). 100µl of this reagent was loaded onto each slide and left for 30 minutes, followed by two washing cycles with TBS for five minutes.

2.6.2.11. *Blotting with 3,3'-diaminobenzidine (DAB)*

DAB peroxidase substrate solution is prepared by mixing DAB chromogen, containing 1.74% (w/v) 3,3'-diaminobenzidine, in a stabiliser solution (product number: RE7105) with Novolink™ DAB substrate buffer polymer, containing ≤0.1% H_2O_2 in a buffered solution and preservative (product number: RE7143). The mixing ratio was 1:20, and it was prepared freshly each time it was needed. The mixed solution was always kept in the dark until use. The application of this solution was at 100µl for five minutes. Excessive chromogen was washed off with TBS twice for five minutes.

2.6.2.12. *Counterstaining with haematoxylin and tissue dehydration*

Although the counterstaining step might seem optional, it provides clear and sharp nucleus staining images that facilitate semi-quantification of the IHC-stained protein. 100µl of haematoxylin reagent, containing <0.1% haematoxylin (product number: RE7107), was applied to each section and left for six minutes to react. The slide coverplates were disassembled inside the initial water tank and excessive haematoxylin staining was removed. The slides were re-assorted into the Leica AXL rack, which was put into its plastic vessel filled with distilled water. The vessel was then placed inside the Leica AXL system and the dehydration programme was activated. The dehydration step included washing the slides in three baths of ethanol (EtOH), for two minutes each, followed by two washes in xylene for five minutes each. The stained tissues were then mounted in DPX

medium and sealed with a transparent cover glass, or coverslip. DPX is a synthetic resin mountant mixture of distyrene (a polystyrene), a plasticiser (tricresyl phosphate), and xylene.

2.7. Imaging of tissue sections and histological scoring (H-score) system

2.7.1. Whole slide imaging

The quality of IHC and H&E staining intensity is not permanent. Staining details can fade due to long periods of storage, direct exposure to light or, simply, the slide itself can be damaged for various reasons. Digitalising histological tissue sections is not only a fancy terminology accompanying modern clinical practices; in fact, it has helped pathologists to keep their patients' highly valued FFPE specimens for a longer time, in a smaller space, with convenient and private sharing and editing options for patients' clinical data.

In this study, all IHC-stained tissue sections were sent to an electron microscopy and imaging unit, at the Cellular Pathology Department, Queen's Medical Centre (QMC), A Floor, West Block, NUH. A specific request form for slides scanning and imaging was requested and filled in appropriately. It includes the total number of slides intended for scanning, name of stained protein, and slide labelling data. A non-specimen request form was completed for billing purposes. High-resolution digital images of the whole slides were generated with the Nanozoomer Digital Pathology scanner (NDP C9600series; Hamamatsu Photonics, Welwyn Garden City, Hertfordshire, UK). The scanning at 20-times magnification power was applied on all slides, and NDP.view2 software (Hamamatsu Photonics, Systems Division) was used to facilitate the storage of the TMA cores using a high-resolution screen (24 inches, 1920x1080 pixels). All scanned images of the studied biomarkers were duplicated and one copy of each TMA set was kept in a labelled folder in the Breast Cancer Research Group (BCRG) local drive; *R:\BreastCancer\ServedImages\Breast EQA 13-04-16\RESEARCH\P14-01 Sultan Sonbul.*

2.7.2. H-score

The selection of the scoring methods used to statistically interpret IHC staining results requires credibility and reproducibility. Hence, the semi-quantitative scoring method in IHC also requires expert and well-trained observers who score the TMA cores with high degrees of scoring concordance. H-score is a common method that was applied to the IHC-stained tissue sections in this study. The H-score is generated by setting four levels of scoring intensities, negative=0,

weak=1, moderate=2, and strong=3, by examining various and randomly selected TMA cores' scanned images from different slides. The sum of individual intensity multiplied by the total percentage of cells expressing it is the mathematical definition of an H-score. An H-score is calculated according to this formula: [0 × (% cells expressing intensity 0)] + [1 × (% cells expressing intensity 1)] + [2 × (% expressing intensity 2)] + [3 × (% expressing intensity 3)].

First, two pathologists besides the main researcher were asked to be the second and the third scorers for at least 25% of the whole cohort under investigation. The selection and scoring of the TMA slides that represented the previous percentage was done blindly and individually. The scoring results were concealed until the main researcher completed the H-scoring for all slides. Intra-class correlation among the scores was calculated and the threshold of H-score agreement was $\geq$75%. Any slide with a lower scoring concordance was re-scored and the discrepancies were discussed. Heterogeneous protein expression may occur and all of the staining intensities could be presented in one TMA core; these biomarkers with this pattern of expression are not suitable for TMA scoring in this scenario. For this reason, a panel of a small number of full-face tissue sections was IHC stained to assess the applicability of TMA.

2.8. Differential gene expression (DGE) and gene selection

2.8.1. Overview

A DGE experiment is defined as the statistical quantification of possible alterations in gene expression levels between two experimental cohorts. Hence, microarray technology provides robust data for prognostic candidates that, potentially, can be translated into BC research and enhance the clinical management of BC patients. In this study, the $LVI^{+/-}$ cohort was constructed meticulously, as previously described in section 2.1.1. The linear models for microarray data (limma) for the DGE were obtained from a total of 170 specimens by applying RNA microarray methodologies. Of these samples, 90 were LVI^{+} and 80 were LVI^{-}/LN^{-}.

2.8.2. Nucleic acids extraction and microarray hybridisation

Malignant features were determined by expert pathologists who examined the tissue sections from the surgical biopsies of all LVI^{+} and LVI^{-} selected specimens. The extraction and purification of nucleic acids were carried out before the beginning of this study utilising archival FFPE blocks and fresh frozen biopsies.

The blocks were sectioned at a thickness of 30μm and each sample provided this study with ten sections.

As described in [38, 171], DNA and RNA were extracted and purified from each specimen with the DNeasy Blood and Tissue Kit and the miRNeasy Kit (Qiagen, Crawley, UK). The extraction/purification process was performed using the fully-automated platform QIAcube (Qiagen). The extracted and purified nucleic acids were quantified with a NanoDrop ND-8000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), then the extracted samples were run on the agarose gel. The extracted and purified DNA probes were hybridised to Affymetrix SNP 6.0 arrays (Affymetrix, Santa Clara, CA, USA) at AROS Applied Biotechnology (Aarhus, Denmark), whereas Agilent 2100 Bioanalyser Nanochip (Agilent Technologies, Wokingham, UK) was used to assess the quality of the RNA probes. Biotin-labelled cRNA was produced from the total extracted and purified RNA. The Illumina Totalprep RNA amplification kit (Ambion, Warrington, UK) was purchased, then the labelled cRNA was hybridised onto Illumina Human HT-12 v3 Expression BeadChips from the latter manufacturer.

2.8.3. Profile analysis of differentially expressed genes

The raw generated data from the scanned BeadChips was introduced into the beadarray package, a bioconductor package using R programming language to analyse the Illumina bead-based data, R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org/). The background noise of the raw gene expression data was corrected and quantiles were normalised. The summarised and normalised data were fitted to a linear model using the limma data and RNA-seq package. Probes' IDs, probes' quality, student t-test, average expression (AvE), the logarithm to base 2 fold change (log2FC), p-value, adjusted p-value, and the log of odds ratio (B-value) were the limma analysis outputs. The considered statistical significance for the differentially expressed genes in this analysis was set to the adjusted p-value. A total number of 34,362 transcripts were found. A list of highly significant DEGs between the interrogated LVI categories was created by specific gene-selection criteria. The same approach was applied to an external validation cohort under the same conditions. Both discovery and validation limma data files were received, on 6 September 2014, as txt. files, which were converted into .csv format.

2.8.4. Gene selection based on LIMMA and CNA results

Gene identifiers in both files were sorted ascendingly based on their p-value using Microsoft Excel 2013. All non-significant identifiers were eliminated. In a separate Excel worksheet, only significant gene identifiers from discovery and validation sets were aligned into two parallel columns. The duplicated identifiers were found and labelled. The overlapping identifiers of the DEGs were sorted in ascending order based on their p-value. Our selection was restricted to those top 100 gene identifiers, which were sent to Dr Abhik Mukherjee on 24 October 2014, who forwarded them to Dr Roslin Russell to retrieve their CNA status. The requested data was received later as PDF files for each gene. Each gene identifier was examined in both LVI categories and data on gain/loss CNA was presented as positive hits with each patient's number. An equation was generated, as in the following formula, to calculate the CNA gain ratio for each gene identifier.

(Gain hits of LVI^+/total number of patients) ÷ (Gain hits of LVI^-/total number patients)

Only those genes with CNA gain ratios >1.00 were considered, and genes with a gain ratio $\leq$1.00 were excluded. Due to the lack of individual expression data for each patient at this stage of study, the CNA for DEGs was suggested as a prioritising method for the selection of LVI-DEGs.

2.8.5. Gene selection based on accessible online resources

Freely and accessible web-based resources were utilised to explore the chosen LVI-DEGs one by one. The list of browsed resources included:

1- (http://cbioportal.org),

2- (https://www.phosphosite.org),

3- (https://reactome.org),

4- (http://www.uniprot.org/),

5- (https://www.proteinatlas.org),

6- (https://string-db.org),

7- (http://software.broadinstitute.org/gsea/msigdb/index.jsp),

8- (http://pantherdb.org),

9- (https://www.ncbi.nlm.nih.gov/pubmed/),

10- (http://bcgenex.centregauducheau.fr), and

11- (https://www.ebi.ac.uk/gxa/home).

2.8.6. Gene selection based on transcriptomic and clinicopathological profiles of the LVI$^{+/-}$ cohort

As focusing only on 100 DEGs can lead to a potential dismissing of LVI key players, we followed a different exhaustive scheme. The clinicopathological and whole transcriptome datasets for each patient in the complete MB (n=1,980) were received on 5 September 2016, courtesy of Dr Oscar M. Rueda, Cambridge Institute, University of Cambridge. The complete transcriptomic profiles for all patients were examined; as a result, more LVI promising candidates were added to this study.

Each normalised quintile for the microarray experiment entries was allocated to each patient in both of the LVI subgroups. The list of gene identifiers was scrutinised for the log2FC values, as well as p-value. DEGs with positive log2FC, positive AvE, and p-value <0.05 were selected. Final refinement of this list was carried out by exclusively involving oligonucleotide probes with a sequence precisely matching their complementarity to a specific gene; therefore, these probes reflect accurate measurements for their corresponding genes. Clinical relevance of the selected gene identifiers was evaluated. The expression mean was statistically investigated against patients' clinicopathological and molecular data, such as LVI, tumour grade, LN stage, PAM50, IntClusts, and BCSS.

2.9. Characterisation of protein expression patterns on full-face FFPE tissue sections

2.9.1. Overview

Protein expression is one of the crucial cellular processes that allow translation of coded genes from mRNA format to functional protein in a living cell. However, the regulation of this crucial mechanism in malignant diseases, such as invasive BC, could be disrupted severely. Tumour heterogeneity/homogeneity, molecular subtypes, histological phenotypes, and varieties of morphological feature in invasive BC are a few elements that modify protein expression levels and protein localisation among different cellular compartments, or simply alter the observed patterns of an expressed protein under a microscope. Studying invasive BC on large cohorts requires efficient time- and cost-management strategies; therefore, all studied specimens were embedded into TMA blocks. Reviewing different protein expression patterns was a robust assessment tool for the use of TMA rather than full-face sections in this study.

2.9.2. Sampling and the characterisation methodology

A total of 30 FFPE blocks were randomly selected from NPS of invasive BC and requested for retrieval from NHSB, as previously described in section 2.2.1. For each candidate biomarker, 10 to 15 specimens were included. Tissue sectioning was performed as mentioned in section 2.2.2. Describing expression patterns for proteins of interest was performed with IHC methods applying specific optimisation conditions for each investigated protein. As well as observational descriptions of protein expression homogeneity or heterogeneity on the BC tissues, the description of protein expression patterns was also engaged with tumour size, tumour grade, reactivity with immune cells, and stromal background staining when applicable.

2.10. Total protein extraction and quantification from FFPE

2.10.1. Overview

Archival clinicopathological records embedded within FFPE biological specimens are precious sources for many proteomic experiments. One of the robust techniques that requires pure protein extracts is HPLC/MS. Recovering peptides from FFPE for MS applications is a challenging task due to the denaturation and crosslinking modifications of formaldehyde on the protein contents of the studied samples. In contrast to protein identification by IHC, where specific Abs and Ag retrieval methods are available, HPLC/MS requires meticulous preparative steps. Tissue homogenisation, clearing with inorganic solvents, rehydration, protein extraction, protein quantification, tryptic digestion, and purification were the applied methods in this MS study. Different optimisation strategies were followed in this study for tissue harvesting and protein quantification from FFPE blocks to verify satisfactory protein concentrations prior to applying the MS procedures.

2.10.2. Tissue harvesting and collection

The specimens of this study were derived from a cohort of luminal-A BC patients. Corresponding H&E stained sections, 4μm, for each specimen were available. Tissue harvesting with a surgical scalpel was applied to each sample. Areas with high tumour burden were marked on the H&E sections and the exact areas on the matching section were dissected at four to six sections, and 8μm of thickness each. Harvested tissues were accordingly collected in 1.5ml collection tubes supplied with the Qproteome FFPE Tissue Kit (catalogue number: 37623, Qiagen, United Kingdom). It was crucial to harvest tissues from the slides while they were still wet after microtome cutting. If the tissues dried on the slide, the amount of collected tissue would decrease significantly.

2.10.3. **Deparaffinisation of harvested tissues**

The removal of the paraffin wax was performed by the addition of 500µl of HPLC-grade n-heptane (99%) (product code: 10664912, Fisher Scientific, UK), as an inorganic solvent. The tubes were sealed tightly and vortexed thoroughly for 10 seconds. The mixture was incubated for 60 minutes at room temperature. Using another micropipette, aliquots of 25µL of methanol (MeOH) were added to each tube and the tubes were re-sealed and vortexed as in the previous step. The dewaxed and dehydrated tissues were micro-centrifuged at 9000 × g for two minutes and, consequently, precipitated as semi-solid white pellets. Supernatant was removed from each tube very carefully without disrupting the pellets at the bottom.

2.10.4. **Protein extraction using Qproteome protein extraction buffer**

Prior to the addition of the protein extraction buffer (PEB), provided by the Qproteome FFPE Tissue Kit, onto protein pellets, the PEB was supplemented with 6% of β-mercaptoethanol as a denaturing agent that reduced the protein's disulphide bonds. The complete volume for each vial of the PEB is 1,000µl; hence, 60µl of PEB was replaced with the exact volume of β-mercaptoethanol. 100µl of supplemented PEB was added to each tube containing the protein pellet. The tubes were mixed vigorously for one minute, then properly sealed with tube sealing clips. The tubes were incubated for 20 minutes at 100°C on a heating block. On another heating block, which had a shaking function, the tubes were incubated for 120 minutes at 80°C with agitation at 750 rounds per minute (rpm). Then, all tubes were placed in the fridge at 4°C for one minute. The sealing clips were removed carefully and the tubes were placed into a refrigerated centrifuge with an angled rotor that holds reaction vials of 2.2–1.5ml, and centrifuged for 15 minutes at 14.000 relative centrifugal force (×g). The supernatant was transferred carefully into new and labelled 1.5ml collection tubes. 10µl aliquots were taken from each protein containing supernatant for the determination of the total protein concentration in each sample.

2.10.4.1. *Bradford protein assay and the determination of total protein concentration*

Bradford protein assay is a rapid and sensitive colorimetric procedure to determine unknown protein concentration in a sample by plotting its absorbance against a standardised curve of serially diluted protein solution. The dye, Coomassie Blue G250, binds to –COOH groups of proteins with Van der Waals forces, forming a stable blue-coloured protein-dye complex [172]. Measurement

sensitivity of the Bradford protein assay allowed us to determine protein concentrations in µg units. The Quick Start™ Bradford Protein Assay kit 1 (product code: 5000201, Bio-Rad, UK) was purchased and kept at 4°C until needed. The chilled working solution dye was equilibrated to room temperature, then mixed properly before use. A standard solution of BSA was provided with the kit. Serial dilution was applied to the BSA stock (2mg/ml) to prepare a standard solutions set ranging from 0.0 to 20.0 µg/ml. The diluent was HPLC-grade sterilised water (product number: 270733-1L-M, EMD Millipore).

The Bradford dye reagent was added to standard solutions and protein extracts at a ratio of 1:1. Due to the small amounts of protein extract, 10µl, obtained in section 2.10.4., a further 10 × dilution was applied to these protein aliquots. Then, 100µl of Bradford dye reagent was added to its equivalent volume of the diluted protein extracts. The dilution factor of 10 was considered in the calculations. The added mixtures were incubated at room temperature for no less than five minutes. A spectrophotometer was set at 595nm, and the instrument was zeroed with a blank sample with no protein contents. All measurements were recorded, noting that the actual readings were obtained after blank subtraction.

2.11. Preparation of protein sample for tryptic digestion

The remaining 90µl of each protein supernatant was mixed with 360µl of MeOH and the tubes were closed tightly and vortexed for 10 seconds. Another centrifugation round was applied at 9000 × g for 10 seconds. Then, 100µl of chloroform was added to all tubes, after which all tubes were re-sealed, vortexed thoroughly, and centrifuged again at the aforementioned conditions for centrifugation. 300µl of HPLC-grade sterilised water (product number: 270733-1L-M, EMD Millipore) was added to each tube, then the tubes were sealed, vortexed for 10 seconds, and centrifuged at 9000 × g for one minute. The samples were visually examined and three layers of different phases were observed: a lower layer containing transparent organic chloroform phase, an interphase layer containing protein as a pale phase, and a higher clear layer representing the aqueous phase. The latter phase was removed slowly with extreme care so as not to disturb the two bottom layers. Then, 300µl of MeOH was added to all tubes; the tubes were sealed, vortexed for 10 seconds, and centrifuged at 9000 × g for two minutes. After this centrifugation, protein pellets were seen as gel-like pellets or round sheets of a whitish colour. The supernatant was discarded carefully and the protein pellets were washed softly with 1ml of absolute EtOH. A final centrifugation round was applied at 9000 × g for two minutes, followed by the

removal of the supernatant. The tubes were sealed with sealing clips and stored at -80°C in an ultra-freezer.

2.12. HPLC/MS procedures

2.12.1. Tryptic digestion of extracted proteins for mass spectrometry

Sample preparation and protein extraction were carried out as per the Qproteome FFPE Handbook 01/2011 (Qiagen) with the moderate alterations arising from extensive optimisation. Extracted protein (100µg) was diluted in 8M Urea + 25mM tri-ethyl ammonium bicarbonate (TEAB) before reduction (50mM DTT at 37°C for one hour), alkylation (100mM iodoacetamide at 37°C for one hour in darkness), and then digestion for 18 hours using Trypsin (Sigma) at 37°C at a 20:1 protein:protease ratio (w/w) in a thermomixer (950 rpm). The reaction was ceased using 5% trifluoroacetic acid (TFA) and further incubated at 37°C for one hour. Samples were then subject to a 20,783×g centrifugation for 30 minutes at room temperature, after which the supernatant was removed and lyophilised before suspension in 5% acetonitrile + 0.1% formic acid for data acquisition with a Sciex TripleTOF 6600 mass spectrometer in both SWATH® and information dependent acquisition (IDA) modes [173, 174].

2.12.2. Reverse-phase high-performance liquid chromatography electrospray ionisation tandem mass spectrometry (RP-HPLC-ESI-MS/MS) in SWATH and IDA modes

Each sample was analysed on a Sciex TripleTOF 6600 mass spectrometer coupled in line with an Eksigent ekspert nano LC 425 system running in micro flow (5µl/minute) in both a data-dependent (IDA) and data-independent acquisition mode (SWATH). In brief, 9ul of each sample was injected and trapped onto a YMC Triart-C18 pre-column (0.3 x 5mm 300 µm ID) at a flow rate of 6µl/min for two minutes. The sample was then eluted off the pre-column and onto a YMC Triart-C18 analytical column (15cm, 3µm, 300µm ID) that was in line with the electrospray ionisation source of the 6600 MS. The following linear gradients were used: Solvent B (2% acetonitrile + 0.1% formic acid) over solvent A (0.1% formic acid), increasing from 3% to 30% over 38 minutes, then increasing to 40% after 43 minutes, followed by a five-minute wash at 80% B before re-equilibration for the remainder of the total 57-minute run (SWATH). Solvent B (2% acetonitrile + 0.1% formic acid) over solvent A (0.1% formic acid), increasing from 3% to 30% over 68 minutes, then increasing to 40% after 73 minutes, followed by a five-minute wash at 80% B before re-equilibration for the remainder of the total 87-minute run (IDA). Electrospray ionisation was performed by applying +5500 V via

a 50µm electrode using the Duospray source (Sciex). Data independent acquisition (SWATH) was performed using 100 variable SWATH windows (optimised via a Sciex variable window tool; 25 milliseconds (msec) accumulation time, rolling collisions energy) and IDA with a top 30 ion fragmentation, followed by 15 seconds exclusion using rolling collision energy.

2.12.3. Construction of peptides library, spectral normalisation, and fold change analysis

From the IDA acquisition of each sample, a library file was generated using ProteinPilot 5.0 (Sciex). The following stringent search criteria were applied: biological modifications, thorough ID, searching the UniProt Swissprot database (January 2015 release) for human species, protein-grouping to count the total number of proteins identified in each run, and a 1% false discovery rate (FDR). The overlap of proteins identified in each library was compared and considered; the two top-performing libraries with the most protein identities, one from each of the LVI positive and negative groups, were combined, again using Protein Pilot 5.0 (Sciex). The resulting library file was aligned to the SWATH data using PeakView 2.0 software SWATH microapp (Sciex), after 1% FDR filtering and the exclusion of shared peptides. SWATH data were processed with an extraction window of five minutes and application of the following parameters: 30 peptides/proteins, six transitions, peptide confidence of >99%, excluding shared peptides and extracted ion chromatogram (XIC) width set at 75 ppm. Quantitation and fold change analysis was carried out to identify differentially expressed proteins (DEPs) between different patient subsets using SCIEX OneOmics cloud processing software (Sciex).

2.13. Statistics

2.13.1. Introduction

Several statistical parameters were accredited in this study. Nominal variables, including categorical and ordinal variables, continuous scale data, and independent or dependent variables exemplify this variety. Each statistical test requires specific conditions and variable type(s); therefore, multiple statistical analyses were conducted. All statistical tests were performed with the Statistical Package for the Social Sciences (SPSS).

2.13.2. Descriptive statistics

The description of the available data in this study was essential to identify specific features of a particular dataset. For example, scale H-score data for a population, mean, median, and mode were investigated and depicted into

histogram charts illustrating the normal distribution curve. These three statistics characterise the central tendency statistics for datasets of interest. Cut-off points for continuous variables were decided according to their normal distribution status. If the data was normally distributed, the cut-off value was determine around the mean. Otherwise, median was designated as a proper cut-off point. Other figures, such as percentage and total number of cases, were also employed in this study using descriptive analyses.

2.13.3. Intra-class correlation coefficient (ICC) and Cohen's kappa coefficient

As part of the quality control and H-score assessment procedures, the continuous H-score data for an immune-stained protein generated by two individual observers were subjected to the ICC test. This test quantitatively measures the resemblance between two datasets and assigns a degree of scoring reliability between the two scorers. The ICC agreement test was applied to all stained protein, and for at least 30% of all valid TMA cores. On the other hand, Cohen's kappa coefficient test was applied to the scoring data that was given a categorical format of positive and negative.

2.13.4. Chi-square test (goodness to fit)

The chi-square test was one of the non-parametric statistical tests that was used in this study to determine the statistical relationships between two, or more, nominal independent variables. A chi-square test could expect that the difference of rates distribution for a selected and categorised biomarker of interest against valid equivalent data belongs to the other nominal variable(s), LVI, tumour grade, LN stage, or other previously stained biomarkers, mathematically. Output data was presented in cross-tabulation matrices of rows and columns. The threshold of significance (p-value) was set at <0.05.

2.13.5. Mann-Whitney U test (median analysis)

As with the chi-square test, the Mann-Whitney-Wilcoxon (MWW) test is another non-parametric test that was used for univariate analysis with clinicopathological variable and selected biomarkers. This test allowed us to evaluate the statistical associations between the nominal data and the distributed median of the continuous data. The postulation of normalised scaling data for this test is not required, in contrast with the student's t-test. The p-value was set at <0.05.

2.13.6. Student's t-test (mean analysis)

The student's t-test was used prior to this study to assess the statistical difference between the means of two normally distributed variables. Basically, the

student's t-test was used in the DGE analysis of limma. The p-value was set at <0.05.

2.13.7. Pearson rank test (model/line of best fit)

The Pearson correlation coefficient (r) was used in this study to determine the strength of the linear association between two normalised transcripts of continuous variables. The value of r takes a range of values from +1 to -1. The value of 0 indicates that there is no association between the two interrogated variables; the closest value to +1 indicates a strong co-expression in the same direction, whereas the negative values suggest an inverse correlation of co-expression. The correlations were considered to be significant if the p-value was <0.05. The correlations were considered to be strong, in either direction, if the values of r were between (-1.0 to -0.5) or (1.0 to 0.5), moderate if the values of r were between (-0.5 to -0.3) or (0.3 to 0.5), weak if the values of r were between (-0.3 to -0.1) or (0.1 to 0.3), and non-existent if the values of r were between (-0.1 to 0.1).

2.13.8. Kaplan-Meier survival analysis and log-rank test

The log-rank test was applied to estimate the survival distribution for two populations of BC patients statistically, when survival observations were censored. The associated estimations of BC patients' survival were plotted onto Kaplan-Meier curves. The survival status for a patient who had died from BC was numerated as (1) in SPSS to indicate this event. The p-value was set at <0.05.

2.13.9. Proportional hazards model (Cox regression model)

The prognostic independence of an investigated differentially expressed gene or protein was assessed against standardised prognostic factors in Cox's proportional hazard model. Tumour grade, LN stage, and tumour size are examples of the used predictive factors. The p-value was set at <0.05.

2.14. Ethical approval

This study was approved by the Nottingham Research Ethics Committee 2 under the title "Development of a molecular genetics classification of breast cancer" under the ethical approval number (RC2020313) and in compliance with the current ethical and legal guidelines of the United Kingdom. All basic and operational ethical principles for this study were conducted in accordance with the 1975 Helsinki Declaration, and its later amendments or comparable ethical standards. The identification and health-sensitive information for all patients in this study were anonymous.

Chapter 3: Integrated Differential Gene Expression and Bioinformatics Analyses

3.1. Introduction

We illustrated how invasive BC is a highly malicious and heterogeneous disease that diverges substantially between different patients, and shows inter-tumour heterogeneity, in the introduction to this study. The sophistication of the morphological, genomic, transcriptomic, and proteomic expression profiles for invasive BC were confirmed in individual tumour-borne patients [175]. The dynamic molecular alterations in invasive BC patients necessitate the improvement of the current molecular classification strategies and targeted therapeutic methods for subsets of patients vulnerable to invasion mechanisms. The successful discovery of the pivotal genomic features, including the aberrant expressions of genes, mRNAs, or proteins, could improve the medical therapeutics; hence, BC management. The recent advancements in the high-throughput techniques have facilitated the identification of several aberrantly expressed biomarkers and pathways that motivate BC invasiveness and may affect the survival of invasive BC patients. Regarding other clinicopathological features of invasive BC, LVI could be histologically detected and molecularly confirmed by applying H&E and IHC techniques.

Recently, integrated differential gene/protein expression and bioinformatics analyses have been utilised in myriad cancer studies in order to reveal clusters of DEGs that are enriched in cancer-related pathways. A recent study imported the gene expression profiles from the Gene Expression Omnibus (GEO) and investigated the key upregulated and downregulated genes between 113 BC and 38 adjacent normal-like tissues. A cluster of 82 genes was identified as being differentially expressed in their datasets, and their functions could be related to cell division, proteinaceous ECM, ECM structural constituents, and ECM-receptor interaction [176]. Another study used downloaded gene expression profiles of the luminal-A phenotype of invasive BC from The Cancer Genome Atlas (TCGA) database and compared the DEGs between the invasive BC patients showing both poor and good outcomes. The study concluded that a cluster of 18 prognostic biomarkers was associated with nine different pathways [177]. Regarding LVI in invasive BC, a recent study on 79 LVI^+ and 29 LVI^- invasive BC patients identified DEGs that are associated with LVI by combining the significance analysis of the microarrays (SAM) method and the limma package in R. The study concluded that the use of gene set enrichment analysis (GSEA), gene ontology (GO), protein-protein interaction (PPI), and pathways analysis proposed that the cytokine-receptor interaction pathway and the disparity between the components of ECM and TME could trigger LVI mechanisms in invasive BC [178].

The employment of LVI implications in invasive BC prognosis could lead to the identification of subsets of high-risk invasive BC patients that require critical consideration. The molecular code for LVI is still ambiguous, and the conclusive gene/protein signatures have still not been deciphered clearly. There are several factors that could challenge the precise molecular signature of LVI in invasive BC. Artefacts in the FFPE archival tissue processing and staining, the sample size of representative $LVI^{+/-}$ specimens, the different molecular phenotypes and clinicopathological characteristics of an LVI prospective cohort, and the possible false detection of LVI^{-} when LNM^{+} are a few examples of obstacles that may attenuate the accurate identification of potential LVI drivers.

The aim of this study is to identify novel DEGs modulating LVI mechanisms in invasive BC patients and to pinpoint potential prognostic and drug-targeted mRNAs, and possibly proteins, for future examination. The utilisation of CNA for selected DEGs and the complete transcriptome for the rest of the DEGs allowed us to modify the DEGs selection with regard to LVI. The availability of a large, robust, and well-characterised cohort of invasive BC such as the METABRIC and NPS cohorts, the accessibility of cutting-edge molecular pathology facilities and techniques in our department, and the contributions from many enthusiastic collaborators at CUH and NTU encouraged us to study the unseen molecular stages that may underlie LVI mechanisms in invasive BC patients.

3.2. Materials and methods

3.2.1. **Study cohorts**

3.2.1.1. *LVI-supervised cohort*

As stated in the materials and methods chapter of this study, the construction of the $LVI^{+/-}$ cohort of invasive BC patients was performed by collecting samples from a subset of the METABRIC cohort. The tissue specimens of the latter cohort were transferred from the NHSB at the NUH to the METABRIC group at CUH, and were utilised in the METABRIC studies on the genomic and transcriptomic landscapes of invasive BC [38, 171]. The qualified samples in the construction of the $LVI^{+/-}$ cohort of invasive BC patients for this study were FFPE tissues that exhibited: 1) robust clinicopathological profiles with long-term follow-up data; 2) histological observations of tumour emboli in the vascular lumen using H&E; 3) biological detection of tumour emboli in the lymphatic cavities and LVI, using PDPN,CD31,CD34/IHC; (Figure 3.1) 4) no LN^{+} in patients with LVI^{-} status; and 5) integrated profiles for genome-wide expression and CNAs. The strict

criteria for sample selection yielded 170 qualified samples: (LVI^+ = 90) and (LVI^- /LN^- = 80).

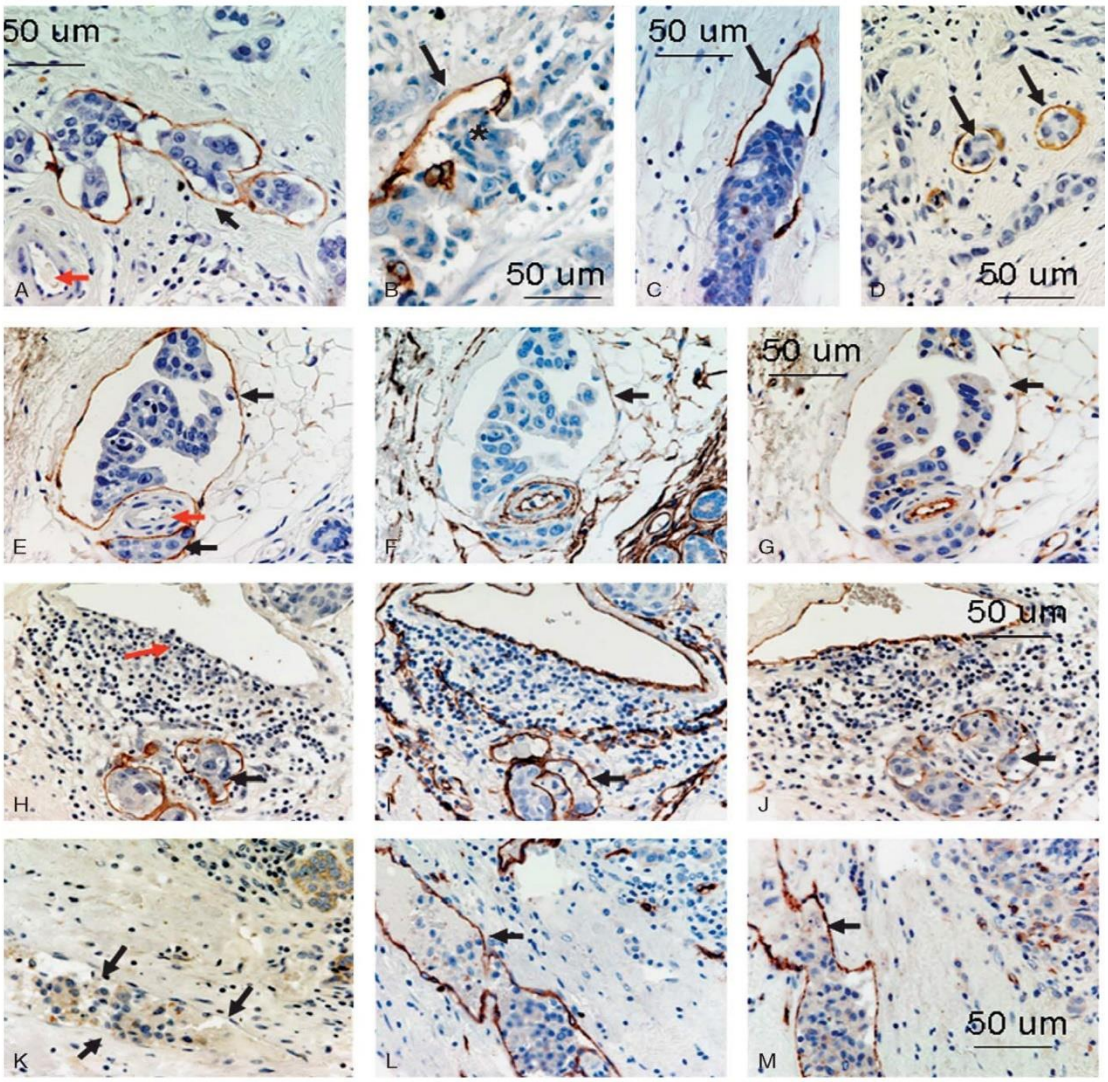


Figure 3. 1: "The first panels (A–D) are multiple forms of lymphatic invasions in sections stained with podoplanin, (A) large emboli in a large lymphatic. The adjacent blood capillaries are negative for podoplanin (red arrow). B, Tumour cells (asterisk) appear to invade toward the lymphatic lumen through a wide gap between lymphatic endothelial cells. C, A large sheet of tumour cells appear to invade part of a lymphatic with few cells detached at the top. D, Two small tumour emboli consists of few tumour cells but easily detected in podoplanin-positive vessels. Scale bar 50 mm. The second panels (E–G) show tumour cell clusters within lymphatics that are nicely demarcated with podoplanin (black arrows) (E), note negative small blood capillary (red arrow). The same lymphatic is negative for CD34 (F) and CD31 (G) indicating the lymphatic nature of that vessel. The third panel shows invasion of lymphatics in (H) (arrow) that are also positive for CD34 (I) and CD31 (J). The lower panel shows tumour cell clusters within a podoplanin-negative vessel (arrows) (K) that is positive for both CD34 (L) and CD31 (M), indicating that this type of VI is a BVI. Scale bar 50 µm."

This figure was reused according to the agreement between Mr. Sultan Sonbul and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc."), License Number: 4444781031820, provided by Wolters Kluwer Health, Inc. and Copyright Clearance Centre. See Appendix B, page 391.

Reference:

MOHAMMED, R. A., MARTIN, S. G., GILL, M. S., GREEN, A. R., PAISH, E. C. & ELLIS, I. O. 2007. Improved methods of detection of lymphovascular invasion demonstrate that it is the predominant method of vascular invasion in breast cancer and has important clinical consequences. Am J Surg Pathol, 31, 1825-33.

The dominant histological type of invasive BC in the $LVI^{+/-}$ cohort of patients was the invasive ductal BC, 71.2%. The LVI^{+} status in the $LVI^{+/-}$ cohort was confirmed for 52.9% of patients. A total of 61.2% of patients were older than 50 years, 41.2% of tumours were larger than two centimetres, 51.8% of tumours were grade-III, and 15.3% were grade-I. The negative nodal involvement, according to the NPI staging system, labelled 61.8% of patients, while the advanced LNM stages were recorded for 9.4% of patients. Therefore, 40.6% of patients belonged to the good NPI index; 51.8% and 7.6% of patients were indexed into moderate and poor NPI indices, respectively.

At the molecular level, $ER^{+/IHC}$ status was observed in 75.3% of patients. The expression of HER2 was positively detected with IHC and validated by CISH/FISH methods, and it was confirmed in 12.9% of patients. The absence of the hormonal receptors ER, PR, and HER2, (triple-negative^{+}) was observed in 17.6% of patients in the $LVI^{+/-}$ cohort. The intrinsic subtyping of the specimens classified the patients of this cohort as 7.1% normal-like, 37.6% luminal-A, 26.5% luminal-B, 8.2% HER2-enriched, and 20.6% containing a basal-like tumour. The integrated classification clusters of METABRIC (IntClusts) revealed that 11.2% of patients fit into IntClust-1, 3.5% fit into IntClust-2, 17.6% fit into IntClust-3, 14.7% fit into IntClust-4, 11.8% fit into IntClust-5, 1.8% fit into IntClust-6, 8.8% fit into IntClust-7, 11.2% fit into IntClust-8, 4.7% fit into IntClust-9, and 14.7% fit into IntClust-10.

3.2.1.2. *The METABRIC cohort*

A large number of transcriptomic and CNA profiles for 1,980 patients were provided by the METABRIC group for this study, supported by validated clinicopathological data. The essential applications for this cohort were to: 1) validate limma output for the DGE experiment performed on the $LVI^{+/-}$ cohort; 2) provide this study with subsets of invasive BC patients with known $LVI^{+/-}/LN^{+/-}$, transcriptomic profile, and CNA data; and 3) perform comprehensive gene expression correlation analysis on a selected cluster of DEGs utilising a large-size cohort of invasive BC patients. The construction of the METABRIC cohort utilised in this study was based on the previous combination of CUH invasive BC patients (n=997), with hundreds of external invasive BC samples obtained from different university hospitals and biobanks, including NHSB. In this cohort, the $LVI^{-}/LN^{+/-}$ patients of the NPS cohort were included to assess the potential differences between the inclusion and exclusion of the LN status in the LVI^{-} patients in this study. For CUH patients, the pathological review of the LVI status was performed locally on subsets of invasive BC patients (n=521), and followed the same sample

selection criteria applied to Nottingham's $LVI^{+/-}$ cohort of invasive BC patients. The DGE from both cohorts was then compared for reciprocal confirmation.

The percentage of patients with LVI^{-} status in the METBARIC cohort was 55.4%, while the LVI^{+} patients accounted for 44.6%. The patients' age at diagnosis was older than 50 years for 78.3% of patients. The invasive BC patients with grade-III tumours accounted for 50.3% of the sample size. The nodal involvement was positive in 47.5% of patients, while slightly more the half of the patients in this cohort were given a moderate NPI. The common histological type in this cohort was the invasive ductal BC, which was seen in 82.6% of the patients. The BCSS for each patient was calculated in months, up to 300 months, and the date recording began on the day of the primary surgical treatment to the time of death caused by invasive BC. The latest follow-up status showed that 25.5% of patients had died due to invasive BC.

At the molecular level, the $ER^{+/IHC}$ status was observed in 75.7% of patients. The expression of HER2 was overly expressed in 14.5% of patients, whereas 3.3% were on the borderline. The intrinsic subtyping of the specimens classified the patients of this cohort as 10.4% normal-like, 36.3% luminal-A, 24.6% luminal-B, 12.1% HERS-enriched, and 16.6% having a basal-like tumour. The IntClusts of METABRIC revealed that 7.0% of patients fit into IntClust-1, 3.6% fit into IntClust-2, 14.6% fit into IntClust-3, 17.3% fit into IntClust-4, 9.6% fit into IntClust-5, 4.3% fit into IntClust-6, 9.6% fit into IntClust-7, 15.1% fit into IntClust-8, 7.4% fit into IntClust-9, and 11.4% fit into IntClust-10.

3.2.2. **DGE and retrieval of copy number aberration**

3.2.2.1. *Nucleic acids extraction*

Prior to this study, the extraction and purification of nucleic acids was performed on 30μm × 10 sections of fresh frozen biopsies, as described in [38]. DNA and RNA were extracted and purified from each specimen using the DNeasy blood and tissue kit and the miRNeasy kit on the QIAcube (Qiagen, Crawley, UK).

For the genomic DNA extraction, the FFPE tissues required special pre-treatment steps before they met the same DNA extraction conditions of the fresh frozen tumour biopsies. All FFPE tissues were sectioned and weighed individually to be no more than 25mg before being placed in 2ml microcentrifuge tubes. The measured FFPE tissues were de-waxed with 1200μl of xylene and vortexed vigorously. The tissue suspensions were centrifuged at maximum speed for five minutes. The supernatants were discarded and 1200μl of 96~100% EtOH was added to each pellet-containing tube using gentle mixing. A second round of

centrifugation was performed on all tubes as in the previous step. The xylene residuals were removed with EtOH by pipetting. The EtOH addition and centrifugation steps were then repeated once more. The pellets were incubated at 37°C for 10–15 minutes in order to evaporate the excessive EtOH. Each pellet was then resuspended in 180µl of the ATL buffer (Cat No. 19076, Qiagen, Crawley, UK). Next, 20µl of proteinase K (Cat No. 19131, Qiagen, Crawley, UK) was added to each FFPE and fresh frozen tumour tissues, mixed thoroughly, and incubated in a shaking water bath at 56°C until the tissues were completely lysed.

In order to prepare RNA-free genomic DNA, 4µl of RNase A (Cat No. 19101, Qiagen, Crawley, UK) at (100mg/ml) was added to each sample, mixed by vortexing, and incubated for two minutes at room temperature. A quick vortex for 15 seconds was applied to each tube before the addition of 200µl of the AL buffer (Cat No. 19075, Qiagen, Crawley, UK). All samples were then mixed thoroughly. Next, 200µl of 96–100% EtOH was added and the samples were mixed again by vortexing. The entire mixture was then pipetted into the DNeasy Mini spin column and placed in a 2ml collection tube provided with the DNeasy kit. All tissue mixtures were then centrifuged at 6000 x g for one minute. The DNeasy Mini spin columns containing the tissues were placed in new 2ml collection tubes, and the flow-through liquids and the used collection tubes were discarded.

After this, 500µl of the AW1 buffer (Cat No. 19081, Qiagen, Crawley, UK) was added to each sample and centrifuged for one minute at 6000 x g. Again, the DNeasy Mini spin columns were placed in new 2ml collection tubes and 500µl of the AW2 buffer (Cat No. 19072, Qiagen, Crawley, UK) was added, before being centrifuged for three minutes at 20,000 x g in order to dry the DNeasy membranes. The DNeasy Mini spin columns were placed in clean 2ml microcentrifuge tubes and the used collection tubes were discarded. Next, 2 × 100µl of the AE buffer (Cat No. 19077, Qiagen, Crawley, UK) were pipetted directly onto the DNeasy membranes to increase DNA elution. They were then incubated for one minute at room temperature before being centrifuged for one minute at 6000 x g after each addition of the AE buffer. The genomic DNA eluates were then obtained.

For the total RNA extraction, an optimum weight of fresh frozen tissues that would yield a total RNA concentration between 50–200ng was set and dissected from each sample accordingly. The dissected tissues were placed into a tissue homogenisation vessel and 700µl of the QIAzol lysis reagent (Cat No. 79306, Qiagen, Crawley, UK) was added. The tissues were disrupted for 20–40 seconds or until they were totally homogenous. The homogenates were placed on the

benchtop and left at room temperature for five minutes. Next, 140µl of chloroform was added to the tubes containing the homogenate, the caps were sealed securely, and all the tubes were rocked intensely for 15 seconds. All the tubes were then kept on the benchtop for two to three minutes before being centrifuged for 15 minutes at 12,000 x g at 4°C. The upper aqueous phase from each specimen was then transferred to a new collection tube. The addition of absolute EtOH at 1.5 volumes was performed and each tube was mixed thoroughly by several up-and-down pipetting cycles. A total of 700µl of each sample was instantly pipetted into an RNeasy Mini spin column in a 2ml collection tube. All the tube lids were gently secured before being centrifuged at 8000 x g for 15 seconds at room temperature. All flow-through liquids were then discarded. Next, 700µl of the RWT buffer (Cat No. 1067933, Qiagen, Crawley, UK) was added to the RNeasy Mini spin column, the lids were closed, and the tubes centrifuged for 15 seconds at 8000 x g in a column washing step. The flow-through was again discarded. After this, 500µl of the RPE buffer (Cat No. 1018013, Qiagen, Crawley, UK) was pipetted into the RNeasy Mini spin column, the lids were closed, and the tubes centrifuged for a further 15 seconds at 8000 x g in a second washing step. The flow-through was discarded. The final washing step with the RPE buffer was repeated, but using two minutes of centrifugation under the same conditions. The elongation of the centrifugation time was intended to dry the RNeasy Mini spin column membranes. The drying step in this protocol was essential in order to completely remove any residual EtOH that may inhibit the RNA downstream reactions. The RNeasy Mini spin column was added to new 1.5ml collection tubes. A volume of 30–50µl of RNase-free water (Cat No. 129112, Qiagen, Crawley, UK) was pipetted directly onto each RNA-containing RNeasy Mini spin column membrane, the lids were closed gently, and the tubes were centrifuged for one minute at 8000 x g in order to elute the RNA. The total RNA eluates were then obtained.

3.2.2.2. *Microarray hybridisation and scanning of Illumina HumanHT-12 v3 expression beadchips*

The extraction/purification process was performed prior to this study using the fully automated platform QIAcube (Qiagen). The extracted and purified nucleic acids were quantified using a NanoDrop ND-8000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) before the extracted samples were run on agarose gel. The extracted and purified DNA probes were hybridised to Affymetrix SNP 6.0 arrays (Affymetrix, Santa Clara, CA) at AROS Applied Biotechnology (Aarhus, Denmark), while an Agilent 2100 Bioanalyser Nanochip (Agilent

Technologies, Wokingham, UK) was used to assess the quality of the RNA probes. Biotin-labelled cRNA was produced from the total extracted and purified RNA. The Illumina Totalprep RNA amplification kit (Ambion, Warrington, UK) was purchased, then the labelled cRNA were hybridised onto Illumina Human HT-12 v3 Expression BeadChips provided by the latter manufacturer. All samples in this study were subjected to quality control check tests for proper optimisation, i.e. total RNA extracts must be between 50–200ng. Any sample that failed to satisfy the minimum requirements for the microarray experiment was eliminated from the study.

The raw generated data from the scanned Illumina HumanHT-12 v3 expression beadchips were introduced into the bead array package, a bioconductor package using R programming language to analyse the Illumina bead-based data. The background noise of the raw gene expression data was corrected and quantiles were normalised. The summarised and normalised mRNA data were fitted to the linear statistical model for microarray data and the alterations in gene expression were quantified between the defined LVI^+ and LVI^-/LN^- test and validation subgroups. The statistical RNA microarray outputs for the LVI^+ and LVI^-/LN^- of the test and validation cohorts were saved as .txt files and converted to .csv format for ease of access.

3.2.2.3. *Analysis of CNA and obtaining the complete transcriptomic data of the METABRIC cohort*

The CNAs refer to the structural alterations that may lead to gene copy gains or losses on a DNA segment of at least one kilobase (Kb) compared to a reference genome. The fundamental difference between the CNA and the copy number variance (CNV) is that the CNA shows any somatic changes in a gene copy number. The CNV denotes variations in germline cells' copy numbers [179, 180].

In the early stages of this study, the complete CNA dataset for the METABRIC cohort was not accessible. Pragmatically, we required a retrieval for CNA data for the top DEGs according to the results of the DGE analysis on the LVI+/- cohort of patients. We agreed on this approach to minimise the time consumption by reducing the number of inquired genes for the interrogated patients of the LVI+/- cohort. The retrieved CNA data was received as a PDF file of CNA distribution box plots. The statistical test for the ordered alternatives, Jonckheere-Trepstra, was performed to assess the linear trend of distribution for a gene CNA status and its mRNA expression across the two LVI categories in the LVI+/- patients. The number of hits in the hetero-deletion, homo-deletion, neutral, gain, and

amplification categories were calculated for each gene among the LVI+/- groups. An adjusted p-value of significance was given for the inquired genes and their CNA categories with the LVI+ or LVI- patients. The Jonckheere-Trepstra statistics were combined with the calculated ratios of CNA-gain hits for the LVI+ and LVI- invasive BC patients in order to prioritise the selection of the LVI candidates. The selected DEGs, MCFD2 and CCT8, were the first patch of targets that we included in the protein expression study, using IHC methods, in the following chapter. However, and in independent approach, we searched the GSEA databases for a co-enriched gene with MCFD2 and CCT8. SNAP23, synaptosomal-associated protein 23, was only investigated at the protein level. Further details are presented in Chapter 4 (section 4.1.3.).

At a later date, we managed to gain full access to the complete DNA-CNA dataset, clinicopathological profiles, and the full transcriptomic database of the METBARIC patients, courtesy of Dr Oscar Rueda at the University of Cambridge. By utilising the full transcriptomic data, we thoroughly analysed a specific cluster of highly significant DEGs for their upregulated or downregulated gene expressions against the LVI^{+} and LVI^{-} categories. The interrogated DEGs were those of positive log2FC, positive AvE, and a perfect DNA probe binding, according to the output of limma analysis. The transcriptomic/LVI analyses were performed on the patients of the $LVI^{+/-}$ cohort. We also included the LVI^{-}/LN^{+} patients with transcriptomic data in a separate investigation to assess the efficacy of our selection of LVI patients in this study. The selection of DEGs with upregulated expression levels in the LVI^{+} patients from this analysis were prioritised for protein expression study, using IHC methods, as presented in the following chapter.

The gene expression gain and loss from Affymetrix SNP 6.0-arrays for the whole transcriptome of the METABRIC cohort (n=1,980) were received in separate data files: one file was for the gene expression gain, while the second was for the gene expression loss. The data in each file were represented with the binary codes of 0 and 1 (0=No, 1=Yes). We checked all of the gene expression profiles in both CNA gain and CNA loss datasets for any repeated 0/1 statuses, but we found no discrepancies. In addition, we utilised the accurate data in both files to generate the neutral CNA status of (0/1) when an inquired gene exhibited neither gain nor loss CNA status.

3.2.3. **Strategies of gene selection**

3.2.3.1. *Profile analysis of DEGs*

As microarray is a large-scale experiment that could result in a massive list of DEGs, and to avoid any potential disposal of key regulatory genes of LVI, we aimed to select an accurate number of LVI candidates based on several selection criteria.

First, we cross-validated the gene identifiers of the LVI-DEGs from $LVI^{+/-}$ patients (n=170) against those from CUH (n=521). We gave these two lists of LVI-DEGs the codes of LVI-NT2225 and LVI-AD3000, respectively. We denoted the identified and cross-validated DEGs from this method as the LVI-NT1154 cluster. Second, we followed a basic approach and inclusively selected the most significant of the cross-validated DEGs from the LVI-NT1154 cluster by setting the adjusted p-value of <0.05 as the threshold of statistical significance for gene selection. We restricted our identified list to the first 100 DEGs ranked by their ascending adjusted p-values. We limited our selection to the first 100 DEGs because this was the maximum doable number of gene identifiers for the retrieval of CNA that the METABRIC group could manage. However, we added more substitutive genes because we identified a few pseudogene/old nomenclature identifiers in the latter list. This group was named the LVI-NT100 cluster. The DEGs that were studied later at the protein level were those originating from epithelial cells, which showed statistical significance with LVI, CNA, CNA gain ratios of >2.50, have commercially available primary Abs specific for WB and IHC methods, and have no records of published studies regarding their regulation of LVI in invasive BC. The required time for CNA retrieval was two months and this was performed before we gained full access to the complete CNA dataset.

Third, we refined the validated cluster of LVI-NT1154 by exclusively pursuing genes with the most significant adjusted p-values <0.05, positive log2FC, positive AvE, and a perfect probe binding in the $LVI^{+/-}$ patients. We named this unique cluster of DEGs the LVI-NT283 cluster. We statistically analysed the mRNA expression levels for the obtained DEGs in this cluster against the histologically-defined LVI status of the $LVI^{+/-}$ patients. Each single DEG from the LVI-NT283 cluster that statistically showed upregulated levels with the LVI^{+} patients was nominated to the later protein studies.

3.2.3.2. *Pathway analysis and clinicopathological interrogations*

The interrogation of the MFs, BPs, and CCs, and the clinicopathological impacts for the selected DEGs in this study were all considered. The exhaustive analysis of the selected clusters of the protein-encoding DEGs was carried out using the Protein Analysis Through Evolutionary Relationships (PANTHER) classification system [181]. The PANTHER database converges annotated data

from the GO phylogenetic annotation project [182] and supports pathway enrichment analysis and pathway classifications from the Reactome peer-reviewed KB [183]. The genetic and protein interactions and the network visualisations for each selected DEG in this study were examined using the Biological General Repository for Interaction Datasets (BioGRID) [184]. We also applied further in silico analysis by utilising other KBs in this study to increase our basic understanding of the selected DEGs. These extra resources were: 1) the Universal Protein Resource (UniProt) KB and the central repository of protein data created by combining the Swiss-Prot, TrEMBL, and PIR-PSD databases [185]; 2) PhosphoSitePlus®, an exploratory KB for the experimentally observed post-translational modifications (PTMs) in cellular signalling [186]; 3) GeneCards®, a genomic, transcriptomic, proteomic, clinical, and functional integrative KB [187] [https://www.genecards.org]; and 4) several databases of molecular information at the National Centre for Biotechnology Information (NCBI). Finally, the associations between the altered expressions of the selected DEGs in this study were assessed against robust clinicopathological profiles, different molecular subtypes, and the BCSS of invasive BC. We used the breast cancer Gene-Expression Miner (bcGenExMiner) version 4.1 to examine the impact of altered expression of the selected DEGs on BCSS in a larger external pooled cohort of invasive BC patients [188, 189].

3.2.4. **Statistics**

Prior to this study, the statistical analyses were performed using the limma R/Bioconductor software package and, in this study, SPSS statistical software v24.0 (IBM, Armonk, NY, USA). The identification of upregulated and downregulated gene expressions against LVI status was achieved by data processing of the normalised gene expression arrays obtained from the limma R script. The list of DEGs from the microarray DGE experiment was generated by performing the limma R/Bioconductor inclusive statistics. The summarised statistics were computed using the moderated t-statistics, which measured the ratio of the log2FC of a gene to its standard error, before the Bonferroni adjustment was used to correct p-values for multiple testing; and the B-statistics, which calculated the odds for a gene to be differentially up- or down-expressed, as described previously in the supplementary information file in [38].

The characteristics of the $LVI^{+/-}$ and the METABRIC cohorts were determined by utilising the descriptive analytical commands of SPSS statistical software v24.0. The chi-squared test was used to test the distribution of the categorised mRNA expression for each highly ranked DEG against robustly defined LVI categories,

age at diagnosis, tumour grade, nodal involvement, NPI, PAM50, and IntClusts of METABRIC. Kaplan–Meier (KM) survival curves for BCSS were plotted and the log-rank test was conducted. The BCSS in this study was determined as the time from the day of surgery to the date of death caused by BC. The Cox proportional hazards regression model and 95% confidence intervals (CI) were applied in order to investigate the relationships between the survival of patients in this study against several clinicopathological and molecular parameters including the mRNA expression for each highly ranked DEG. Finally, we applied comprehensive gene expression analyses to assess the gene-gene interrelations for the selected LVI candidates. The Pearson correlation test was used to evaluate the gene expression correlations between the identified DEGs from the LVI-NT100 and LVI-NT283 clusters. The median mRNA expression value was set as a cut-off point in the mRNA-based categorised statistics.

3.3. Results

3.3.1. DGE profile of the LVI-NT1154

The total number of identifiers for the DEGs with statistical significance against the two categories of LVI in the $LVI^{+/-}$ was 2,225. The external validation cohort of the CUH patients yielded 3,000 DEGs under the same experimental conditions. The cross-validated results for the two RNA microarray experiments of invasive BC patients revealed a cluster of 1,154 DEGs that had their gene expressions altered significantly against LVI. The mapping identifiers for these LVI-DEGs were checked for any updates on their nomenclatures, and whether they were protein-encoding genes, pseudogenes, uncharacterised protein-encoding genes, or RNA genes. Our inspection of these 1,154 DEGs revealed 28 mapping identifiers with no annotations for protein encoding in the (Gene) database on NCBI. The refined cluster of LVI-NT1154 offered a total of 1,126 protein-encoding DEGs from which to select.

3.3.2. DGE profile of the LVI-NT100

The presented results of the gene/protein classification for the LVI-NT1154 cluster in the previous sections showed how the study of LVI mechanisms can be extremely challenging. We aimed to minimise the number of studied DGEs in order to identify focused MFs, BPs, CCs, PWs, and PCs GO terms/child terms that could be enriched in LVI and with its regulatory mechanisms. The listed genes in the LVI-NT100 showed one pseudogene, *C14orf19*, and three other genes with no mapping identifiers in the PANTHER database, *C4orf41*, *HIAT1*, and *KIAA1279*. We

excluded two more gene identifiers, *APPL1* and *FAM162A*, from the PANTHER gene/protein classification analysis as their CNA data were not retrieved.

3.3.2.1. *CNA data analysis of the LVI-NT100 cluster of DEGs*

The results of Jonckheere-statistics of CNA in the $LVI^{+/-}$ cohort of invasive BC patients revealed a limited number of statistically significant genes (n=12, p-value <0.05). These genes were: *ARFGEF1*, *ASCC3*, *CCT8*, *CD46*, *CUL4A*, *DDX59*, *EDEM3*, *FAM103A1*, *HRSP12*, *MCFD2*, *MED23,* and *PPP2CB*. The quantitative relationship between the CNA gain hits among the LVI^{+} and LVI^{-} invasive BC patients showed a larger list of LVI potential key players when the CNA gain ratio was >1. These genes were: *ALG13*, *ANK3*, *ARFGEF1*, *ARHGAP18*, *ASCC3*, *C7orf11*, *CAPZA1*, *CCT8*, *CD164*, *CD46*, *CREM*, *CUL4A*, *DDX59*, *DSTN*, *EIF2AK3*, *FAM179B*, *FYTTD1*, *GPBP1L1*, *HEATR5A*, *HERPUD2*, *KLHL7*, *MCFD2*, *MED7*, *NQO1*, *OMA1*, *PAPOLA*, *PLEKHF2*, *PMS1*, *POT1*, *RANBP6*, *RAPH1*, *RG9MTD1*, *SENP7*, *SOCS4*, *SPHAR*, *SPOPL*, *TRMT6*, *U2AF1*, *WSB2*, *ZBTB11*, and *ZNF181*. The combination between the CNA statistics and the CNA gain ratio calculations produced a promising and concise list of DEGs that could be involved in the LVI mechanisms in invasive BC. The p-values, CNA gain ratios, and gene IDs for these DEGs were: (1E-8, 1.07) for *ARFGEF1*, (0.011, 1.37) for *ASCC3*, (0.038, 2.74) for *CCT8*, (0.018, 1.03) for *CD46*, (0.002, 1.60) for *CUL4A*, (0.0001, 1.11) for *DDX59*, and (0.021, 4.56) for *MCFD2*. The credible best choices were *MCFD2* and *CCT8*. (Figure 3.1). The altered expression of *CUL4A* was investigated in a separate study and the results were presented in the Joint Meeting of Pathological Society of Great Britain and Ireland and the British Division of International Academy of Pathology, Nottingham, UK.

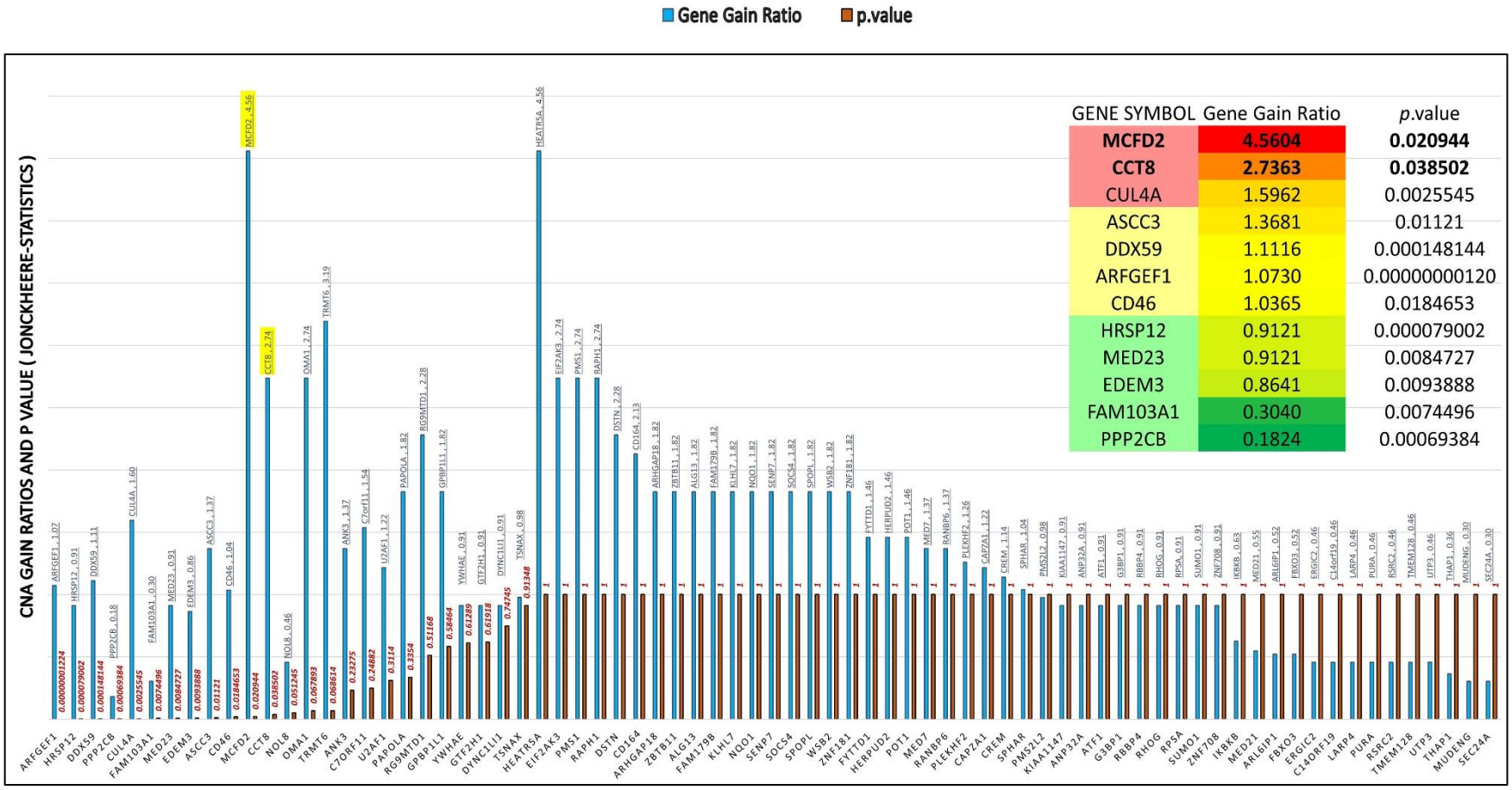


Figure 3. 2: Bar charts for differentially expressed genes of the LVI-NT100 sorted by their statistical significance

3.3.2.2. *Clinicopathological impacts for the DEGs in the LVI-NT100 cluster*

The clinicopathological relevance of the *MCFD2*- and *CCT8*-altered gene expressions was assessed in the LVI$^{+/-}$ and METABRIC cohorts against histological LVI status, age of patient, tumour size, tumour grade, LN involvement according to the NPI, and NPI groups. We also evaluated the distribution of the up- and downregulated levels of *MCFD2* and *CCT8* among the intrinsic subtypes of PAM50 and the IntClusts. Univariate statistical analysis was performed to estimate BCSS for invasive BC patients in both cohorts. Cox proportional hazards regression model and 95% CI were applied in order to assess the prognostic independence features of each transcript.

3.3.2.2.1. ***MCFD2*: Multiple coagulation factor deficiency protein-2**

The clinicopathological impact for *MCFD2* was assessed in the invasive BC patients of the LVI$^{+/-}$ and METABRIC cohorts. There were no significant relationships between the altered expression levels of *MCFD2* against all tested clinicopathological parameters in the LVI$^{+/-}$ patients. However, the highest percentage of patients with *MCFD2*$^{+}$ expression belonged to the basal intrinsic subtype, whereas patients of *MCFD2*$^{-}$ expression scored the highest percentage with the luminal-A intrinsic subtype (p-value $\leq$0.023). Similarly, the *MCFD2*$^{+}$ patients fitted into the IntClust-10 (p-value $\leq$0.013). (Table 3.1) The same statistical trend for the altered expression of *MCFD2* was found in the larger cohort of the METABRIC. The associations with the basal subtype for *MCFD2*$^{+}$ and the luminal-A subtype were confirmed by PAM50 (p-value $\leq$0.000002), and IntClusts of the METABRIC (p-value $\leq$1.7E-9) (Table 3.2)

According to the KM estimates of the BCSS for patients in the LVI$^{+/-}$ cohort, the altered expression of *MCFD2* has no correlation with BC-related death (p-value $\leq$0.070); however, patients with downregulated levels of *MCFD2* in the METABRIC cohort demonstrated a shorter BCSS (p-value $\leq$0.029). (Figure 3.2) The Cox regression model of multivariate analysis involving categorical *MCFD2* levels of expression was applied to the larger cohort of the METABRIC against clinicopathological parameters that were found to be significant in the univariate analysis of BCSS. Higher levels of *MCFD2* had an HR of 1.45 (p-value ≤0.005), and this suggests that upregulated *MCFD2* gene expression is an independent predictor of poor outcome progression (Table 3.3).

3.3.2.2.2. CCT8: T-complex protein 1 subunit theta

The invasive BC patients of the *CCT8*$^{+}$ profiles were positively associated with the high tumour grade exclusively in the LVI$^{+/-}$ cohort (Table 3.1). This relationship was absent in the METABRIC patients (Table 3.2). No other significant correlations were found. In the LVI$^{+/-}$ cohort, the altered expression of *CCT8* was not associated with a specific intrinsic subtype or IntClust. In contrast, the expression of *CCT8* in the METABRIC cohort was preferentially overexpressed in the luminal-B subtype (p-value $\leq$9.1E-13), and in the IntClust-10 and IntClust-1 (p-value $\leq$0.000004). There was no clear impact for the aberrant expression of *CCT8* on BCSS in the LVI$^{+/-}$ and the METABRIC cohorts (Figure 3.3). The altered expression of CCT8 did not show an independent predictive quality (Table 3.3).

	MCFD2 mRNA			*CCT8* mRNA		
	- n (%)	+ n (%)	χ2 (*p*-value<)	- n (%)	+ n (%)	χ2 (*p*-value<)
Lymphovascular invasion (LVI), (H&E/IHC)						
Negative/LN$^-$	42 (52.5)	38 (47.5)	0.378 (0.539)	45 (56.3)	35 (43.8)	2.361 (0.124)
Positive/LN$^{+/-}$	43 (47.8)	47 (52.2)		40 (44.4)	50 (55.6)	
Age (years)						
≤ 50	39 (59.1)	27 (40.9)	3.566 (0.059)	27 (40.9)	39 (59.1)	3.566 (0.059)
> 50	46 (44.2)	58 (55.8)		58 (55.8)	46 (44.2)	
Tumour size						
≤ 2.0cm	52 (52.0)	48 (48.0)	0.389 (0.533)	51 (51.0)	49 (49.0)	0.097 (0.755)
> 2.0cm	33 (47.1)	37 (52.9)		34 (48.6)	36 (51.4)	
Tumour Grade						
Grade-I	14 (53.8)	12 (46.2)	0.212 (0.900)	19 (73.1)	7 (26.9)	6.833 (**0.033**)
Grade-II	28 (50.9)	27 (49.1)		26 (47.3)	29 (52.7)	
Grade-III	43 (48.9)	45 (51.1)		39 (44.3)	49 (55.7)	
Nodal involvement, (NPI staging system)*						
Stage-I	57 (54.3)	48 (45.7)	2.424 (0.298)	57 (54.3)	48 (45.7)	2.424 (0.298)
Stage-II	20 (40.8)	29 (59.2)		20 (40.8)	29 (59.2)	
Stage-III	8 (50.0)	8 (50.0)		8 (50.0)	8 (50.0)	
Nottingham Prognostic Index (NPI)						
good	37 (53.6)	32 (46.4)	0.848 (0.654)	40 (58.0)	29 (42.0)	3.173 (0.205)
moderate	41 (46.6)	47 (53.4)		40 (45.5)	48 (54.5)	
poor	7 (53.8)	6 (46.2)		5 (38.5)	8 (61.5)	
Prediction analysis of microarray 50, PAM50						
Normal-like	4 (33.3)	8 (66.7)	11.380 (**0.023**)	9 (75.0)	3 (25.0)	5.677 (0.225)
Luminal-A	39 (60.9)	25 (39.1)		35 (54.7)	29 (45.3)	
Luminal-B	25 (55.6)	20 (44.4)		18 (40.0)	27 (60.0)	
Her2-enriched	7 (50.0)	7 (50.0)		6 (42.9)	8 (57.1)	
Basal	10 (28.6)	25 (71.4)		17 (48.6)	18 (51.4)	
Integrated classification clusters of METABRIC						
cluster-1	10 (52.6)	9 (47.4)	20.965 (**0.013**)	9 (47.4)	10 (52.6)	8.132 (0.521)
cluster-2	4 (66.7)	2 (33.3)		4 (66.7)	2 (33.3)	
cluster-3	14 (46.7)	16 (53.3)		17 (56.7)	13 (43.3)	
cluster-4	6 (24.0)	19 (76.0)		15 (60.0)	10 (40.0)	
cluster-5	13 (65.0)	7 (35.0)		8 (40.0)	12 (60.0)	
cluster-6	1 (33.3)	2 (66.7)		3 (100.0)	0 (0.0)	
cluster-7	12 (80.0)	3 (20.0)		7 (46.7)	8 (53.3)	
cluster-8	13 (68.4)	6 (31.6)		9 (47.4)	10 (52.6)	
cluster-9	4 (50.0)	4 (50.0)		4 (50.0)	4 (50.0)	
cluster-10	8 (32.0)	17 (68.0)		9 (36.0)	16 (64.0)	

Table 3. 1: Clinicopathological relevance for MCFD2 and CCT8 in the LVI-supervised patients, n=170. Significant values are in bold

	MCFD2 mRNA			*CCT8* mRNA		
	-	+	$\chi 2$	-	+	$\chi 2$
	n (%)	n (%)	(*p*-value<)	n (%)	n (%)	(*p*-value<)
Lymphovascular invasion (LVI), (H&E/IHC)						
Negative/LN$^{+/-}$	55 (33.5)	109 (66.5)	1.087 (0.297)	33 (20.1)	131 (79.9)	0.577 (0.447)
Positive/LN$^{+/-}$	52 (39.4)	80 (60.6)		22 (16.7)	110 (83.3)	
Age (years)						
≤ 50	230 (53.6)	199 (46.4)	2.860 (0.091)	201 (46.9)	228 (53.1)	0.954 (0.329)
> 50	760 (49.0)	791 (51.0)		768 (49.5)	783 (50.5)	
Tumour size						
≤ 2.0cm	417 (48.6)	441 (51.4)	0.947 (0.330)	428 (49.9)	430 (50.1)	0.499 (0.480)
> 2.0cm	560 (50.8)	542 (49.2)		532 (48.3)	570 (51.7)	
Tumour Grade						
Grade-I	77 (45.0)	94 (55.0)	3.436 (0.179)	79 (46.2)	92 (53.8)	3.063 (0.216)
Grade-II	403 (52.3)	367 (47.7)		392 (50.9)	378 (49.1)	
Grade-III	471 (49.5)	481 (50.5)		447 (47.0)	505 (53.0)	
Nodal involvement, (NPI staging system)*						
Stage-I	509 (49.2)	526 (50.8)	3.284 (0.194)	505 (48.8)	530 (51.2)	0.148 (0.929)
Stage-II	306 (49.2)	316 (50.8)		304 (48.9)	318 (51.1)	
Stage-III	173 (54.7)	143 (45.3)		158 (50.0)	158 (50.0)	
Nottingham Prognostic Index (NPI)						
good	342 (50.3)	338 (49.7)	1.135 (0.567)	342 (50.3)	338 (49.7)	0.783 (0.676)
moderate	542 (49.2)	559 (50.8)		532 (48.3)	569 (51.7)	
poor	106 (53.3)	93 (46.7)		95 (48.9)	104 (52.3)	
Prediction analysis of microarray 50, PAM50						
Normal-like	97 (47.3)	108 (52.7)	32.436 (**0.000002**)	139 (67.8)	66 (32.2)	62.402 (**9.1E-13**)
Luminal-A	378 (52.6)	340 (47.4)		377 (52.5)	341 (47.5)	
Luminal-B	252 (51.6)	236 (48.4)		179 (36.7)	309 (63.3)	
Her2-enriched	141 (58.8)	99 (41.3)		117 (48.8)	123 (51.3)	
Basal	122 (37.1)	207 (62.9)		157 (47.7)	172 (52.3)	
Integrated classification clusters of METABRIC						
cluster-1	70 (50.4)	69 (49.6)	59.475 (**1.7E-9**)	56 (40.3)	83 (59.7)	41.498 (**0.000004**)
cluster -2	32 (44.4)	40 (55.6)		34 (47.2)	38 (52.8)	
cluster -3	144 (49.7)	146 (50.3)		142 (49.0)	148 (51.0)	
cluster -4	140 (40.8)	203 (59.2)		211 (61.5)	132 (38.5)	
cluster -5	114 (60.0)	76 (40.0)		84 (44.2)	106 (55.8)	
cluster -6	51 (60.0)	34 (40.0)		43 (50.6)	42 (49.4)	
cluster -7	100 (52.6)	90 (47.4)		104 (54.7)	86 (45.3)	
cluster -8	172 (57.5)	127 (42.5)		132 (44.1)	167 (55.9)	
cluster -9	89 (61.0)	57 (39.0)		74 (50.7)	72 (49.3)	
cluster -10	78 (34.5)	148 (65.5)		89 (39.4)	137 (60.6)	

Table 3. 2: Clinicopathological relevance for MCFD2 and CCT8 in the METABRIC patients, n=1980. Significant values are in bold

Breast cancer specific survival (BCSS), n= 1869[0]

LVI-NT100 candidates	Hazard ratio (95% CI)	Lower	Upper	*p*-value
MCFD2 mRNA	1.450	1.118	1.882	**0.005**
Tumour grade	1.359	1.047	1.765	0.021
Nodal staging*	1.797	1.509	2.139	0.000
Tumour size	1.522	1.149	2.017	0.003
Estrogen expression**	0.679	0.503	0.917	0.012
Her2 expression***	1.231	1.041	1.454	0.015
***CCT8* mRNA**	0.848	0.656	1.097	0.209
Tumour grade	1.372	1.057	1.780	0.017
Nodal staging*	1.789	1.503	2.129	0.000
Tumour size	1.519	1.146	2.013	0.004
Estrogen expression**	0.714	0.530	0.962	0.027
Her2 expression***	1.245	1.054	1.472	0.010

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
[0] Sum of patients does not equal to 1980 because of missing data among different categories.

Table 3. 3: Multivariate Cox regression analysis for prediction of breast cancer specific survival.

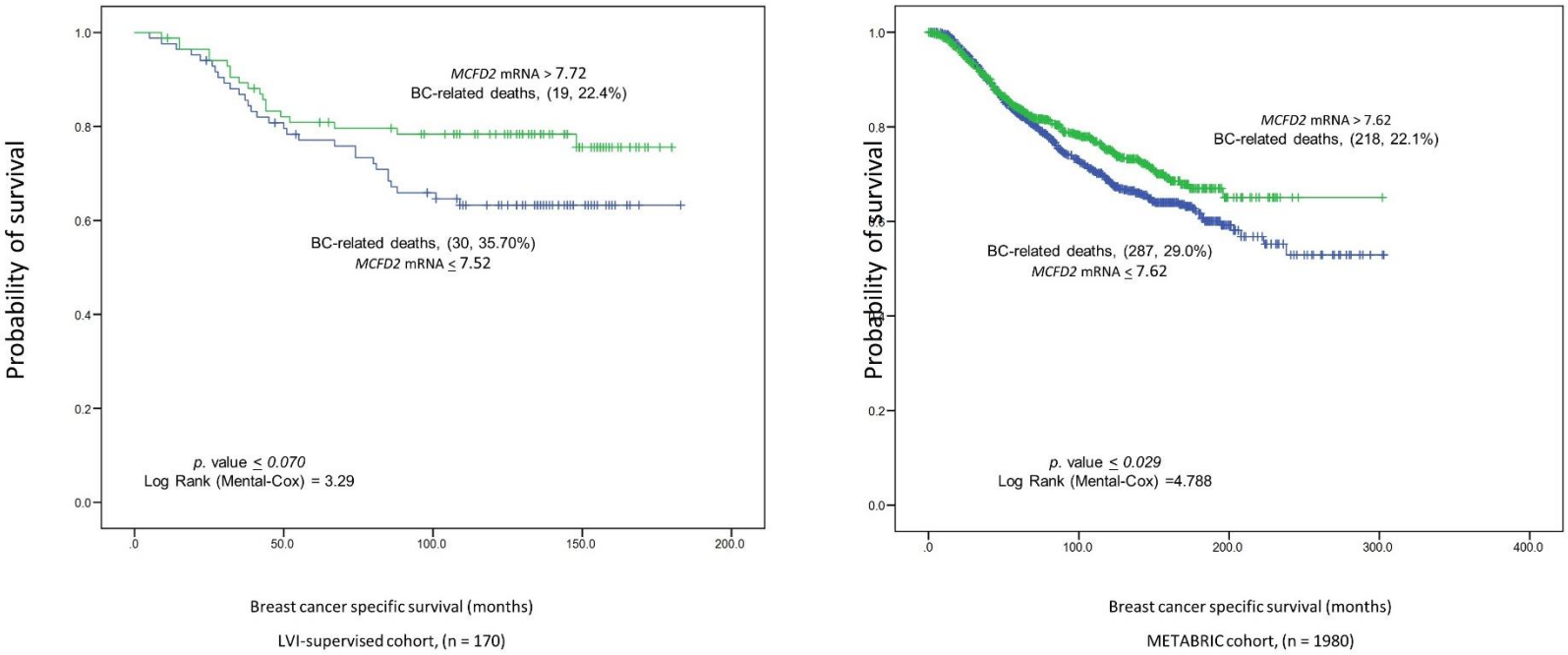


Figure 3. 3: Associations between the altered expression of MCFD2 mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

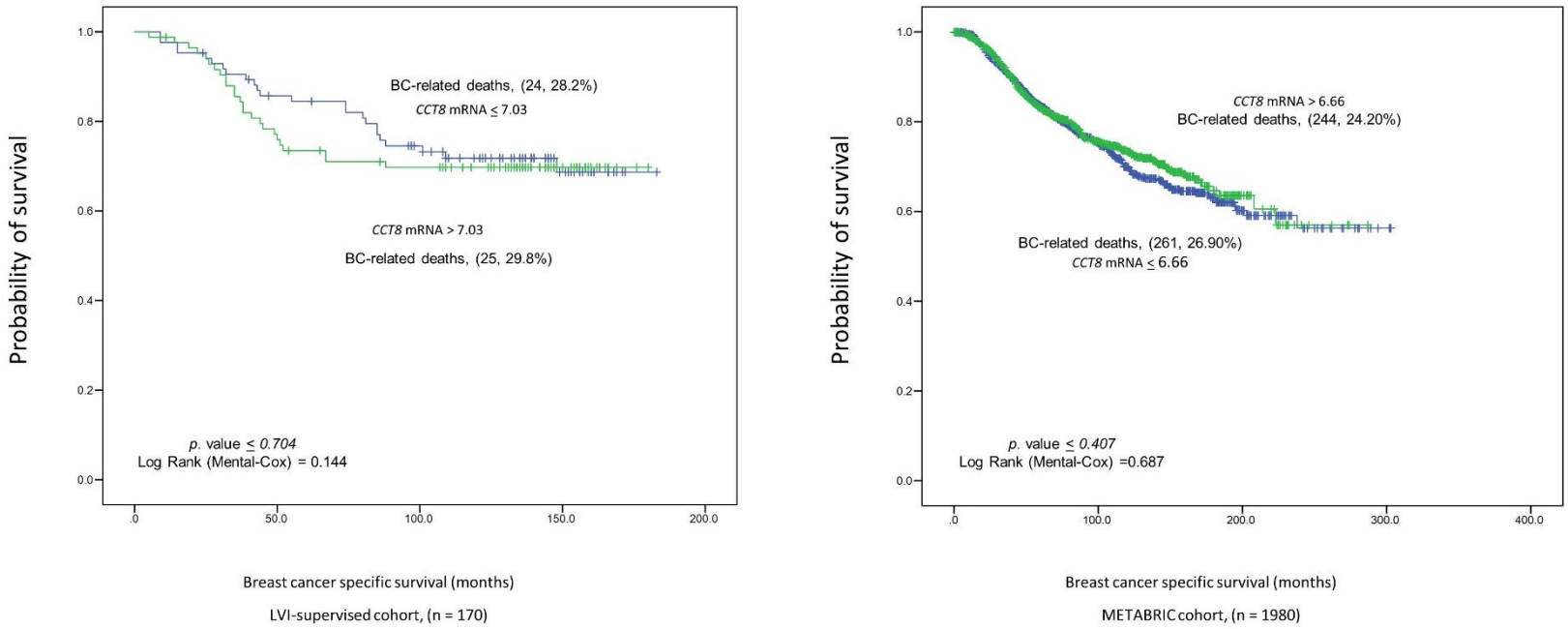


Figure 3. 4: Associations between the altered expression of CCT8 mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

3.3.3. **DGE profile of the LVI-NT283**

After obtaining full access to the METABRIC data, we attempted to enhance our search for LVI candidate genes. The transcriptomic data for each DEG in the LVI-NT283 cluster was assessed against the histologically defined LVI status. The tested cohorts included the $LVI^{+/-}$ cohort of patients and a subset of METABRIC patients with $LVI^{-}/LN^{+/-}$. The DEGs in this refined cluster were selected based on their significant associations with LVI, positive log2FC and AvE, and perfect base sequence pairing.

3.3.3.1. *Transcriptomic/LVI analysis of the LVI-NT283 cluster of DEGs*

The comprehensive statistical analysis of the LVI-NT283 cluster of DEGs revealed statistical relevance for 35 LVI-related transcripts with LVI status. The differential mean rank scores, u-test, against the histologically defined LVI categories for each identified transcript of the 35 split them into two distinctive classes of LVI candidates. The first class was the upregulated transcripts of the LVI^{+} invasive BC patients. These transcripts were denoted as (UP-13/LVI^{+}), and they were: **1)** *PPM1G* (p-value $\leq$0.0001), **2)** *SEC14L1* (p-value $\leq$0.005), **3)** *P4HA2* (p-value $\leq$0.007), **4)** *CTSD* (p-value $\leq$0.009), **5)** *ITPRIP* (p-value $\leq$0.016), **6)** *PILRA* (p-value $\leq$0.031), **7)** *ZDHHC9* (p-value $\leq$0.031), **8)** *PDLIM7* (p-value $\leq$0.032), **9)** *SLC3A2* (p-value $\leq$0.032), **10)** *RNASE1* (p-value $\leq$0.032), **11)** *IKZF3* (p-value $\leq$0.036), **12)** *SIGLEC1* (p-value $\leq$0.039), and **13)** *SLC6A9* (p-value $\leq$0.047) (Table 3.4).

The second class of transcripts that completed the identified transcript of the 35 LVI-related transcripts was designated as (UP-22/LVI^{-}), and they were: **1)** *LPAR1* (p-value $\leq$0.0002), **2)** *CXCL12* (p-value $\leq$0.001), **3)** *C17orf81* (p-value $\leq$0.002), **4)** *ECM2* (p-value $\leq$0.004), **5)** *FBLN1* (p-value $\leq$0.005), **6)** *APCDD1* (p-value $\leq$0.005), **7)** *ANXA5* (p-value $\leq$0.008), **8)** *INKA1* (p-value $\leq$0.011), **9)** *TMEM98* (p-value $\leq$0.012), **10)** *FEZ1* (p-value $\leq$0.016), **11)** *THAP4* (p-value $\leq$0.016), **12)** *TNFRSF19* (p-value $\leq$0.016), **13)** *NBR1* (p-value $\leq$0.018), **14)** *WISP2* (p-value $\leq$0.019), **15)** *CNRIP1* (p-value $\leq$0.020), **16)** *ITSN1* (p-value $\leq$0.021), **17)** *C1S* (p-value $\leq$0.022), **18)** *TNC* (p-value $\leq$0.028), **19)** *PRRX1* (p-value $\leq$0.030), **20)** *STAT3* (p-value $\leq$0.035), **21)** *TSC1* (p-value $\leq$0.036), and **22)** *TWIST2* (p-value $\leq$0.047) (Table 3.5). Detailed protein information for the potential LVI-associated transcripts are listed from page 393 to page 413, Appendix B.

Gene Identifier	LVI status	N	Mean Rank	Sum of Ranks	*p*. value
CTSD	Negative	80	75.05	6004.00	***0.009***
	Definite	90	94.79	8531.00	
IKZF3	Negative	80	77.09	6167.00	***0.036***
	Definite	90	92.98	8368.00	
ITPRIP	Negative	80	75.83	6066.00	***0.016***
	Definite	90	94.10	8469.00	
P4HA2	Negative	80	74.78	5982.00	***0.007***
	Definite	90	95.03	8553.00	
PDLIM7	Negative	80	76.90	6152.00	***0.032***
	Definite	90	93.14	8383.00	
PILRA	Negative	80	76.85	6148.00	***0.031***
	Definite	90	93.19	8387.00	
PPM1G	Negative	80	70.23	5618.00	***0.0001***
	Definite	90	99.08	8917.00	
RNASE1	Negative	80	76.94	6155.00	***0.032***
	Definite	90	93.11	8380.00	
SEC14L1	Negative	80	74.38	5950.00	***0.005***
	Definite	90	95.39	8585.00	
SIGLEC1	Negative	80	77.25	6180.00	***0.039***
	Definite	90	92.83	8355.00	
SLC3A2	Negative	80	76.93	6154.00	***0.032***
	Definite	90	93.12	8381.00	
SLC6A9	Negative	80	77.56	6204.50	***0.047***
	Definite	90	92.56	8330.50	
ZDHHC9	Negative	80	76.86	6149.00	***0.031***
	Definite	90	93.18	8386.00	

Table 3. 4: Differential mean rank scores for upregulated mRNAs in the LVI-definite patients

Gene Identifier	LVI status	N	Mean Rank	Sum of Ranks	*p*. value
ANXA5	Negative	80	96.07	7685.50	**0.008**
	Definite	90	76.11	6849.50	
APCDD1	Negative	80	96.76	7741.00	**0.005**
	Definite	90	75.49	6794.00	
C17orf81	Negative	80	98.14	7851.00	**0.002**
	Definite	90	74.27	6684.00	
C1S	Negative	80	94.66	7573.00	***0.022***
	Definite	90	77.36	6962.00	
CNRIP1	Negative	80	94.83	7586.00	**0.02**
	Definite	90	77.21	6949.00	
CXCL12	Negative	80	98.55	7884.00	**0.001**
	Definite	90	73.90	6651.00	
ECM2	Negative	80	97.11	7769.00	**0.004**
	Definite	90	75.18	6766.00	
FBLN1	Negative	80	96.87	7749.50	**0.005**
	Definite	90	75.39	6785.50	
FEZ1	Negative	80	95.16	7613.00	**0.016**
	Definite	90	76.91	6922.00	
INKA1	Negative	80	95.74	7659.00	**0.011**
	Definite	90	76.40	6876.00	
ITSN1	Negative	80	94.74	7579.00	**0.021**
	Definite	90	77.29	6956.00	
LPAR1	Negative	80	100.64	8051.00	**0.0002**
	Definite	90	72.04	6484.00	
NBR1	Negative	80	95.01	7601.00	**0.018**
	Definite	90	77.04	6934.00	
PRRX1	Negative	80	94.20	7536.00	**0.03**
	Definite	90	77.77	6999.00	
STAT3	Negative	80	93.93	7514.00	**0.035**
	Definite	90	78.01	7021.00	
THAP4	Negative	80	95.11	7609.00	**0.016**
	Definite	90	76.96	6926.00	
TMEM98	Negative	80	95.51	7641.00	**0.012**
	Definite	90	76.60	6894.00	
TNC	Negative	80	94.33	7546.00	**0.028**
	Definite	90	77.66	6989.00	
TNFRSF19	Negative	80	95.11	7609.00	**0.016**
	Definite	90	76.96	6926.00	
TSC1	Negative	80	93.91	7513.00	**0.036**
	Definite	90	78.02	7022.00	
TWIST2	Negative	80	93.44	7475.00	**0.047**
	Definite	90	78.44	7060.00	
WISP2	Negative	80	94.93	7594.00	**0.019**
	Definite	90	77.12	6941.00	

Table 3. 5: Differential mean rank scores for upregulated mRNAs in the LVI-absent patients

3.3.3.2. *Clinicopathological impact for transcripts in the UP-13/LVI⁺ cluster*

For all transcripts in the UP-13/LVI^{+} cluster, detailed clinicopathological and molecular classification results are illustrated in (Table 3.6) for the LVI$^{+/-}$ patients and in (Table 3.7) for the METABRIC patients. Details on multivariate analysis are illustrated in (Table 3.8) for UP-13/LVI^{+} transcripts. KM plots are depicted for each

transcript individually and in alphabetic order, (Figure 3.5) to (Figure 3.17), for easier tracking. All tables and figures are placed in the end of this section.

3.3.3.2.1. PPM1G: Protein phosphatase 1G

Remarkably, the altered expression of *PPM1G* showed strong significant correlations against all of the tested clinicopathological factors in the $LVI^{+/-}$ cohort. The invasive BC patients with $PPM1G^{+}$ expression profiles were positively associated with definite LVI (p-value ≤0.002), younger patients (p-value ≤0.012), large tumour size (p-value ≤0.002), high tumour grade (p-value ≤0.000007), positive nodal involvement (p-value ≤0.011), and poor NPI (p-value ≤0.0008). The association between *PPM1G* aberrant expression and LVI was diminished upon the inclusion of the $LVI^{-}/LN^{+/-}$ subset of the METABRIC patients (p-value ≤0.131). The rest of the statistical significance for the altered expression of *PPM1G* was maintained in the METABRIC patients. The METABRIC patients with $PPM1G^{+}$ expression profiles were positively associated with younger patients (p-value ≤3.5E-10), large tumour size (p-value ≤0.008), high tumour grade (p-value ≤6.7E-42), positive nodal involvement (p-value ≤0.0001), and poor NPI (p-value ≤2.8E-25).).

The distribution of invasive BC patients in the $LVI^{+/-}$ cohort with $PPM1G^{+}$ expression profiles among the different intrinsic subtypes of invasive BC showed that the overexpressed *PPM1G* was associated with the patients of the HER2-enriched and basal subtypes (p-value ≤4.2E-8). The exact results were obtained from the IntClust of the METABRIC. The same statistical trend for the $PPM1G^{+}$ expression profiles was identified in the subclass of invasive BC patients fitted into IntClust-5, a cluster that is molecularly defined with ERBB2 amplification, and IntClust-10 (p-value ≤0.000003). In the METABRIC cohort, the overexpression of *PPM1G* also showed strong relationships with the HER2-enriched and basal subtypes (p-value ≤7.5E-80), and IntClust-5 and IntClust-10 (p-value ≤3.0E-65)

The KM estimates of the BCSS for patients showed that the overexpression of *PPM1G* was related to more incidences of death caused by BC in the $LVI^{+/-}$ (p-value ≤0.003) and in the METABRIC (p-value ≤6.7E-12) cohorts. The overexpression of *PPM1G* in invasive BC patients had a hazard ratio of 1.677 (p-value ≤0.001), signifying that the upregulated *PPM1G* could be an independent predictor of BC progression.

3.3.3.2.2. SEC14L1: SEC14-like protein 1

The altered expression of *SEC14L1* was positively associated with definite LVI occurrences in the $LVI^{+/-}$ cohort of invasive BC patients in this study (p-value

≤0.031). In addition, the *SEC14L1*$^{+}$ expression profile was correlated with high tumour grade (p-value ≤0.000003), positive nodal involvement (p-value ≤0.010), and poor NPI (p-value ≤0.000008). For the aberrant expression of *PPM1G*, the relationship between *SEC14L1* and LVI was lost completely in the LVI^{-}/LN$^{+/-}$ subset of the METABRIC patients (p-value ≤0.403). The positive expression of *SEC14L1* in the METABRIC cohort was associated with younger patients (p-value ≤0.008), high tumour grade (p-value ≤1.2E-27), positive nodal involvement (p-value ≤1.9E-9), and poor NPI (p-value ≤1.4E-26).

The distribution of invasive BC patients in the LVI$^{+/-}$ cohort with *SEC14L1*$^{+}$ expression profiles among the different intrinsic subtypes of invasive BC showed that the overexpressed *SEC14L1* was associated with the patients of the HER2-enriched and basal subtypes (p-value ≤8.3E-8). The classification of invasive BC according to the IntClusts of the METABRIC allocated the *SEC14L1*$^{+}$ patients into IntClust-9 and IntClust-10 (p-value ≤3.8E-7). In the larger cohort of the METABRIC, the majority of *SEC14L1*$^{+}$ patients fitted into the HER2-enriched and basal subtypes (p-value ≤7.5E-45), and IntClust-10 (p-value ≤1.4E-55).

The KM estimates of the BCSS for patients showed that the overexpression of *SEC14L1* was related to worse survival rates in the LVI$^{+/-}$ (p-value ≤0.00007), and in the METABRIC (p-value ≤0.0004) cohorts. The expression of *SEC14L1* in invasive BC patients had a hazard ratio of 1.096 (p-value ≤0.513), suggesting that the upregulated *SEC14L1* could be a dependent predictor of BC progression.

3.3.3.2.3. P4HA2: Prolyl 4-hydroxylase subunit alpha-2

In the LVI$^{+/-}$ cohort, the positive expression of *P4HA2* showed significant and positive correlations only against definite LVI (p-value ≤0.006), and the early stage for nodal involvement (p-value ≤0.022). The inclusion of the LVI^{-}/LN$^{+/-}$ subset of the METABRIC patients had no impact on the relationship between LVI and *P4HA2* as it remained significant (p-value ≤0.009). However, the association with LNM was totally absent (p-value ≤0.243). The positive expression of *P4HA2* was associated with larger tumour size (p-value ≤0.012), and the percentage of patients with *P4HA2*$^{+}$ expression profile in the METABRIC cohort was distributed equally among the intermediate and poor NPI categories (p-value ≤0.032).

There was no specific distribution for the invasive BC patients with altered *P4HA2* expression profiles in the LVI$^{+/-}$ cohort among the different intrinsic subtypes or the molecular clusters of the METABRIC. Alternatively, a preferential upregulation for *P4HA2* was calculated for the HER2-enriched subtype of invasive

BC patients (p-value ≤0.015). The majority of patients with *P4HA2*$^{+}$ expression profiles were fitted into IntClust-5 and IntClust-6 (p-value ≤2.3E-9).

The overexpression of *P4HA2* was strongly associated with the poor survival rates in the patients of LVI$^{+/-}$ (p-value ≤0.0004), and METABRIC (p-value ≤0.0003). The overexpression of *P4HA2* in invasive BC patients had a hazard ratio of 1.390 (p-value ≤0.016) suggesting that the upregulated expression of *P4HA2* could be an independent predictor of BC progression.

3.3.3.2.4. CTSD: Cathepsin D

In the LVI$^{+/-}$ cohort, the altered expression of *CTSD* was associated exclusively and positively with LVI categories (p-value ≤0.014). A trend toward statistical significance for the positive expression of *CTSD* was recorded against the positive nodal involvement (p-value ≤0.060). In the METABRIC cohort, there were no associations between *CTSD* and LVI (p-value ≤0.082). Also, there was no relationship between a patient's age and the altered expression of *CTSD*. Instead, the majority of patients in the tumour grade-I category had a *CTSD*$^{-}$ expression profile (p-value ≤0.012). The low expression of *CTSD* was associated with smaller tumour size (p-value ≤0.002). The positive expression profiles for *CTSD* in the invasive BC patients of the METABRIC were associated with positive nodal involvement (p-value ≤0.00002) and the worse NPI categories (p-value ≤0.002).

The altered expression of *CTSD* in the invasive BC patients of the LVI$^{+/-}$ cohort had no specific trend of distribution among the different intrinsic subtypes or molecular classes of the METABRIC. On the other hand, the majority of patients of the HER2-enriched subtype in the METABRIC cohort showed an overexpression of *CTSD* (p-value ≤0.00003). A larger group of patients in IntClust-5 positively expressed *CTSD* in comparison to the other IntClusts (p-value ≤5.5E-7).

The KM estimates of the BCSS for the LVI$^{+/-}$ and METABRIC patients showed no impacts for the aberrant expression of *CTSD* on the survival of invasive BC patients (p-value ≤0.750 and p-value ≤0.250, respectively). The altered expression of *CTSD* in invasive BC patients had a hazard ratio of 0.835 (p-value ≤0.180), and accordingly suggested that the expression of *CTSD* had no independent prediction qualities of BC progression.

3.3.3.2.5. ITPRIP: Inositol 1,4,5-trisphosphate receptor-interacting protein

In the LVI$^{+/-}$ cohort, the altered expression of *ITPRIP* showed a trend toward significance against LVI status (p-value ≤0.065). The majority of grade-II patients were labelled with *ITPRIP*$^{-}$ expression profiles, whereas most of the grade-III patients showed elevated expression levels of *ITPRIP* (p-value ≤0.039). The

positive expression of *ITPRIP* was associated with poorer NPI (p-value ≤0.029). In the METABRIC cohort, younger BC patients overexpressed *ITPRIP* (p-value ≤0.001). The positive expression of *ITPRIP* was detected only in the tumour grade-III patients (p-value ≤1.2E-14). The positive nodal status and the worse NPI indices were also associated with the patients of *ITPRIP*$^{+}$ expression profiles (p-value ≤0.006) and (p-value ≤9.3E-11), accordingly.

The intrinsic subtyping and molecular classification analyses revealed that the majority of patients with *ITPRIP*$^{+}$ expression profiles fitted into the basal subtype (p-value ≤0.00009) and IntClust-10 (p-value ≤0.003) in the LVI$^{+/-}$ patients. Comparably, the positive expression of *ITPRIP* was associated with the patients of the basal subtype (p-value ≤6.9E-79) and IntClust-10 (p-value ≤9.0E-52), respectively.

The KM estimates of the BCSS for the LVI$^{+/-}$ showed no associations between the altered expression of *ITPRIP* and patients' survival (p-value ≤0.169). On the other hand, the upregulated ITPRIP showed a negative impact on the BCSS of the invasive BC patients (p-value ≤0.002). The altered expression of *ITPRIP* in invasive BC patients had a hazard ratio of 1.150 (p-value ≤0.340). The latter finding may suggest that the expression of *ITPRIP* had no independent prediction qualities of BC progression.

3.3.3.2.6. PILRA: Paired immunoglobulin-like type 2 receptor alpha

For the altered expression of *ITPRIP*, the altered expression of *PIRLA* showed a trend toward significance against LVI status (p-value ≤0.065) in the LVI$^{+/-}$ cohort. The positive expression of *PILRA* was positively associated with the majority of invasive BC patients who were labelled with tumour grade-III (p-value ≤0.000007), positive nodal involvement (p-value ≤0.011), and poorer NPI (p-value ≤0.00008). In the METABRIC cohort, the patients with a *PILRA*$^{+}$ expression profile were accordingly associated with tumour grade-III (p-value ≤2.3E-13), positive nodal involvement (p-value ≤0.002), and poorer NPI (p-value ≤5.6E-11).

The overexpressed *PLIRA* was statistically correlated with the intrinsic subtypes of HER2-enriched and basal (p-value ≤0.000005). The molecular integrative clustering analysis showed that the patients of *PILRA*$^{+}$ expression profiles were majorly associated with IntClust-10 (p-value ≤0.001). In the METABRIC cohort, the *PILRA*$^{+}$ patients fitted into the basal and HER2-enriched subtypes (p-value ≤3.2E-25) and into IntClust-10 (p-value ≤8.0E-23).

The KM estimates of the BCSS for patients showed that the overexpression of *PILRA* was related to worse patient survival rates in the LVI$^{+/-}$ (p-value ≤0.015)

and the METABRIC (p-value ≤0.0003). The expression of *PILRA* in invasive BC patients had a hazard ratio of 1.242 (p-value ≤0.118), suggesting that the upregulated *PILRA* could be a dependent predictor of BC progression.

3.3.3.2.7. ZDHHC9: Palmitoyltransferase ZDHHC9

In the LVI$^{+/-}$ cohort, the altered expression of *ZDHHC9* showed no relevant results against all of the tested clinicopathological parameters. In the METABRIC cohort, the upregulation of *ZDHHC9* was associated with large tumour size (p-value ≤0.020), high tumour grade (p-value ≤2.8E-11), and poorer NPI (p-value ≤0.0004).

In the LVI+/- cohort, the upregulated *ZDHHC9* was associated mainly with the HER2-enriched subtype (p-value ≤0.001) and the molecular clusters of IntClust-2, IntClust-5, IntClust-10, and IntClust-9 (p-value ≤4.0E-31). In the METABRIC cohort, the overexpression of *ZDHHC9* was statistically correlated with the intrinsic subtypes of HER2-enriched and basal (p-value ≤2.4E-18) and the molecular classes of IntClust-10, IntClust-9, IntClust-5, and IntClust-6 (p-value ≤4.0E-31).

The KM estimates of the BCSS for patients showed that the overexpression of *ZDHHC9* had a significant impact on shortening the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.040) and the METABRIC (p-value ≤0.013) cohorts. The expression of *ZDHHC9* in invasive BC patients had a hazard ratio of 0.998 (p-value ≤0.990), suggesting that the altered expression of *ZDHHC9* could be a weak predictor of BC progression.

3.3.3.2.8. PDLIM7: PDZ and LIM domain protein 7

The altered expression of *PDLIM7* was positively associated with definite LVI occurrences in the LVI$^{+/-}$ cohort (p-value ≤0.006). In addition, the patients of the *PDLIM7*$^{+}$ expression profiles were significantly associated with large tumour size (p-value ≤0.013), high tumour grade (p-value ≤0.001), and poorer NPI (p-value ≤0.0003). In the METABRIC cohort, the overexpression of *PDLIM7* was associated with tumour grade (p-value ≤0.016), positive nodal involvement (p-value ≤0.042), and worse NPI (p-value ≤0.001).

The overexpressed *PDLIM7* was statistically correlated with the intrinsic subtypes of HER2-enriched (p-value ≤0.005). The molecular integrative clustering analysis showed no specific distribution for the aberrantly expressed *PDLIM7* within any IntClust (p-value ≤0.080). In the METABRIC cohort, the *PDLIM7*$^{+}$ patients fitted into the HER2-enriched subtypes (p-value ≤1.1E-8) and IntClust-5 (p-value ≤2.1E-7).

The KM estimates of the BCSS for patients showed that the overexpression of *PDLIM7* had a significant impact on shortening the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value $\leq$0.005) and the METABRIC (p-value $\leq$0.00005) cohorts. The overexpression of *PDLIM7* in invasive BC patients had a hazard ratio of 1.290 (p-value $\leq$0.062), suggesting that the altered expression of *PDLIM7* could be a dependent predictor of BC progression.

3.3.3.2.9. SLC3A2: 4F2 cell-surface antigen heavy chain

In the LVI$^{+/-}$ cohort, the elevated expression levels of *SLC3A2* did show a relevant correlation with LVI occurrence (p-value $\leq$0.014), high tumour grade (p-value $\leq$0.001), and worse NPI (p-value $\leq$0.002). In the METABRIC cohort, the downregulation of *SLC3A2* was associated with LVI absence (p-value $\leq$0.048), small tumour size (p-value $\leq$0.002), low tumour grade (p-value $\leq$5.8E-7), negative nodal involvement (p-value $\leq$0.00005), and poorer NPI (p-value $\leq$8.2E-10).

The overexpressed *SLC3A2* was statistically correlated with the intrinsic subtypes of HER2-enriched, and the molecular clusters of IntClust-5 and IntClust-9 in the LVI$^{+/-}$ cohort of patients (p-value $\leq$0.002 and $\leq$0.013, respectively). In the METABRIC cohort, the expression of *SLC3A2* was positively correlated with more patients of the basal subtype (p-value $\leq$1.6E-12). The analysis of molecular clustering revealed that the patients of the *SLC3A2*$^{+}$ expression profiles fitted primarily into the clusters of IntClust-10 and IntClust-5 (p-value $\leq$3.5E-14).

The KM estimates of the BCSS for patients showed that the overexpression of *SLC3A2* had a significant impact on the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value $\leq$0.015). For the METABRIC patients, there was no association between the altered expression of *SLC3A2* and BCSS (p-value $\leq$0.358). The overexpression of *SLC3A2* in invasive BC patients had a hazard ratio of 0.733 (p-value $\leq$0.027), suggesting that the altered expression of *SLC3A2* could be an independent predictor of BC progression.

3.3.3.2.10. RNASE1: Ribonuclease pancreatic

In the LVI$^{+/-}$ cohort, the altered expression of *RNASE1* showed a trend toward significance with LVI status (p-value $\leq$0.065). The patients of early-stage nodal metastasis, stage-II, were associated with elevated levels of *RNASE1* (p-value $\leq$0.008). The positive expression of *RNASE1* was also associated with worse NPI (p-value $\leq$0.028). For the METABRIC patients, the altered expression of *RNASE1* was correlated with tumour grade (p-value $\leq$0.016), nodal involvement (p-value $\leq$0.009), and NPI (p-value $\leq$0.0001).

The analysis of the intrinsic subtyping of invasive BC showed that the basal subtype of invasive BC patients in the LVI+/- cohort was overly expressed *RNASE1* (p-value ≤0.018). No associations were identified for *RNASE1* against the molecular clusters of METABRIC. In the METABRIC cohort, the upregulation of *RNASE1* was associated with patients fitted into the normal-like and basal intrinsic subtypes (p-value ≤8.4E-12). The molecular clustering analysis showed that the overexpression of *RNASE1* was associated with the patients of IntClust-4 and IntClust-10 (p-value ≤0.027).

The KM estimates of the BCSS for patients showed that the overexpression of *RNASE1* had a significant impact on the BCSS of the invasive BC patients in the $LVI^{+/-}$ (p-value ≤0.005). For the METABRIC patients, there was no association between the altered expression of *RNASE1* and BCSS (p-value ≤0.143). The overexpression of *RNASE1* in invasive BC patients had a hazard ratio of 1.036 (p-value ≤0.793), suggesting that the altered expression of *RNASE1* could be a dependent predictor of BC progression.

3.3.3.2.11. IKZF3: Zinc finger protein Aiolos

In the $LVI^{+/-}$ cohort, the altered expression of *IKZF3* showed a trend toward significance with LVI status (p-value ≤0.065). The overly expressed *IKZF3* was related to larger tumour size (p-value ≤0.013), positive nodal involvement (p-value ≤0.047), and worse NPI (p-value ≤0.012). In the METABRIC cohort, the low expression of *IKZF3* was associated with LVI absence (p-value ≤0.023), tumour grade-I (p-value ≤5.9E-12), nodal involvement (p-value ≤0.0003), and NPI (p-value ≤1.2E-9).

The analysis of the intrinsic subtyping of invasive BC showed that the overexpression of *IKZF3* was correlated with more patients fitted into the aggressive intrinsic subtypes of HER2-enriched, basal, and luminal-B in the LVI+/- cohort (p-value ≤0.020). No relationships were identified between the altered expression of *IKZF3* and the IntClusts. In the METABRIC cohort, the overexpression of *IKZF3* was associated majorly with the HER2-enriched and basal subtypes (p-value ≤6.3E-29) and IntClust-5 (p-value ≤3.9E-28).

The KM estimates of the BCSS for patients showed that the overexpression of *IKZF3* had no relevant impact on the BCSS of the invasive BC patients in the $LVI^{+/-}$ (p-value ≤0.080). For the METABRIC patients, there was a significant association between the altered expression of *IKZF3* and BCSS (p-value ≤0.021). The aberrant expression of *IKZF3* in invasive BC patients had a hazard ratio of

0.945 (p-value ≤0.680), suggesting that the altered expression of *IKZF3* had no qualities as an independent predictor of BC progression.

3.3.3.2.12. SIGLEC1: Sialoadhesin

In the LVI$^{+/-}$ cohort, the altered expression of *SIGLEC1* showed no significance against LVI status, patients' age, and tumour size. The positive expression of *SIGLEC1* showed positive relationships with tumour grade-III (p-value ≤0.003), stage-II (p-value ≤0.005), and poorer NPI (p-value ≤0.011). In the METABRIC cohort, the overly expressed *SIGLEC1* was accordingly correlated with younger patients (p-value ≤0.033), tumour grade-III (p-value ≤5.3E-11), positive nodal involvement (p-value ≤0.0003), and poorer NPI (p-value ≤3.6E-9).

The analysis of the intrinsic subtyping of invasive BC showed that the overexpression of *SIGLEC1* was correlated with more patients fitted into the intrinsic subtypes of luminal-B and basal in the LVI+/- cohort (p-value ≤0.0005). The upregulation of *SIGLEC1* was associated with IntClust-10 (p-value ≤0.004). In the METABRIC cohort, the overexpression of *SIGLEC1* was associated majorly with the luminal-B and basal subtypes (p-value ≤8.5E-18) and IntClust-10, IntClust-1, and IntClust-2 (p-value ≤4.2E-21).

The KM estimates of the BCSS for patients showed that the overexpression of *SIGLEC1* had a relevant impact on the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.014) and the METABRIC patients (p-value ≤0.000002). The overexpression of *SIGLEC1* in invasive BC patients had a hazard ratio of 1.286 (p-value ≤0.066), suggesting that the altered expression of *SIGLEC1* could be an independent predictor of BC progression.

3.3.3.2.13. SLC6A9: Sodium- and chloride-dependent glycine transporter 1

In the LVI$^{+/-}$ cohort, the elevated expression levels of *SLC6A9* showed a significant correlation with LVI (p-value ≤0.014), high tumour grade (p-value ≤0.010), positive nodal involvement (p-value ≤0.028), and worse NPI (p-value ≤0.005). In the METABRIC cohort, the altered expression of *SLC6A9* was not associated with LVI (p-value ≤0.112). The positive expression of *SLC6A9* was significantly associated with larger tumour size (p-value ≤0.00008), high tumour grade (p-value ≤5.8E-9), positive nodal involvement (p-value ≤0.027), and poorer NPI (p-value ≤0.000001).

The analysis of the intrinsic subtyping of invasive BC showed that the overexpression of *SLC6A9* was correlated with more patients fitted into the HER2-enriched in the LVI$^{+/-}$ cohort (p-value ≤0.000021). The *SLC6A9*$^{+}$ patients mainly

fitted into the molecular clusters of IntClust-10 and IntClust-5 (p-value $\leq$0.0003). In the METABRIC cohort, the overexpression of *SLC6A9* was associated majorly with the basal subtype (p-value $\leq$4.0E-28) and IntClust-10 and IntClust-5 (p-value <1.4E-33).

The KM estimates of the BCSS for patients showed that the overexpression of *SLC6A9* had a relevant impact on the BCSS of the invasive BC patients in the $LVI^{+/-}$ (p-value $\leq$0.029) and the METABRIC patients (p-value $\leq$0.021). The overexpression of *SIGLEC1* in invasive BC patients had a hazard ratio of 0.949 (p-value $\leq$0.713), suggesting that the altered expression of *SLC6A9* could be a dependent predictor of BC progression.

Table 3. 6: Clinicopathological significance of positively correlated genes with LVI in the LVI-supervised cohort (n=170). Significant values are in bold

	CTSD			*IKZF3*			*ITPRIP*			*P4HA2*		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN^{-}	48 (60.0)	32 (40.0)	6.044 (**0.014**)	46 (57.5)	34 (42.5)	3.400 (0.065)	46 (57.5)	34 (42.5)	3.400 (0.065)	49 (61.3)	31 (38.8)	7.650 (**0.006**)
Definite/LN$^{+/-}$	37 (41.1)	53 (58.9)		39 (43.3)	51 (56.7)		39 (43.3)	51 (56.7)		36 (40.0)	54 (60.0)	
Age (years)												
≤ 50	34 (51.5)	32 (48.5)	0.099 (0.753)	33 (50.0)	33 (50.0)	0.000 (1.000)	32 (48.5)	34 (51.5)	0.099 (0.753)	29 (43.9)	37 (56.1)	1.585 (0.208)
> 50	51 (49.0)	53 (51.0)		52 (50.0)	52 (50.0)		53 (51.0)	51 (49.0)		56 (53.8)	48 (46.2)	
Tumour size												
≤ 2.0cm	55 (55.0)	45 (45.0)	2.429 (0.119)	58 (58.0)	42 (42.0)	6.217 (**0.013**)	50 (50.0)	50 (50.0)	0.000 (1.000)	53 (53.0)	47 (47.0)	0.874 (0.350)
> 2.0cm	30 (42.9)	40 (57.1)		27 (38.6)	43 (61.4)		35 (50.0)	35 (50.0)		32 (45.7)	38 (54.3)	
Tumour Grade												
Grade-I	16 (61.5)	10 (38.5)	2.897 (0.235)	18 (69.2)	8 (30.8)	5.640 (0.060)	16 (61.5)	10 (38.5)	6.488 (**0.039**)	14 (53.8)	12 (46.2)	2.757 (0.252)
Grade-II	23 (41.8)	32 (58.2)		29 (52.7)	26 (47.3)		33 (60.0)	22 (40.0)		32 (58.2)	23 (41.8)	
Grade-III	45 (51.1)	43 (48.9)		38 (43.2)	50 (56.8)		36 (40.9)	52 (59.1)		39 (44.3)	49 (55.7)	
Nodal involvement, (NPI staging system)												
Stage-I	60 (57.1)	45 (42.9)	5.612 (0.060)	59 (56.2)	46 (43.8)	6.120 (**0.047**)	60 (57.1)	45 (42.9)	5.612 (0.060)	61 (58.1)	44 (41.9)	7.594 (**0.022**)
Stage-II	19 (38.8)	30 (61.2)		22 (44.9)	27 (55.1)		19 (38.8)	30 (61.2)		17 (34.7)	32 (65.3)	
Stage-III	6 (37.5)	10 (62.5)		4 (25.0)	12 (75.0)		6 (37.5)	10 (62.5)		7 (43.8)	9 (56.3)	
Nottingham Prognostic Index (NPI)												
good	38 (55.1)	31 (44.9)	1.196 (0.550)	41 (59.4)	28 (40.6)	8.862 (**0.012**)	43 (62.3)	26 (37.7)	7.108 (**0.029**)	40 (58.0)	29 (42.0)	2.967 (0.227)
moderate	41 (46.6)	47 (53.4)		42 (47.7)	46 (52.3)		37 (42.0)	51 (58.0)		39 (44.3)	49 (55.7)	
poor	6 (46.2)	7 (53.8)		2 (15.4)	11 (84.6)		5 (38.5)	8 (61.5)		6 (46.2)	7 (53.8)	

	CTSD			***IKZF3***			***ITPRIP***			***P4HA2***		
	-	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50												
Normal-like	6 (50.0)	6 (50.0)		8 (66.7)	4 (33.3)		4 (33.3)	8 (66.7)		6 (50.0)	6 (50.0)	
Luminal-A	31 (48.4)	33 (51.6)		41 (64.1)	23 (35.9)		38 (59.4)	26 (40.6)		37 (57.8)	27 (42.2)	
Luminal-B	21 (46.7)	24 (53.3)	3.720 (0.445)	18 (40.0)	27 (60.0)	11.653 (**0.020**)	31 (68.9)	14 (31.1)	23.748 (**0.00009**)	17 (37.8)	28 (62.2)	6.109 (0.191)
Her2-enriched	5 (35.7)	9 (64.3)		5 (35.7)	9 (64.3)		5 (35.7)	9 (64.3)		5 (35.7)	9 (64.3)	
Basal	22 (62.9)	13 (37.1)		13 (37.1)	22 (62.9)		7 (20.00)	28 (80.0)		20 (57.1)	15 (42.9)	
Integrated classification clusters of METABRIC												
cluster-1	10 (52.6)	9 (47.4)		5 (26.3)	14 (73.7)		14 (73.7)	5 (26.3)		8 (42.1)	11 (57.9)	
cluster -2	4 (66.7)	2 (33.3)		3 (50.0)	3 (50.0)		3 (50.0)	3 (50.0)		4 (66.7)	2 (33.3)	
cluster -3	15 (50.0)	15 (50.0)		17 (56.7)	13 (43.3)		14 (46.7)	16 (53.3)		14 (46.7)	16 (53.3)	
cluster -4	15 (60.0)	10 (40.0)		15 (60.0)	10 (40.0)		8 (32.0)	17 (68.0)		16 (64.0)	9 (36.0)	
cluster -5	5 (20.0)	15 (80.0)	11.793 (0.225)	6 (30.00)	14 (70.0)	13.930 (0.125)	9 (45.0)	11 (55.0)	25.200 (**0.003**)	4 (20.0)	16 (80.0)	12.386 (0.192)
cluster -6	3 (100.0)	0 (0.0)		3 (100.0)	0 (0.0)		2 (66.7)	1 (33.3)		2 (66.7)	1 (33.3)	
cluster -7	6 (40.0)	9 (60.0)		9 (60.0)	6 (40.0)		11 (73.3)	4 (26.7)		7 (46.7)	8 (53.3)	
cluster -8	8 (42.1)	11 (57.9)		11 (57.9)	8 (42.1)		14 (73.7)	5 (26.3)		10 (52.6)	9 (47.4)	
cluster -9	4 (50.0)	4 (50.0)		5 (62.5)	3 (37.5)		5 (62.5)	3 (37.5)		5 (62.5)	3 (37.5)	
cluster -10	15 (60.0)	10 (40.0)		11 (44.0)	14 (56.0)		5 (20.0)	20 (80.0)		15 (60.0)	10 (40.0)	

	PDLIM7			***PILRA***			***PPM1G***			***RNASE1***		
	-	**+**	**χ^2 (*p*-value≤)**	**-**	**+**	**χ^2 (*p*-value≤)**	**-**	**+**	**χ^2 (*p*-value≤)**	**-**	**+**	**χ^2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN$^-$	49 (61.3)	31 (38.8)	7.650 (**0.006**)	46 (57.5)	34 (42.5)	3.400 (0.65)	50 (62.5)	30 (37.5)	9.444 (**0.002**)	46 (57.5)	34 (42.5)	3.400 (0.065)
Definite/LN$^{+/-}$	36 (40.0)	54 (60.0)		39 (43.3)	51 (56.7)		35 (38.9)	55 (61.1)		39 (43.3)	51 (56.7)	
Age (years)												
≤ 50	36 (54.5)	30 (45.5)	0.892 (0.345)	29 (43.9)	37 (56.1)	1.585 (0.208)	25 (37.9)	41 (62.1)	6.340 (**0.012**)	35 (53.0)	31 (47.0)	0.396 (0.529)
> 50	49 (47.1)	55 (52.9)		56 (53.8)	48 (46.2)		60 (57.7)	44 (42.3)		50 (48.1)	54 (51.9)	
Tumour size												
≤ 2.0cm	58 (58.0)	42 (42.0)	6.217 (**0.013**)	53 (53.0)	47 (47.0)	0.874 (0.350)	60 (60.0)	40 (40.0)	9.714 (**0.002**)	52 (52.0)	48 (48.0)	0.389 (0.533)
> 2.0cm	27 (38.6)	43 (61.4)		32 (45.7)	38 (54.3)		25 (35.7)	45 (64.3)		33 (47.1)	37 (52.9)	
Tumour Grade												
Grade-I	18 (69.2)	8 (30.8)		21 (80.8)	5 (19.2)		19 (73.1)	7 (26.9)		18 (69.2)	8 (30.8)	
Grade-II	35 (63.6)	20 (36.4)	14.477 (**0.001**)	36 (65.5)	19 (34.5)	26.732 (**0.000002**)	37 (67.3)	18 (32.7)	23.733 (**0.000007**)	28 (50.9)	27 (49.1)	4.995 (0.082)
Grade-III	32 (36.4)	56 (63.6)		28 (31.8)	60 (68.2)		28 (31.8)	60 (68.2)		39 (44.3)	49 (55.7)	
Nodal involvement, (NPI staging system)0												
Stage-I	59 (56.2)	46 (43.8)		63 (60.0)	42 (40.0)		62 (59.0)	43 (41.0)		62 (59.0)	43 (41.0)	
Stage-II	21 (42.9)	28 (57.1)	4.860 (0.088)	16 (32.7)	33 (67.3)	11.098 (**0.004**)	17 (34.7)	32 (65.3)	9.030 (**0.011**)	16 (32.7)	33 (67.3)	9.586 (**0.008**)
Stage-III	5 (31.3)	11 (68.8)		6 (37.5)	10 (62.5)		6 (37.5)	10 (62.5)		7 (43.8)	9 (56.3)	
Nottingham Prognostic Index (NPI)												
good	47 (68.1)	22 (31.9)		50 (72.5)	19 (27.5)		48 (69.6)	21 (30.4)		43 (62.3)	26 (37.7)	
moderate	35 (39.8)	53 (60.2)	16.509 (**0.0003**)	30 (34.1)	58 (65.9)	23.529 (**0.000008**)	34 (38.6)	54 (61.4)	18.880 (**0.00008**)	36 (40.9)	52 (59.1)	7.174 (**0.028**)
poor	3 (21.0)	10 (76.9)		5 (38.5)	8 (61.5)		3 (23.1)	10 (76.9)		6 (46.2)	7 (53.8)	

	PDLIM7			***PILRA***			***PPM1G***			***RNASE1***		
	-	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50												
Normal-like	8 (66.7)	4 (33.3)		6 (50.0)	6 (50.0)		10 (83.3)	2 (16.7)		5 (41.7)	7 (58.3)	
Luminal-A	40 (62.5)	24 (37.5)		48 (75.0)	16 (25.0)		43 (67.2)	21 (32.8)		42 (65.6)	22 (34.4)	
Luminal-B	22 (48.9)	23 (51.1)	14.813 (**0.005**)	19 (42.2)	26 (57.8)	29.975 (**0.000005**)	25 (55.6)	20 (44.4)	40.051 (**4.2E-8**)	20 (44.4)	25 (55.6)	11.967 (**0.018**)
Her2-enriched	2 (14.3)	12 (85.7)		4 (28.6)	10 (71.4)		0 (0.0)	14 (100.0)		7 (50.0)	7 (50.0)	
Basal	13 (37.1)	22 (62.9)		8 (22.9)	27 (77.1)		7 (20.0)	28 (80.0)		11 (31.4)	24 (68.6)	
Integrated classification clusters of METABRIC												
cluster-1	7 (36.8)	12 (63.2)		8 (42.1)	11 (57.9)		8 (42.1)	11 (57.9)		10 (52.6)	9 (47.4)	
cluster -2	2 (33.3)	4 (66.74)		3 (50.0)	3 (50.0)		2 (33.3)	4 (66.7)		4 (66.7)	2 (33.3)	
cluster -3	16 (53.3)	14 (46.7)		22 (73.3)	8 (26.7)		22 (73.3)	8 (26.7)		20 (66.7)	10 (33.3)	
cluster -4	14 (56.0)	11 (44.0)		12 (48.0)	13 (52.0)		14 (56.0)	11 (44.0)		11 (44.0)	14 (56.0)	
cluster -5	6 (30.0)	14 (70.0)	15.409 (0.080)	7 (35.0)	13 (65.0)	29.275 (**0.001**)	4 (20.0)	16 (80.0)	42.636 (**0.000003**)	7 (35.0)	13 (65.0)	9.962 (0.354)
cluster -6	2 (66.7)	1 (33.3)		1 (33.3)	2 (66.7)		2 (66.7)	1 (33.3)		2 (66.7)	1 (33.3)	
cluster -7	13 (86.7)	2 (13.3)		10 (66.7)	5 (33.3)		10 (66.7)	5 (33.3)		6 (40.0)	9 (60.0)	
cluster -8	11 (57.9)	8 (42.1)		15 (78.9)	4 (21.1)		17 (89.5)	2 (10.5)		12 (63.2)	7 (36.8)	
cluster -9	3 (37.5)	5 (62.5)		3 (37.5)	5 (62.5)		2 (25.0)	6 (75.0)		3 (37.5)	5 (62.5)	
cluster -10	11 (44.0)	14 (56.0)		4 (16.0)	21 (84.0)		4 (16.0)	21 (84.0)		10 (40.0)	15 (60.0)	

	SEC14L1			*SIGLEC1*			*SLC3A2*			*SLC6A9*			*ZDHHC9*		
	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value<)
Lymphovascular invasion (LVI), (H&E/IHC)															
Negative/LN^{-}	47 (58.8)	33 (41.3)	4.628 (**0.031**)	44 (55.0)	36 (45.0)	1.511 (0.219)	48 (60.0)	32 (40.0)	6.044 (**0.014**)	48 (60.0)	32 (40.0)	6.044 (**0.014**)	44 (55.0)	36 (45.0)	1.511 (0.219)
Definite/LN$^{+/-}$	38 (42.2)	52 (57.8)		41 (45.6)	49 (54.4)		37 (41.1)	53 (58.9)		37 (41.1)	53 (58.9)		41 (45.6)	49 (54.4)	
Age (years)															
≤ 50	33 (50.0)	33 (50.0)	0.000 (1.000)	27 (40.9)	39 (59.1)	3.566 (0.059)	34 (51.5)	32 (48.5)	0.099 (0.753)	37 (56.1)	29 (43.9)	1.585 (0.208)	30 (45.5)	36 (54.5)	0.892 (0.345)
> 50	52 (50.0)	52 (50.0)		58 (55.8)	46 (44.2)		51 (49.0)	53 (51.0)		48 (46.2)	56 (53.8)		55 (52.9)	49 (47.1)	
Tumour size															
≤ 2.0cm	53 (53.0)	47 (47.0)	0.874 (0.350)	52 (52.0)	48 (48.0)	0.389 (0.533)	55 (55.0)	45 (45.0)	2.429 (0.119)	54 (54.0)	46 (46.0)	1.554 (0.213)	55 (55.0)	45 (45.0)	2.429 (0.119)
> 2.0cm	32 (45.7)	38 (54.3)		33 (47.1)	37 (52.9)		30 (42.9)	40 (57.1)		31 (44.3)	39 (55.7)		30 (42.9)	40 (57.1)	
Tumour Grade															
Grade-I	19 (73.1)	7 (26.9)		20 (76.9)	6 (23.1)		16 (61.5)	10 (38.5)		17 (65.4)	9 (34.6)		17 (65.4)	9 (34.6)	
Grade-II	38 (69.1)	17 (30.9)	25.188 (**0.000003**)	29 (52.7)	26 (47.3)	11.378 (**0.003**)	36 (65.5)	19 (34.5)	13.179 (**0.001**)	33 (60.0)	22 (40.0)	9.201 (**0.010**)	29 (52.7)	26 (47.3)	4.256 (0.119)
Grade-III	28 (31.8)	60 (68.2)		35 (39.8)	53 (60.2)		32 (36.4)	56 (63.6)		34 (38.6)	54 (61.4)		38 (43.2)	50 (56.6)	
Nodal involvement, (NPI staging system)															
Stage-I	62 (59.0)	43 (41.0)		62 (59.0)	43 (41.0)		60 (57.1)	45 (42.9)		60 (57.1)	45 (42.9)		58 (55.2)	47 (44.8)	
Stage-II	18 (36.7)	31 (63.3)	9.137 (**0.010**)	15 (30.6)	34 (69.4)	10.805 (**0.005**)	19 (38.8)	30 (61.2)	5.612 (0.060)	21 (42.9)	28 (57.1)	7.143 (**0.028**)	20 (40.8)	29 (59.2)	3.055 (0.217)
Stage-III	5 (31.3)	11 (68.8)		8 (50.0)	8 (50.0)		6 (37.5)	10 (62.5)		4 (25.0)	12 (75.0)		7 (43.8)	9 (56.3)	
Nottingham Prognostic Index (NPI)															
good	50 (72.5)	19 (27.5)		44 (63.8)	25 (36.2)		46 (66.7)	23 (33.3)		44 (63.8)	25 (36.2)		41 (59.4)	28 (40.6)	
moderate	31 (35.2)	57 (64.8)	23.532 (**0.000008**)	35 (39.8)	53 (60.2)	8.991 (**0.011**)	34 (38.6)	54 (61.4)	12.904 (**0.002**)	38 (43.2)	50 (56.8)	10.637 (**0.005**)	39 (44.3)	49 (55.7)	4.278 (0.118)
poor	4 (30.8)	9 (69.2)		6 (46.2)	7 (53.8)		5 (38.5)	8 (61.5)		3 (23.1)	10 (76.9)		5 (38.5)	8 (61.5)	

	SEC14L1			***SIGLEC1***			***SLC3A2***			***SLC6A9***			**ZDHHC9**		
	-	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50															
Normal-like	7 (58.3)	5 (41.7)		8 (66.7)	4 (33.3)		7 (58.3)	5 (41.7)		9 (75.0)	3 (25.0)		9 (75.0)	3 (25.0)	
Luminal-A	49 (76.6)	15 (23.4)		44 (68.8)	20 (31.3)		44 (68.8)	20 (31.3)		42 (65.6)	22 (34.4)		40 (62.5)	24 (37.5)	
Luminal-B	20 (44.4)	25 (55.6)	38.637 (**8.3E-8**)	14 (31.1)	31 (68.9)	20.213 (**0.0005**)	17 (37.8)	28 (62.2)	16.908 (**0.002**)	24 (53.3)	21 (46.7)	26.907 (**0.000021**)	22 (48.9)	23 (51.1)	19.622 (**0.001**)
Her2-enriched	3 (21.4)	11 (78.6)		7 (50.0)	7 (50.0)		4 (28.6)	10 (71.4)		2 (14.3)	12 (85.7)		1 (7.1)	13 (92.9)	
Basal	6 (17.1)	29 (82.9)		12 (34.3)	23 (65.7)		13 (37.1)	22 (62.9)		8 (22.9)	27 (71.1)		13 (37.1)	22 (62.9)	
Integrated classification clusters of METABRIC															
cluster-1	5 (26.3)	14 (73.7)		7 (36.8)	12 (63.2)		6 (31.6)	13 (68.4)		8 (42.1)	11 (57.9)		11 (57.9)	8 (42.1)	
cluster -2	2 (33.3)	4 (66.7)		4 (66.7)	2 (33.3)		5 (83.3)	1 (16.7)		5 (83.3)	1 (16.7)		1 (16.7)	5 (83.3)	
cluster -3	25 (83.3)	5 (16.7)		21 (7.0)	9 (30.0)		20 (66.7)	10 (33.3)		21 (70.0)	9 (30.0)		19 (63.3)	11 (36.7)	
cluster -4	13 (52.0)	12 (48.0)		18 (72.0)	7 (28.0)		16 (64.0)	9 (36.0)		17 (68.0)	8 (32.0)		17 (68.0)	8 (32.0)	
cluster -5	7 (35.0)	13 (65.0)	47.058 (**3.8E-7**)	8 (40.0)	12 (60.0)	24.172 (**0.004**)	5 (25.0)	15 (75.0)	20.906 (**0.013**)	5 (25.0)	15 (75.0)	30.860 (**0.0003**)	6 (30.0)	14 (70.0)	24.299 (**0.004**)
cluster -6	2 (66.7)	1 (33.3)		1 (33.3)	2 (66.7)		2 (66.7)	1 (33.3)		3 (100.0)	0 (0.0)		3 (100.0)	0 (0.0)	
cluster -7	10 (66.7)	5 (33.3)		6 (40.0)	9 (60.0)		9 (60.0)	6 (40.0)		8 (53.3)	7 (46.7)		11 (73.3)	4 (26.7)	
cluster -8	16 (84.2)	3 (15.8)		12 (63.2)	7 (36.8)		11 (57.9)	8 (42.1)		10 (52.6)	9 (47.4)		6 (31.6)	13 (68.4)	
cluster -9	1 (12.5)	7 (87.5)		3 (37.5)	5 (62.5)		2 (25.0)	6 (75.0)		4 (50.0)	4 (50.0)		3 (37.5)	5 (62.5)	
cluster -10	4 (16.0)	21 (84.0)		5 (20.0)	20 (80.0)		9 (36.0)	16 (64.0)		4 (16.0)	21 (84.0)		8 (32.0)	17 (68.0)	

Table 3. 7: Clinicopathological significance of positively correlated genes with LVI in the METABRIC cohort (n=1980). Significant values are in bold

	CTSD			*IKZF3*			*ITPRIP*			*P4HA2*		
	-	+	χ2	-	+	χ2	-	+	χ2	-	+	χ2
	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN$^{+/-}$	85 (51.8)	79 (48.2)	3.030 (0.082)	121 (73.8)	43 (26.2)	5.203 (**0.023**)	92 (56.1)	72 (43.9)	0.278 (0.598)	86 (52.4)	78 (47.6)	6.917 (**0.009**)
Definite/LN$^{+/-}$	55 (41.7)	77 (58.3)		81 (61.4)	51 (38.6)		70 (53.0)	62 (47.0)		49 (37.1)	83 (62.9)	
Age (years)												
≤ 50	225 (52.4)	204 (47.6)	1.312 (0.252)	211 (49.2)	218 (50.8)	0.146 (0.703)	183 (42.7)	246 (57.3)	11.811 (**0.001**)	203 (47.5)	224 (52.5)	1.290 (0.256)
> 50	765 (49.3)	786 (50.7)		779 (50.2)	772 (49.8)		807 (52.0)	744 (48.0)		785 (50.6)	765 (49.4)	
Tumour size												
≤ 2.0cm	463 (54.0)	395 (46.0)	9.834 (**0.002**)	448 (52.2)	410 (47.8)	2.594 (0.107)	447 (52.1)	411 (47.9)	2.819 (0.093)	456 (53.1)	402 (46.9)	6.350 (**0.012**)
> 2.0cm	516 (46.8)	586 (53.2)		535 (48.5)	567 (51.5)		532 (48.3)	570 (51.7)		521 (47.4)	578 (52.6)	
Tumour Grade												
Grade-I	104 (60.8)	67 (39.2)		113 (66.1)	58 (33.9)		97 (56.7)	74 (43.3)		98 (57.3)	73 (42.7)	
Grade-II	375 (48.7)	395 (51.3)	8.794 (**0.012**)	431 (56.0)	339 (44.0)	51.690 (**5.9E-12**)	460 (59.7)	310 (40.3)	64.116 (**1.2E-14**)	388 (50.5)	380 (49.5)	4.883 (0.87)
Grade-III	468 (49.2)	484 (50.8)		402 (42.2)	550 (57.8)		389 (40.9)	563 (59.1)		459 (48.3)	492 (51.7)	
Nodal involvement, (NPI staging system)												
Stage-I	563 (54.4)	472 (45.6)		555 (53.6)	480 (46.4)		554 (53.5)	481 (46.5)		536 (51.8)	498 (48.2)	
Stage-II	297 (47.7)	325 (52.3)	21.425 (**0.00002**)	306 (49.2)	316 (50.8)	16.216 (**0.0003**)	292 (46.9)	330 (53.1)	10.306 (**0.006**)	302 (48.7)	318 (51.3)	2.827 (0.243)
Stage-III	127 (40.2)	189 (59.8)		129 (40.8)	187 (59.2)		143 (45.3)	173 (54.7)		149 (47.2)	167 (52.8)	
Nottingham Prognostic Index (NPI)												
good	375 (55.1)	305 (44.9)		400 (58.8)	280 (41.2)		411 (60.4)	269 (39.6)		367 (54.1)	312 (45.9)	
moderate	529 (48.0)	572 (52.0)	12.549 (**0.002**)	519 (47.1)	582 (52.9)	41.108 (**1.2E-9**)	497 (45.1)	604 (54.9)	46.207 (**9.3E-11**)	526 (47.9)	573 (52.1)	6.872 (**0.032**)
poor	86 (43.2)	113 (56.8)		71 (35.7)	128 (64.3)		82 (41.2)	117 (58.8)		95 (47.7)	104 (52.3)	

	CTSD			*IKZF3*			*ITPRIP*			*P4HA2*		
	-	+	χ^2	-	+	χ^2	-	+	χ^2	-	+	χ^2
	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)
Prediction analysis of microarray 50, PAM50												
Normal-like	126 (61.5)	79 (38.5)		89 (43.4)	116 (56.6)		53 (25.9)	152 (74.1)		115 (56.1)	90 (43.9)	
Luminal-A	367 (51.1)	351 (48.9)		456 (63.5)	262 (36.5)		463 (64.5)	255 (35.5)		375 (52.3)	342 (47.7)	
Luminal-B	224 (45.9)	264 (54.1)	25.911 (**0.00003**)	266 (54.5)	222 (45.5)	138.366 (**6.3E-29**)	346 (70.9)	142 (29.1)	370.402 (**6.9E-79**)	238 (48.9)	249 (51.1)	12.275 (**0.015**)
Her2-enriched	96 (40.0)	144 (60.0)		73 (30.4)	167 (69.6)		67 (27.9)	173 (72.1)		99 (41.3)	141 (58.8)	
Basal	177 (53.8)	152 (46.2)		106 (32.2)	223 (67.8)		61 (18.5)	268 (81.5)		161 (49.1)	167 (50.9)	
Integrated classification clusters of METABRIC												
cluster-1	66 (47.5)	73 (52.5)		73 (52.5)	66 (47.5)		93 (66.9)	46 (33.1)		72 (51.8)	67 (48.2)	
cluster -2	38 (52.8)	34 (47.2)		35 (48.6)	37 (51.4)		49 (68.1)	23 (31.9)		31 (43.1)	41 (56.9)	
cluster -3	148 (51.0)	142 (49.0)		183 (63.1)	107 (36.9)		162 (55.9)	128 (44.1)		159 (54.8)	131 (45.2)	
cluster -4	185 (53.9)	158 (46.1)		139 (40.5)	204 (59.5)		108 (31.5)	235 (68.5)		184 (53.8)	158 (46.2)	
cluster -5	70 (36.8)	120 (63.2)	46.211 (**5.5E-7**)	46 (24.2)	144 (75.8)	151.726 (**3.9E-28**)	56 (29.5)	134 (70.5)	264.409 (**9.0E-52**)	60 (31.6)	130 (68.4)	58.803 (**2.3E-9**)
cluster -6	65 (76.5)	20 (23.5)		47 (55.3)	38 (44.7)		49 (57.6)	36 (42.4)		27 (31.8)	58 (68.2)	
cluster -7	85 (44.7)	105 (55.3)		109 (57.4)	81 (42.6)		134 (70.5)	56 (28.8)		122 (64.2)	68 (35.8)	
cluster -8	144 (48.2)	155 (51.8)		209 (43.8)	90 (30.1)		213 (71.2)	86 (28.8)		146 (49.2)	151 (50.8)	
cluster -9	65 (44.5)	81 (55.5)		64 (43.8)	82 (56.2)		75 (51.4)	71 (48.6)		74 (50.7)	72 (49.3)	
cluster -10	124 (54.9)	102 (45.1)		85 (37.6)	141 (62.4)		51 (22.6)	175 (77.4)		113 (50.0)	113 (50.0)	

	PDLIM7			***PILRA***			***PPM1G***			***RNASE1***		
	-	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN$^{+/-}$	84 (51.2)	80 (48.8)	1.894 (0.169)	77 (47.0)	87 (53.0)	0.267 (0.605)	112 (68.3)	52 (31.7)	2.278 (0.131)	56 (34.1)	108 (65.9)	0.969 (0.325)
Definite/LN$^{+/-}$	57 (43.2)	75 (56.8)		58 (43.9)	74 (56.1)		79 (59.8)	53 (40.2)		38 (28.8)	94 (71.2)	
Age (years)												
≤ 50	204 (47.6)	225 (52.4)	1.312 (0.252)	225 (52.4)	204 (47.6)	1.312 (0.252)	157 (36.6)	272 (63.4)	39.354 (**3.5E-10**)	203 (47.3)	226 (52.7)	1.574 (0.210)
> 50	786 (50.7)	765 (49.3)		768 (49.3)	786 (50.7)		833 (53.7)	718 (46.3)		787 (50.7)	764 (49.3)	
Tumour size												
≤ 2.0cm	449 (52.3)	409 (47.7)	3.767 (0.052)	447 (52.1)	411 (47.9)	3.093 (0.079)	459 (53.5)	399 (46.8)	7.033 (**0.008**)	435 (50.7)	423 (49.3)	0.443 (0.505)
> 2.0cm	528 (47.9)	574 (52.1)		530 (48.1)	572 (51.9)		523 (47.5)	579 (52.5)		542 (49.2)	560 (50.8)	
Tumour Grade												
Grade-I	93 (54.4)	78 (45.6)		111 (64.9)	60 (35.1)		135 (78.9)	36 (21.1)		89 (52.0)	82 (48.0)	
Grade-II	402 (52.2)	368 (47.8)	8.240 (**0.016**)	437 (56.8)	333 (43.2)	58.141 (**2.3E-13**)	477 (61.9)	293 (38.1)	189.602 (**6.7E-42**)	410 (53.2)	360 (46.8)	8.327 (**0.016**)
Grade-III	439 (46.1)	513 (53.9)		393 (41.3)	559 (58.7)		331 (34.8)	621 (65.2)		442 (46.4)	510 (53.6)	
Nodal involvement, (NPI staging system)												
Stage-I	545 (52.7)	490 (47.3)		557 (53.8)	478 (46.2)		563 (54.4)	472 (45.6)		553 (53.4)	482 (46.6)	
Stage-II	297 (47.7)	325 (52.3)	6.322 (**0.042**)	288 (46.3)	334 (53.7)	12.672 (**0.002**)	287 (46.1)	335 (53.9)	18.397 (**0.0001**)	293 (47.1)	329 (52.9)	9.410 (**0.009**)
Stage-III	145 (45.9)	171 (54.1)		142 (44.9)	174 (55.1)		135 (42.7)	181 (57.3)		144 (45.6)	172 (54.4)	
Nottingham Prognostic Index (NPI)												
good	380 (55.9)	300 (44.1)		410 (60.3)	270 (39.7)		448 (65.9)	232 (34.1)		382 (56.2)	298 (43.8)	
moderate	519 (47.1)	582 (52.9)	14.469 (**0.001**)	503 (45.7)	598 (54.3)	47.197 (**5.6E-11**)	478 (43.4)	623 (56.6)	113.040 (**2.8E-25**)	524 (47.6)	577 (52.4)	17.757 (**0.0001**)
poor	91 (45.7)	108 (54.3)		77 (38.7)	122 (61.3)		64 (32.2)	135 (67.8)		84 (42.2)	115 (57.8)	

	PDLIM7			*PILRA*			*PPM1G*			*RNASE1*		
	-	+	χ^2	-	+	χ^2	-	+	χ^2	-	+	χ^2
	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)
Prediction analysis of microarray 50, PAM50												
Normal-like	111 (54.1)	94 (45.9)		125 (61.0)	80 (39.0)		131 (63.9)	74 (36.1)		78 (38.0)	127 (62.0)	
Luminal-A	371 (51.7)	347 (48.3)		448 (62.4)	270 (37.6)		505 (70.3)	213 (29.7)		425 (59.2)	293 (40.8)	
Luminal-B	280 (57.4)	208 (42.6)	42.812 (**1.1E-8**)	228 (46.7)	260 (53.3)	121.023 (**3.2E-25**)	254 (52.0)	234 (48.0)	374.856 (**7.5E-80**)	232 (47.5)	256 (52.5)	57.796 (**8.4E-12**)
Her2-enriched	80 (33.3)	160 (66.7)		85 (35.4)	155 (64.6)		40 (16.7)	200 (83.3)		130 (54.2)	110 (45.8)	
Basal	148 (45.0)	181 (55.0)		104 (31.6)	225 (68.4)		60 (18.2)	269 (81.8)		125 (38.0)	204 (62.0)	
Integrated classification clusters of METABRIC												
cluster-1	77 (55.4)	62 (44.6)		56 (40.3)	83 (59.7)		50 (36.0)	89 (64.0)		73 (52.5)	66 (47.5)	
cluster -2	42 (58.3)	30 (41.7)		33 (45.8)	39 (54.2)		42 (58.3)	30 (41.7)		38 (52.8)	34 (47.2)	
cluster -3	155 (53.4)	135 (46.6)		189 (65.2)	101 (34.8)		213 (73.4)	77 (26.6)		153 (52.8)	137 (47.2)	
cluster -4	164 (47.8)	179 (52.2)		144 (42.0)	199 (58.0)		194 (56.6)	149 (43.4)		132 (38.5)	211 (61.5)	
cluster -5	57 (30.0)	133 (70.0)		75 (39.5)	115 (60.5)		40 (21.1)	150 (78.9)		78 (41.1)	112 (58.9)	
cluster -6	31 (36.5)	54 (63.5)	48.441 (**2.1E-7**)	48 (56.5)	37 (43.5)	125.981 (**8.0E-23**)	34 (40.0)	51 (60.00)	327.987 (**3.0E-65**)	46 (54.1)	39 (45.9)	70.246 (**1.4E-11**)
cluster -7	105 (55.3)	85 (44.7)		101 (53.2)	89 (46.8)		120 (63.2)	70 (36.8)		110 (57.9)	80 (42.1)	
cluster -8	165 (55.2)	134 (44.8)		204 (68.2)	95 (31.8)		208 (69.6)	91 (30.4)		195 (65.2)	104 (34.8)	
cluster -9	78 (53.4)	68 (46.6)		71 (48.6)	75 (51.4)		55 (37.7)	91 (62.3)		77 (52.7)	69 (47.3)	
cluster -10	116 (51.3)	110 (48.7)		69 (30.5)	157 (69.5)		34 (15.0)	192 (85.0)		88 (38.9)	138 (61.1)	

	SEC14L1			*SIGLEC1*			*SLC3A2*			*SLC6A9*			*ZDHHC9*		
	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)
Lymphovascular invasion (LVI), (H&E/IHC)															
Negative/LN$^{+/-}$	90 (54.9)	74 (45.1)	0.698 (0.403)	80 (48.8)	84 (51.2)	0.689 (0.407)	102 (62.2)	62 (37.8)	3.906 (**0.048**)	102 (62.2)	62 (37.8)	2.524 (0.112)	110 (67.1)	54 (32.9)	0.121 (0.728)
Definite/LN$^{+/-}$	66 (50.0)	66 (50.0)		58 (43.9)	74 (56.1)		67 (50.8)	65 (49.2)		70 (53.0)	62 (47.0)		86 (65.2)	46 (34.8)	
Age (years)															
≤ 50	190 (44.3)	239 (55.7)	7.145 (**0.008**)	195 (45.5)	234 (54.5)	4.526 (**0.033**)	214 (49.9)	215 (50.1)	0.003 (0.956)	227 (52.9)	202 (47.1)	1.860 (0.173)	206 (48.0)	223 (52.0)	0.860 (0.354)
> 50	800 (51.6)	751 (48.4)		795 (51.3)	756 (48.7)		776 (50.0)	775 (50.0)		763 (49.2)	788 (50.8)		784 (50.5)	767 (49.5)	
Tumour size															
≤ 2.0cm	451 (52.6)	407 (47.4)	3.400 (0.065)	438 (51.0)	420 (49.0)	0.882 (0.348)	462 (53.8)	396 (46.2)	9.515 (**0.002**)	471 (54.9)	387 (45.1)	15.556 (**0.00008**)	455 (53.0)	403 (47.0)	5.418 (**0.020**)
> 2.0cm	533 (48.4)	569 (51.6)		539 (48.9)	563 (51.1)		516 (46.8)	586 (53.2)		506 (45.9)	596 (54.1)		526 (47.7)	576 (52.3)	
Tumour Grade															
Grade-I	115 (67.3)	56 (32.7)		112 (65.5)	59 (34.5)		100 (58.5)	71 (41.5)		103 (60.2)	68 (39.8)		117 (68.4)	54 (31.6)	
Grade-II	475 (61.7)	295 (38.3)	123.947 (**1.2E-27**)	423 (54.9)	347 (45.1)	47.336 (**5.3E-11**)	426 (55.3)	344 (44.7)	28.735 (**5.8E-7**)	434 (56.4)	336 (43.6)	37.938 (**5.8E-9**)	422 (54.8)	348 (45.2)	48.612 (**2.8E-11**)
Grade-III	355 (37.3)	597 (62.7)		401 (42.1)	551 (57.9)		416 (43.7)	536 (56.3)		410 (43.1)	542 (56.9)		410 (43.1)	542 (56.9)	
Nodal involvement, (NPI staging system)															
Stage-I	575 (55.6)	460 (44.4)		563 (54.4)	472 (45.6)		568 (54.9)	467 (45.1)		545 (52.7)	490 (47.3)		518 (50.0)	517 (50.0)	
Stage-II	301 (48.4)	321 (51.6)	40.201 (**1.9E-9**)	280 (45.0)	342 (55.0)	16.316 (**0.0003**)	280 (45.0)	342 (55.0)	19.682 (**0.00005**)	301 (48.4)	321 (51.6)	7.224 (**0.027**)	322 (51.8)	300 (48.2)	2.601 (0.272)
Stage-III	112 (35.4)	204 (64.6)		145 (45.9)	171 (54.1)		141 (44.6)	175 (55.4)		141 (44.6)	175 (55.4)		146 (46.2)	170 (53.8)	
Nottingham Prognostic Index (NPI)															
good	444 (65.3)	236 (34.7)		405 (59.6)	275 (40.4)		407 (59.9)	273 (40.1)		394 (57.9)	286 (42.1)		375 (55.1)	305 (44.9)	
moderate	493 (44.8)	608 (55.2)	119.098 (**1.4E-26**)	502 (45.6)	599 (54.4)	38.871 (**3.6E-9**)	502 (45.6)	599 (54.4)	41.831 (**8.2E-10**)	512 (46.5)	589 (53.5)	27.367 (**0.000001**)	535 (48.6)	566 (51.4)	15.722 (**0.0004**)
poor	53 (26.6)	146 (73.4)		83 (41.7)	116 (58.3)		81 (40.7)	118 (59.3)		84 (42.2)	115 (57.8)		80 (40.2)	119 (59.8)	

	SEC14L1			*SIGLEC1*			*SLC3A2*			*SLC6A9*			*ZDHHC9*		
	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)
Prediction analysis of microarray 50, PAM50															
Normal-like	105 (51.2)	100 (48.8)		117 (57.1)	88 (42.9)		102 (49.8)	103 (50.2)		125 (61.0)	80 (39.0)		129 (62.9)	76 (37.1)	
Luminal-A	481 (67.0)	237 (33.0)		441 (61.4)	277 (38.6)		432 (60.2)	286 (39.8)		417 (58.1)	301 (41.9)		427 (59.5)	291 (40.5)	
Luminal-B	252 (51.6)	236 (48.4)	212.543 (**7.5E-45**)	194 (39.8)	294 (60.2)	86.197 (**8.5E-18**)	214 (43.9)	274 (56.1)	61.213 (**1.6E-12**)	279 (57.2)	209 (42.8)	134.631 (**4.0E-28**)	232 (47.5)	256 (52.5)	88.786 (**2.4E-18**)
Her2-enriched	78 (32.5)	162 (67.5)		118 (49.2)	122 (50.8)		122 (50.8)	118 (49.2)		80 (33.3)	160 (66.7)		82 (34.2)	158 (65.8)	
Basal	74 (22.5)	255 (77.5)		120 (36.5)	209 (63.5)		120 (36.5)	209 (63.5)		89 (27.1)	240 (72.9)		120 (36.5)	209 (63.5)	
Integrated classification clusters of METABRIC															
cluster-1	41 (29.5)	98 (70.5)		48 (34.5)	91 (65.5)		56 (40.3)	83 (59.7)		82 (59.0)	57 (41.0)		63 (45.3)	76 (54.7)	
cluster -2	38 (52.8)	34 (47.2)		25 (34.7)	47 (65.3)		33 (45.8)	39 (54.2)		38 (52.8)	34 (47.2)		39 (54.2)	33 (45.8)	
cluster -3	204 (70.3)	86 (29.7)		187 (64.5)	103 (35.5)		189 (65.2)	101 (34.8)		186 (64.1)	104 (35.9)		193 (66.6)	97 (33.4)	
cluster -4	133 (38.8)	210 (61.2)		151 (44.0)	192 (56.0)		167 (48.7)	176 (51.3)		217 (63.3)	126 (36.7)		224 (65.3)	119 (34.7)	
cluster -5	65 (34.2)	125 (65.8)	282.400 (**1.4E-55**)	91 (47.9)	99 (52.1)	117.571 (**4.2E-21**)	73 (38.4)	117 (61.6)	83.319 (**3.5E-14**)	58 (30.5)	132 (69.5)	177.882 (**1.4E-33**)	62 (32.6)	128 (67.4)	166.120 (**4.0E-31**)
cluster -6	41 (48.2)	44 (51.8)		43 (50.6)	42 (49.4)		41 (48.2)	44 (51.8)		55 (39.5)	30 (35.3)		29 (34.1)	56 (65.9)	
cluster -7	128 (67.4)	62 (32.6)		103 (54.2)	87 (45.8)		95 (50.0)	95 (50.0)		75 (39.5)	115 (60.5)		125 (65.8)	65 (34.2)	
cluster -8	228 (76.3)	71 (23.7)		204 (68.2)	95 (31.8)		189 (63.2)	110 (36.8)		170 (56.9)	129 (43.1)		135 (45.2)	164 (54.8)	
cluster -9	59 (40.4)	87 (59.6)		61 (41.8)	85 (58.2)		66 (45.2)	80 (54.8)		59 (40.4)	87 (59.6)		50 (34.2)	96 (68.8)	
cluster -10	53 (23.5)	173 (76.5)		77 (34.1)	149 (65.9)		81 (35.8)	145 (64.2)		50 (22.1)	176 (77.9)		70 (31.0)	156 (69.0)	

Table 3. 8: Multivariate Cox regression analysis for prediction of breast cancer specific survival, UP-13/LVI^{+}

	Breast cancer specific survival (BCSS), n= 1869$^{\lozenge}$								
	Hazard ratio (95% CI)	**Lower**	**Upper**	***p*-value**		**Hazard ratio (95% CI)**	**Lower**	**Upper**	***p*-value**
CTSD mRNA	0.835	0.642	1.087	0.180	***IKZF3* mRNA**	0.945	0.724	1.235	0.680
Tumour grade	1.379	1.063	1.789	0.016	**Tumour grade**	1.376	1.060	1.785	0.016
Nodal staging*	1.831	1.533	2.186	0.000	**Nodal staging***	1.795	1.507	2.137	0.000
Tumour size	1.520	1.147	2.014	0.004	**Tumour size**	1.515	1.142	2.008	0.004
Estrogen expression**	0.721	0.535	0.971	0.031	**Estrogen expression****	0.714	0.530	0.964	0.028
Her2 expression***	1.253	1.060	1.481	0.008	**Her2 expression*****	1.250	1.055	1.481	0.010
	Hazard ratio (95% CI)	**Lower**	**Upper**	***p*-value**		**Hazard ratio (95% CI)**	**Lower**	**Upper**	***p*-value**
***ITPRIP* mRNA**	1.150	0.863	1.534	0.340	***P4HA2* mRNA**	1.390	1.063	1.819	**0.016**
Tumour grade	1.365	1.052	1.772	0.019	**Tumour grade**	1.349	1.039	1.752	0.025
Nodal staging*	1.782	1.496	2.123	0.000	**Nodal staging***	1.790	1.501	2.133	0.000
Tumour size	1.514	1.143	2.007	0.004	**Tumour size**	1.542	1.164	2.044	0.003
Estrogen expression**	0.763	0.553	1.053	0.099	**Estrogen expression****	0.722	0.537	0.971	0.031
Her2 expression***	1.232	1.042	1.457	0.015	**Her2 expression*****	1.207	1.021	1.428	0.028

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
$^{\lozenge}$ Sum of patients does not equal to 1980 because of missing cases among different categories.

Breast cancer specific survival (BCSS), n= 1869[◊]

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***PDLIM7* mRNA**	1.290	0.988	1.684	0.062
Tumour grade	1.342	1.033	1.744	0.027
Nodal staging*	1.788	1.502	2.128	0.000
Tumour size	1.516	1.145	2.008	0.004
Estrogen expression**	0.728	0.540	0.980	0.037
Her2 expression***	1.218	1.030	1.442	0.021

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***PILRA* mRNA**	1.242	0.947	1.628	0.118
Tumour grade	1.343	1.034	1.745	0.027
Nodal staging*	1.792	1.504	2.134	0.000
Tumour size	1.513	1.141	2.004	0.004
Estrogen expression**	0.747	0.552	1.009	0.058
Her2 expression***	1.232	1.043	1.456	0.014

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***PPM1G* mRNA**	1.677	1.248	2.252	**0.001**
Tumour grade	1.254	0.960	1.637	0.097
Nodal staging*	1.770	1.489	2.105	0.000
Tumour size	1.497	1.131	1.981	0.005
Estrogen expression**	0.836	0.615	1.138	0.255
Her2 expression***	1.225	1.037	1.449	0.017

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***RNASE1* mRNA**	1.036	0.796	1.349	0.793
Tumour grade	1.370	1.056	1.778	0.018
Nodal staging*	1.792	1.505	2.134	0.000
Tumour size	1.512	1.141	2.004	0.004
Estrogen expression**	0.723	0.535	0.976	0.034
Her2 expression***	1.240	1.049	1.465	0.012

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
[◊] Sum of patients does not equal to 1980 because of missing cases among different categories.

Breast cancer specific survival (BCSS), n= 1869[◊]

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***SEC14L1* mRNA**	1.096	0.832	1.445	0.513
Tumour grade	1.356	1.043	1.763	0.023
Nodal staging*	1.788	1.502	2.130	0.000
Tumour size	1.508	1.138	1.999	0.004
Estrogen expression**	0.735	0.541	0.996	0.047
Her2 expression***	1.238	1.047	1.463	0.012

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***SIGLEC1* mRNA**	1.286	0.983	1.682	0.066
Tumour grade	1.335	1.027	1.735	0.031
Nodal staging*	1.791	1.503	2.135	0.000
Tumour size	1.522	1.148	2.017	0.004
Estrogen expression**	0.747	0.553	1.009	0.057
Her2 expression***	1.244	1.052	1.471	0.011

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***SLC3A2* mRNA**	0.733	0.557	0.965	**0.027**
Tumour grade	1.400	1.080	1.817	0.011
Nodal staging*	1.845	1.548	2.199	0.000
Tumour size	1.563	1.179	2.072	0.002
Estrogen expression**	0.672	0.496	0.912	0.011
Her2 expression***	1.265	1.070	1.496	0.006

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***SLC6A9* mRNA**	0.949	0.717	1.255	0.713
Tumour grade	1.373	1.059	1.782	0.017
Nodal staging*	1.795	1.507	2.138	0.000
Tumour size	1.509	1.139	2.000	0.004
Estrogen expression**	0.703	0.511	0.968	0.031
Her2 expression***	1.239	1.049	1.465	0.012

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
[◊] Sum of patients does not equal to 1980 because of missing cases among different categories.

Breast cancer specific survival (BCSS), n= 1869[◊]

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***ZDHHC9* mRNA**	0.998	0.767	1.299	0.990
Tumour grade	1.372	1.055	1.783	0.018
Nodal staging*	1.793	1.505	2.136	0.000
Tumour size	1.510	1.139	2.001	0.004
Estrogen expression**	0.719	0.532	0.970	0.031
Her2 expression***	1.242	1.051	1.467	0.011

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
[◊] Sum of patients does not equal to 1980 because of missing cases among different categories.

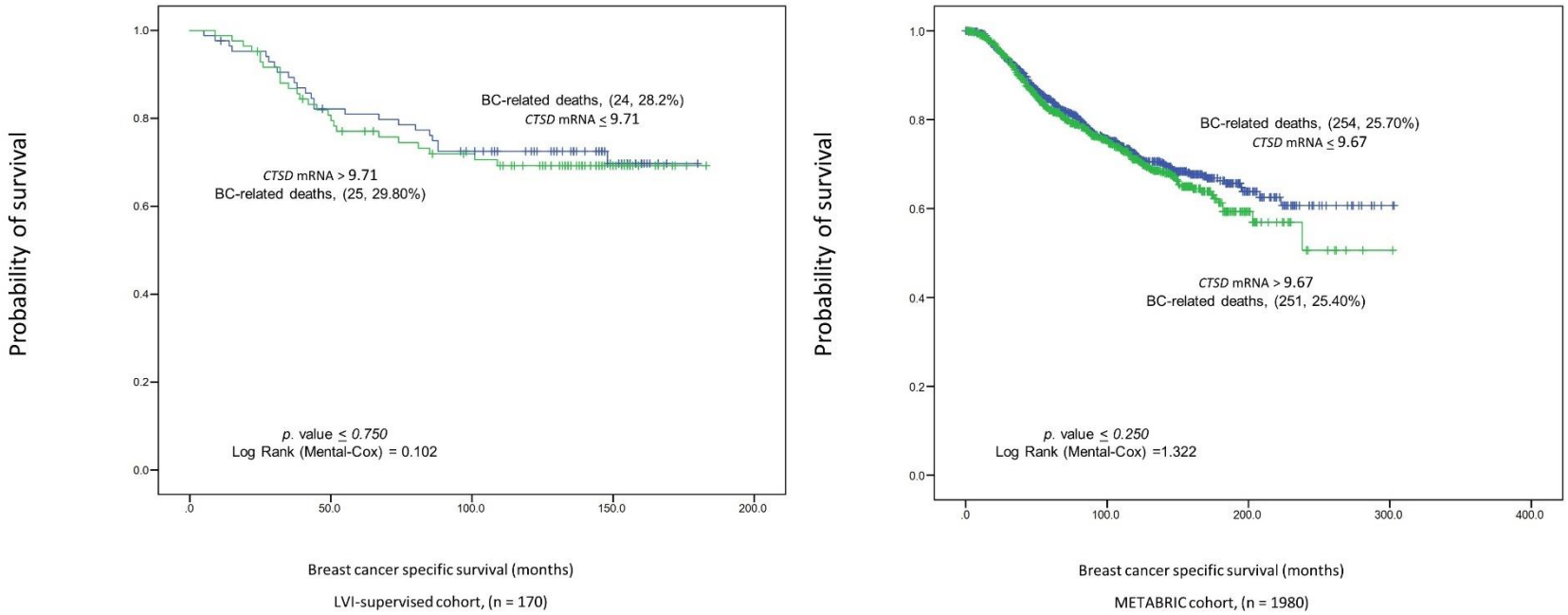


Figure 3. 5: Associations between the altered expression of *CTSD* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

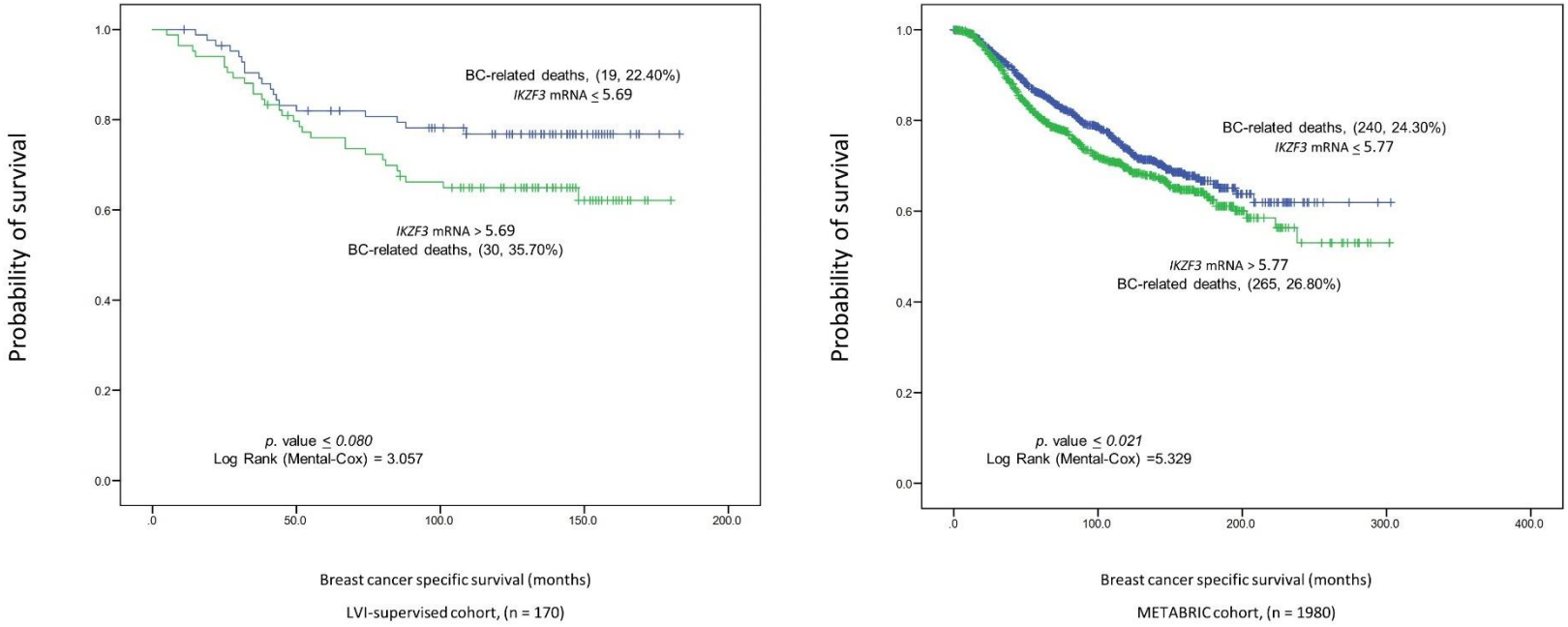


Figure 3. 6: Associations between the altered expression of *IKZF3* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

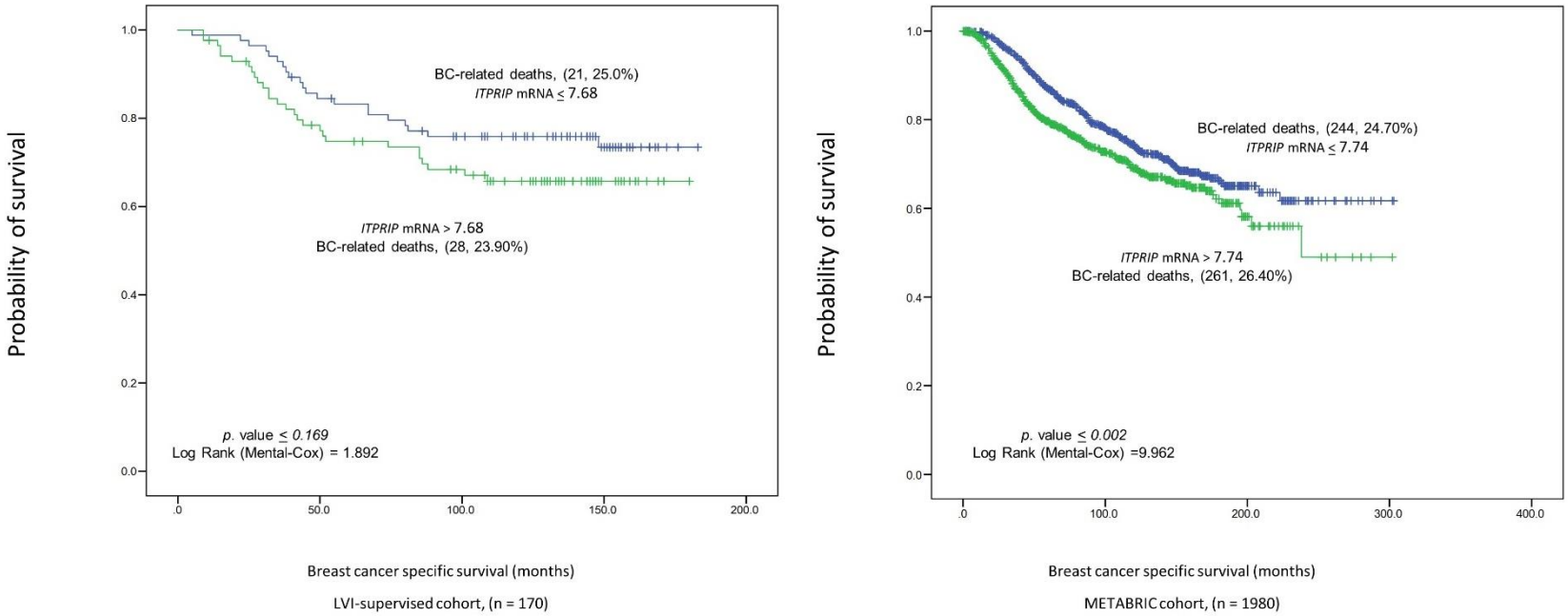


Figure 3. 7: Associations between the altered expression of *ITPRIP* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

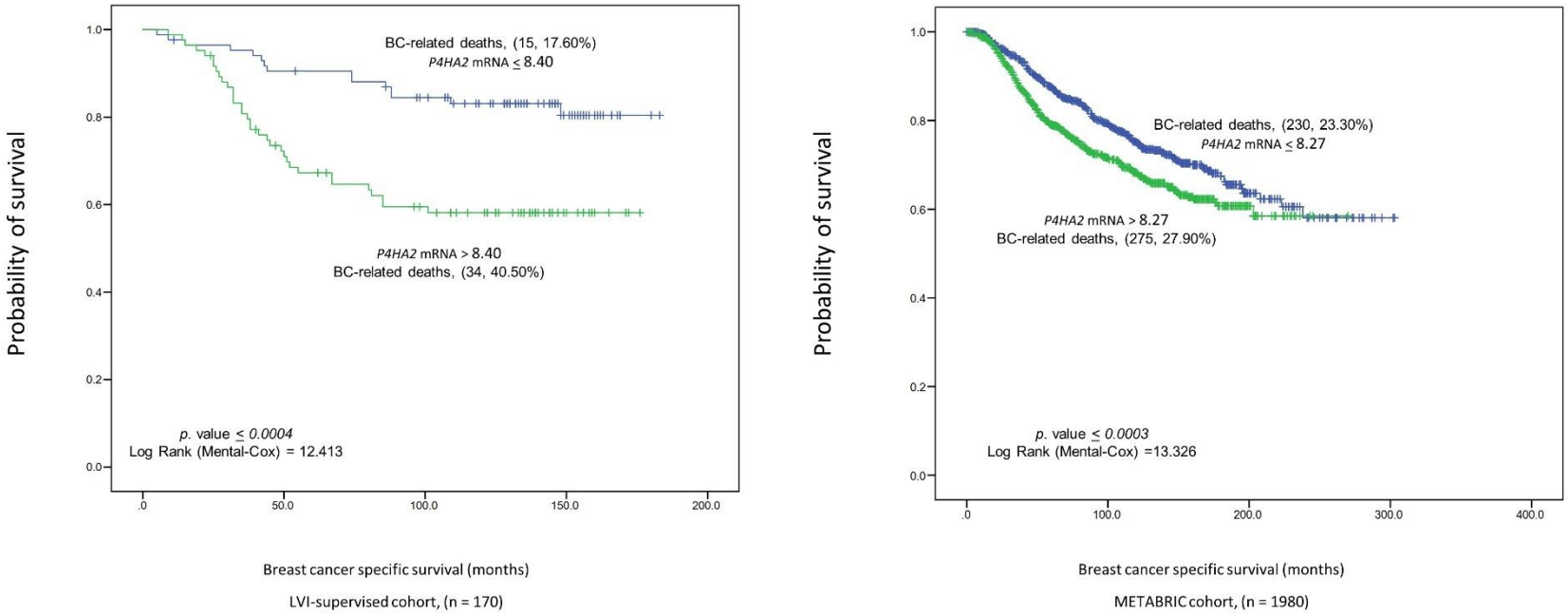


Figure 3. 8: Associations between the altered expression of *P4HA2* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

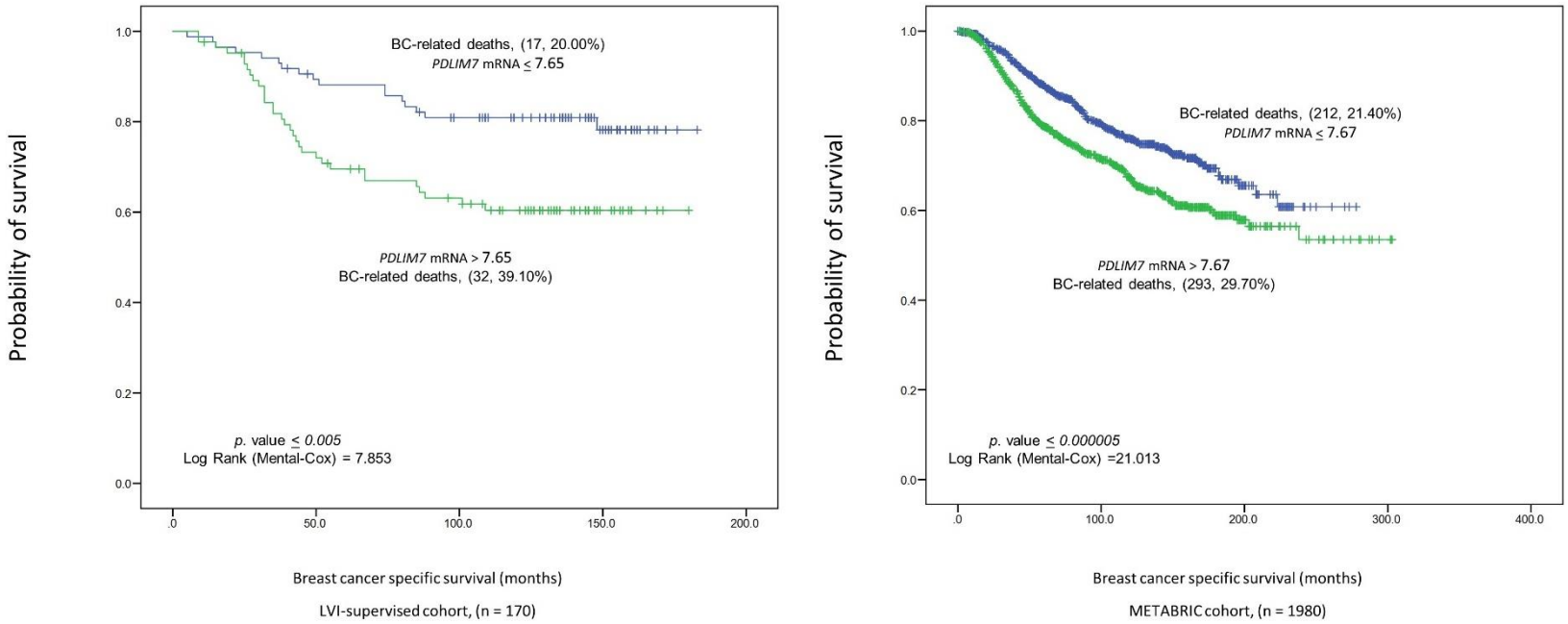


Figure 3. 9: Associations between the altered expression of *PDLIM7* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

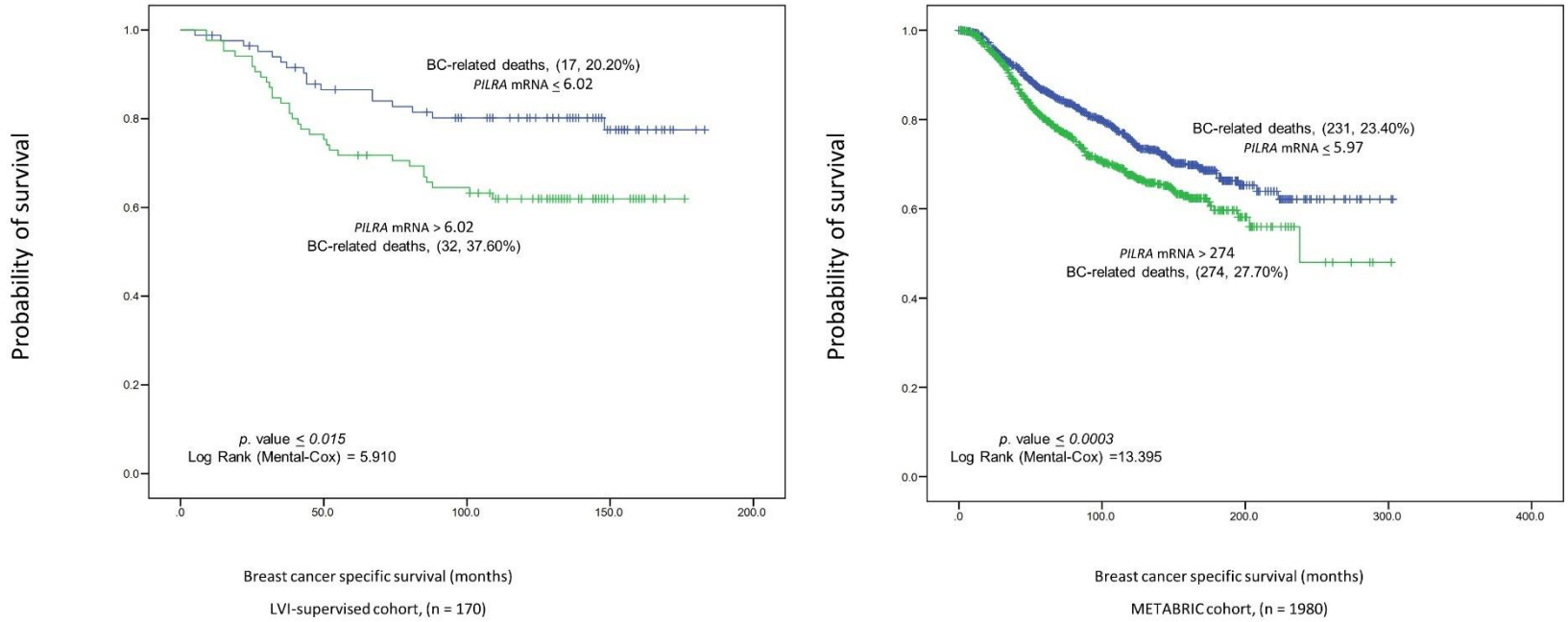


Figure 3. 10: Associations between the altered expression of *PILRA* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

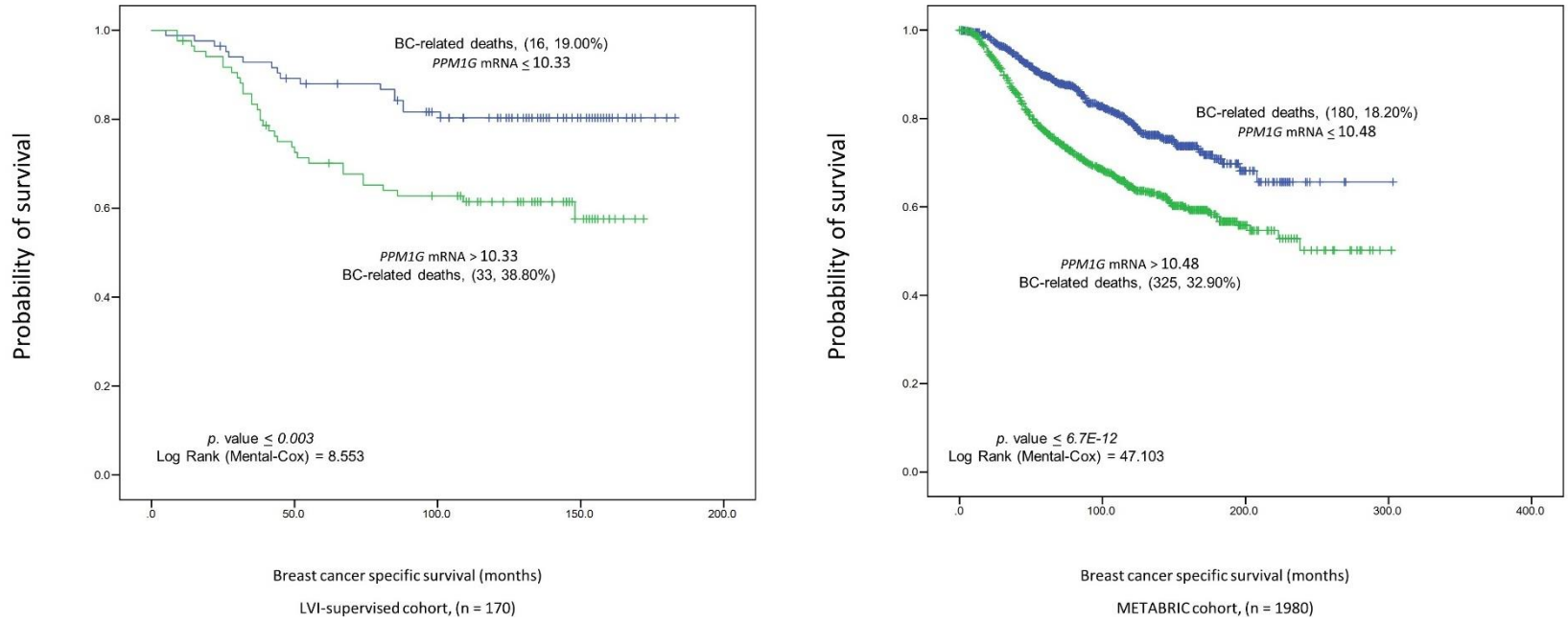


Figure 3. 11: Associations between the altered expression of *PPM1G* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

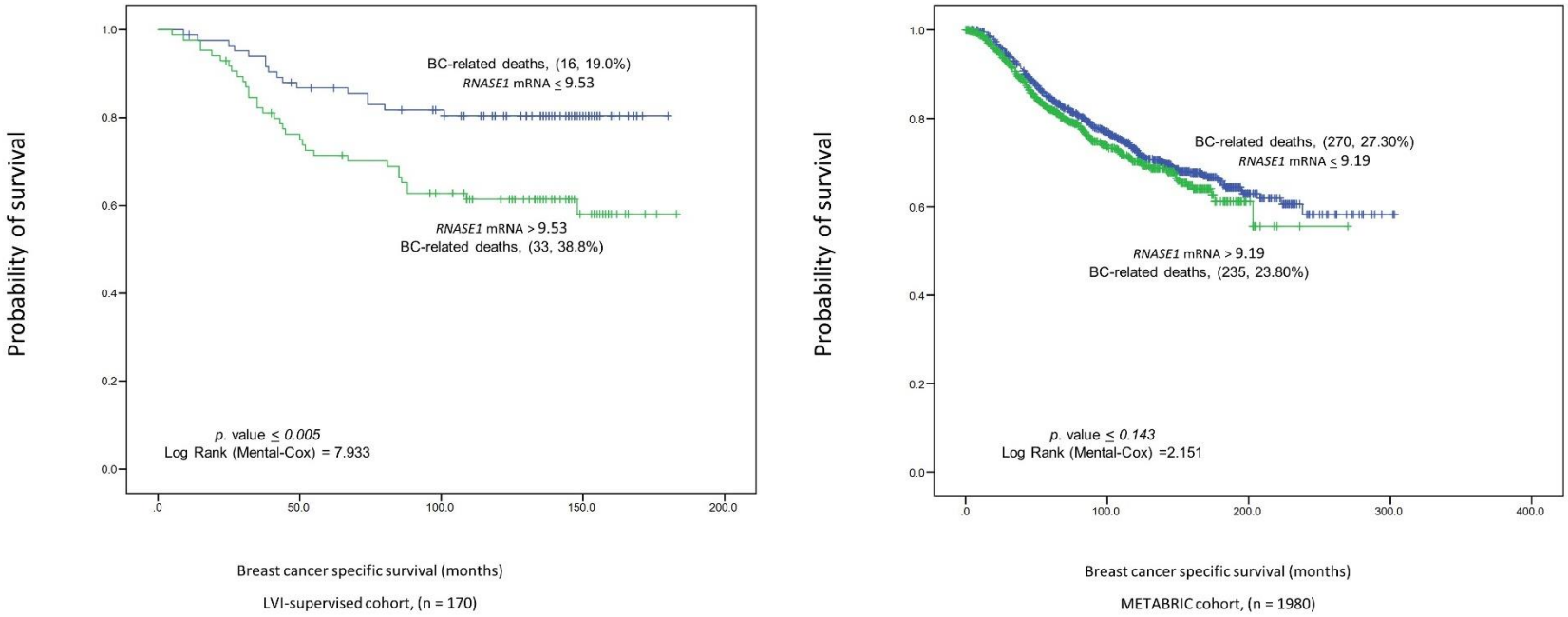


Figure 3. 12: Associations between the altered expression of *RNASE1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

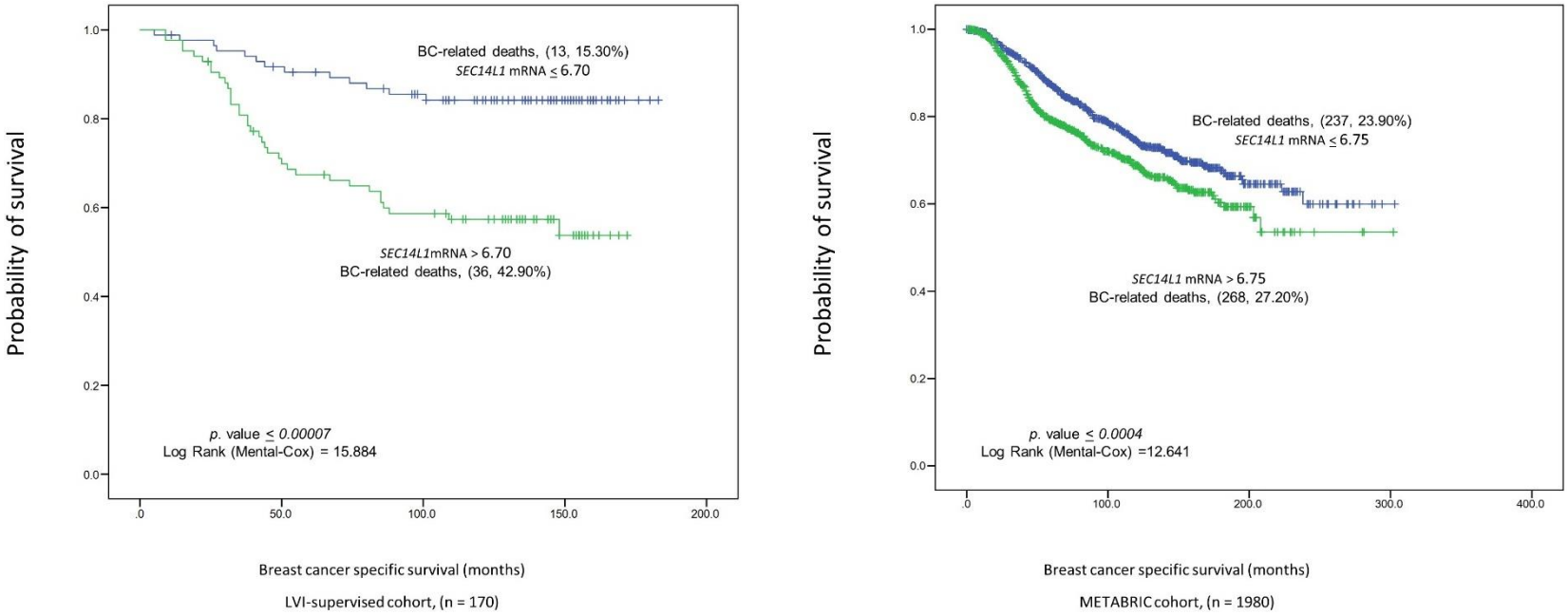


Figure 3. 13: Associations between the altered expression of *SEC14L1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

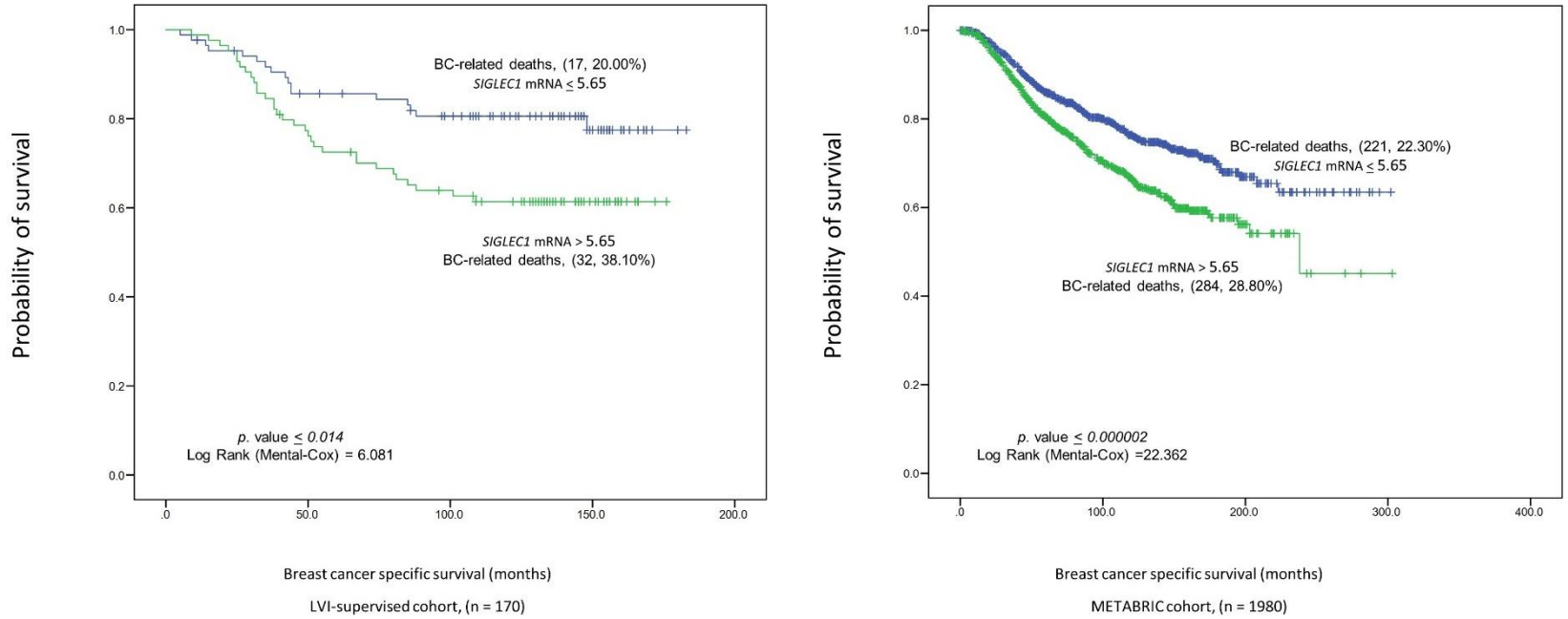


Figure 3. 14: Associations between the altered expression of *SIGLEC1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

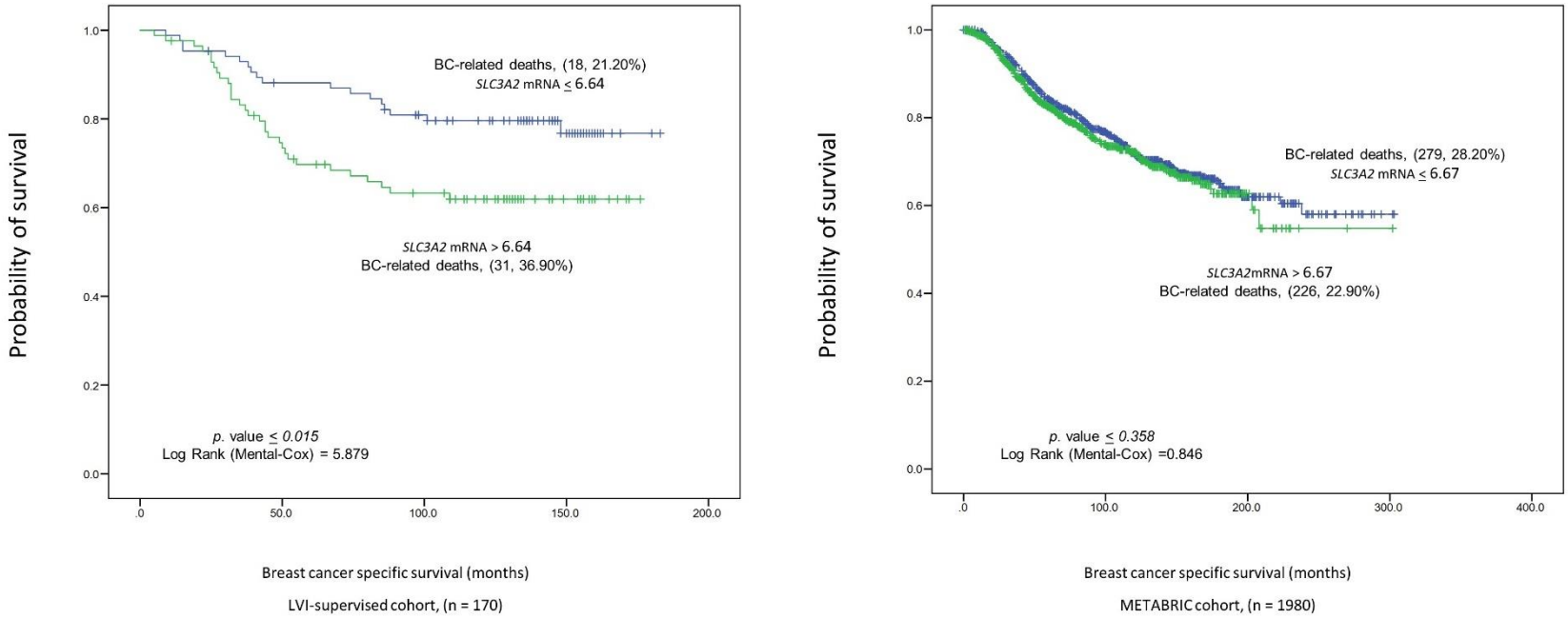


Figure 3. 15: Associations between the altered expression of *SLC3A2* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

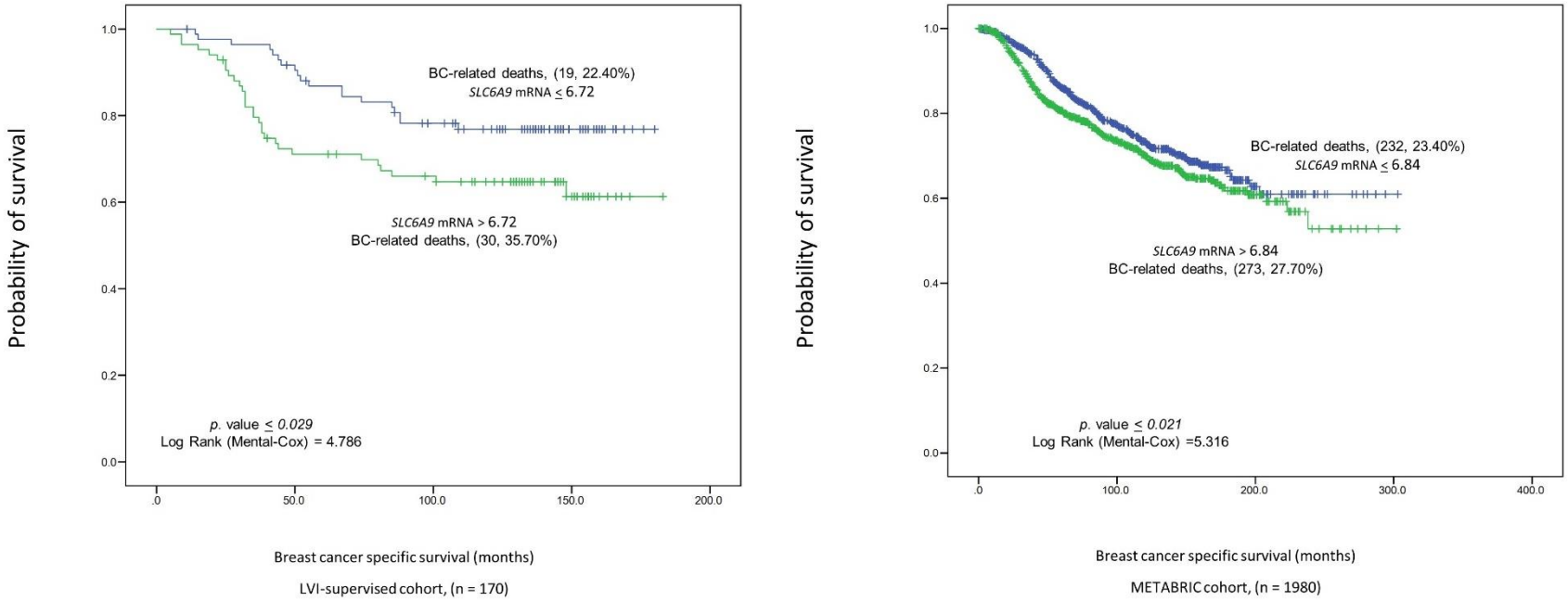


Figure 3. 16: Associations between the altered expression of *SLC6A9* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

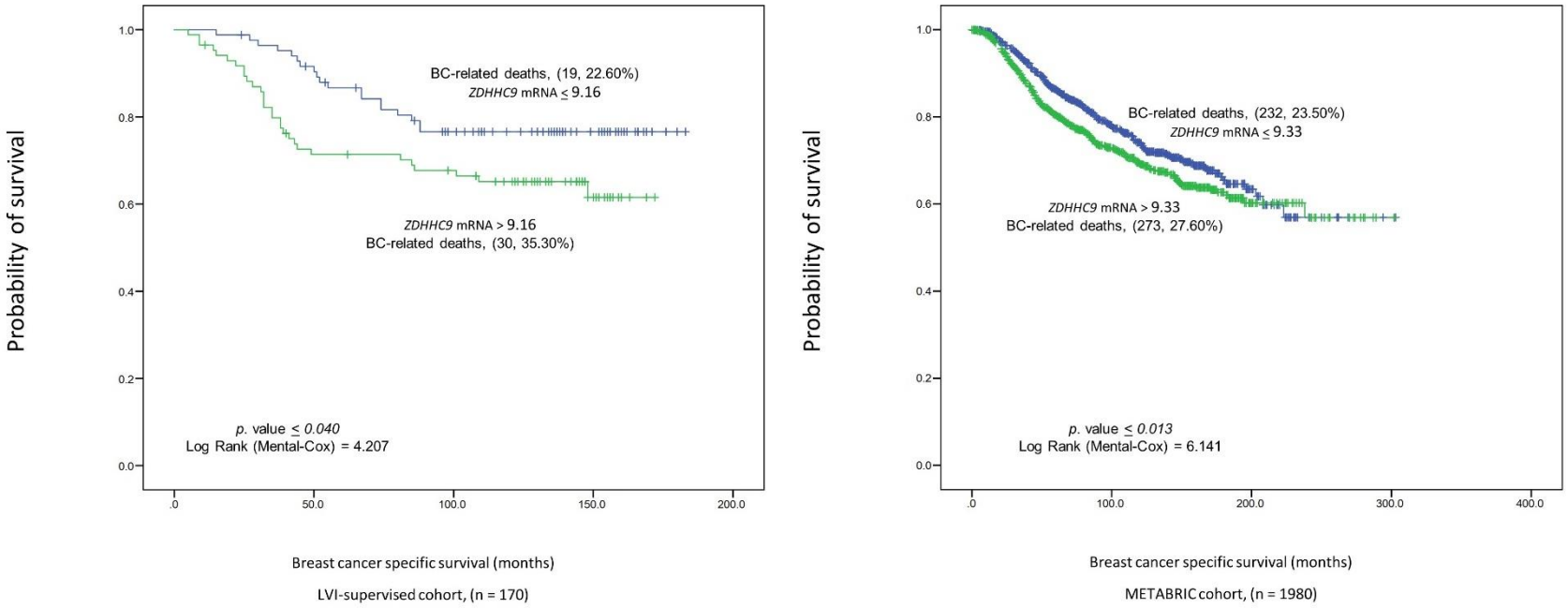


Figure 3. 17: Associations between the altered expression of *ZDHHC9* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

3.3.3.3. *Clinicopathological impact for transcripts in the UP-22/LVI$^-$ cluster*

For all transcripts in the UP-22/LVI$^-$ cluster, detailed clinicopathological and molecular classification results are illustrated in (Table 3.9) for the LVI$^{+/-}$ patients and in (Table 3.10) for the METABRIC patients. Details on multivariate analysis are illustrated in (Table 3.11) for UP-22/LVI$^-$ transcripts. KM plots are depicted for each transcript individually and in alphabetic order, (Figure 3.18) to (Figure 3.39), for easier tracking. All tables and figures are placed in the end of this section.

3.3.3.3.1. **LPAR1: Lysophosphatidic acid receptor 1**

The overexpression of *LPAR1* was strongly associated with the LVI$^-$ patients (p-value $\leq$0.0002), tumour grade-I (p-value $\leq$0.00003), stage-I (p-value $\leq$0.0001), and good NPI (p-value $\leq$0.00003) in the LVI$^{+/-}$ cohort. In the METABRIC cohort, the good prognostic indications for *LPAR1* were maintained. The patients of the *LAPR1*$^+$ expression profiles were categorised into LVI$^-$ (p-value $\leq$0.024), small tumour size (p-value $\leq$0.001), tumour grade-I (p-value $\leq$3.2E-16), and favourable NPI (p-value $\leq$7.2E-8).

The intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *LAPR1*$^+$ invasive BC patients were the normal-like and luminal-A subtypes (p-value $\leq$0.001) and IntClust-4 and IntClust-3 (p-value $\leq$0.0001) in the LVI$^{+/-}$ cohort. The same findings were obtained for the *LAPR1*$^+$ patients in the METABRIC cohort. The positive expression of *LPAR1* was strongly associated with the normal-like and luminal-A subtypes (p-value $\leq$5.7E-32) and IntClust-4 and IntClust-3 (p-value $\leq$4.2E-56).

The KM estimates of the BCSS for patients showed that the overexpression of *LPAR1* had a positive impact on extending the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value $\leq$0.000003) and the METABRIC patients (p-value $\leq$1.4E-7). The expression of *LPAR1* in invasive BC patients had a hazard ratio of 0.659 (p-value $\leq$0.002), suggesting that the downregulated expression of *LPAR1* could be a strong independent predictor of outcome for invasive BC patients.

3.3.3.3.2. **CXCL12: Stromal cell-derived factor 1**

In the LVI$^{+/-}$ cohort, the overexpression of *CXCL12* was strongly associated with the LVI$^-$ patients (p-value $\leq$0.0002), tumour grade-I (p-value $\leq$0.000001), stage-I (p-value $\leq$0.001), and good NPI (p-value $\leq$6.3E-7). In the METABRIC cohort, the good prognostic indications for *CXCL12* were maintained. The patients of the *CXCL12*$^+$ expression profiles were characterised with LVI$^-$ profiles (p-value $\leq$0.015), older age (p-value $\leq$0.004), small tumour size (p-value $\leq$0.00002), tumour grade-I (p-value $\leq$1.0E-14), and favourable NPI (p-value $\leq$6.5E-7).

The intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *CXCL12*$^{+}$ invasive BC patients were the normal-like and luminal-A subtypes (p-value ≤0.00001) and IntClust-7 and IntClust-4 (p-value ≤0.002) in the LVI$^{+/-}$ cohort. The same findings were obtained for the *LAPR1*$^{+}$ patients in the METABRIC cohort. The positive expression of *LPAR1* was strongly associated with the normal-like and luminal-A subtypes (p-value ≤6.4E-45) and IntClust-4 and IntClust-3 (p-value ≤2.7E-39).

The KM estimates of the BCSS for patients showed that the overexpression of *CXCL12* had a positive impact on prolonging the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.006) and the METABRIC patients (p-value ≤0.0002). The expression of *CXCL12* in invasive BC patients had a hazard ratio of 0.787 (p-value ≤0.075), suggesting that the downregulated expression of *CXCL12* could be a dependent predictor of outcome for invasive BC patients.

3.3.3.3.3. C17orf81/ELP: Elongator acetyltransferase complex protein 5

In the LVI$^{+/-}$ cohort, the overexpression of *ELP5* was also strongly correlated with the LVI^{-} patients (p-value ≤0.002), tumour grade-I (p-value ≤0.0003), stage-I (p-value ≤0.001), and good NPI (p-value ≤0.0003). In the METABRIC cohort, the good prognostic indications for *ELP5* were maintained; however, the association with LVI was lost (p-value ≤0.132). The patients of the *ELP5*$^{+}$ expression profiles were characterised by small tumour size (p-value ≤0.006), tumour grade-I (p-value ≤2.2E-12), negative nodal status (p-value ≤0.002), and good NPI (p-value ≤5.6E-8).

In the LVI$^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *ELP5*$^{+}$ invasive BC patients were the luminal-A subtype (p-value ≤0.018) and IntClust-8 and IntClust-7 (p-value ≤0.005). The *ELP5*$^{+}$ patients in the METABRIC cohort fitted into the normal-like and luminal-A subtypes (p-value ≤2.2E-19) and IntClust-8 (p-value ≤8.1E-21).

The KM estimates of the BCSS for patients showed that the overexpression of *ELP5* had a positive impact on extending the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.033) and the METABRIC patients (p-value ≤0.000005). The expression of *ELP* in invasive BC patients had a hazard ratio of 0.888 (p-value ≤0.386), suggesting that the downregulated expression of *ELP5* could be a dependent predictor of outcome for moderately invasive BC patients.

3.3.3.3.4. ECM2: Extracellular matrix protein 2

The overexpression of *ECM2* was correlated with the LVI^{-} patients (p-value ≤0.014), small tumour size (p-value ≤0.029), tumour grade-I (p-value ≤0.00003),

stage-I (p-value ≤0.022), and good NPI (p-value ≤0.0001) in the LVI$^{+/-}$ cohort. In the METABRIC cohort, the patients of the *ECM2*$^{+}$ expression profiles fitted into the groups of small tumour size (p-value ≤0.0001), tumour grade-I (p-value ≤6.2E-15), and favourable NPI (p-value ≤0.000004).

In the LVI$^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *ECM2*$^{+}$ invasive BC patients were the normal-like and luminal-A subtypes (p-value ≤0.00001) and IntClust-3 and IntClust-4 (p-value ≤0.0003). In the METABRIC cohort, the positive expression of *ECM2* was strongly associated with the normal-like and luminal-A subtypes (p-value ≤3.5E-37) and IntClust-3 and IntClust-4 (p-value ≤2.1E-51).

The KM estimates of the BCSS for patients showed that the overexpression of *ECM2* had a positive impact on extending the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.009) and the METABRIC patients (p-value ≤0.002). The expression of *ECM2* in invasive BC patients had a hazard ratio of 0.821 (p-value ≤0.139), suggesting that the downregulated expression of *ECM2* could be a strong independent predictor of outcome for invasive BC patients.

3.3.3.3.5. FBLN1: Fibulin-1

In the LVI$^{+/-}$ cohort, the overexpression of *FBLN1* showed a strong association with the LVI^{-} patients (p-value ≤0.006), tumour grade-I (p-value ≤0.00003), stage-I (p-value ≤0.028), and favourable NPI (p-value ≤0.00008). In the METABRIC cohort, the patients of the *FBLN1*$^{+}$ expression profiles were correlated with the categories of younger patients (p-value ≤0.010), small tumour size (p-value ≤0.00001), tumour grade-I (p-value ≤4.4E-15), and favourable NPI (p-value ≤4.0E-7).

The intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *FBLN1*$^{+}$ invasive BC patients were the normal-like and luminal-A subtypes (p-value ≤8.2E-7) and IntClust-3 (p-value ≤0.00002) in the LVI$^{+/-}$ cohort. Similarly, the positive expression of *FBLN1* was strongly associated with the normal-like and luminal-A subtypes (p-value ≤1.5E-57) and IntClust-3 and IntClust-4 (p-value ≤2.7E-57).

The KM estimates of the BCSS for patients showed that the overexpression of *FBLN1* had a positive impact on prolonging the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.002) and the METABRIC patients (p-value ≤0.001). The expression of *FBLN1* in invasive BC patients had a hazard ratio of 0.914 (p-value ≤0.503), suggesting that the downregulated expression of *FBLN1* could be a dependent predictor of outcome for invasive BC patients.

3.3.3.3.6. APCDD1: APC downregulated 1

In the LVI$^{+/-}$ cohort, the overexpression of *APCDD1* showed a relevant correlation with the LVI^{-} patients (p-value ≤0.031), small tumour size (p-value ≤0.002), and tumour grade-I (p-value ≤0.009). In the METABRIC cohort, the patients of the *APCDD1*$^{+}$ expression profiles were correlated with LVI absence (p-value ≤0.014), younger patients (p-value ≤2.0E-8), small tumour size (p-value ≤0.002), and tumour grade-III (p-value ≤0.038).

The intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *APCDD1*$^{+}$ invasive BC patients were the normal-like followed by the basal subtype (p-value ≤0.019) and IntClust-6 and IntClust-4 (p-value ≤0.002) in the LVI$^{+/-}$ cohort. Similarly, the positive expression of *APCDD1* in the METABRIC cohort was strongly associated with the normal-like and basal subtypes (p-value ≤3.7E-41) and IntClust-10 and IntClust-4 (p-value ≤1.0E-26).

The KM estimates of the BCSS for patients showed that the overexpression of *APCDD1* had no impact on the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.420) and the METABRIC patients (p-value ≤0.504). The expression of *APCDD1* in invasive BC patients had a hazard ratio of 1.062 (p-value ≤0.652), suggesting that the upregulated expression of *APCDD1* has no independent predictive features of outcome for invasive BC patients.

3.3.3.3.7. ANXA5: Annexin A5

In the LVI$^{+/-}$ cohort, the overexpression of *ANXA5* showed a relevant association with the LVI^{-} patients (p-value ≤0.014). However, there were no other significant results for the altered expression of *ANXA5* against the rest of the scrutinised clinicopathological parameters. In the METABRIC cohort, the patients of the *APCDD1*$^{+}$ expression profiles were correlated with LVI absence (p-value ≤0.041), younger patients (p-value ≤0.00002), and small tumour size (p-value ≤0.011).

There were no associations between the altered expression of *ANXA5* and the intrinsic subtypes and molecular clusters of invasive BC in the LVI$^{+/-}$ cohort. The intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *ANXA5*$^{+}$ invasive BC patients were the basal followed by the normal-like subtypes (p-value ≤1.7E-20) and IntClust-10, IntClust-4, and IntClust-3 (p-value ≤6.6E-30).

The KM estimates of the BCSS for patients showed that the overexpression of *ANXA5* had no impact on the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.226) and the METABRIC patients (p-value ≤0.540). The expression of *ANXA5* in

invasive BC patients had a hazard ratio of 0.986 (p-value ≤0.916), suggesting that the expression of *ANXA5* has no independent predictive features of outcome for invasive BC patients.

3.3.3.3.8. INKA1: PAK4-inhibitor INKA1

In the LVI$^{+/-}$ cohort, the overexpression of *INKA1* was strongly associated with the LVI^{-} patients (p-value ≤0.002), tumour grade-I (p-value ≤0.0003), stage-I (p-value ≤0.049), and good NPI (p-value ≤0.0003). In the METABRIC cohort, the patients of the *INKA1*$^{+}$ expression profiles fitted into the categories of small tumour size (p-value ≤0.00001), tumour grade-I (p-value ≤2.4E-16), and favourable NPI (p-value ≤2.1E-10).

In the LVI$^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *INKA1*$^{+}$ invasive BC patients were the normal-like and luminal-A subtypes (p-value ≤0.000002) and IntClust-7, IntClust-4, and IntClust-3 (p-value ≤2.8E-7). In the METABRIC patients, the positive expression of *INKA1* was strongly associated with the normal-like and luminal-A subtypes (p-value ≤1.0E-49) and IntClust-4 and IntClust-3 (p-value ≤5.8E-52).

The KM estimates of the BCSS for patients showed that the overexpression of *INKA1* had a positive impact on prolonging the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.003) and the METABRIC patients (p-value ≤0.001). The expression of *INKA1* in invasive BC patients had a hazard ratio of 0.858 (p-value ≤0.253), suggesting that the downregulated expression of *INKA1* has no independent predictive features of outcome for invasive BC patients.

3.3.3.3.9. TMEM98: Transmembrane protein 98

In the LVI$^{+/-}$ cohort, the overexpression of *TMEM98* showed a significant correlation with the LVI^{-} patients (p-value ≤0.014) and tumour grade-I (p-value ≤0.044). In the METABRIC cohort, the patients of the *TMEM98*$^{+}$ expression profiles fitted into the categories of younger patients (p-value ≤0.025), tumour grade-I (p-value ≤2.2E-9), and favourable NPI (p-value ≤0.00001).

In the LVI$^{+/-}$ cohort, the altered expression of *TMEM98* had no relationships with the intrinsic subtypes and the molecular clusters of invasive BC. In the METABRIC patients, the positive expression of *TMEM98* was strongly associated with the normal-like and luminal-A subtypes (p-value ≤4.2E-22) and IntClust-8 and IntClust-3 (p-value ≤4.1E-17).

The KM estimates of the BCSS for patients showed that the overexpression of *TMEM98* had a positive impact on prolonging the BCSS of the invasive BC patients

in the LVI$^{+/-}$ (p-value ≤0.023) and the METABRIC patients (p-value ≤0.000007). The expression of *TMEM98* in invasive BC patients had a hazard ratio of 0.698 (p-value ≤0.009), suggesting that the downregulated expression of *TMEM98* could be a strong independent predictor of outcome for invasive BC patients.

3.3.3.3.10. FEZ1: Fasciculation and elongation protein zeta-1

In the LVI$^{+/-}$ cohort, the overexpression of *FEZ1* was significantly associated with the LVI^{-} patients (p-value ≤0.014). There were no other significant results for the altered expression of *FEZ1* against the remaining clinicopathological factors. In the METABRIC cohort, the patients of *FEZ1*$^{+}$ expression profiles were associated with small tumour size (p-value ≤0.00001), tumour grade-I (p-value ≤0.000002), and favourable NPI (p-value ≤0.009).

In the LVI$^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *FEZ1*$^{+}$ invasive BC patients were the normal-like and luminal-A subtypes (p-value ≤0.0002) and IntClust-3 and IntClust-4 (p-value ≤0.007). In the METABRIC patients, the positive expression of *FEZ1* was associated with the normal-like and luminal-A subtypes (p-value ≤1.3E-35) and IntClust-4 (p-value ≤3.7E-45).

The KM estimates of the BCSS for patients showed that the aberrant expression of *FEZ1* had no impact on the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.151) and the METABRIC patients (p-value ≤0.236). The expression of *FEZ1* in invasive BC patients had a hazard ratio of 1.147 (p-value ≤0.306), suggesting that the altered expression of *FEZ1* has no independent predictive features of outcome for invasive BC patients.

3.3.3.3.11. THAP4: THAP domain-containing protein 4

In the LVI$^{+/-}$ cohort, the overexpression of *THAP4* was relevantly associated with the LVI^{-} patients (p-value ≤0.031), tumour grade-I (p-value ≤0.00004), stage-I (p-value ≤0.007), and favourable NPI (p-value ≤0.0001). In the METABRIC cohort, the patients of the *THAP4*$^{-}$ expression profiles were associated with younger patients (p-value ≤0.00001), tumour grade-I (p-value ≤2.0E-22), and favourable NPI (p-value ≤2.5E-13).

In the LVI$^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *THAP4*$^{+}$ invasive BC patients were the luminal-A and luminal-B subtypes (p-value ≤0.00002) and IntClust-6 and IntClust-7 (p-value ≤0.00002). In the METABRIC patients, the positive expression of *THAP4* was associated with the luminal-A subtype (p-value ≤1.3E-44) and IntClust-8 (p-value ≤1.2E-46).

The KM estimates of the BCSS for patients showed that the overexpression of *THAP4* had a strong impact on extending the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.001) and the METABRIC patients (p-value ≤0.002). The expression of *THAP4* in invasive BC patients had a hazard ratio of 0.872 (p-value ≤0.318), suggesting that the altered expression of *THAP4* has no independent predictive features of outcome for invasive BC patients.

3.3.3.3.12. TNFRSF19: Tumour necrosis factor receptor superfamily member 19

In the LVI$^{+/-}$ cohort, the overexpression of *TNFRSF19* had no associations with the LVI status (p-value ≤0.219). The positive expression of *TNFRSF19* was correlated with tumour grade-I (p-value ≤0.002), small tumour size (p-value ≤0.029), and good NPI (p-value ≤0.042). In the METABRIC cohort, the patients of the *TNFRSF19*$^{+}$ expression profiles were fitted into the categories of small tumour size (p-value ≤0.001), tumour grade-I (p-value ≤2.1E-12), nodal stage-I (p-value ≤0.003), and favourable NPI (p-value ≤6.4E-11).

In the LVI$^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of *TNFRS19F*$^{+}$ invasive BC patients were the normal-like and luminal-A subtypes (p-value ≤0.0004) and IntClust-3 (p-value ≤0.002). In the METABRIC patients, the positive expression of *TNFRSF19* was strongly associated with the luminal-A subtype (p-value ≤4.6E-28) and IntClust-3 (p-value ≤1.4E-20).

The KM estimates of the BCSS for patients showed that the overexpression of *TNFRSF19* had a positive impact on prolonging the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.022) and the METABRIC patients (p-value ≤0.028). The expression of *TNFRSF19* in invasive BC patients had a hazard ratio of 1.186 (p-value ≤0.200), suggesting that the upregulated expression of *TNFRSF19* could be a dependent predictor of outcome for invasive BC patients.

3.3.3.3.13. NBR1: Next to BRCA1 gene 1 protein

In the LVI$^{+/-}$ cohort, the overexpression of *NBR1* was only associated with tumour grade-I (p-value ≤0.006) and favourable NPI (p-value ≤0.011). In the METABRIC cohort, the patients of the *NBR1*$^{-}$ expression profiles were associated with the categories of patients' age (p-value ≤0.008), tumour grade-I (p-value ≤3.6E-8), and favourable NPI (p-value ≤0.00001).

In the LVI$^{+/-}$ cohort, the altered expression of *NBR1* had no associations with the intrinsic subtypes or the molecular clusters of invasive BC. In the METABRIC patients, the positive expression of *NBR1* was strongly associated with the normal-

like and luminal-A subtypes (p-value ≤1.8E-22) and IntClust-2 and IntClust-7 (p-value ≤4.4E-16).

The KM estimates of the BCSS for patients showed that the overexpression of *NBR1* had a positive impact on elongating the BCSS of the invasive BC patients in the $LVI^{+/-}$ (p-value ≤0.024) and the METABRIC patients (p-value ≤0.001). The expression of *NBR1* in invasive BC patients had a hazard ratio of 0.743 (p-value ≤0.032), suggesting that the downregulated expression of *NBR1* could be an independent predictor of outcome for invasive BC patients.

3.3.3.3.14. WISP2: WNT1-inducible-signalling pathway protein 2

In the $LVI^{+/-}$ cohort, the overexpression of *WISP2* was relevantly associated with LVI^{-} patients (p-value ≤0.014), tumour grade-I (p-value ≤0.001), and favourable NPI (p-value ≤0.001). In the METABRIC cohort, the patients of the $WISP2^{+}$ expression profiles were associated with small tumour size (p-value ≤0.043) and tumour grade-I (p-value ≤0.000003).

In the $LVI^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of $WISP2^{+}$ invasive BC patients were the luminal-A subtype (p-value ≤0.007) and IntClust-8, IntClust-2, and IntClust-7 (p-value ≤0.025). In the METABRIC patients, the positive expression of *WISP2* was associated with the luminal-A and normal-like subtypes (p-value ≤2.6E-10) and IntClust-4 (p-value ≤4.8E-13).

The KM estimates of the BCSS for patients showed that the expression of *WISP2* had no impact on the BCSS of the invasive BC patients in the $LVI^{+/-}$ (p-value ≤0.772) and the METABRIC patients (p-value ≤0.095). The expression of *WISP2* in invasive BC patients had a hazard ratio of 0.989 (p-value ≤0.931), suggesting that the altered expression of *WISP2* has no independent predictive features of outcome for invasive BC patients.

3.3.3.3.15. CNRIP1: CB1 cannabinoid receptor-interacting protein 1

In the $LVI^{+/-}$ cohort, the overexpression of *CNRIP1* was strongly associated with LVI^{-} patients (p-value ≤0.006), small tumour size (p-value ≤0.001), tumour grade-I (p-value ≤0.00001), and favourable NPI (p-value ≤0.00003). In the METABRIC cohort, the patients of the $CNRIP1^{+}$ expression profiles were associated with small tumour size (p-value ≤0.00003), tumour grade-I (p-value ≤2.3E-7), and good NPI (p-value ≤0.004).

In the $LVI^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of $CNRIP1^{+}$ invasive BC patients were the normal-

like subtype (p-value ≤0.0001) and IntClust-4 and IntClust-3 (p-value ≤0.010). In the METABRIC patients, the positive expression of *CNRIP1* was associated with the normal-like subtype (p-value ≤3.9E-30) and IntClust-4 and IntClust-3 (p-value ≤9.4E-44).

The KM estimates of the BCSS for patients showed that the overexpression of *CNRIP1* had a positive impact on extending the BCSS of invasive BC patients in the $LVI^{+/-}$ (p-value ≤0.015), but not the METABRIC patients (p-value ≤0.135). The expression of *CNRIP1* in invasive BC patients had a hazard ratio of 1.063 (p-value ≤0.645), suggesting that the altered expression of *CNRIP1* has dependent predictive features of outcome for invasive BC patients.

3.3.3.3.16. ITSN1: Intersectin-1

In the $LVI^{+/-}$ cohort, the overexpression of *ITSN1* was statistically associated with LVI^{-} patients (p-value ≤0.031) and favourable NPI (p-value ≤0.028). In the METABRIC cohort, the patients of the $ITSN1^{-}$ expression profiles were associated with definite LVI occurrences (p-value ≤0.034) and tumour grade-III (p-value ≤0.003).

In the $LVI^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of $ITSN1^{+}$ invasive BC patients were the normal-like subtype (p-value ≤0.014) and IntClust-2, IntClust-3, and IntClust-7 (p-value ≤0.033). In the METABRIC patients, the positive expression of *ITSN1* was associated with the normal-like subtype (p-value ≤1.0E-49) and IntClust-4 (p-value ≤5.8E-52).

The KM estimates of the BCSS for patients showed that the overexpression of *ITSN1* had a trend toward significance on extending the BCSS of invasive BC patients in the $LVI^{+/-}$ (p-value ≤0.063), but not the METABRIC patients (p-value ≤0.979). The expression of *ITSN1* in invasive BC patients had a hazard ratio of 1.155 (p-value ≤0.276), suggesting that the altered expression of *ITSN1* has dependent predictive features of outcome for invasive BC patients.

3.3.3.3.17. C1S: Complement C1s subcomponent

In the $LVI^{+/-}$ cohort, the overexpression of *C1S* was only associated with the LVI^{-} patients (p-value ≤0.014). In the METABRIC cohort, the patients of the $C1S^{+}$ expression profiles were associated with younger patients (p-value ≤2.7E-9) and small tumour size (p-value ≤0.00004).

In the $LVI^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of $C1S^{+}$ invasive BC patients were the normal-like

subtype (p-value ≤0.0002) and IntClust-4 and IntClust-3 (p-value ≤0.015). In the METABRIC patients, the positive expression of *C1S* was associated with the normal-like subtype (p-value ≤1.1E-44) and IntClust-4 (p-value ≤2.1E-47).

The KM estimates of the BCSS for patients showed that the overexpression of *C1S* had a significant impact on prolonging the BCSS of invasive BC patients in the $LVI^{+/-}$ (p-value ≤0.023), but not the METABRIC patients (p-value ≤0.667). The expression of *C1S* in invasive BC patients had a hazard ratio of 1.045 (p-value ≤0.746), suggesting that the altered expression of *C1S* has dependent predictive features of outcome for invasive BC patients.

3.3.3.3.18. TNC: Tenascin

In the $LVI^{+/-}$ cohort, the overexpression of *TNC* showed a trend toward significance with the LVI^{-} patients (p-value ≤0.065) and a significant association with tumour grade-I (p-value ≤0.003). In the METABRIC cohort, the patients of the TNC^{+} expression profiles were associated with younger patients (p-value ≤1.4E-9), small tumour size (p-value ≤0.001), and negative nodal involvement (p-value ≤0.048).

In the $LVI^{+/-}$ cohort, the intrinsic subtypes and the molecular clusters that exhibited the highest percentage of TNC^{+} invasive BC patients were the normal-like subtype (p-value ≤0.005) and IntClust-6, IntClust-4, and IntClust-3 (p-value ≤0.038). In the METABRIC patients, the positive expression of *TNC* was associated with the normal-like subtype (p-value ≤6.3E-28) and IntClust-5 (p-value ≤1.7E-25).

The KM estimates of the BCSS for patients showed that the overexpression of *TNC* had no significant impact on the BCSS of invasive BC patients in the $LVI^{+/-}$ (p-value ≤0.753) and the METABRIC patients (p-value ≤0.832). The expression of *TNC* in invasive BC patients had a hazard ratio of 0.899 (p-value ≤0.427), suggesting that the altered expression of *TNC* has no predictive features of outcome for invasive BC patients.

3.3.3.3.19. PRRX1: Paired mesoderm homeobox protein 1

In the $LVI^{+/-}$ cohort, the overexpression of *PRRX1* showed a trend toward significance with the LVI^{-} patients (p-value ≤0.065) and a significant correlation with small tumour size (p-value ≤0.013). In the METABRIC cohort, the patients of the $PRRX1^{+}$ expression profiles were associated with younger patients (p-value ≤0.008), small tumour size (p-value ≤0.00002), and tumour grade-I (p-value ≤0.005).

In the LVI$^{+/-}$ cohort, the altered expression of the *PRRX1* was not associated with any intrinsic subtypes or molecular clusters. In the METABRIC patients, the positive expression of *PRRX1* was associated with the normal-like subtype (p-value ≤1.1E-12) and IntClust-4 and IntClust-3 (p-value ≤4.4E-24).

The KM estimates of the BCSS for patients showed that the overexpression of *PRRX1* had no significant impact on the BCSS of invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.225) and the METABRIC patients (p-value ≤0.636). The expression of *PRRX1* in invasive BC patients had a hazard ratio of 1.034 (p-value ≤0.802), suggesting that the altered expression of *PRRX1* has no predictive features of outcome for invasive BC patients.

3.3.3.3.20. STAT3: Signal transducer and activator of transcription 3

In the LVI$^{+/-}$ cohort, the overexpression of *STAT3* showed significant correlations with small tumour size (p-value ≤0.029), tumour grade-I (p-value ≤0.00001), negative nodal involvement (p-value ≤0.028), and favourable NPI (p-value ≤0.0001). In the METABRIC cohort, the patients of the *STAT3*$^{-}$ expression profiles were associated with younger patients (p-value ≤0.044). The patients of the *STAT3*$^{+}$ expression profiles were associated with small tumour size (p-value ≤0.014), tumour grade-I (p-value ≤9.3E-9), negative nodal involvement (p-value ≤ .00002), and good NPI (p-value ≤1.0E-10).

In the LVI$^{+/-}$ cohort, the altered expression of *STAT3* exhibited relationships with the intrinsic subtypes of normal-like and luminal-A (p-value ≤0.00001) and the molecular clusters of IntClust-3 and IntClust-7 (p-value ≤0.0003). In the METABRIC cohort, the positive expression of *STAT3* was mainly associated with the luminal-A subtype (p-value ≤1.4E-29) and IntClust-3 and IntClust-8 (p-value ≤3.6E-24).

The KM estimates of the BCSS for patients showed that the overexpression of *STAT3* had a positive impact on prolonging the BCSS of invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.004) and the METABRIC patients (p-value ≤0.000002). The expression of *STAT3* in invasive BC patients had a hazard ratio of 0.797 (p-value ≤0.095), suggesting that the downregulated expression of *STAT3* could be a dependent predictor of outcome for invasive BC patients.

3.3.3.3.21. TSC1: Hamartin

In the LVI$^{+/-}$ cohort, the overexpression of *TSC1* showed relevant correlation with tumour grade-I (p-value ≤0.0004) and favourable NPI (p-value ≤0.0004). In the METABRIC cohort, the patients of the *TSC1*$^{+}$ expression profiles were associated

with small tumour size (p-value ≤0.001), tumour grade-I (p-value ≤1.2E-13), negative nodal involvement (p-value ≤0.0001), and good NPI (p-value ≤1.4E-12).

In the LVI$^{+/-}$ cohort, the altered expression of *TSC1* exhibited relationships with the intrinsic subtypes of luminal-A and normal-like (p-value ≤0.0001) and the molecular clusters of IntClust-4, IntClust-3, and IntClust-7 (p-value ≤0.0004). In the METABRIC cohort, the positive expression of *TSC1* was associated with the luminal-A and normal-like subtypes (p-value ≤1.4E-29) and IntClust-3 and IntClust-8 (p-value ≤9.8E-30).

The KM estimates of the BCSS for patients showed that the overexpression of *TSC1* had a positive impact on extending the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.008), but not the METABRIC patients (p-value ≤0.070). The expression of *TSC1* in invasive BC patients had a hazard ratio of 0.914 (p-value ≤0.509), suggesting that the downregulated expression of *TSC1* could be a dependent predictor of outcome for invasive BC patients.

3.3.3.3.22. TWIST2: Twist-related protein 2

In the LVI$^{+/-}$ cohort, the overexpression of *TWIST2* showed a single relevant correlation with tumour grade-I (p-value ≤0.027). In the METABRIC cohort, the patients of the *TWIST2*$^{+}$ expression profiles were associated with younger patients (p-value ≤0.00007), small tumour size (p-value ≤0.001), and tumour grade-I (p-value ≤0.006).

In the LVI$^{+/-}$ cohort, the altered expression of *TWIST2* exhibited relationships with the intrinsic subtypes of normal-like and luminal-A (p-value ≤0.006), but there were no associations between *TWIST2* and the molecular clusters. In the METABRIC cohort, the positive expression of *TWIST2* was majorly associated with the normal-like subtype (p-value ≤1.6E-35) and IntClust-4 and IntClust-3 (p-value ≤2.8E-25).

The KM estimates of the BCSS for patients showed that the altered expression of *TWIST2* had no impact on the BCSS of the invasive BC patients in the LVI$^{+/-}$ (p-value ≤0.290) and the METABRIC patients (p-value ≤0.064). The expression of *TWIST2* in invasive BC patients had a hazard ratio of 0.835 (p-value ≤0.174), suggesting that the downregulated expression of *TSC1* has no predictive features of outcome for invasive BC patients.

Table 3. 9: Clinicopathological significance of upregulated genes with LVI^{-} patients in the LVI-supervised cohort (n=170). Significant values are in bold.

	ANXA5			*APCDD1*			*C17orf81/ELP5*			*C1S*		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN^{-}	32 (40.0)	48 (60.0)	6.044 (**0.014**)	33 (41.3)	47 (58.8)	4.628 (**0.031**)	30 (37.5)	50 (62.5)	9.444 (**0.002**)	32 (40.0)	48 (60.0)	6.044 (**0.014**)
Positive/LN$^{+/-}$	53 (58.9)	37 (41.1)		52 (57.8)	38 (42.2)		55 (61.1)	35 (38.9)		53 (58.9)	37 (41.1)	
Age (years)												
≤ 50	31 (47.0)	35 (53.0)	0.396 (0.529)	36 (54.5)	30 (45.5)	0.892 (0.345)	31 (47.0)	35 (53.0)	0.396 (0.529)	33 (50.0)	33 (50.0)	0.000 (1.000)
> 50	54 (51.9)	50 (48.1)		49 (47.1)	55 (52.9)		54 (51.9)	50 (48.1)		52 (50.0)	52 (50.0)	
Tumour size												
≤ 2.0cm	46 (46.0)	54 (54.0)	1.554 (0.213)	40 (40.0)	60 (60.0)	9.714 (**0.002**)	44 (44.0)	56 (56.0)	3.497 (0.061)	47 (47.0)	53 (53.0)	0.874 (0.350)
> 2.0cm	39 (55.7)	31 (44.3)		45 (64.3)	25 (35.7)		41 (58.6)	29 (41.4)		38 (54.3)	32 (45.7)	
Tumour Grade												
Grade-I	10 (38.5)	16 (61.5)		8 (30.8)	18 (69.2)		4 (15.4)	22 (84.6)		13 (50.0)	13 (50.0)	
Grade-II	26 (47.3)	29 (52.7)	2.679 (0.262)	36 (65.5)	19 (34.5)	9.504 (**0.009**)	24 (43.6)	31 (56.4)	21.029 (**0.00003**)	23 (41.8)	32 (58.2)	2.603 (0.272)
Grade-III	49 (55.7)	39 (44.3)		41 (46.6)	47 (53.4)		57 (64.8)	31 (35.2)		49 (55.7)	39 (44.3)	
Nodal involvement, (NPI staging system)												
Stage-I	49 (46.7)	56 (53.3)		48 (45.7)	57 (54.3)		41 (39.0)	64 (61.0)		47 (44.8)	58 (55.2)	
Stage-II	28 (57.1)	21 (42.9)	1.467 (0480)	26 (53.1)	23 (46.9)	3.205 (0.201)	32 (65.3)	17 (34.7)	13.630 (**0.001**)	28 (57.1)	21 (42.9)	3.152 (0.207)
Stage-III	8 (50.0)	8 (50.0)		11 (68.8)	5 (31.3)		12 (75.0)	4 (25.0)		10 (62.5)	6 (37.5)	
Nottingham Prognostic Index (NPI)												
good	27 (39.1)	42 (60.9)		35 (50.7)	34 (49.3)		22 (31.9)	47 (68.1)		31 (44.9)	38 (55.1)	
moderate	51 (58.0)	37 (42.0)	5.565 (0.062)	41 (46.6)	47 (53.4)	2.347 (0.309)	53 (60.2)	35 (39.8)	16.509 (**0.0003**)	45 (51.1)	43 (48.9)	2.679 (0.262)
poor	7 (53.8)	6 (46.2)		9 (69.2)	4 (30.8)		10 (76.9)	3 (23.1)		9 (69.2)	4 (30.8)	

	ANXA5			*APCDD1*			*C17orf81/ELP5*			*C1S*		
	-	+	χ2	-	+	χ2	-	+	χ2	-	+	χ2
	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)	n (%)	n (%)	(*p*-value≤)
Prediction analysis of microarray 50, PAM50												
Normal-like	4 (33.3)	8 (66.7)		2 (16.7)	10 (83.3)		6 (50.0)	6 (50.0)		1 (8.3)	11 (91.7)	
Luminal-A	30 (46.9)	34 (53.1)		32 (50.0)	32 (50.0)		24 (37.5)	40 (62.5)		32 (50.0)	32 (50.0)	
Luminal-B	29 (64.4)	16 (35.6)	8.796 (0.066)	30 (66.7)	15 (33.3)	11.733 (**0.019**)	23 (51.1)	22 (48.9)	11.879 (**0.018**)	32 (71.1)	13 (28.9)	22.327 (**0.0002**)
Her2-enriched	9 (64.3)	5 (35.7)		7 (50.0)	7 (50.0)		12 (85.7)	2 (14.3)		9 (64.3)	5 (35.7)	
Basal	13 (37.1)	22 (62.9)		14 (40.0)	21 (60.0)		20 (57.1)	15 (42.9)		11 (31.4)	24 (68.6)	
Integrated classification clusters of METABRIC												
cluster-1	13 (68.4)	6 (31.6)		15 (78.9)	4 (21.1)		9 (47.4)	10 (52.6)		13 (68.4)	6 (31.6)	
cluster -2	3 (50.0)	3 (50.0)		3 (50.0)	3 (50.0)		4 (66.7)	2 (33.3)		3 (50.0)	3 (50.0)	
cluster -3	13 (43.3)	17 (56.7)		14 (46.7)	16 (53.3)		13 (43.3)	17 (56.7)		9 (30.0)	21 (70.0)	
cluster -4	7 (28.0)	18 (72.0)		5 (20.0)	20 (80.0)		10 (40.0)	15 (60.0)		7 (28.0)	18 (72.0)	
cluster -5	10 (50.0)	10 (50.0)	12.886 (0.168)	9 (45.0)	11 (55.0)	25.677 (**0.002**)	16 (80.0)	4 (20.0)	23.776 (**0.005**)	13 (65.0)	7 (35.0)	20.558 (**0.015**)
cluster -6	1 (33.3)	2 (66.7)		0 (0.0)	3 (100.0)		1 (33.3)	2 (66.7)		3 (100.0)	0 (0.0)	
cluster -7	11 (73.3)	4 (26.7)		7 (46.7)	8 (53.3)		4 (26.7)	11 (73.3)		9 (60.0)	6 (40.0)	
cluster -8	11 (57.9)	8 (42.1)		15 (78.9)	4 (21.1)		5 (26.3)	14 (73.7)		13 (68.4)	6 (31.6)	
cluster -9	5 (62.5)	3 (37.5)		5 (62.5)	3 (37.5)		7 (87.5)	1 (12.5)		4 (50.0)	4 (50.)	
cluster -10	11 (44.0)	14 (56.0)		12 (48.0)	13 (52.0)		16 (64.0)	9 (36.0)		11 (44.0)	14 (56.0)	

	CNRIP1			*CXCL12*			*ECM2*			*FBLN1*		
	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN^{-}	31 (38.8)	49 (61.3)	7.650 (**0.006**)	28 (35.0)	52 (65.0)	13.600 (**0.0002**)	32 (40.0)	48 (60.0)	6.044 (**0.014**)	31 (38.8)	49 (61.3)	7.650 (**0.006**)
Positive/LN$^{+/-}$	54 (60.0)	36 (40.0)		57 (63.3)	33 (36.7)		53 (58.9)	37 (41.1)		54 (60.0)	36 (40.0)	
Age (years)												
≤ 50	30 (45.5)	36 (54.5)	0.892 (0.345)	31 (47.0)	35 (53.0)	0.396 (0.529)	32 (48.5)	34 (51.5)	0.099 (0.753)	32 (48.5)	34 (51.5)	0.099 (0.753)
> 50	55 (52.9)	49 (47.1)		54 (51.9)	50 (48.1)		53 (51.0)	51 (49.0)		53 (51.0)	51 (49.0)	
Tumour size												
≤ 2.0cm	39 (39.0)	61 (61.0)	11.754 (**0.001**)	44 (44.0)	56 (56.0)	3.497 (0.061)	43 (43.0)	57 (57.0)	4.760 (**0.029**)	44 (44.0)	56 (56.0)	3.497 (0.061)
> 2.0cm	46 (65.7)	24 (34.3)		41 (58.6)	29 (41.4)		42 (60.0)	28 (40.0)		41 (58.6)	29 (41.4)	
Tumour Grade												
Grade-I	5 (19.2)	21 (80.0)		4 (15.4)	22 (84.6)		7 (26.9)	19 (73.1)		6 (23.1)	20 (76.9)	
Grade-II	21 (38.2)	34 (61.6)	23.141 (**0.00001**)	21 (38.2)	34 (61.8)	27.166 (**0.000001**)	19 (34.5)	36 (65.5)	21.015 (**0.00003**)	22 (40.0)	33 (60.0)	17.415 (**0.0002**)
Grade-III	59 (67.0)	29 (33.0)		60 (68.2)	26 (31.8)		59 (67.0)	29 (33.0)		57 (64.8)	31 (35.2)	
Nodal involvement, (NPI staging system)												
Stage-I	46 (43.8)	59 (56.2)		41 (39.0)	64 (61.0)		44 (41.9)	61 (58.1)		45 (42.9)	60 (57.1)	
Stage-II	30 (61.2)	19 (38.8)	4.329 (0.115)	32 (65.3)	17 (34.7)	13.630 (**0.001**)	32 (65.3)	17 (34.7)	7.594 (**0.022**)	28 (57.1)	21 (42.9)	7.143 (**0.028**)
Stage-III	9 (56.2)	7 (43.8)		12 (75.0)	4 (25.0)		9 (56.2)	7 (43.8)		12 (75.0)	4 (25.0)	
Nottingham Prognostic Index (NPI)												
good	20 (29.0)	49 (71.0)		18 (26.1)	51 (73.9)		21 (30.4)	48 (69.6)		21 (30.4)	48 (69.6)	
moderate	56 (63.6)	32 (36.4)	20.657 (**0.00003**)	56 (63.6)	32 (36.4)	28.559 (**6.3E-7**)	56 (63.6)	32 (36.4)	17.803 (**0.0001**)	54 (61.4)	34 (38.6)	18.880 (**0.00008**)
poor	9 (69.2)	4 (30.8)		11 (84.6)	2 (15.4)		8 (61.5)	5 (38.5)		10 (76.9)	3 (23.1)	

	CNRIP1			***CXCL12***			***ECM2***			***FBLN1***		
	-	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50												
Normal-like	2 (16.7)	10 (83.3)		2 (16.7)	10 (83.3)		3 (25.0)	9 (75.0)		1 (8.3)	11 (91.7)	
Luminal-A	21 (32.8)	43 (67.2)		19 (29.7)	45 (70.3)		18 (28.1)	46 (71.9)		19 (29.7)	45 (70.3)	
Luminal-B	28 (62.2)	17 (37.8)	23.613 (**0.0001**)	30 (66.7)	15 (33.3)	28.924 (**0.00001**)	27 (60.0)	18 (40.0)	29.879 (**0.00001**)	32 (71.1)	13 (28.9)	33.804 (**8.2E-7**)
Her2-enriched	11 (78.6)	3 (21.4)		11 (78.6)	3 (21.4)		11 (78.6)	3 (21.4)		11 (78.6)	3 (21.4)	
Basal	23 (65.7)	12 (34.3)		23 (65.7)	12 (34.3)		26 (74.3)	9 (25.7)		22 (62.9)	13 (37.1)	
Integrated classification clusters of METABRIC												
cluster-1	11 (57.9)	8 (42.1)		11 (57.9)	8 (42.1)		11 (57.9)	8 (42.1)		13 (68.4)	6 (31.6)	
cluster -2	3 (50.0)	3 (50.0)		3 (50.0)	3 (50.0)		3 (50.0)	3 (50.0)		3 (50.0)	3 (50.0)	
cluster -3	9 (30.0)	21 (70.0)		10 (33.3)	20 (66.7)		5 (16.7)	25 (83.3)		4 (13.3)	26 (86.7)	
cluster -4	6 (24.0)	19 (76.0)		7 (28.0)	18 (72.0)		9 (36.0)	16 (64.0)		8 (32.0)	17 (68.0)	
cluster -5	13 (65.0)	7 (35.0)	21.546 (**0.010**)	13 (65.0)	7 (35.0)	25.526 (**0.002**)	14 (70.0)	6 (30.0)	30.953 (**0.0003**)	15 (75.0)	5 (25.0)	37.428 (**0.00002**)
cluster -6	2 (66.7)	1 (33.3)		3 (100.0)	0 (0.0)		2 (66.7)	1 (33.3)		2 (66.7)	1 (33.3)	
cluster -7	7 (46.7)	8 (53.3)		4 (26.7)	11 (73.3)		6 (40.0)	10 (52.6)		8 (53.3)	7 (46.7)	
cluster -8	10 (52.6)	9 (47.4)		9 (47.4)	10 (52.6)		9 (47.4)	10 (52.6)		7 (36.8)	12 (63.2)	
cluster -9	5 (62.5)	3 (37.5)		6 (75.0)	2 (25.0)		6 (75.0)	2 (25.0)		6 (75.0)	2 (25.0)	
cluster -10	19 (76.0)	6 (24.0)		19 (76.0)	6 (24.0)		20 (80.0)	5 (20.0)		19 (76.0)	6 (24.0)	

	FEZ1			*INKA1*			*ITSN1*			*LPAR1*		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN^{-}	32 (40.0)	48 (60.0)	6.044 (**0.014**)	30 (37.5)	50 (62.5)	9.444 (**0.002**)	33 (41.3)	47 (58.8)	4.628 (**0.031**)	28 (35.0)	52 (65.0)	13.600 (**0.0002**)
Positive/LN$^{+/-}$	53 (58.9)	37 (41.1)		55 (61.1)	35 (38.9)		52 (57.8)	38 (42.2)		57 (63.3)	33 (36.7)	
Age (years)												
≤ 50	34 (51.5)	32 (48.5)	0.099 (0.753)	37 (56.1)	29 (43.9)	1.585 (0.208)	33 (50.0)	33 (50.0)	0.000 (1.000)	33 (50.0)	33 (50.0)	0.000 (1.000)
> 50	51 (49.0)	53 (51.0)		48 (46.2)	56 (53.8)		52 (50.0)	52 (50.0)		52 (50.0)	52 (50.0)	
Tumour size												
≤ 2.0cm	47 (47.0)	53 (53.0)	0.874 (0.350)	44 (44.0)	56 (56.0)	3.497 (0.061)	47 (47.0)	53 (53.0)	0.874 (0.350)	44 (44.0)	56 (56.0)	3.497 (0.061)
> 2.0cm	38 (54.3)	32 (45.7)		41 (58.6)	29 (41.4)		38 (54.3)	32 (45.7)		41 (58.6)	29 (41.4)	
Tumour Grade												
Grade-I	8 (30.8)	18 (69.2)		7 (26.9)	19 (73.1)		11 (42.3)	15 (57.7)		8 (30.8)	18 (69.2)	
Grade-II	27 (49.1)	28 (50.9)	5.495 (0.064))	21 (38.2)	34 (61.8)	16.288 (**0.0003**)	22 (40.0)	33 (60.0)	5.719 (0.057)	18 (32.7)	37 (67.3)	20.632 (**0.00003**)
Grade-III	50 (56.8)	38 (43.2)		57 (64.8)	31 (35.2)		52 (59.1)	36 (40.9)		59 (67.0)	29 (33.0)	
Nodal involvement, (NPI staging system)												
Stage-I	47 (44.8)	58 (55.2)		45 (42.9)	60 (57.1)		46 (43.8)	59 (56.2)		39 (37.1)	66 (62.9)	
Stage-II	26 (53.1)	23 (46.9)	5.336 (0.069)	29 (59.2)	20 (40.8)	6.046 (**0.049**)	30 (61.2)	19 (38.8)	4.329 (0.115)	33 (67.3)	16 (32.7)	19.091 (**0.0001**)
Stage-III	12 (75.0)	4 (25.0)		11 (68.8)	5 (31.3)		9 (56.2)	7 (43.8)		13 (81.3)	3 (18.8)	
Nottingham Prognostic Index (NPI)												
good	28 (40.6)	41 (59.4)		22 (31.9)	47 (68.1)		26 (37.7)	43 (62.3)		19 (27.5)	50 (72.5)	
moderate	47 (53.4)	41 (46.6)	6.628 (0.036)	53 (60.2)	35 (39.8)	16.509 (**0.0003**)	52 (59.1)	36 (40.9)	7.174 (**0.028**)	55 (62.5)	33 (37.5)	25.658 (**0.00003**)
poor	10 (76.9)	3 (23.1)		10 (76.9)	3 (23.1)		7 (53.8)	6 (46.2)		11 (84.6)	2 (15.4)	

	FEZ1			*INKA1*			*ITSN1*			*LPAR1*		
	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Prediction analysis of microarray 50, PAM50												
Normal-like	1 (8.3)	11 (91.7)		0 (0.0)	12 (100.0)		3 (25.0)	9 (75.0)		2 (16.7)	10 (83.3)	
Luminal-A	24 (37.5)	40 (62.5)		23 (35.9)	41 (64.1)		24 (37.5)	40 (62.5)		23 (35.9)	41 (64.1)	
Luminal-B	31 (68.9)	14 (31.1)	21.584 (**0.0002**)	28 (62.2)	17 (37.8)	31.437 (**0.000002**)	27 (60.0)	18 (40.0)	12.543 (**0.014**)	27 (60.0)	18 (40.0)	19.082 (**0.001**)
Her2-enriched	10 (71.4)	4 (28.6)		13 (92.9)	1 (7.1)		8 (57.1)	6 (42.9)		11 (78.6)	3 (21.4)	
Basal	19 (54.3)	16 (45.7)		21 (60.0)	14 (40.0)		23 (65.7)	12 (34.3)		22 (62.9)	13 (37.1)	
Integrated classification clusters of METABRIC												
cluster-1	12 (63.2)	7 (36.8)		10 (52.6)	9 (47.4)		14 (73.7)	5 (26.3)		12 (63.2)	7 (36.8)	
cluster -2	3 (50.0)	3 (50.0)		3 (50.0)	3 (50.0)		2 (33.3)	4 (66.7)		3 (50.0)	3 (50.0)	
cluster -3	7 (23.3)	23 (76.7)		8 (26.7)	22 (73.3)		10 (33.3)	20 (66.7)		7 (23.3)	23 (76.7)	
cluster -4	7 (28.0)	18 (72.0)		5 (20.0)	20 (80.0)		9 (36.0)	16 (64.0)		5 (20.0)	20 (80.0)	
cluster -5	13 (65.0)	7 (35.0)	22.715 (**0.007**)	17 (85.0)	3 (15.0)	47.760 (**2.8E-7**)	11 (55.0)	9 (45.0)	18.237 (**0.033**)	15 (75.0)	5 (25.0)	34.016 (**0.0001**)
cluster -6	1 (33.3)	2 (66.7)		2 (66.7)	1 (33.3)		2 (66.7)	1 (33.3)		1 (33.3)	2 (66.7)	
cluster -7	9 (60.0)	6 (40.0)		2 (13.3)	13 (86.7)		5 (33.3)	10 (66.7)		6 (40.0)	9 (60.0)	
cluster -8	10 (52.6)	9 (47.4)		11 (57.9)	8 (42.1)		11 (57.9)	8 (42.1)		11 (57.9)	8 (42.1)	
cluster -9	6 (75.0)	2 (25.0)		7 (87.5)	1 (12.5)		3 (37.5)	5 (62.5)		6 (75.0)	2 (25.0)	
cluster -10	17 (68.0)	8 (32.0)		20 (80.0)	5 (20.0)		18 (72.0)	7 (28.0)		19 (76.0)	6 (24.0)	

	NBR1			*PRRX1*			*STAT3*			*THAP4*		
	-	+	**χ^2 (*p*-value≤)**	-	+	**χ^2 (*p*-value≤)**	-	+	**χ^2 (*p*-value≤)**	-	+	**χ^2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN$^-$	34 (42.5)	46 (57.5)	3.400 (0.065)	34 (42.5)	46 (57.5)	3.400 (0.065)	35 (43.8)	45 (56.3)	2.361 (0.124)	33 (41.3)	47 (58.8)	4.628 (**0.031**)
Positive/LN$^{+/-}$	51 (56.7)	39 (43.3)		51 (56.7)	39 (43.3)		50 (55.6)	40 (44.4)		52 (57.8)	38 (42.2)	
Age (years)												
≤ 50	32 (48.5)	34 (51.5)	0.099 (0.753)	31 (47.0)	35 (53.0)	0.396 (0.529)	38 (57.6)	28 (42.4)	2.477 (0.116)	39 (59.1)	27 (40.9)	3.566 (0.059)
> 50	53 (51.0)	51 (49.0)		54 (51.9)	50 (48.1)		47 (45.2)	57 (54.8)		46 (44.2)	58 (55.8)	
Tumour size												
≤ 2.0cm	48 (48.0)	52 (52.0)	0.389 (0.533)	42 (42.0)	58 (58.0)	6.217 (**0.013**)	43 (43.0)	57 (57.0)	4.760 (**0.029**)	47 (47.0)	53 (53.0)	0.874 (0.350)
> 2.0cm	37 (52.9)	33 (47.1)		43 (61.4)	27 (38.6)		42 (60.0)	28 (40.0)		38 (54.3)	32 (45.7)	
Tumour Grade												
Grade-I	9 (34.6)	17 (65.4)		10 (38.5)	16 (61.5)		7 (26.9)	19 (73.1)		6 (23.1)	20 (76.9)	
Grade-II	21 (38.2)	34 (61.8)	10.074 (**0.006**)	26 (47.3)	29 (52.7)	2.679 (0.262)	18 (32.7)	37 (67.3)	22.324 (**0.00001**)	20 (36.4)	35 (63.6)	20.533 (**0.00004**)
Grade-III	54 (61.4)	34 (38.6)		49 (55.7)	39 (44.3)		59 (67.0)	29 (33.0)		58 (65.9)	30 (34.1)	
Nodal involvement, (NPI staging system)												
Stage-I	45 (42.9)	60 (57.1)		47 (44.8)	58 (55.2)		45 (42.9)	60 (57.1)		43 (41.0)	62 (59.0)	
Stage-II	31 (63.3)	18 (36.7)	5.842 (0.054)	28 (57.1)	21 (42.9)	3.152 (0.207)	28 (57.1)	21 (42.9)	7.143 (**0.028**)	30 (61.2)	19 (38.8)	9.907 (**0.007**)
Stage-III	9 (56.2)	7 (43.8)		10 (62.5)	6 (37.5)		12 (75.0)	4 (25.0)		12 (75.0)	4 (25.0)	
Nottingham Prognostic Index (NPI)												
good	25 (36.2)	44 (63.8)		30 (43.5)	39 (56.5)		23 (33.3)	46 (66.7)		21 (30.4)	48 (69.6)	
moderate	53 (60.2)	35 (39.8)	8.991 (**0.011**)	46 (52.3)	42 (47.7)	3.279 (0.194)	50 (56.8)	38 (43.2)	18.611 (**0.0001**)	54 (61.4)	34 (38.6)	18.880 (**0.0001**)
poor	7 (53.8)	6 (46.2)		9 (69.2)	4 (30.8)		12 (92.3)	1 (7.7)		10 (76.9)	3 (23.1)	

	NBR1			***PRRX1***			***STAT3***			***THAP4***		
	-	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50												
Normal-like	6 (50.0)	6 (50.0)		3 (25.0)	9 (75.0)		3 (25.0)	9 (75.0)		6 (50.0)	6 (50.0)	
Luminal-A	26 (40.6)	38 (59.4)		27 (42.2)	37 (57.8)		18 (28.1)	46 (71.9)		20 (31.3)	44 (68.8)	
Luminal-B	21 (46.7)	24 (53.3)	7.336 (0.119)	25 (55.6)	20 (44.4)	7.661 (0.105)	31 (68.9)	14 (31.1)	28.558 (**0.00001**)	20 (44.4)	25 (55.6)	27.013 (**0.00002**)
Her2-enriched	10 (71.4)	4 (28.6)		9 (64.3)	5 (35.7)		11 (78.6)	3 (21.4)		12 (85.7)	2 (14.3)	
Basal	22 (62.9)	13 (37.1)		21 (60.0)	14 (40.0)		22 (62.9)	13 (37.1)		27 (77.1)	8 (22.9)	
Integrated classification clusters of METABRIC												
cluster-1	10 (52.6)	9 (47.7)		13 (68.4)	6 (31.6)		14 (73.7)	5 (26.3)		11 (57.9)	8 (42.1)	
cluster -2	3 (50.0)	3 (50.0)		2 (33.3)	4 (66.7)		5 (83.3)	1 (16.7)		3 (50.0)	3 (50.0)	
cluster -3	12 (40.0)	18 (60.0)		7 (23.3)	23 (76.7)		7 (23.3)	23 (76.7)		11 (36.7)	19 (63.3)	
cluster -4	14 (56.0)	11 (44.0)		10 (40.0)	15 (60.0)		11 (44.0)	14 (56.0)		14 (56.0)	11 (44.0)	
cluster -5	13 (65.0)	7 (35.0)	11.609 (0.236)	11 (55.0)	9 (45.0)	16.655 (0.054)	14 (70.0)	6 (30.0)	30.702 (**0.0003**)	16 (80.0)	4 (20.0)	37.502 (**0.00002**)
cluster -6	1 (33.3)	2 (66.7)		2 (66.7)	1 (33.3)		2 (66.7)	1 (33.3)		1 (6.7)	14 (93.3)	
cluster -7	6 (40.0)	9 (60.0)		8 (53.3)	7 (46.7)		4 (26.7)	11 (73.3)		3 (15.8)	16 (84.2)	
cluster -8	5 (26.3)	14 (73.7)		12 (63.2)	7 (36.8)		6 (31.6)	13 (68.4)		8 (75.0)	2 (25.0)	
cluster -9	6 (75.0)	2 (25.0)		4 (50.0)	4 (50.0)		7 (87.5)	1 (12.5)		6 (75.0)	2 (25.0)	
cluster -10	15 (60.0)	10 (40.0)		16 (64.0)	9 (36.0)		15 (60.0)	10 (40.0)		18 (50.0)	18 (50.0)	

	TMEM98			*TNC*			*TNFRSF19*			*TSC1*		
	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN^{-}	32 (40.0)	48 (60.0)	6.044 (**0.014**)	34 (42.5)	46 (57.5)	3.400 (0.065)	36 (45.0)	44 (55.0)	1.511 (0.219)	36 (45.0)	44 (55.0)	1.511 (0.219)
Positive/LN$^{+/-}$	53 (58.9)	37 (41.1)		51 (56.7)	39 (43.3)		49 (54.5)	41 (45.6)		49 (54.4)	41(45.6)	
Age (years)												
≤ 50	33 (50.0)	33 (50.0)	0.000 (1.000)	30 (45.5)	36 (54.5)	0.892 (0.345)	32 (48.5)	34 (51.5)	0.099 (0.753)	37 (56.1)	29 (43.9)	1.585 (0.208)
> 50	52 (50.0)	52 (50.0)		55 (52.9)	49 (47.1)		53 (51.0)	51 (49.0)		48 (46.2)	56 (53.8)	
Tumour size												
≤ 2.0cm	48 (48.0)	52 (52.0)	0.389 (0.533)	49 (49.0)	51 (51.0)	0.097 (0.755)	43 (43.0)	57 (57.0)	4.730 (**0.029**)	47 (47.0)	53 (53.0)	0.874 (0.350)
> 2.0cm	37 (52.9)	33 (47.1)		36 (51.4)	34 (48.6)		42 (60.0)	28 (40.0)		38 (54.3)	32 (45.7)	
Tumour Grade												
Grade-I	8 (30.8)	18 (69.2)		6 (23.1)	20 (76.9)		5 (19.2)	21 (80.8)		8 (30.8)	18 (69.2)	
Grade-II	26 (47.3)	29 (52.7)	6.231 (**0.044**)	35 (63.6)	20 (36.4)	11.624 (**0.003**)	28 (50.9)	27 (49.1)	12.768 (**0.002**)	20 (36.4)	35 (63.6)	15.614 (**0.0004**)
Grade-III	51 (58.0)	37 (42.0)		44 (50.0)	44 (50.0)		52 (59.1)	36 (40.9)		57 (64.8)	31 (35.2)	
Nodal involvement, (NPI staging system)												
Stage-I	45 (42.9)	60 (57.1)		49 (46.7)	56 (53.3)		47 (44.8)	58 (55.2)		46 (43.8)	59 (56.2)	
Stage-II	31 (63.3)	18 (36.7)	5.842 (0.054)	26 (53.1)	23 (46.9)	1.650 (0.438)	28 (57.1)	21 (42.9)	3.152 (0.207)	29 (59.2)	20 (40.8)	4.263 (0.119)
Stage-III	9 (56.3)	7 (43.8)		10 (62.5)	6 (37.5)		10 (62.5)	6 (37.5)		10 (62.5)	6 (37.5)	
Nottingham Prognostic Index (NPI)0												
good	29 (42.0)	40 (58.0)		33 (47.8)	36 (52.2)		27 (39.1)	42 (60.9)		22 (31.9)	47 (68.1)	
moderate	48 (54.5)	40 (45.5)	3.173 (0.205)	44 (50.0)	44 (50.0)	0.823 (0.663)	49 (55.7)	39 (44.3)	6.320 (**0.042**)	54 (61.4)	34 (38.6)	15.527 (**0.0004**)
poor	8 (61.5)	5 (38.5)		8 (61.5)	5 (38.5)		9 (69.2)	4 (30.8)		9 (69.2)	4 (30.8)	

	TMEM98			*TNC*			*TNFRSF19*			*TSC1*		
	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50												
Normal-like	4 (33.3)	8 (66.7)		2 (16.7)	10 (83.3)		4 (33.3)	8 (66.7)		4 (33.3)	8 (66.7)	
Luminal-A	27 (42.2)	37 (57.8)		30 (46.9)	34 (53.1)		21 (32.8)	43 (67.2)		20 (31.3)	44 (68.8)	
Luminal-B	23 (51.1)	22 (48.9)	6.889 (0.142)	32 (71.1)	13 (28.9)	15.006 (**0.005**)	25 (55.6)	20 (44.4)	20.051 (**0.0004**)	24 (53.3)	21 (46.7)	23.419 (**0.0001**)
Her2-enriched	10 (71.4)	4 (28.6)		7 (50.0)	7 (50.0)		12 (85.7)	2 (14.3)		10 (71.4)	4 (28.6)	
Basal	21 (60.0)	14 (40.0)		14 (40.0)	21 (60.0)		23 (65.7)	12 (34.3)		27 (77.1)	8 (22.9)	
Integrated classification clusters of METABRIC												
cluster-1	10 (52.6)	9 (47.4)		15 (78.9)	4 (21.1)		14 (73.7)	5 (26.3)		9 (47.4)	10 (52.6)	
cluster -2	3 (50.0)	3 (50.0)		3 (50.0)	3 (50.0)		2 (33.3)	4 (66.7)		4 (66.7)	2 (33.3)	
cluster -3	10 (33.3)	20 (66.7)		10 (33.3)	20 (66.7)		7 (23.3)	23 (76.7)		10 (33.3)	20 (66.7)	
cluster -4	15 (60.0)	10 (40.0)		8 (32.0)	17 (68.0)		9 (36.0)	16 (64.0)		7 (28.0)	18 (72.0)	
cluster -5	11 (55.0)	9 (45.0)	14.109 (0.118)	11 (55.0)	9 (45.0)	17.755 (**0.038**)	12 (60.0)	8 (40.0)	25.732 (**0.002**)	13 (65.0)	7 (35.0)	30.0235 (**0.0004**)
cluster -6	2 (66.7)	1 (33.3)		0 (0.0)	3 (100.0)		1 (33.3)	2 (66.7)		3 (100.0)	0 (0.0)	
cluster -7	8 (53.3)	7 (46.7)		9 (60.0)	6 (40.0)		9 (60.0)	6 (40.0)		5 (33.3)	10 (66.7)	
cluster -8	5 (26.3)	14 (73.7)		11 (57.9)	8 (42.1)		7 (36.8)	12 (63.2)		7 (36.8)	12 (63.2)	
cluster -9	7 (87.5)	1 (12.5)		5 (62.5)	3 (37.5)		5 (62.5)	3 (37.5)		6 (75.0)	2 (25.0)	
cluster -10	14 (56.0)	11 (44.0)		13 (52.0)	12 (48.0)		19 (76.0)	6 (24.0)		21 (84.0)	4 (16.0)	

	TWIST2			*WISP2*		
	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**	
Lymphovascular invasion (LVI), (H&E/IHC)						
Negative/LN^{-}	37 (46.3)	43 (53.8)	0.850 (0.357)	32 (40.0)	48 (60.0)	6.044 (**0.014**)
Positive/LN$^{+/-}$	48 (53.3)	42 (46.7)		53 (58.9)	37 (41.1)	
Age (years)						
≤ 50	31 (47.0)	35 (53.0)	0.396 (0.529)	30 (45.5)	36 (54.5)	0.892 (0.345)
> 50	54 (51.9)	50 (48.1)		55 (52.9)	49 (47.1)	
Tumour size						
≤ 2.0cm	50 (50.0)	50 (50.0)	0.000 (1.000)	44 (44.0)	56 (56.0)	3.497 (0.061)
> 2.0cm	35 (50.0)	35 (50.0)		41 (58.6)	29 (41.4)	
Tumour Grade						
Grade-I	10 (38.5)	16 (61.5)	7.261 (**0.027**)	5 (19.2)	21 (80.8)	18.995 (**0.001**)
Grade-II	22 (40.0)	33 (60.0)		23 (41.8)	32 (58.2)	
Grade-III	53 (60.2)	35 (39.8)		57 (64.8)	31 (35.2)	
Nodal involvement, (NPI staging system)*						
Stage-I	51 (48.6)	54 (51.4)	1.106 (0.575)	48 (45.7)	57 (54.3)	2.282 (0.320)
Stage-II	24 (49.0)	25 (51.0)		27 (55.1)	22 (44.9)	
Stage-III	10 (62.5)	6 (37.5)		10 (62.5)	6 (37.5)	
Nottingham Prognostic Index (NPI)						
good	27 (39.1)	42 (60.9)	5.590 (0.061)	23 (33.3)	46 (66.7)	13.272 (**0.001**)
moderate	50 (56.8)	38 (43.2)		53 (60.2)	35 (39.8)	
poor	8 (61.5)	5 (38.5)		9 (69.2)	4 (30.8)	

	TWIST2			WISP2		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)	
Prediction analysis of microarray 50, PAM50						
Normal-like	2 (16.7)	10 (83.3)		5 (41.7)	7 (58.3)	
Luminal-A	25 (39.1)	39 (60.9)		22 (34.4)	42 (65.6)	
Luminal-B	30 (66.7)	15 (33.3)	14.396 (**0.006**)	24 (53.3)	21 (46.7)	14.183 (**0.007**)
Her2-enriched	8 (57.1)	6 (42.9)		10 (71.4)	4 (28.6)	
Basal	20 (57.1)	15 (42.9)		24 (68.6)	11 (31.4)	
Integrated classification clusters of METABRIC						
cluster-1	13 (68.4)	6 (31.6)		9 (47.4)	10 (52.6)	
cluster -2	2 (33.3)	4 (66.7)		2 (33.3)	4 (66.7)	
cluster -3	11 (36.7)	19 (63.3)		13 (43.3)	17 (56.7)	
cluster -4	12 (48.0)	13 (52.0)		10 (40.0)	15 (60.0)	
cluster -5	9 (45.0)	11 (55.0)	15.732 (0.073)	14 (70.0)	6 (30.0)	19.032 (**0.025**)
cluster -6	1 (33.3)	2 (66.7)		2 (66.7)	1 (33.3)	
cluster -7	4 (26.7)	11 (73.3)		5 (33.3)	10 (66.7)	
cluster -8	10 (52.6)	9 (47.4)		6 (31.6)	13 (68.4)	
cluster -9	7 (87.5)	1 (12.5)		4 (50.0)	4 (50.0)	
cluster -10	16 (64.0)	9 (36.0)		20 (80.0)	5 (20.0)	

Table 3. 10: Clinicopathological significance of upregulated genes with LVI^{-} patients in the METABRIC cohort (n=1980), patients of lymph node positive and lymph node negative (LN+/-) were included in both LVI categories. Significant values are in bold

	ANXA5			***APCDD1***			***C17orf81/ELP5***			***C1S***		
	-	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN$^{+/-}$	59 (36.0)	105 (64.0)	4.169 (**0.041**)	66 (40.2)	98 (59.8)	6.011 (**0.014**)	75 (45.7)	89 (54.3)	2.273 (0.132)	50 (30.5)	114 (69.5)	1.140 (0.286)
Definite/LN$^{+/-}$	63 (47.7)	69 (52.3)		72 (54.5)	60 (45.5)		72 (54.5)	60 (45.5)		48 (36.4)	84 (63.6)	
Age (years)												
≤ 50	171 (39.9)	258 (60.1)	22.523 (**0.00002**)	163 (38.0)	266 (62.0)	31.570 (**2.0E-8**)	222 (51.7)	207 (48.3)	0.670 (0.413)	160 (37.3)	269 (62.7)	35.355 (**2.7E-9**)
> 50	819 (52.8)	732 (47.2)		827 (53.3)	724 (46.7)		768 (49.5)	783 (50.5)		830 (53.5)	721 (46.5)	
Tumour size												
≤ 2.0cm	401 (46.7)	457 (53.3)	6.501 (**0.011**)	393 (45.8)	485 (54.2)	9.976 (**0.002**)	397 (46.3)	461 (53.7)	7.587 (**0.006**)	384 (44.8)	474 (55.2)	16.791 (**0.00004**)
> 2.0cm	579 (52.3)	523 (47.5)		584 (53.0)	518 (47.0)		579 (52.5)	523 (47.5)		596 (54.1)	506 (45.9)	
Tumour Grade												
Grade-I	68 (39.8)	103 (60.2)	8.056 (0.018)	87 (50.9)	84 (49.1)	6.518 (**0.038**)	56 (32.7)	115 (67.3)	53.636 (**2.2E-12**)	80 (46.8)	91 (53.2)	2.788 (0.248)
Grade-II	396 (51.4)	374 (48.6)		408 (53.0)	362 (47.0)		346 (44.9)	424 (55.1)		399 (51.8)	371 (48.2)	
Grade-III	484 (50.8)	468 (49.2)		446 (46.8)	506 (53.2)		554 (58.2)	398 (41.8)		459 (48.2)	493 (51.8)	
Nodal involvement, (NPI staging system)												
Stage-I	518 (50.0)	517 (50.0)	0.363 (0.834)	504 (48.7)	531 (51.3)	1.732 (0.421)	496 (47.9)	539 (52.1)	12.659 (**0.002**)	528 (51.0)	507 (49.0)	1.964 (0.375)
Stage-II	308 (49.5)	314 (50.5)		319 (51.3)	303 (48.7)		305 (49.0)	317 (51.0)		312 (50.2)	310 (49.8)	
Stage-III	163 (51.6)	153 (48.4)		165 (52.2)	151 (47.8)		187 (59.2)	129 (40.8)		147 (46.5)	169 (53.3)	
Nottingham Prognostic Index (NPI)												
good	323 (47.5)	357 (52.5)	2.589 (0.274)	356 (52.4)	324 (47.6)	2.294 (0.318)	284 (41.8)	396 (58.2)	33.385 (**5.6E-8**)	351 (51.6)	329 (48.4)	4.398 (0.111)
moderate	565 (51.3)	536 (48.7)		537 (48.8)	564 (51.2)		583 (53.0)	518 (47.0)		553 (50.2)	548 (49.8)	
poor	102 (51.3)	97 (48.7)		97 (48.7)	102 (51.3)		123 (61.8)	76 (38.2)		86 (43.2)	113 (56.8)	

	ANXA5			*APCDD1*			C17orf81/*ELP5*			*C1S*		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Prediction analysis of microarray 50, PAM50												
Normal-like	70 (34.1)	135 (65.9)		46 (22.4)	159 (77.6)		89 (43.4)	116 (56.6)		45 (22.0)	160 (78.0)	
Luminal-A	368 (51.3)	350 (48.7)		370 (51.5)	348 (48.5)		298 (41.5)	420 (58.5)		350 (48.7)	368 (51.3)	
Luminal-B	310 (63.5)	178 (36.5)	98.844 (**1.7E-20**)	333 (68.2)	155 (31.8)	195.353 (**3.7E-41**)	245 (50.2)	243 (49.8)	93.658 (**2.2E-19**)	357 (73.2)	131 (26.8)	211.842 (**1.1E-44**)
Her2-enriched	134 (55.8)	106 (44.2)		145 (60.4)	95 (39.6)		184 (76.7)	56 (23.3)		131 (54.6)	109 (45.4)	
Basal	108 (32.8)	221 (67.2)		96 (29.2)	233 (70.8)		174 (52.9)	155 (47.1)		107 (32.5)	222 (67.5)	
Integrated classification clusters of METABRIC												
cluster-1	82 (59.0)	57 (41.0)		84 (60.4)	55 (39.6)		91 (65.5)	48 (34.5)		91 (65.5)	48 (34.5)	
cluster -2	44 (61.1)	28 (38.9)		40 (55.6)	32 (44.4)		30 (41.7)	42 (58.3)		45 (62.5)	27 (37.5)	
cluster -3	111 (38.3)	179 (61.7)		125 (43.1)	165 (56.9)		123 (42.4)	167 (57.6)		96 (33.1)	194 (66.9)	
cluster -4	121 (35.3)	222 (64.7)		114 (33.2)	229 (66.8)		165 (48.1)	178 (51.9)		90 (26.2)	253 (73.8)	
cluster -5	83 (43.7)	107 (56.3)	160.256 (**6.6E-30**)	89 (46.8)	101 (53.2)	144.903 (**1.0E-26**)	147 (77.4)	43 (22.6)	116.182 (**8.1E-21**)	89 (46.8)	101 (53.2)	243.721 (**2.1E-47**)
cluster -6	49 (57.6)	36 (42.4)		47 (55.3)	38 (44.7)		44 (51.8)	41 (48.2)		55 (64.7)	30 (35.3)	
cluster -7	136 (71.6)	54 (28.4)		118 (62.1)	72 (37.9)		86 (45.3)	104 (54.7)		119 (62.6)	71 (37.4)	
cluster -8	178 (59.5)	121 (40.5)		207 (69.2)	92 (30.8)		102 (34.1)	197 (65.9)		219 (73.2)	80 (26.8)	
cluster -9	108 (74.0)	38 (26.0)		92 (63.0)	54 (37.0)		86 (58.9)	60 (41.1)		99 (67.8)	47 (32.2)	
cluster -10	78 (34.5)	148 (65.5)		74 (32.7)	152 (67.3)		116 (51.3)	110 (48.7)		87 (38.5)	139 (61.5)	

	CNRIP1			*CXCL12*			*ECM2*			*FBLN1*		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN$^{+/-}$	64 (39.0)	100 (61.0)	0.212 (0.645)	53 (32.3)	111 (67.7)	5.963 (**0.015**)	38 (23.2)	126 (76.8)	2.773 (0.096)	68 (41.5)	96 (58.5)	2.975 (0.085)
Definite/LN$^{+/-}$	55 (41.7)	77 (58.3)		61 (46.2)	71 (53.8)		42 (31.8)	90 (68.2)		68 (51.5)	64 (48.5)	
Age (years)												
≤ 50	207 (48.3)	222 (51.7)	0.709 (0.400)	188 (43.8)	241 (56.2)	8.359 (**0.004**)	206 (48.0)	223 (52.0)	0.860 (0.354)	191 (44.5)	238 (55.5)	6.573 (**0.010**)
> 50	784 (50.5)	767 (49.5)		802 (51.7)	749 (48.3)		784 (50.5)	767 (49.5)		799 (51.5)	752 (48.5)	
Tumour size												
≤ 2.0cm	383 (44.6)	475 (55.4)	17.213 (**0.00003**)	382 (44.5)	476 (55.5)	18.659 (**0.00002**)	386 (45.0)	472 (55.0)	15.021 (**0.0001**)	378 (44.1)	480 (55.9)	20.833 (**0.00001**)
> 2.0cm	596 (54.1)	506 (45.9)		599 (54.4)	503 (45.6)		593 (53.8)	509 (46.2)		600 (54.4)	502 (45.6)	
Tumour Grade												
Grade-I	60 (35.1)	111 (64.9)		52 (30.4)	119 (69.6)		56 (32.7)	115 (67.3)		57 (33.3)	114 (66.7)	
Grade-II	362 (47.0)	408 (53.0)	30.581 (**2.3E-7**)	344 (44.7)	426 (55.3)	59.883 (**1.0E-14**)	333 (43.2)	437 (56.8)	65.438 (**6.2E-15**)	327 (42.5)	443 (57.5)	66.110 (**4.4E-15**)
Grade-III	531 (55.8)	421 (44.2)		553 (58.1)	399 (41.9)		562 (59.0)	390 (41.0)		560 (58.8)	392 (41.2)	
Nodal involvement, (NPI staging system)												
Stage-I	522 (50.4)	513 (49.6)		503 (48.6)	532 (51.4)		509 (49.2)	526 (50.8)		491 (47.4)	544 (52.6)	
Stage-II	322 (51.8)	300 (48.2)	3.333 (0.189)	328 (52.7)	294 (47.3)	2.721 (0.257)	321 (51.6)	301 (48.4)	0.922 (0.631)	331 (53.2)	291 (46.8)	5.906 (0.052)
Stage-III	144 (45.6)	172 (54.4)		156 (49.4)	160 (50.6)		159 (50.3)	157 (49.7)		165 (52.2)	151 (47.8)	
Nottingham Prognostic Index (NPI)												
good	306 (45.0)	374 (55.0)		284 (41.8)	396 (58.2)		288 (42.4)	392 (57.6)		283 (41.6)	397 (58.4)	
moderate	585 (53.1)	516 (46.9)	11.127 (**0.004**)	602 (54.7)	499 (45.3)	28.490 (**6.5E-7**)	590 (53.6)	511 (46.4)	24.715 (**0.000004**)	595 (54.0)	506 (46.0)	29.447 (**4.0E-7**)
poor	100 (50.3)	99 (49.7)		104 (52.3)	95 (47.7)		112 (56.3)	87 (43.7)		112 (56.3)	87 (43.7)	

	CNRIP1			*CXCL12*			*ECM2*			*FBLN1*		
	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50												
Normal-like	48 (23.4)	157 (76.6)		37 (18.0)	168 (82.0)		53 (25.9)	152 (74.1)		36 (17.6)	169 (82.4)	
Luminal-A	290 (40.4)	428 (59.6)		278 (38.7)	440 (61.3)		266 (37.0)	452 (63.0)		256 (35.7)	462 (64.3)	
Luminal-B	309 (63.3)	179 (36.7)	144.020 (**3.9E-30**)	336 (68.9)	152 (31.1)	212.862 (**6.4E-45**)	299 (61.3)	189 (38.7)	176.898 (**3.5E-37**)	356 (73.0)	132 (27.0)	271.469 (**1.5E-57**)
Her2-enriched	146 (60.8)	94 (39.2)		152 (63.3)	88 (36.7)		148 (61.7)	92 (38.3)		145 (60.4)	95 (39.6)	
Basal	198 (60.2)	131 (39.8)		187 (56.8)	142 (43.2)		224 (68.1)	105 (31.9)		197 (59.9)	132 (40.1)	
Integrated classification clusters of METABRIC												
cluster-1	85 (61.5)	54 (38.8)		89 (64.0)	50 (36.0)		91 (65.5)	48 (34.5)		108 (77.7)	31 (22.3)	
cluster -2	42 (58.3)	30 (41.7)		39 (54.2)	33 (45.8)		39 (54.2)	33 (45.8)		48 (66.7)	24 (33.3)	
cluster -3	84 (29.0)	206 (71.0)		88 (30.3)	202 (69.7)		70 (24.1)	220 (75.9)		74 (25.5)	216 (74.5)	
cluster -4	84 (24.5)	259 (75.5)		88 (25.7)	255 (74.3)		99 (28.9)	244 (71.1)		91 (26.5)	252 (73.5)	
cluster -5	100 (52.6)	90 (47.4)	226.410 (**9.4E-44**)	107 (56.3)	83 (43.7)	205.180 (**2.7E-39**)	102 (53.7)	88 (46.3)	262.693 (**2.1E-51**)	104 (54.7)	86 (45.3)	290.517 (**2.7E-57**)
cluster -6	49 (57.6)	36 (42.4)		51 (60.0)	34 (40.0)		50 (58.8)	35 (41.2)		48 (56.5)	37 (43.5)	
cluster -7	111 (58.4)	79 (41.6)		98 (51.6)	92 (48.4)		95 (50.0)	95 (50.0)		88 (46.3)	102 (53.7)	
cluster -8	178 (59.5)	121 (40.5)		174 (58.2)	125 (41.8)		162 (54.2)	137 (45.8)		152 (50.8)	147 (49.2)	
cluster -9	94 (64.4)	52 (35.6)		98 (67.1)	48 (32.9)		100 (68.5)	46 (31.5)		103 (70.5)	43 (29.5)	
cluster -10	164 (72.6)	62 (27.4)		158 (69.9)	68 (30.1)		182 (80.5)	44 (19.5)		174 (77.0)	52 (23.0)	

	FEZ1			*INKA1*			*ITSN1*			*LPAR1*		
	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN$^{+/-}$	63 (38.4)	101 (61.6)	1.827 (0.177)	69 (42.1)	95 (57.9)	1.215 (0.270)	92 (56.1)	72 (43.9)	4.510 (**0.034**)	63 (38.4)	101 (61.6)	5.088 (**0.024**)
Definite/LN$^{+/-}$	61 (46.2)	71 (53.8)		64 (48.5)	68 (51.5)		90 (68.2)	42 (31.8)		68 (51.5)	64 (48.5)	
Age (years)												
≤ 50	211 (49.2)	218 (50.8)	0.146 (0.703)	221 (51.5)	208 (48.5)	0.503 (0.478)	214 (49.9)	215 (50.1)	0.003 (0.956)	232 (54.1)	197 (45.9)	3.645 (0.56)
> 50	779 (50.2)	772 (49.8)		769 (49.6)	782 (50.4)		776 (50.0)	775 (50.0)		758 (48.9)	793 (51.1)	
Tumour size												
≤ 2.0cm	379 (44.2)	479 (55.8)	20.368 (**0.00001**)	381 (44.4)	477 (55.6)	20.524 (**0.00001**)	418 (48.7)	440 (51.3)	0.850 (0.357)	391 (45.6)	467 (54.4)	11.428 (**0.001**)
> 2.0cm	600 (54.4)	502 (45.6)		603 (54.7)	499 (45.3)		560 (50.8)	542 (49.2)		587 (53.3)	515 (46.7)	
Tumour Grade												
Grade-I	66 (38.6)	105 (61.4)		57 (33.3)	114 (66.7)		73 (42.7)	98 (57.3)		59 (34.5)	112 (65.5)	
Grade-II	349 (45.3)	421 (54.7)	26.514 (**0.000002**)	332 (43.1)	438 (56.9)	71.887 (**2.4E-16**)	363 (47.1)	407 (52.9)	11.316 (**0.003**)	325 (42.2)	445 (57.8)	71.383 (**3.2E-16**)
Grade-III	527 (55.4)	425 (44.6)		572 (60.1)	380 (39.9)		511 (53.7)	441 (46.3)		569 (59.8)	383 (40.2)	
Nodal involvement, (NPI staging system)												
Stage-I	507 (49.0)	528 (51.0)		491 (47.4)	544 (52.6)		498 (48.1)	537 (51.9)		491 (47.4)	544 (52.6)	
Stage-II	319 (51.3)	303 (48.7)	1.036 (0.596)	330 (53.1)	292 (46.9)	5.845 (0.054)	336 (54.0)	286 (46.0)	5.687 (0.058)	331 (53.2)	291 (46.8)	5.742 (0.057)
Stage-III	162 (51.3)	154 (48.7)		166 (52.5)	150 (47.5)		154 (48.7)	162 (51.3)		164 (51.9)	152 (48.1)	
Nottingham Prognostic Index (NPI)												
good	309 (45.4)	371 (54.6)		271 (39.9)	409 (60.1)		316 (46.5)	364 (53.5)		281 (41.3)	399 (58.7)	
moderate	571 (51.9)	530 (48.1)	9.396 (**0.009**)	600 (54.5)	501 (45.5)	44.551 (**2.1E-10**)	570 (51.8)	531 (48.2)	5.177 (0.075)	592 (53.8)	509 (46.2)	32.889 (**7.2E-8**)
poor	110 (55.3)	89 (44.7)		119 (59.8)	80 (40.2)		104 (52.3)	95 (47.7)		117 (58.8)	82 (41.2)	

	FEZ1			*INKA1*			*ITSN1*			*LPAR1*		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Prediction analysis of microarray 50, PAM50												
Normal-like	45 (22.0)	160 (78.0)		42 (20.5)	163 (79.5)		53 (25.9)	152 (74.1)		49 (23.9)	156 (76.1)	
Luminal-A	293 (40.8)	425 (59.2)		260 (36.2)	458 (63.8)		340 (47.4)	378 (52.6)		286 (39.8)	432 (60.2)	
Luminal-B	339 (69.5)	149 (30.5)	169.510 (**1.3E-35**)	346 (70.9)	142 (29.1)	235.192 (**1.0E-49**)	297 (60.9)	191 (39.1)	76.636 (**9.0E-16**)	282 (57.8)	206 (42.2)	152.589 (**5.7E-32**)
Her2-enriched	126 (52.5)	114 (47.5)		149 (62.1)	91 (37.9)		118 (49.2)	122 (50.8)		155 (64.6)	85 (35.4)	
Basal	187 (56.8)	142 (43.2)		193 (58.7)	136 (41.3)		182 (55.3)	147 (44.7)		218 (66.3)	111 (33.7)	
Integrated classification clusters of METABRIC												
cluster-1	96 (69.1)	43 (30.9)		96 (69.1)	43 (30.9)		91 (65.5)	48 (34.5)		97 (69.8)	42 (30.2)	
cluster -2	46 (63.9)	26 (36.1)		44 (61.1)	28 (38.9)		40 (55.6)	32 (44.4)		27 (37.5)	45 (62.5)	
cluster -3	95 (32.8)	195 (67.2)		105 (36.2)	185 (63.8)		124 (42.8)	166 (57.2)		97 (33.4)	193 (66.6)	
cluster -4	80 (23.3)	263 (76.7)		71 (20.7)	272 (79.3)		122 (35.6)	221 (64.4)		87 (25.4)	256 (74.6)	
cluster -5	84 (44.2)	106 (55.8)	233.064 (**3.7E-45**)	125 (65.8)	65 (34.2)	265.297 (**5.8E-52**)	91 (47.9)	99 (52.1)	76.740 (**7.2E-13**)	119 (62.6)	71 (37.4)	284.865 (**4.2E-56**)
cluster -6	46 (54.1)	39 (45.9)		44 (51.8)	41 (48.2)		45 (52.9)	40 (47.1)		43 (50.6)	42 (49.4)	
cluster -7	105 (55.3)	85 (44.7)		77 (40.5)	113 (59.5)		84 (44.2)	106 (55.8)		85 (44.7)	105 (55.3)	
cluster -8	180 (60.2)	119 (39.8)		161 (53.8)	138 (46.2)		167 (55.9)	(132) 44.1		142 (47.5)	157 (52.5)	
cluster -9	95 (65.1)	51 (34.9)		96 (65.8)	50 (34.2)		80 (54.8)	66 (45.2)		104 (71.2)	42 (28.8)	
cluster -10	163 (72.1)	63 (27.9)		171 (75.7)	55 (24.3)		146 (64.6)	80 (35.4)		189 (83.6)	37 (16.4)	

	NBR1			*PRRX1*			*STAT3*			*THAP4*		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN$^{+/-}$	99 (60.4)	65 (39.6)	0.715 (0.398)	50 (30.5)	114 (69.5)	2.569 (0.109)	106 (64.6)	58 (35.4)	0.052 (0.819)	95 (57.9)	69 (42.1)	0.358 (0.549)
Definite/LN$^{+/-}$	86 (65.2)	46 (34.8)		52 (39.4)	80 (60.6)		87 (65.9)	45 (34.1)		81 (61.4)	51 (38.6)	
Age (years)												
≤ 50	239 (55.7)	190 (44.3)	7.145 (**0.008**)	190 (44.3)	239 (55.7)	7.145 (**0.008**)	233 (54.3)	196 (45.7)	4.074 (**0.044**)	255 (59.4)	174 (40.6)	19.524 (**0.00001**)
> 50	751 (48.4)	800 (51.6)		800 (51.6)	751 (48.4)		757 (48.8)	794 (51.2)		735 (47.4)	816 (52.6)	
Tumour size												
≤ 2.0cm	417 (48.6)	441 (51.4)	1.375 (0.241)	380 (44.3)	478 (55.7)	18.511 (**0.00002**)	401 (46.7)	457 (53.3)	6.101 (**0.014**)	430 (50.1)	428 (49.9)	0.044 (0.833)
> 2.0cm	565 (51.3)	537 (48.7)		596 (54.1)	506 (45.9)		577 (52.4)	525 (47.6)		547 (49.6)	555 (50.4)	
Tumour Grade												
Grade-I	68 (39.8)	103 (60.2)		72 (42.1)	99 (57.9)		68 (39.8)	103 (60.2)		54 (31.6)	117 (68.4)	
Grade-II	347 (45.1)	423 (54.9)	34.291 (**3.6E-8**)	362 (47.0)	408 (53.0)	10.427 (**0.005**)	339 (44.0)	431 (56.0)	36.992 (**9.3E-9**)	313 (40.6)	457 (59.4)	100.00 (**2.0E-22**)
Grade-III	545 (57.2)	407 (42.8)		505 (53.0)	447 (47.0)		543 (57.0)	409 (43.0)		585 (61.4)	367 (38.6)	
Nodal involvement, (NPI staging system)												
Stage-I	504 (48.7)	531 (51.3)		521 (50.3)	514 (49.7)		474 (45.8)	561 (54.2)		496 (47.9)	539 (52.1)	
Stage-II	325 (52.3)	297 (47.7)	2.003 (0.367)	308 (49.5)	314 (50.5)	0.113 (0.945)	324 (52.1)	298 (47.9)	21.357 (**0.00002**)	316 (50.8)	306 (49.2)	5.187 (0.075)
Stage-III	160 (50.6)	156 (49.4)		159 (50.3)	157 (49.7)		190 (60.1)	128 (39.9)		174 (55.1)	142 (44.9)	
Nottingham Prognostic Index (NPI)												
good	290 (42.6)	390 (57.4)		328 (48.2)	352 (51.8)		276 (40.6)	404 (59.4)		261 (38.4)	419 (61.6)	
moderate	588 (53.4)	513 (46.6)	22.956 (**0.00001**)	557 (50.6)	544 (49.4)	1.609 (0.447)	585 (53.1)	516 (46.9)	45.911 (**1.0E-10**)	608 (55.2)	493 (44.8)	58.015 (**2.5E-13**)
poor	112 (56.3)	87 (43.7)		105 (52.8)	94 (47.2)		129 (64.8)	70 (35.2)		121 (60.8)	78 (39.2)	

	NBR1			*PRRX1*			*STAT3*			*THAP4*		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Prediction analysis of microarray 50, PAM50												
Normal-like	85 (41.5)	120 (58.5)		73 (35.6)	132 (64.4)		87 (42.4)	118 (57.6)		118 (57.6)	87 (42.4)	
Luminal-A	286 (39.8)	432 (60.2)		311 (43.3)	407 (56.7)		262 (36.5)	456 (63.5)		242 (33.7)	476 (66.3)	
Luminal-B	236 (48.4)	252 (51.6)	108.175 (**1.8E-22**)	302 (61.9)	186 (38.1)	62.013 (**1.1E-12**)	291 (59.6)	197 (40.4)	141.438 (**1.4E-29**)	214 (43.9)	274 (56.1)	244.496 (**1.3E-44**)
Her2-enriched	173 (72.1)	67 (27.9)		120 (50.0)	120 (50.0)		183 (76.3)	57 (23.8)		168 (70.0)	72 (30.0)	
Basal	210 (63.8)	119 (36.2)		184 (55.9)	145 (44.1)		167 (50.8)	162 (49.2)		248 (75.4)	81 (24.6)	
Integrated classification clusters of METABRIC												
cluster-1	84 (60.4)	55 (39.6)		89 (64.0)	50 (36.0)		102 (73.4)	37 (26.6)		61 (43.9)	78 (56.1)	
cluster -2	26 (36.1)	46 (63.9)		44 (61.1)	28 (38.9)		40 (55.6)	32 (44.4)		32 (44.4)	40 (55.6)	
cluster -3	137 (47.2)	153 (52.8)		99 (34.1)	191 (65.9)		103 (35.5)	187 (64.5)		111 (38.3)	179 (61.7)	
cluster -4	148 (43.1)	195 (56.9)		110 (32.1)	233 (67.9)		158 (46.1)	185 (53.9)		226 (65.9)	117 (34.1)	
cluster -5	130 (68.4)	60 (31.6)	92.850 (**4.4E-16**)	82 (43.2)	108 (56.8)	132.076 (**4.4E-24**)	133 (70.0)	57 (30.0)	132.542 (**3.6E-24**)	135 (71.1)	55 (28.9)	240.145 (**1.2E-46**)
cluster -6	35 (41.2)	50 (58.8)		46 (54.1)	39 (45.9)		56 (65.9)	29 (34.1)		50 (58.8)	35 (41.2)	
cluster -7	73 (38.4)	117 (61.6)		107 (56.3)	83 (43.7)		91 (47.9)	99 (52.1)		59 (31.1)	131 (68.9)	
cluster -8	126 (42.1)	173 (57.9)		177 (59.2)	122 (40.8)		106 (35.5)	193 (64.5)		74 (24.7)	225 (75.3)	
cluster -9	79 (54.1)	67 (45.9)		91 (62.3)	55 (37.7)		92 (63.0)	54 (37.0)		78 (53.4)	68 (46.6)	
cluster -10	152 (67.3)	74 (32.7)		145 (64.2)	81 (35.8)		109 (48.2)	117 (51.8)		164 (72.6)	62 (27.4)	

	TMEM98			*TNC*			*TNFRSF19*			*TSC1*		
	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**	-	+	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Lymphovascular invasion (LVI), (H&E/IHC)												
Negative/LN$^{+/-}$	74 (45.1)	90 (54.9)	2.198 (0.138)	68 (41.5)	96 (58.5)	0.475 (0.491)	73 (44.5)	91 (55.5)	0.178 (0.673)	53 (32.3)	111 (67.7)	0.747 (0.387)
Definite/LN$^{+/-}$	71 (53.8)	61 (46.2)		60 (45.5)	72 (54.5)		62 (47.0)	70 (53.0)		49 (37.1)	83 (62.9)	
Age (years)												
≤ 50	194 (45.2)	235 (54.8)	5.002 (**0.025**)	159 (37.1)	270 (62.9)	36.664 (**1.4E-9**)	203 (47.3)	226 (52.7)	1.574 (0.210)	223 (52.0)	206 (48.0)	0.860 (0.354)
> 50	796 (51.3)	755 (48.7)		831 (53.6)	720 (46.4)		787 (50.7)	764 (49.3)		767 (49.5)	784 (50.0)	
Tumour size												
≤ 2.0cm	413 (48.1)	445 (51.9)	2.008 (0.156)	393 (45.8)	465 (54.2)	10.746 (**0.001**)	393 (45.8)	465 (54.2)	10.230 (**0.001**)	394 (45.9)	464 (54.1)	10.672 (**0.001**)
> 2.0cm	566 (51.4)	536 (48.6)		587 (53.3)	515 (46.7)		585 (53.1)	517 (46.9)		588 (53.4)	514 (46.6)	
Tumour Grade												
Grade-I	63 (36.8)	108 (63.2)		77 (45.0)	94 (55.0)		59 (34.5)	112 (65.5)		61 (35.7)	110 (64.3)	
Grade-II	343 (44.5)	427 (55.5)	39.854 (**2.2E-9**)	398 (51.7)	372 (48.3)	3.126 (0.210)	336 (43.6)	434 (56.4)	53.807 (**2.1E-12**)	331 (43.0)	439 (57.0)	59.460 (**1.2E-13**)
Grade-III	543 (57.0)	409 (43.0)		463 (48.6)	489 (51.4)		553 (58.1)	399 (41.9)		561 (58.9)	391 (41.1)	
Nodal involvement, (NPI staging system)												
Stage-I	494 (47.7)	541 (52.3)		491 (47.4)	544 (52.6)		482 (46.6)	553 (53.4)		468 (45.2)	567 (54.8)	
Stage-II	329 (52.9)	293 (47.1)	5.015 (0.081)	329 (52.9)	293 (47.1)	6.059 (**0.048**)	326 (52.4)	296 (47.6)	11.899 (**0.003**)	340 (54.7)	282 (45.3)	19.443 (**0.0001**)
Stage-III	166 (52.5)	150 (47.5)		168 (53.2)	148 (46.8)		179 (56.6)	137 (43.4)		177 (56.0)	139 (44.0)	
Nottingham Prognostic Index (NPI)												
good	290 (42.6)	390 (57.4)		334 (49.1)	346 (50.9)		269 (39.6)	411 (60.4)		264 (38.8)	416 (61.2)	
moderate	595 (54.0)	506 (46.0)	22.508 (**0.00001**)	554 (50.3)	547 (49.7)	0.382 (0.826)	602 (54.7)	499 (45.3)	46.932 (**6.4E-11**)	604 (54.9)	497 (45.1)	54.551 (**1.4E-12**)
poor	105 (52.8)	94 (47.2)		102 (51.3)	97 (48.7)		119 (59.8)	80 (40.2)		122 (61.3)	77 (38.7)	

	TMEM98			*TNC*			*TNFRSF19*			*TSC1*		
	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)	-	+	χ2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Prediction analysis of microarray 50, PAM50												
Normal-like	78 (38.0)	127 (62.0)		53 (25.9)	152 (74.1)		91 (44.4)	114 (55.6)		88 (42.9)	117 (57.1)	
Luminal-A	293 (40.8)	425 (59.2)		380 (52.9)	338 (47.1)		249 (34.7)	469 (65.3)		282 (39.3)	436 (60.7)	
Luminal-B	255 (52.3)	233 (47.7)	106.431 (**4.2E-22**)	329 (67.4)	159 (32.6)	133.697 (**6.3E-28**)	278 (57.0)	210 (43.0)	134.358 (**4.6E-28**)	241 (49.4)	247 (50.6)	120.780 (**3.6E-25**)
Her2-enriched	183 (76.3)	57 (23.8)		104 (43.3)	136 (56.7)		163 (67.9)	77 (32.1)		133 (55.4)	107 (44.6)	
Basal	181 (55.0)	148 (45.0)		124 (37.7)	205 (62.3)		209 (63.5)	120 (36.5)		246 (74.8)	83 (25.2)	
Integrated classification clusters of METABRIC												
cluster-1	87 (62.6)	52 (37.4)		96 (69.1)	43 (30.9)		100 (71.9)	39 (28.1)		62 (44.6)	77 (55.4)	
cluster -2	39 (54.2)	33 (45.8)		46 (63.9)	26 (36.1)		42 (58.3)	30 (41.7)		49 (68.1)	23 (31.9)	
cluster -3	106 (36.6)	184 (63.4)		144 (49.7)	146 (50.3)		92 (31.7)	198 (68.3)		104 (35.9)	186 (64.1)	
cluster -4	163 (47.5)	180 (52.5)		112 (32.7)	231 (67.3)		156 (45.5)	187 (54.5)		163 (47.5)	180 (52.5)	
cluster -5	122 (64.2)	68 (35.8)	97.948 (**4.1E-17**)	51 (26.8)	139 (73.2)	138.918 (**1.7E-25**)	104 (54.7)	86 (45.3)	114.997 (**1.4E-20**)	102 (53.7)	88 (46.3)	159.424 (**9.8E-30**)
cluster -6	52 (61.2)	33 (38.8)		48 (56.5)	37 (43.5)		44 (51.8)	41 (48.2)		48 (56.5)	37 (43.5)	
cluster -7	105 (55.3)	85 (44.7)		111 (58.4)	79 (41.6)		90 (47.4)	100 (52.6)		80 (42.1)	110 (57.9)	
cluster -8	99 (33.1)	200 (66.9)		184 (61.5)	115 (38.5)		123 (41.1)	176 (58.9)		108 (36.1)	191 (63.9)	
cluster -9	89 (61.0)	57 (39.0)		90 (61.6)	56 (38.4)		86 (58.9)	60 (41.1)		92 (63.0)	54 (37.0)	
cluster -10	128 (56.6)	98 (43.4)		108 (47.8)	118 (52.2)		153 (67.7)	73 (32.3)		182 (80.5)	44 (19.5)	

	TWIST2			*WISP2*		
	-	+	**χ^2 (*p*-value≤)**	-	+	**χ^2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**	
Lymphovascular invasion (LVI), (H&E/IHC)						
Negative/LN$^{+/-}$	65 (39.6)	99 (60.4)	1.016 (0.314)	52 (31.7)	112 (68.3)	3.625 (0.057)
Definite/LN$^{+/-}$	60 (45.5)	72 (54.5)		56 (42.4)	76 (57.6)	
Age (years)						
≤ 50	178 (41.5)	251 (58.5)	15.858 (**0.00007**)	216 (50.3)	213 (49.7)	0.027 (0.870)
> 50	812 (52.4)	739 (47.6)		774 (49.9)	777 (50.1)	
Tumour size						
≤ 2.0cm	392 (45.7)	466 (54.3)	11.622 (**0.001**)	405 (47.2)	453 (52.8)	4.105 (**0.043**)
> 2.0cm	589 (53.4)	513 (46.6)		571 (51.8)	531 (48.2)	
Tumour Grade						
Grade-I	69 (40.4)	102 (59.6)	10.184 (**0.006**)	64 (37.4)	107 (62.6)	25.265 (**0.000003**)
Grade-II	371 (48.2)	399 (51.8)		356 (46.2)	414 (53.8)	
Grade-III	502 (52.7)	450 (47.3)		525 (55.1)	427 (44.9)	
Nodal involvement, (NPI staging system)*						
Stage-I	513 (49.6)	522 (50.4)	0.882 (0.643)	526 (50.8)	509 (49.2)	0.486 (0.784)
Stage-II	310 (49.8)	312 (50.2)		306 (49.2)	316 (50.8)	
Stage-III	166 (52.5)	150 (47.5)		156 (49.4)	160 (50.6)	
Nottingham Prognostic Index (NPI)						
good	320 (47.1)	360 (52.9)	3.642 (0.162)	317 (46.6)	363 (53.4)	4.950 (0.084)
moderate	569 (51.7)	532 (48.3)		567 (51.5)	534 (48.5)	
poor	101 (50.8)	98 (49.2)		106 (53.3)	93 (46.7)	

	TWIST2			*WISP2*		
	-	**+**	**χ2 (*p*-value≤)**	**-**	**+**	**χ2 (*p*-value≤)**
	n (%)	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50						
Normal-like	37 (18.0)	168 (82.0)		91 (44.4)	114 (55.6)	
Luminal-A	324 (45.1)	394 (54.9)		300 (41.8)	418 (58.2)	
Luminal-B	336 (68.9)	152 (31.1)	169.163 (**1.6E-35**)	259 (53.1)	229 (46.9)	50.654 (**2.6E-10**)
Her2-enriched	141 (58.8)	99 (41.3)		130 (54.2)	110 (45.8)	
Basal	152 (46.2)	177 (53.8)		210 (63.8)	119 (36.2)	
Integrated classification clusters of METABRIC						
cluster-1	89 (64.0)	50 (36.0)		87 (62.6)	52 (37.4)	
cluster -2	43 (59.7)	29 (40.3)		39 (54.2)	33 (45.8)	
cluster -3	97 (33.4)	193 (66.6)		121 (41.7)	169 (58.3)	
cluster -4	106 (30.9)	237 (69.1)		132 (38.5)	211 (61.5)	
cluster -5	91 (47.9)	99 (52.1)	137.919 (**2.8E-25**)	97 (44.7)	47 (55.3)	77.598 (**4.8E-13**)
cluster -6	48 (56.5)	37 (43.5)		38 (44.7)	47 (55.3)	
cluster -7	109 (57.4)	81 (42.6)		92 (48.4)	98 (51.6)	
cluster -8	191 (63.9)	108 (36.1)		148 (49.5)	151 (50.5)	
cluster -9	94 (64.4)	52 (35.6)		75 (51.4)	71 (48.6)	
cluster -10	122 (54.0)	104 (46.0)		161 (71.2)	65 (28.8)	

Table 3. 11: Multivariate Cox regression analysis for prediction of breast cancer specific survival, UP-22/LVI$^-$

Breast cancer specific survival (BCSS), n= 1869$^{\lozenge}$

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
ANXA5 mRNA	0.986	0.758	1.282	0.916
Tumour grade	1.370	1.055	1.779	0.018
Nodal staging*	1.793	1.506	2.135	0.000
Tumour size	1.509	1.138	2.000	0.004
Estrogen expression**	0.717	0.530	0.969	0.031
Her2 expression***	1.242	1.051	1.468	0.011

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***APCDD1* mRNA**	1.062	0.817	1.381	0.652
Tumour grade	1.369	1.055	1.777	0.018
Nodal staging*	1.800	1.510	2.145	0.000
Tumour size	1.509	1.138	2.000	0.004
Estrogen expression**	0.726	0.537	0.982	0.037
Her2 expression***	1.241	1.050	1.466	0.011

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***C17orf81* mRNA**	0.888	0.678	1.162	0.386
Tumour grade	1.354	1.041	1.760	0.024
Nodal staging*	1.782	1.496	2.122	0.000
Tumour size	1.512	1.141	2.003	0.004
Estrogen expression**	0.728	0.540	0.982	0.037
Her2 expression***	1.231	1.041	1.457	0.015

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***C1S* mRNA**	1.045	0.802	1.360	0.746
Tumour grade	1.373	1.058	1.781	0.017
Nodal staging*	1.794	1.506	2.136	0.000
Tumour size	1.514	1.142	2.007	0.004
Estrogen expression**	0.725	0.536	0.982	0.038
Her2 expression***	1.240	1.049	1.465	0.012

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
$^{\lozenge}$ Sum of patients does not equal to 1980 because of missing cases among different categories.

Breast cancer specific survival (BCSS), n= 1869[◊]

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***CNRIP1* mRNA**	1.063	0.819	1.380	0.645
Tumour grade	1.377	1.061	1.788	0.016
Nodal staging*	1.791	1.504	2.133	0.000
Tumour size	1.516	1.143	2.011	0.004
Estrogen expression**	0.714	0.529	0.963	0.027
Her2 expression***	1.237	1.046	1.463	0.013

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***CXCL12* mRNA**	0.787	0.604	1.025	0.075
Tumour grade	1.321	1.015	1.719	0.038
Nodal staging*	1.801	1.513	2.145	0.000
Tumour size	1.495	1.128	1.981	0.005
Estrogen expression**	0.727	0.539	0.981	0.037
Her2 expression***	1.248	1.057	1.474	0.009

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***ECM2* mRNA**	0.821	0.632	1.066	0.139
Tumour grade	1.342	1.032	1.744	0.028
Nodal staging*	1.802	1.514	2.146	0.000
Tumour size	1.494	1.127	1.981	0.005
Estrogen expression**	0.733	0.543	0.989	0.042
Her2 expression***	1.242	1.051	1.467	0.011

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***FBLN1* mRNA**	0.914	0.702	1.190	0.503
Tumour grade	1.360	1.047	1.766	0.021
Nodal staging*	1.790	1.503	2.132	0.000
Tumour size	1.497	1.128	1.987	0.005
Estrogen expression**	0.727	0.538	0.981	0.037
Her2 expression***	1.242	1.051	1.467	0.011

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
[◊] Sum of patients does not equal to 1980 because of missing cases among different categories.

Breast cancer specific survival (BCSS), n= 1869[◊]

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***FEZ1* mRNA**	1.147	0.882	1.491	0.306
Tumour grade	1.381	1.064	1.792	0.015
Nodal staging*	1.809	1.518	2.156	0.000
Tumour size	1.526	1.151	2.024	0.003
Estrogen expression**	0.709	0.525	0.955	0.024
Her2 expression***	1.230	1.040	1.455	0.016

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***INKA1* mRNA**	0.858	0.661	1.115	0.253
Tumour grade	1.351	1.040	1.754	0.024
Nodal staging*	1.793	1.506	2.135	0.000
Tumour size	1.505	1.136	1.994	0.004
Estrogen expression**	0.724	0.537	0.976	0.034
Her2 expression***	1.234	1.044	1.459	0.014

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***ITSN1* mRNA**	1.155	0.891	1.497	0.276
Tumour grade	1.376	1.060	1.785	0.016
Nodal staging*	1.799	1.511	2.141	0.000
Tumour size	1.521	1.147	2.016	0.004
Estrogen expression**	0.709	0.526	0.956	0.024
Her2 expression***	1.248	1.056	1.476	0.009

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***LPAR1* mRNA**	0.659	0.505	0.859	**0.002**
Tumour grade	1.331	1.027	1.726	0.031
Nodal staging*	1.785	1.499	2.126	0.000
Tumour size	1.523	1.151	2.015	0.003
Estrogen expression**	0.757	0.561	1.020	0.068
Her2 expression***	1.238	1.048	1.463	0.012

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
[◊] Sum of patients does not equal to 1980 because of missing cases among different categories.

Breast cancer specific survival (BCSS), n= 1869[◊]

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
NBR1 mRNA	0.743	0.566	0.975	**0.032**
Tumour grade	1.341	1.035	1.739	0.027
Nodal staging*	1.791	1.504	2.134	0.000
Tumour size	1.529	1.154	2.025	0.003
Estrogen expression**	0.750	0.556	1.011	0.059
Her2 expression***	1.232	1.043	1.455	0.014

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***PRRX1* mRNA**	1.034	0.797	1.341	0.802
Tumour grade	1.373	1.058	1.782	0.017
Nodal staging*	1.793	1.506	2.135	0.000
Tumour size	1.517	1.142	2.015	0.004
Estrogen expression**	0.719	0.534	0.969	0.030
Her2 expression***	1.241	1.050	1.466	0.011

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***STAT3* mRNA**	0.797	0.611	1.040	0.095
Tumour grade	1.359	1.048	1.762	0.021
Nodal staging*	1.773	1.488	2.112	0.000
Tumour size	1.519	1.146	2.013	0.004
Estrogen expression**	0.727	0.540	0.980	0.036
Her2 expression***	1.220	1.032	1.444	0.020

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***THAP4* mRNA**	0.872	0.666	1.141	0.318
Tumour grade	1.347	1.036	1.752	0.026
Nodal staging*	1.795	1.507	2.137	0.000
Tumour size	1.529	1.153	2.029	0.003
Estrogen expression**	0.737	0.545	0.996	0.047
Her2 expression***	1.235	1.045	1.460	0.013

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
[◊] Sum of patients does not equal to 1980 because of missing cases among different categories.

Breast cancer specific survival (BCSS), n= 1869[◊]

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
TMEM98 mRNA	0.698	0.533	0.913	**0.009**
Tumour grade	1.351	1.042	1.752	0.023
Nodal staging*	1.794	1.508	2.134	0.000
Tumour size	1.559	1.177	2.065	0.002
Estrogen expression**	0.757	0.560	1.023	0.070
Her2 expression***	1.225	1.037	1.447	0.017

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***TNC* mRNA**	0.899	0.690	1.170	0.427
Tumour grade	1.367	1.053	1.774	0.019
Nodal staging*	1.792	1.505	2.134	0.000
Tumour size	1.505	1.136	1.995	0.004
Estrogen expression**	0.704	0.520	0.952	0.023
Her2 expression***	1.251	1.058	1.480	0.009

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***TNFRSF19* mRNA**	1.186	0.914	1.540	0.200
Tumour grade	1.394	1.074	1.811	0.013
Nodal staging*	1.798	1.510	2.141	0.000
Tumour size	1.506	1.136	1.996	0.004
Estrogen expression**	0.704	0.522	0.949	0.021
Her2 expression***	1.240	1.050	1.465	0.011

	Hazard ratio (95% CI)	Lower	Upper	*p*-value
***TSC1* mRNA**	0.914	0.700	1.193	0.509
Tumour grade	1.361	1.048	1.768	0.021
Nodal staging*	1.789	1.502	2.131	0.000
Tumour size	1.513	1.142	2.005	0.004
Estrogen expression**	0.733	0.541	0.992	0.044
Her2 expression***	1.242	1.052	1.468	0.011

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
[◊] Sum of patients does not equal to 1980 because of missing cases among different categories.

Breast cancer specific survival (BCSS), n= 1869◊

	Hazard ratio (95% CI)	Lower	Upper	*p*-value		Hazard ratio (95% CI)	Lower	Upper	*p*-value
TWIST2 mRNA	0.835	0.645	1.082	0.174	***WISP2* mRNA**	0.989	0.763	1.281	0.931
Tumour grade	1.367	1.054	1.773	0.019	**Tumour grade**	1.370	1.055	1.779	0.018
Nodal staging*	1.794	1.508	2.135	0.000	**Nodal staging***	1.794	1.506	2.136	0.000
Tumour size	1.497	1.130	1.984	0.005	**Tumour size**	1.509	1.138	2.001	0.004
Estrogen expression**	0.711	0.528	0.959	0.025	**Estrogen expression****	0.719	0.534	0.969	0.030
Her2 expression***	1.248	1.057	1.475	0.009	**Her2 expression*****	1.241	1.051	1.467	0.011

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
◊ Sum of patients does not equal to 1980 because of missing cases among different categories.

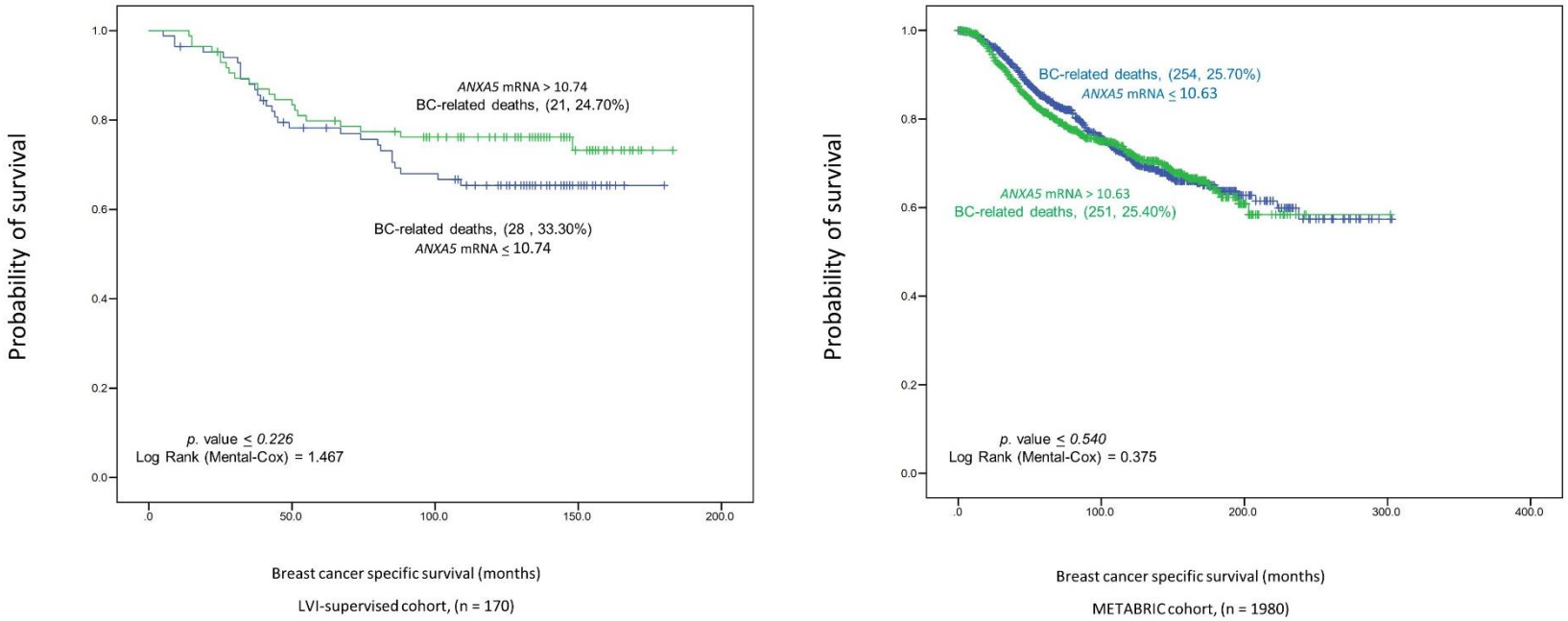


Figure 3. 18: Associations between the altered expression of *ANXA5* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

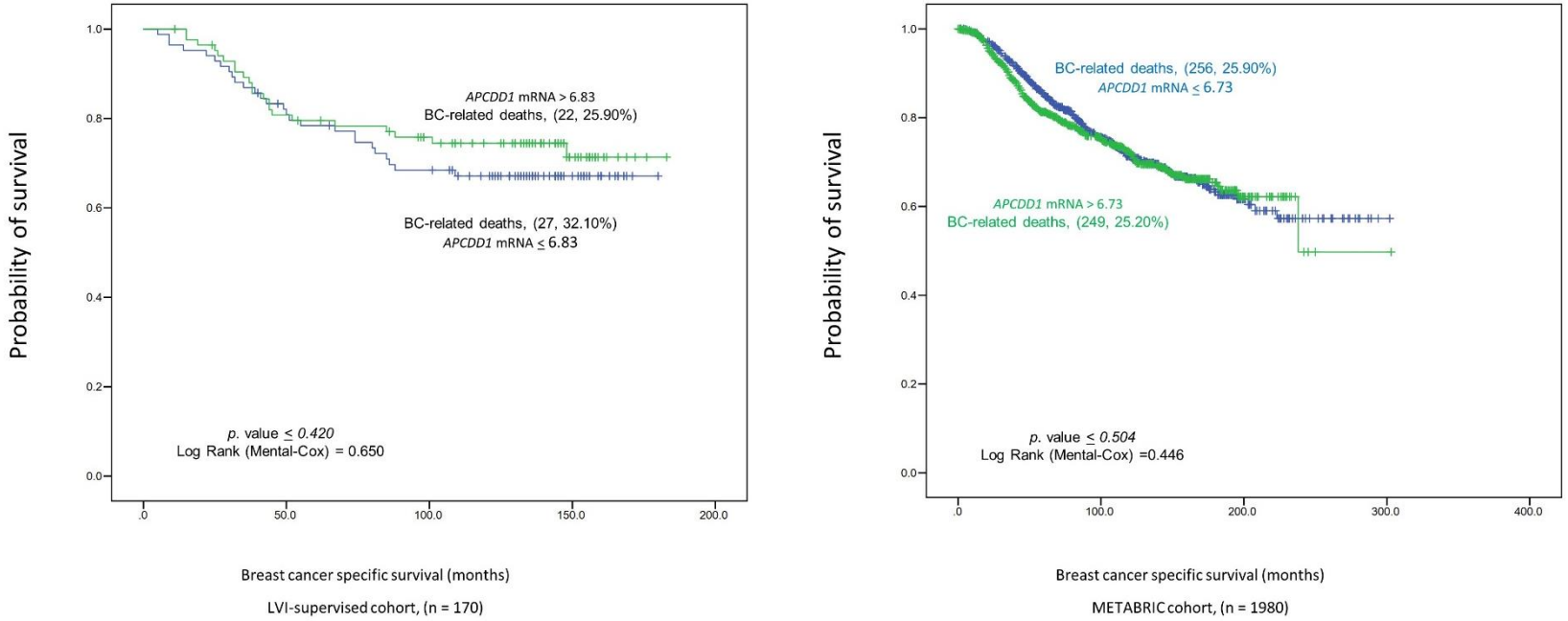


Figure 3. 19: Associations between the altered expression of *APCDD1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

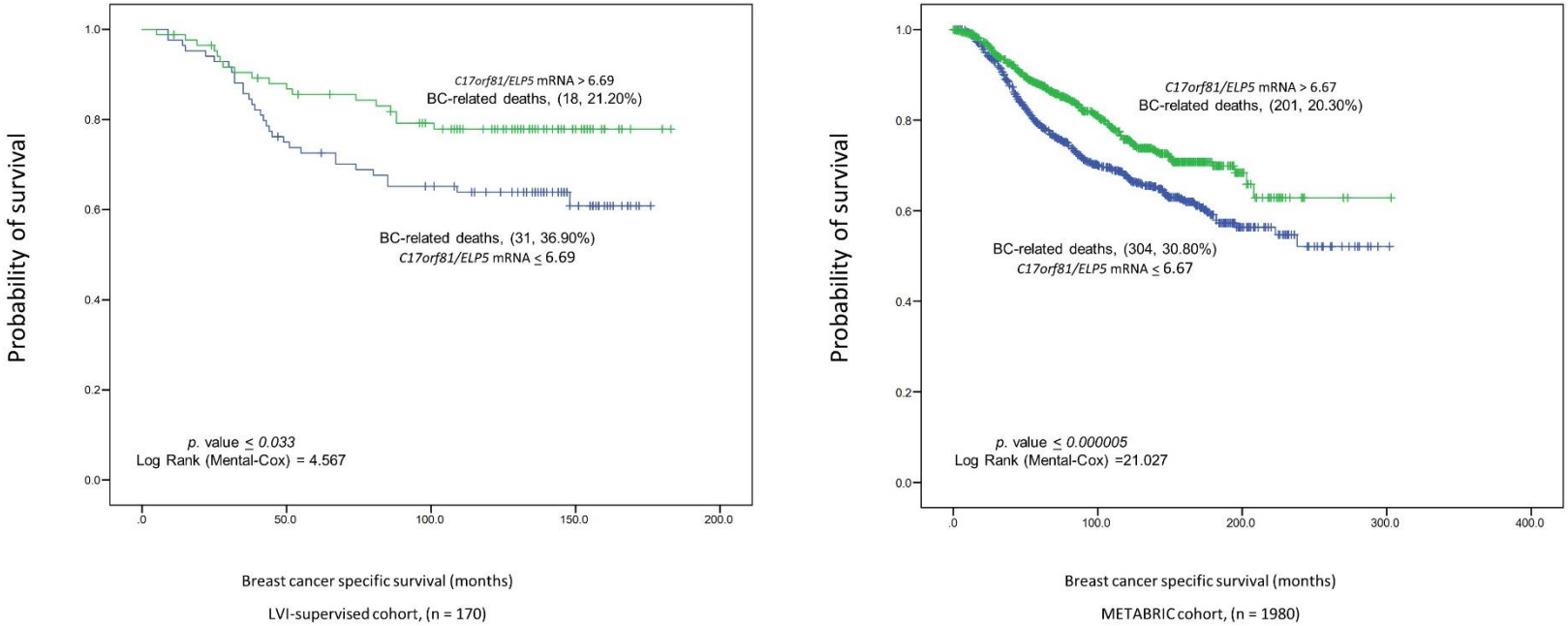


Figure 3. 20: Associations between the altered expression of *C17orf81/ELP5* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

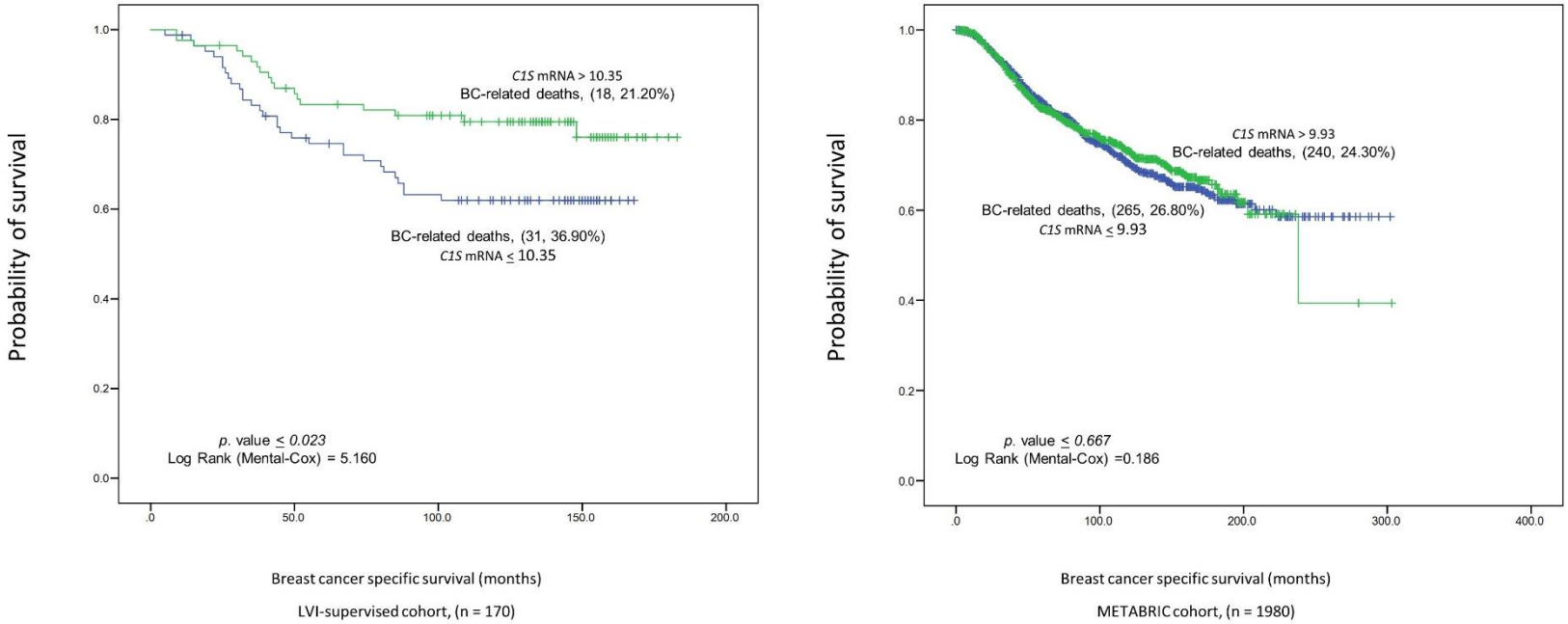


Figure 3. 21: Associations between the altered expression of *C1S* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

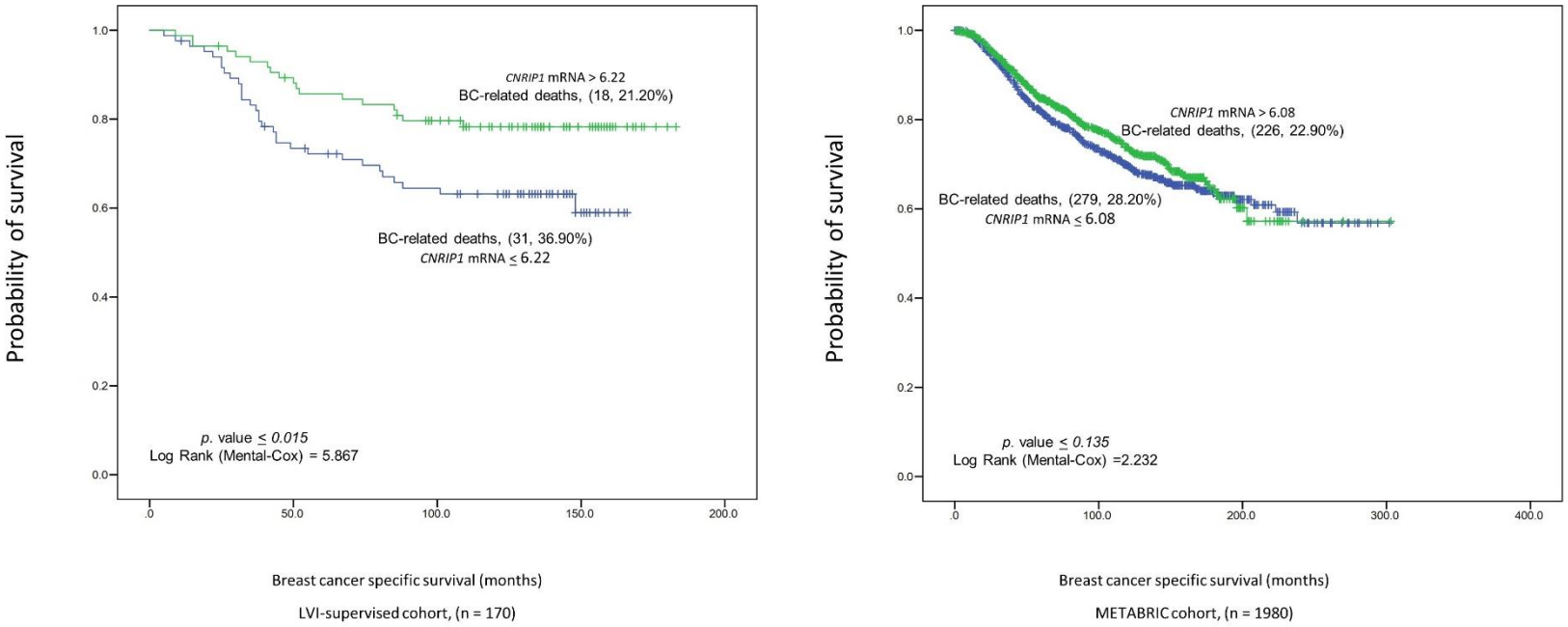


Figure 3. 22: Associations between the altered expression of *CNRIP1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

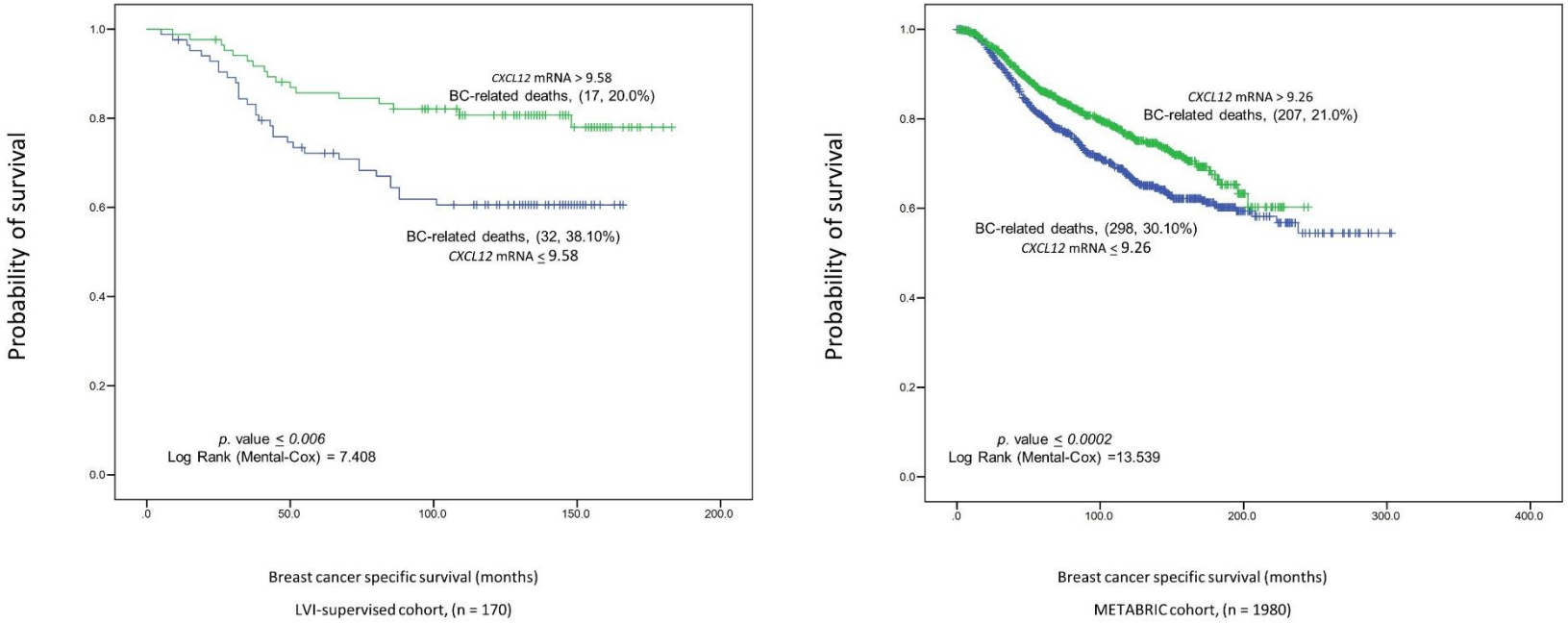


Figure 3. 23: Associations between the altered expression of *CXCL12* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

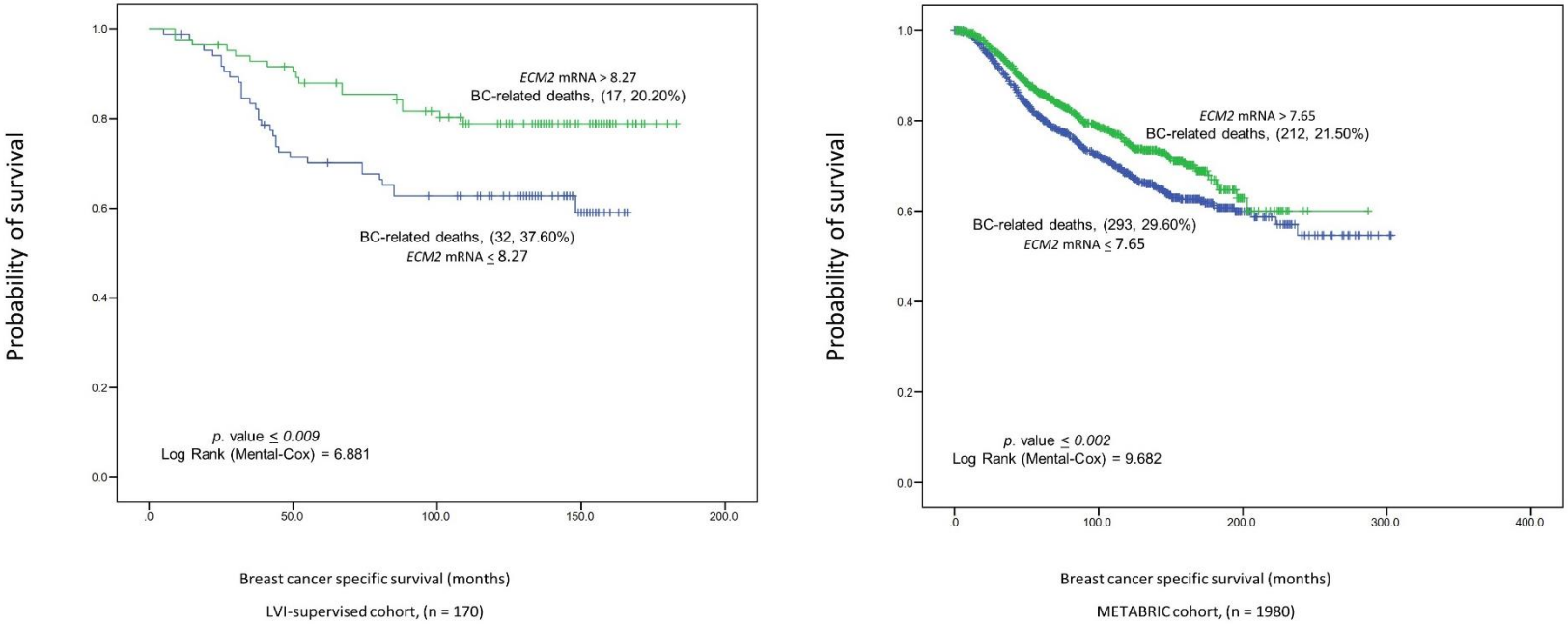


Figure 3. 24: Associations between the altered expression of *ECM2* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

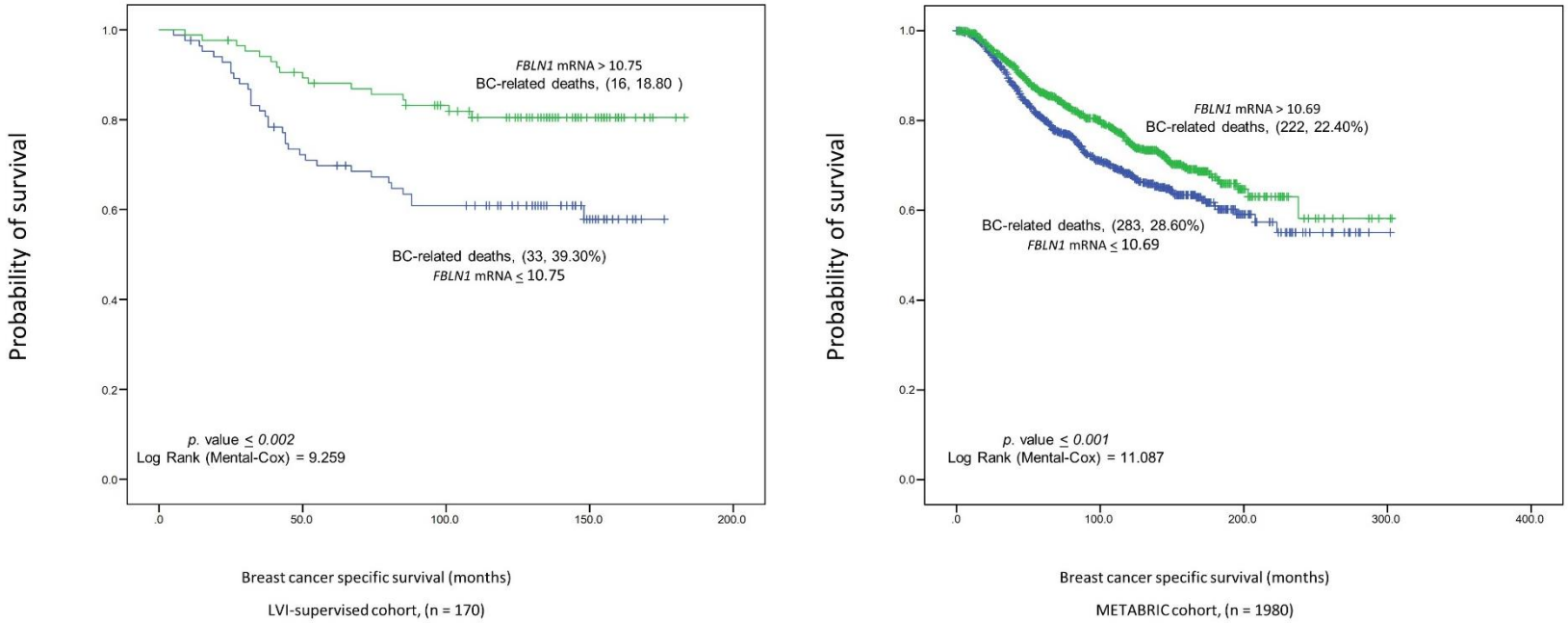


Figure 3. 25: Associations between the altered expression of *FBLN1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

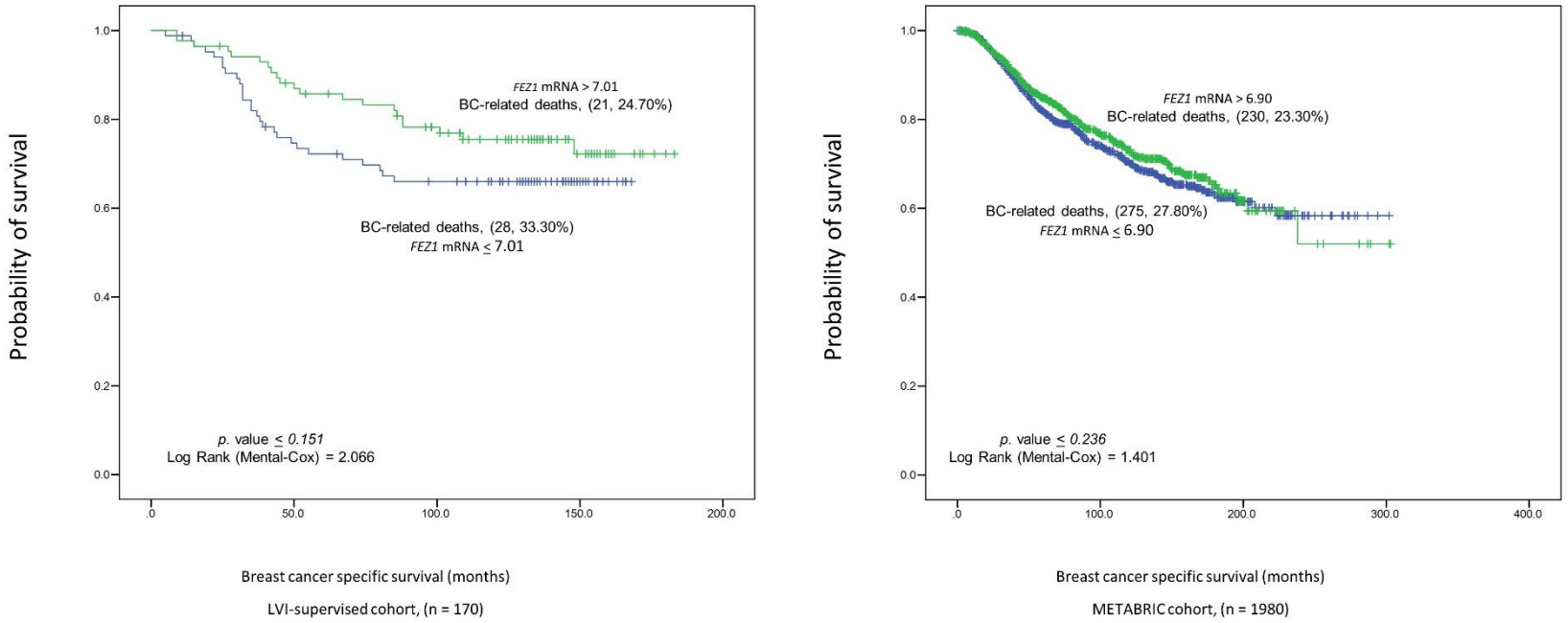


Figure 3. 26: Associations between the altered expression of *FEZ1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

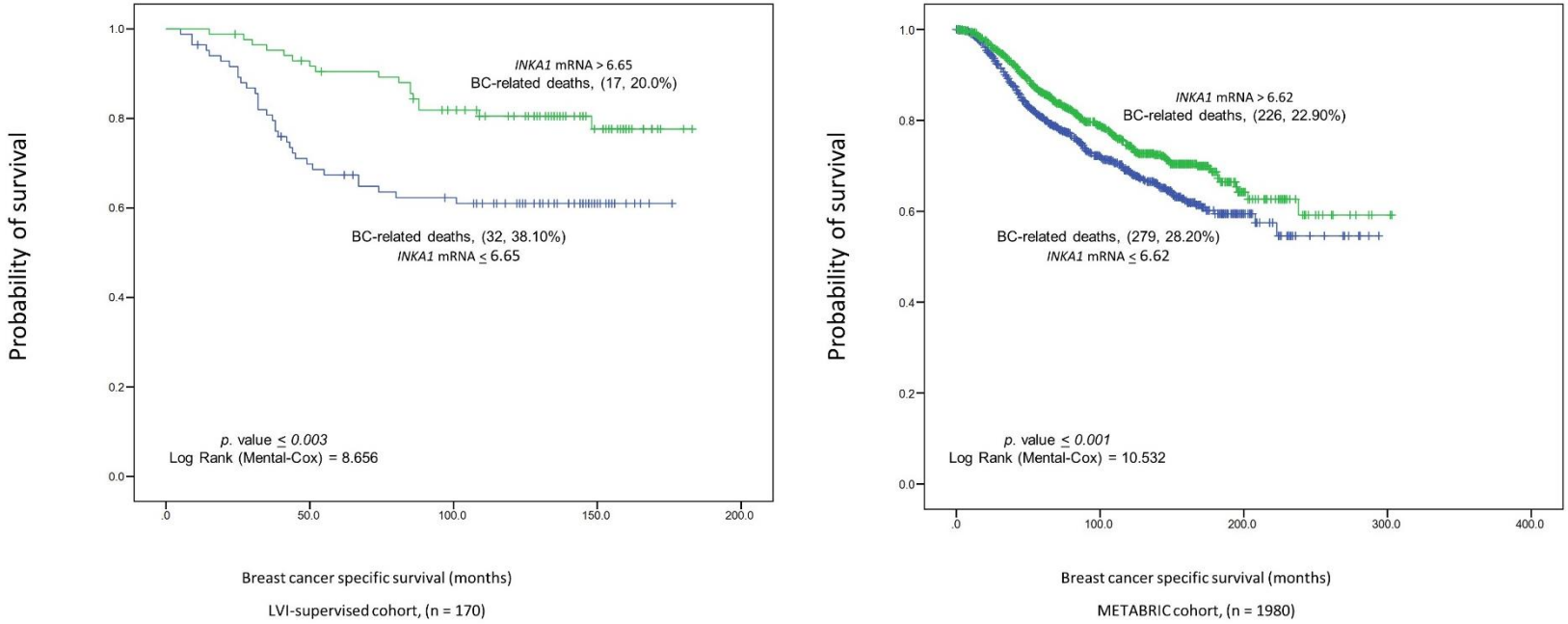


Figure 3. 27: Associations between the altered expression of *INKA1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

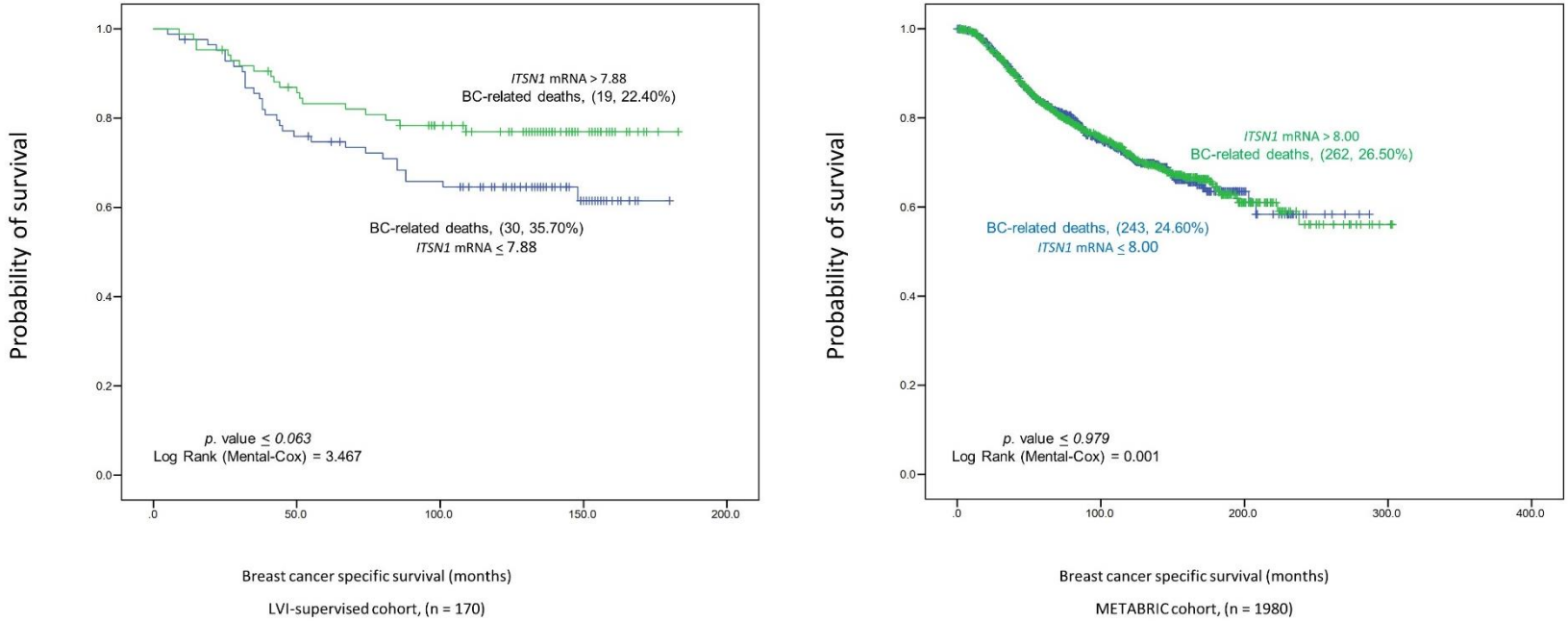


Figure 3. 28: Associations between the altered expression of *ITSN1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

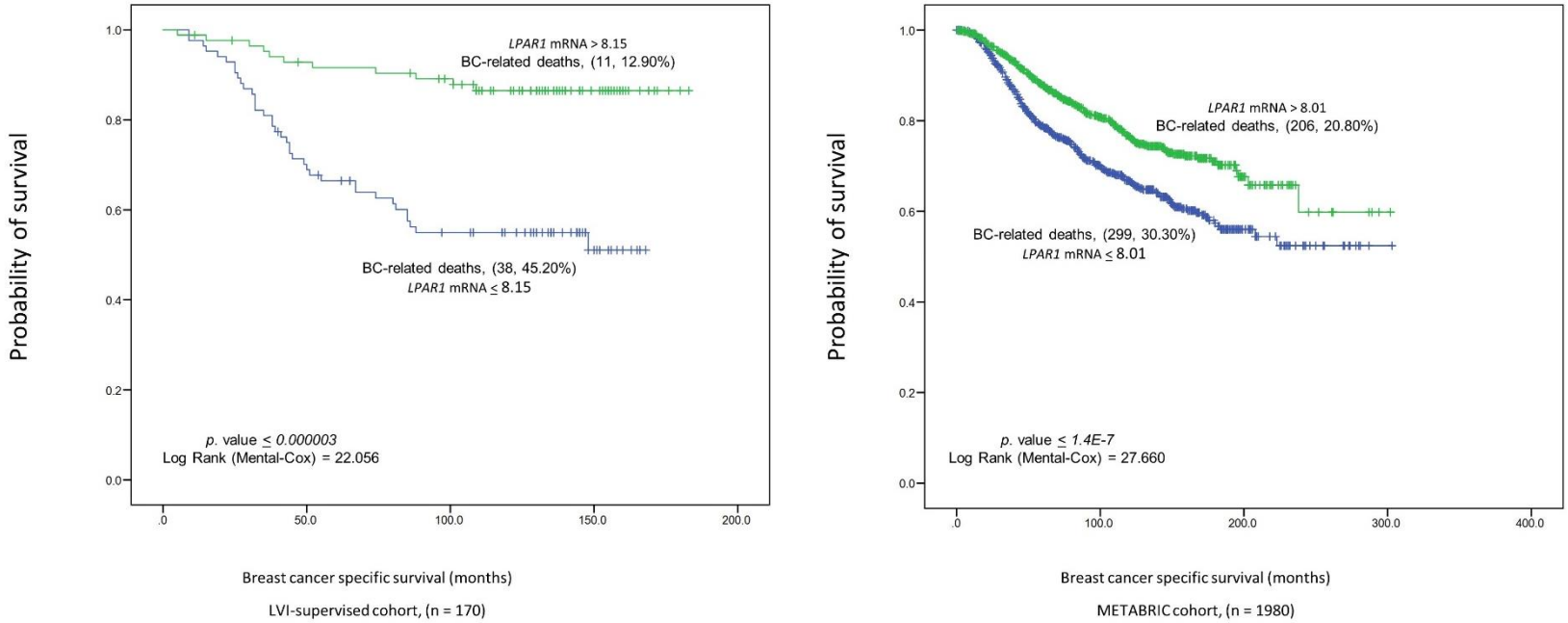


Figure 3. 29: Associations between the altered expression of *LPAR1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

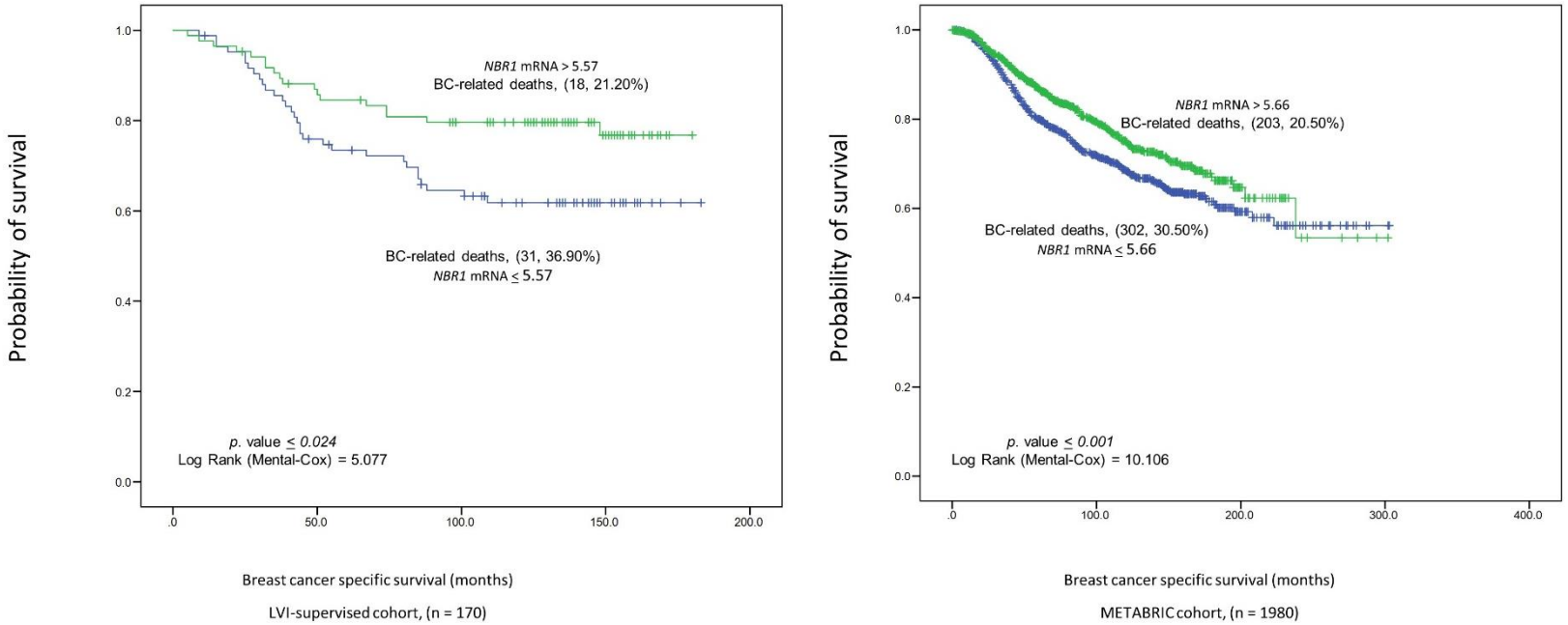


Figure 3. 30: Associations between the altered expression of *NBR1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

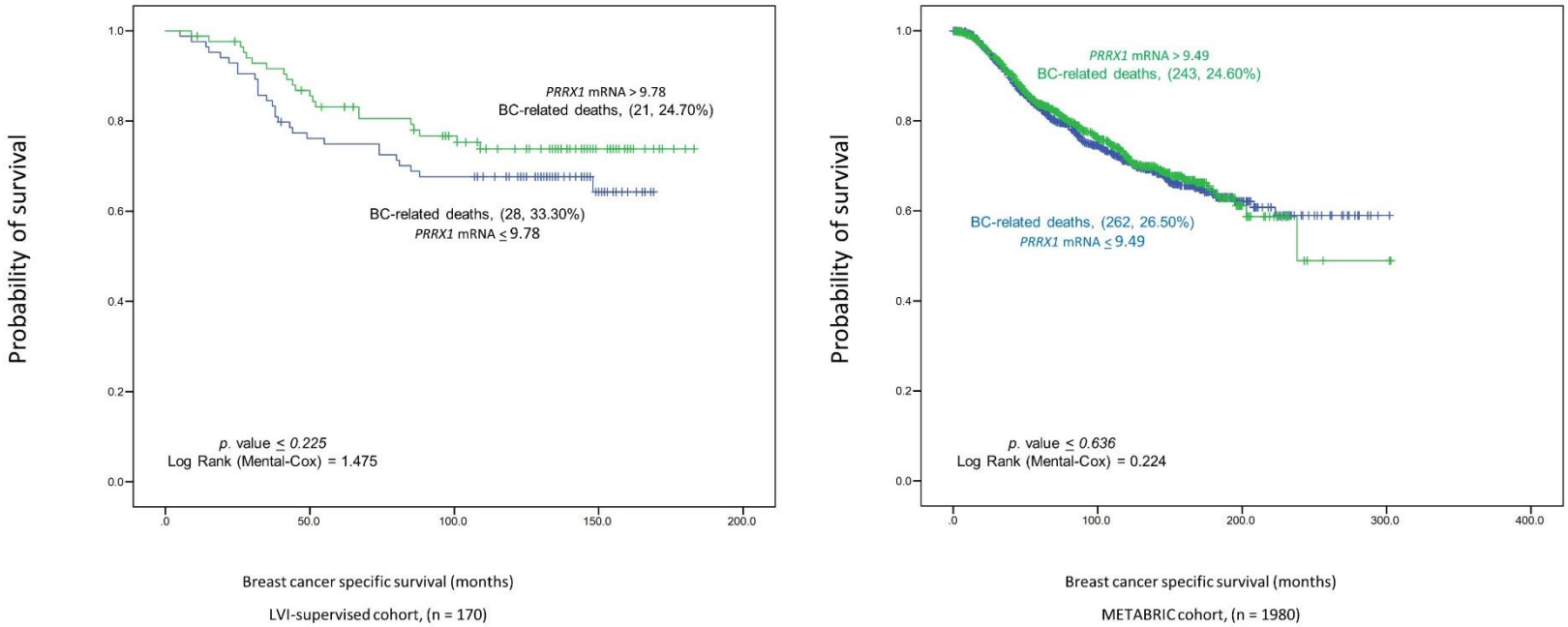


Figure 3. 31: Associations between the altered expression of *PRRX1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

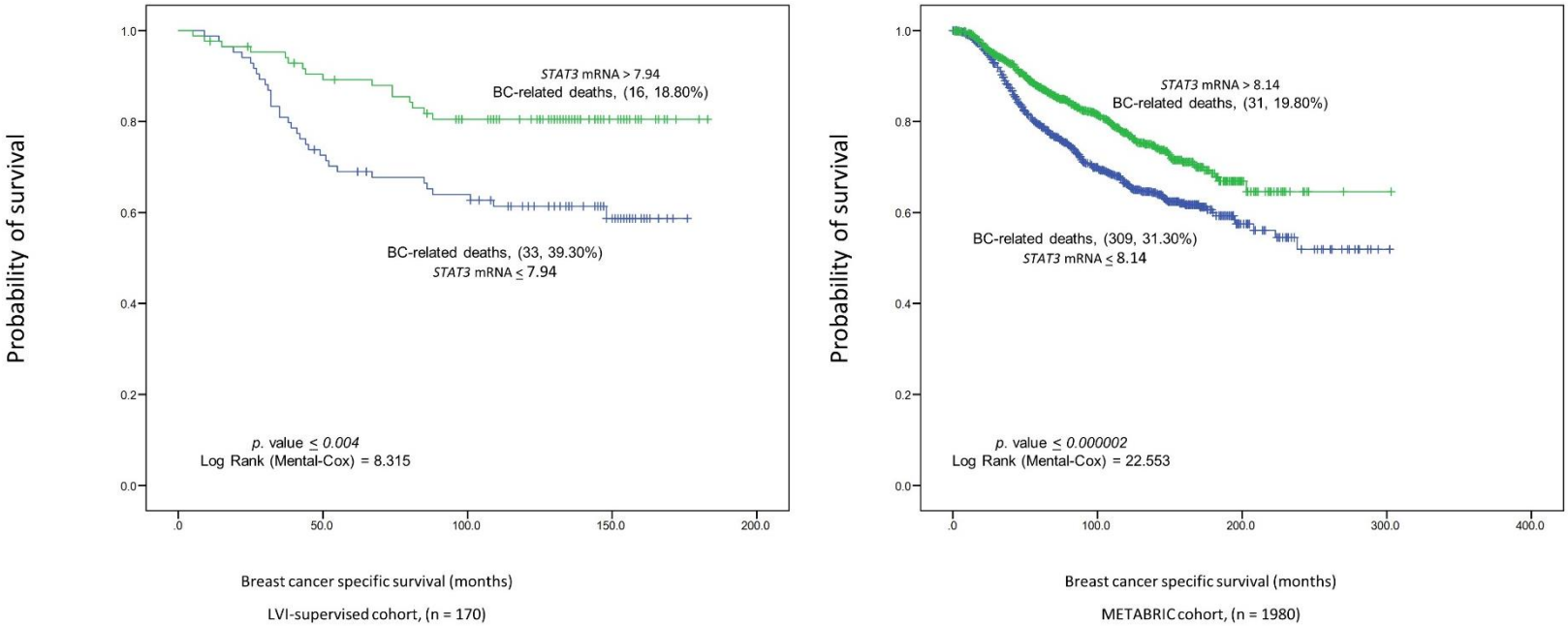


Figure 3. 32: Associations between the altered expression of *STAT3* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

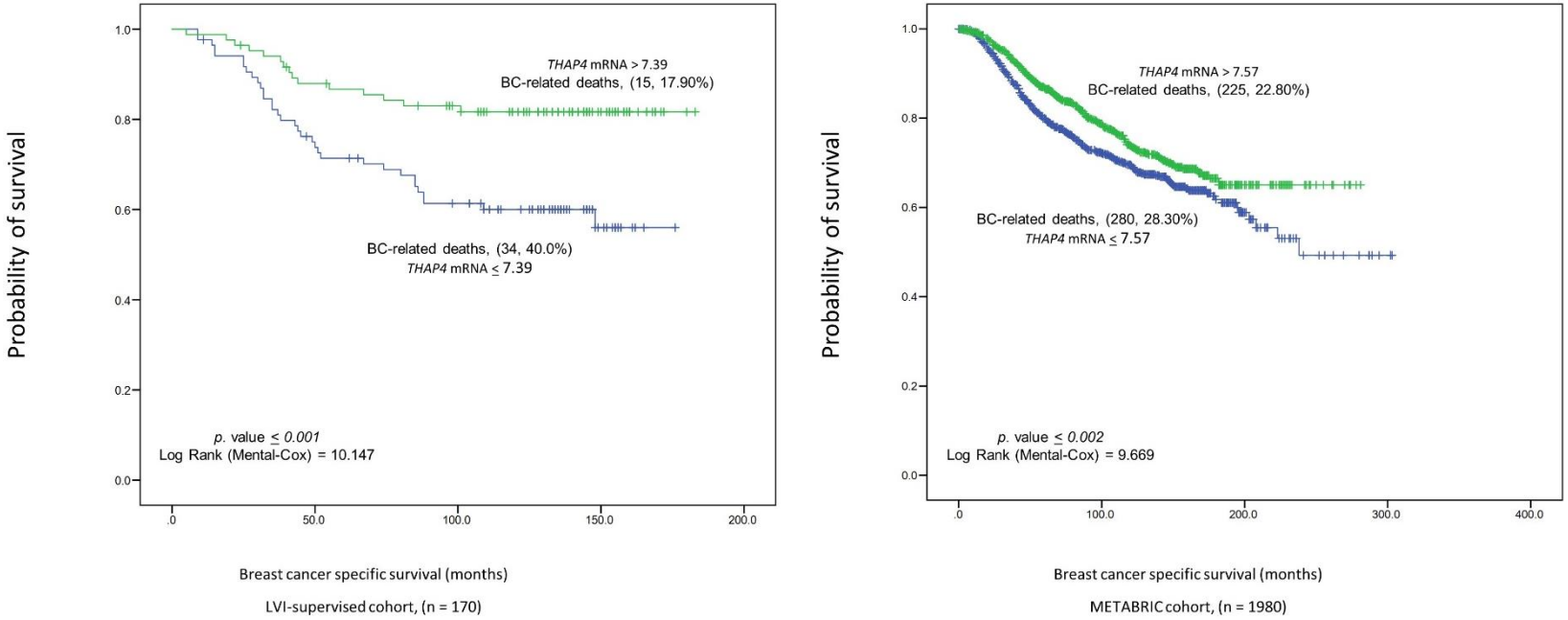


Figure 3. 33: Associations between the altered expression of *THAP4* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

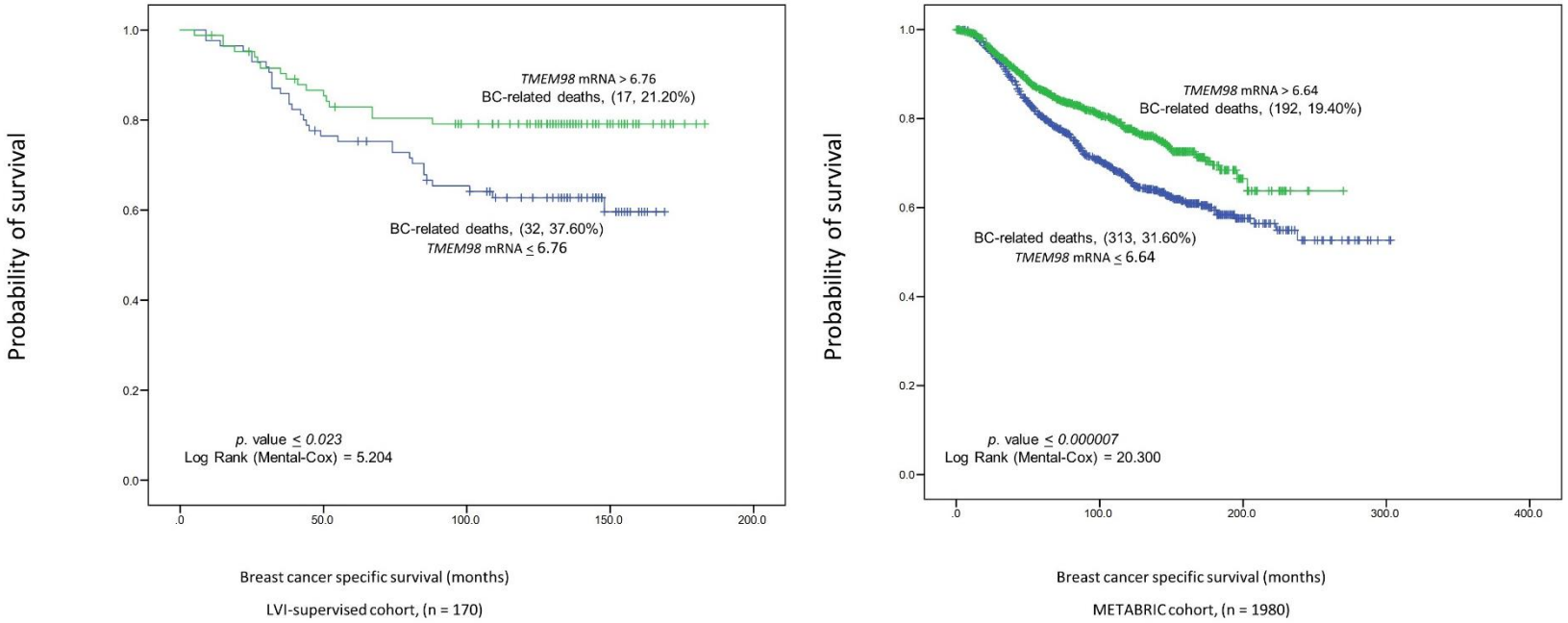


Figure 3. 34: Associations between the altered expression of *TMEM98* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

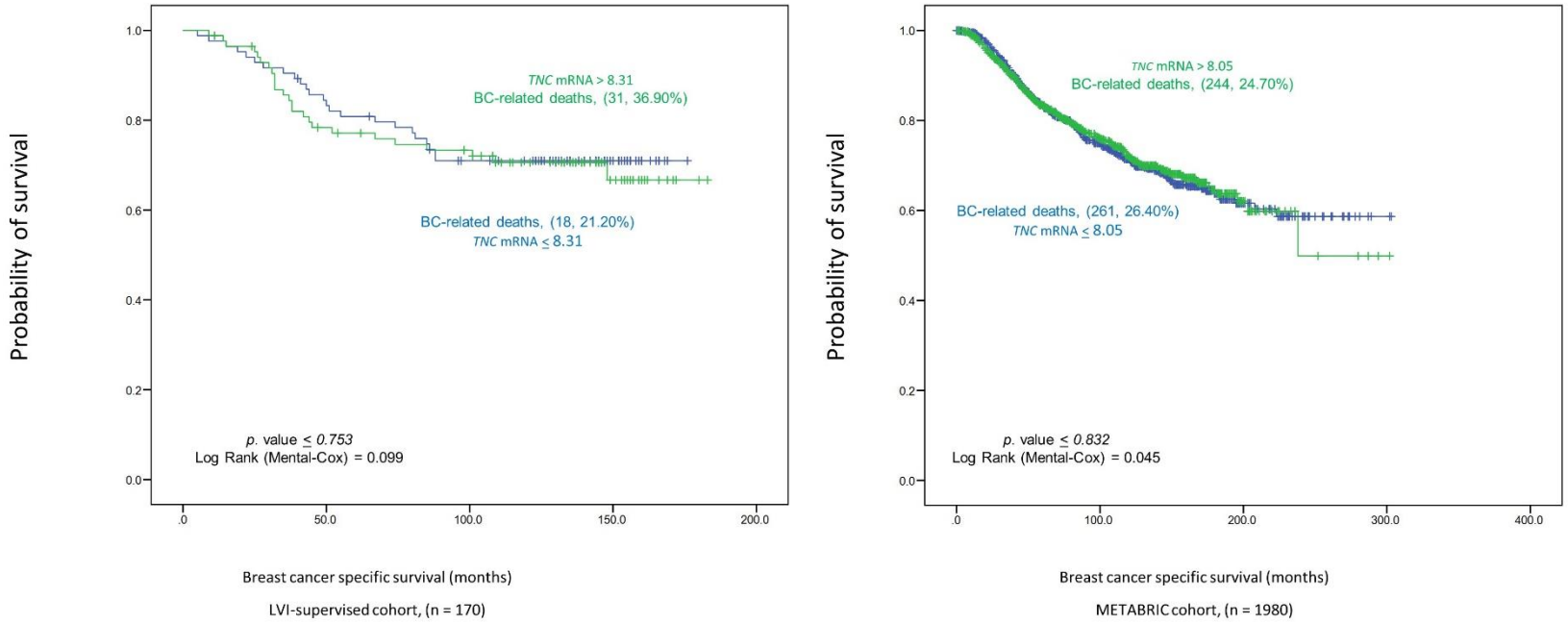


Figure 3. 35: Associations between the altered expression of *TNC* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

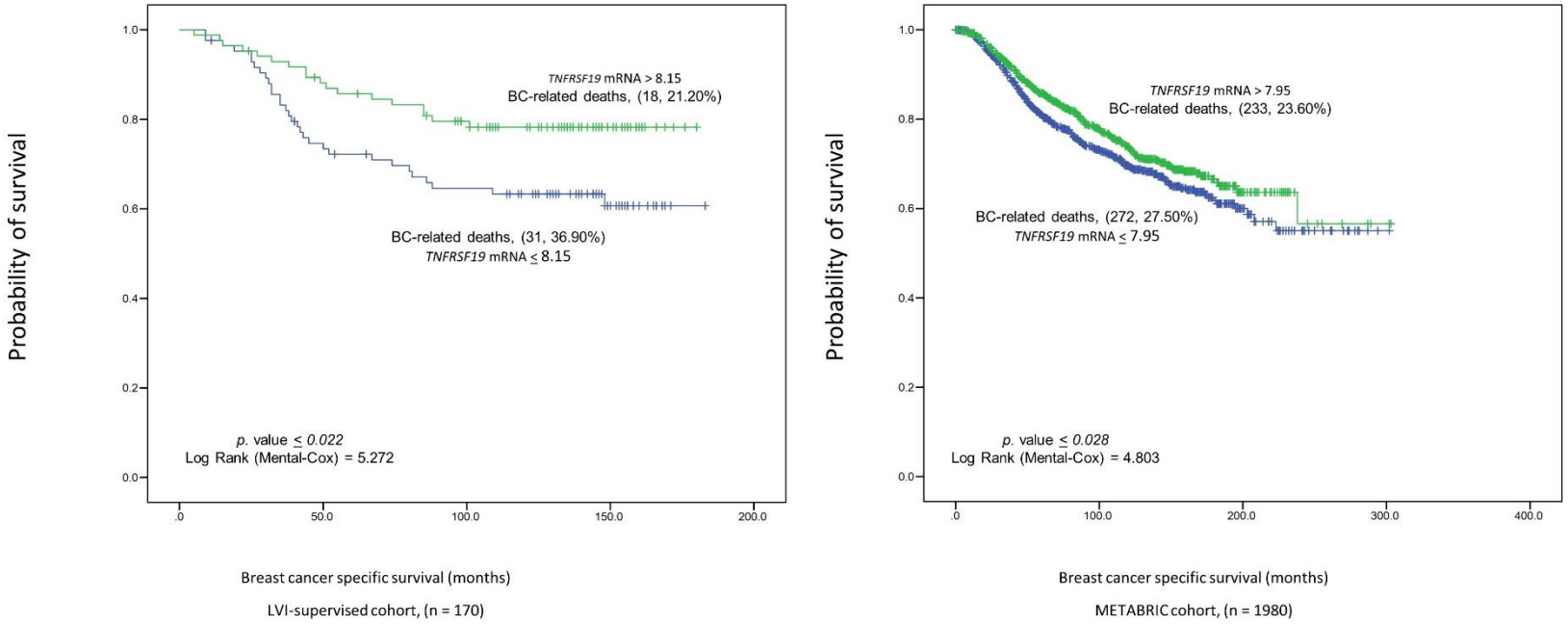


Figure 3. 36: Associations between the altered expression of *TNFRSF19* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

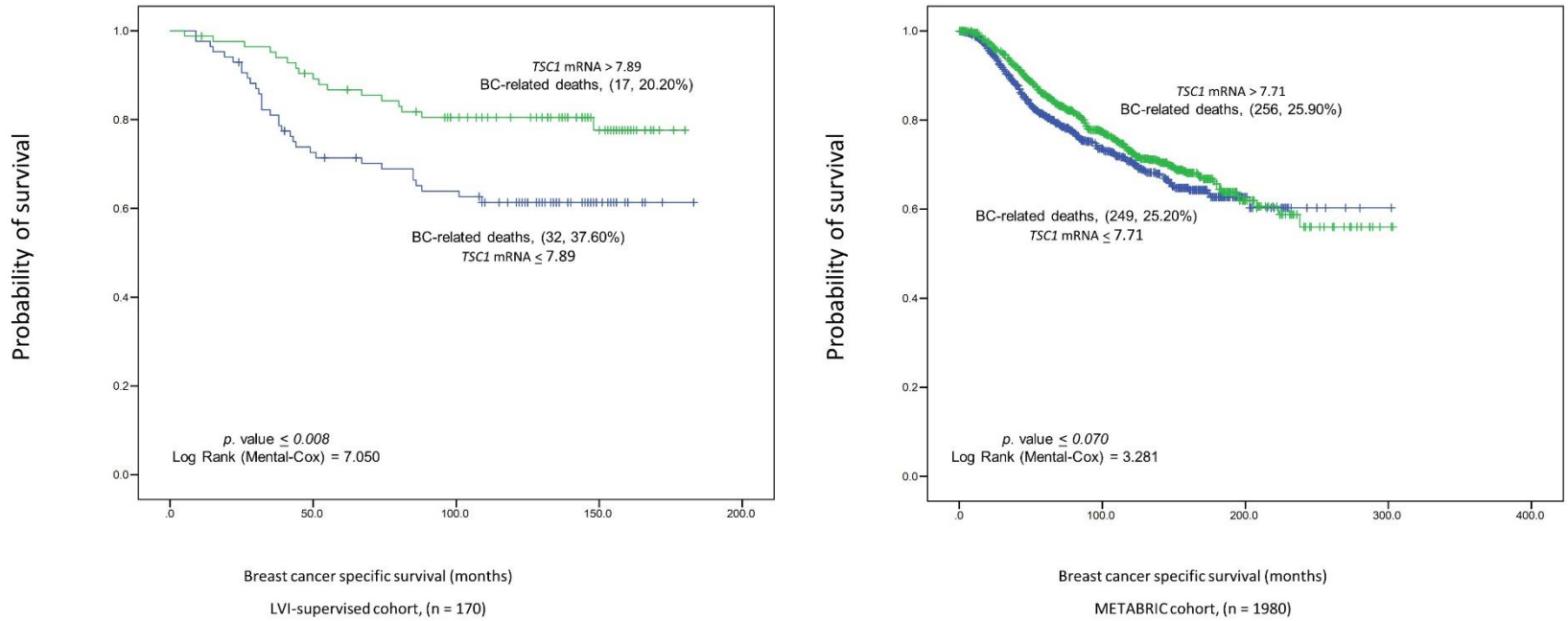


Figure 3. 37: Associations between the altered expression of *TSC1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

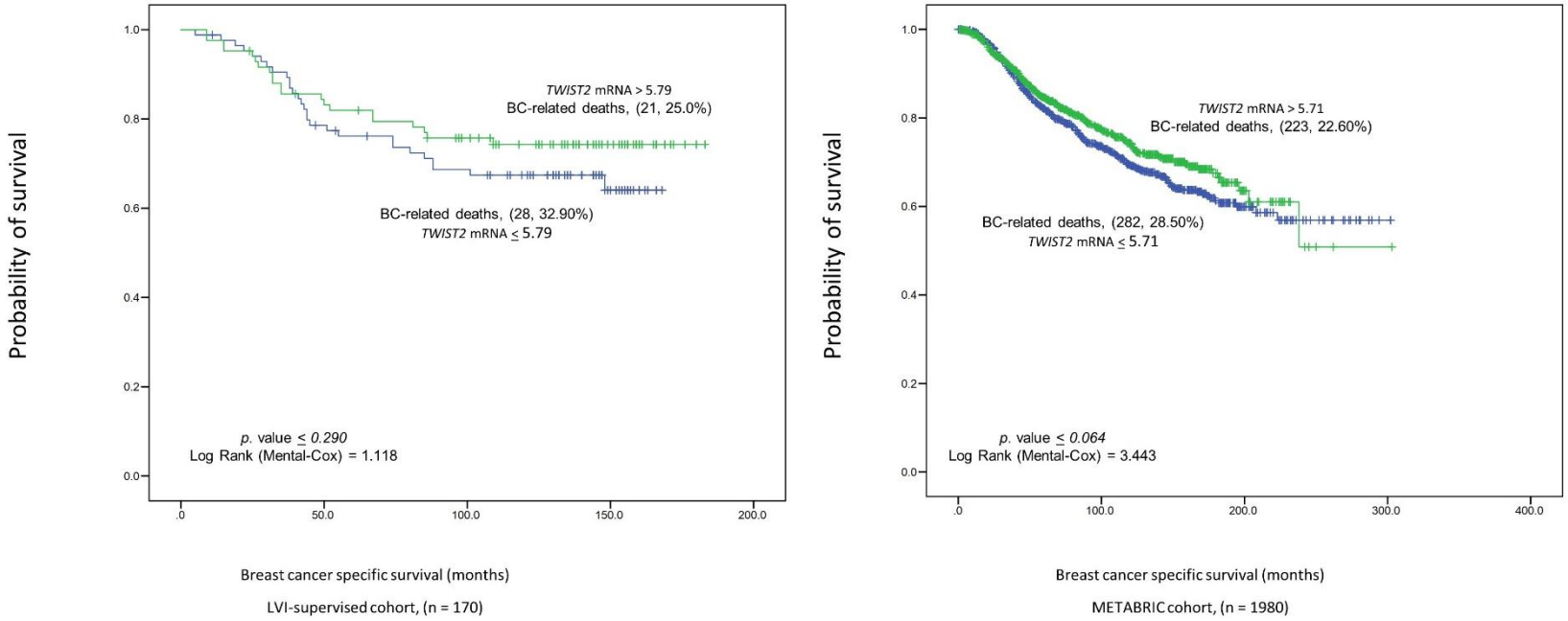


Figure 3. 38: Associations between the altered expression of TWIST2 mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

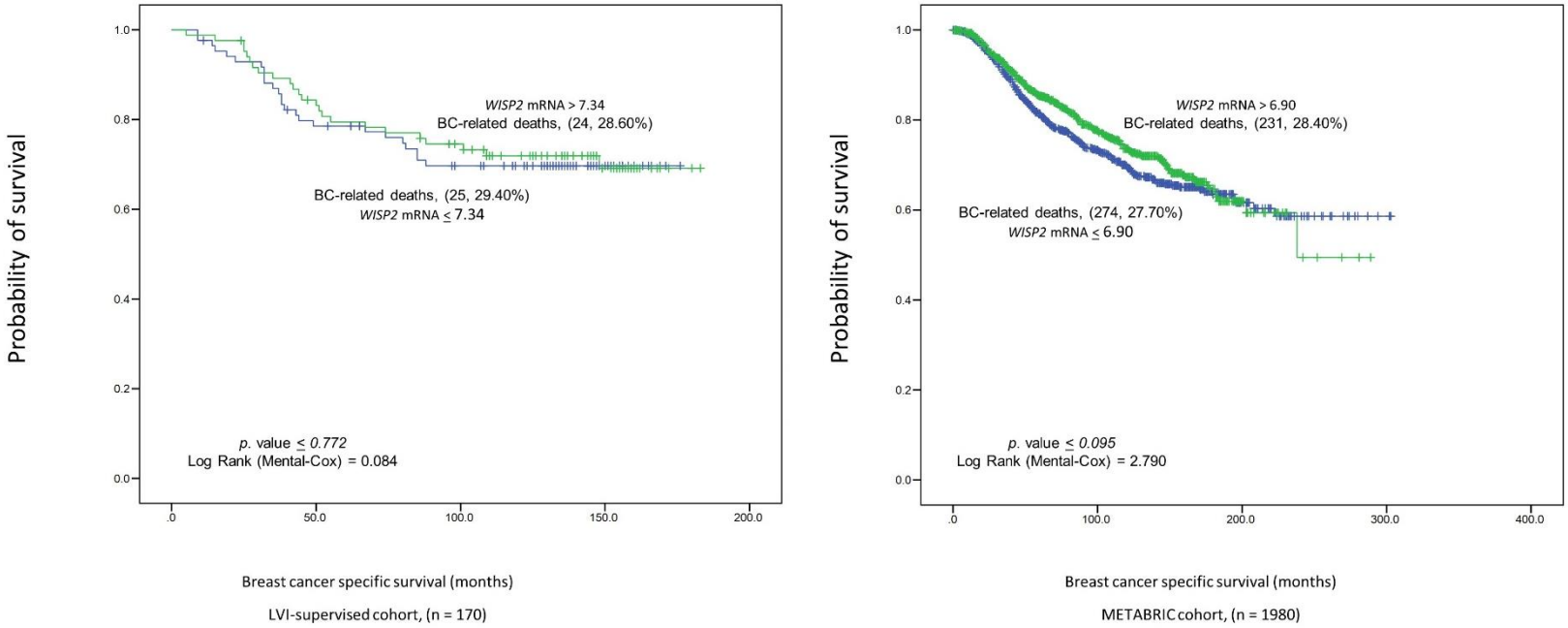


Figure 3. 39: Associations between the altered expression of *WISP2* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

3.4. Discussion

3.4.1. **Overview**

In this study, we aimed to pursue the transcripts that dictate LVI mechanisms in invasive BC patients. Hunting for the main event drivers of LVI was challenging. We applied different strategies to browse and expand the panel of LVI-related transcripts through the combination between different gene selection criteria, different cohorts and analytical tests. As mentioned previously, the agreement on utilising the CNA data in the early stage of this study was persuasive in order to focus our selection on fewer targets. However, the highest-scoring transcripts from the LVI-NT100 cluster did not illustrate conclusive indications of their roles in LVI mechanisms, at least at the mRNA level. The examination of *MCFD2* and *CCT8* protein expressions will be discussed in the following chapter, combining their clinicopathological impacts with the other published molecules produced by this study [43, 190-193]. The amplification of a gene expression may not necessarily produce a rational explanation of how it may be involved in the regulation of the LVI mechanism. Recently, several publications have stated that the CNAs for some genes could act as driver events in the progression of different cancers [194-196]. However, the contributions of the driver genes of high mRNA/CNA correlation in cancer progression could be attenuated or completely lost at the proteome level later. Multiple BPs, such as post-transcriptional and translational modifications and epigenetics, could dictate different scenarios in the biology of invasive BC; hence, LVI.

As the complete transcriptomic and CNA data of the METABRIC cohort were requested, and due to the long period of time over which these files had to be obtained from the METABRIC group at the University of Cambridge, we aimed to utilise that time to try to examine other targets from the LVI-NT100 cluster of DEGs. The examined transcripts/proteins were selected according to their significant correlations with LVI, and descending CNA gain ratios, as for the selection of *MCFD2* and *CCT8*. These targets were: 1) Rho GTPase-activating protein-18 (*ARHGAP18*); 2) kelch-like protein-7 (*KLHL7*); 3) mediator of RNA polymerase II transcription subunit-7 (*MED7*); and 4) Ras-associated and pleckstrin homology domains-containing protein-1 (*RAPH1*). We also aimed to inspect other transcripts/proteins in which their roles in promoting LVI were documented, but in malignancies other than invasive BC. The nomination of the CCR7 protein was based on the published literature on the invasion of lymphatics by melanoma cells [197, 198]. The protein expressions of these five targets were studied using IHC on invasive BC tissue

sections of NPS, and all have been published individually by our group over the last two years [43, 190, 192, 193, 199].

The inclusion of other factors, in addition to the significant correlation with LVI and CNA gain ratios, produced the cluster of LVI-NT283. The obtained scaled transcriptomic data and the robustly categorised histological profiles of LVI for the studied invasive BC patients were utilised to generate the LVI-NT283 cluster of DEGs, and to split them into two distinctive mRNA signatures based on their rankings against LVI categories. One was for the upregulated genes in the LVI^{+} patients (UP-13/LVI^{+}), and the second was for the upregulated genes in the LVI^{-} patients (UP-22/LVI^{-}). Therefore, targeting the DEGs of upregulated log2FC and significant associations with LVI may also identify substantial molecules that could regulate LVI occurrence and show prognostic features. From the UP-13/LVI^{+}, we selected *SEC14L1* to be studied at the protein level as a potential driver of LVI. Our research on *SEC14L1* was also published recently [191].

Over the next few sections, we will discuss some of the DEGs based on our current results for the prominent LVI-related molecules from the cluster of LVI-NT100, and the suggested mRNA signatures of UP-13/LVI^{+} and UP-22/LVI^{-}.

3.4.2. LVI-related transcripts from the LVI-NT100 cluster

The promising indications for the mRNA differential expressions of *MCFD2* and *CCT8* backed up by their significant CNAs in the $LVI^{+/-}$ patients were not translated accordingly when their transcriptomic data were tested against several prognostic parameters of invasive BC.

3.4.2.1. *MCFD2*

The most notable finding for the positive expression of *MCFD2* in the $LVI^{+/-}$ patients was that more *MCFD2*$^{+}$ patients fitted into the normal-like and basal subtypes of PAM50, and into IntClust-4 and IntClust-10 of the METABRIC IntClusts. Although the normal-like tumours were similar to the good-prognosis luminal-A tumours, and they were $ER^{+}/PR^{+/-}/HER2^{-}$, their prognosis may not be as good as luminal-A tumours; although they were still better than the basal-like tumours that were characterised with a worse prognosis. The distribution of the *MCFD2*$^{+}$ patients in the different molecular subtypes and the molecular clusters of the METABRIC in the larger cohort of the METABRIC confirmed the preferential upregulation of *MCFD2* in the invasive BC patients of basal and normal-like subtypes as the *MCFD2*$^{+}$ patients aggregated in IntClust-4 and IntClust-10. IntClust-10 comprised 89.38% of invasive BC patients with the basal subtype, and IntClust-4 comprised 31.87% of invasive BC patients with the normal-like subtype.

The results of the BCSS estimates revealed that the invasive BC patients with *MCFD2*$^{+}$ profiles had better outcomes in the METABRIC cohort, and a trend toward favourable outcomes in the LVI$^{+/-}$ cohort. The results of multivariate analyses confirmed that the overexpression of *MCFD2* exhibited strong characteristics of independent prognosis for invasive BC patients. Principally, the dominant molecular phenotype in both cohorts was the luminal-A phenotype. The different distribution of the *MCFD2*$^{+/-}$ expression profiles could, possibly, point at modified interactions for *MCFD2* in invasive BC according to their molecular phenotypes. These findings may suggest that a subclass of *MCFD2*$^{+}$/basal or normal-like patients may suffer from unfavourable cancer prognosis in comparison to other patients of different phenotypes. Studies on the effects of *MCFD2* and its protein product on cancer progression are limited. In a recent cancer study, the *MCFD2* knockdown significantly reduced tumour invasion and migration, and increased cellular adhesion in oral squamous cell carcinoma (OSCC) [200]. Another recent study stated that the gene deletions in the chromosomal region of 2p16-p21 may cause Lynch syndrome: a high-risk autosomal genetic condition associated with several cancers. The genetic alterations in the *MCFD2* gene, located at 2p21, may include splice site, frameshift, and missense mutations, but not loss of function [201]. An earlier study identified positive MCFD2 expression in testicular seminoma for the first time. The authors stated that the expression level of *MCFD2* mRNA in the tested tissues was downregulated according to the microarray procedures, but it was highly detectable on the protein level by IHC [202]. There was a lack of invasive BC studies that questioned the altered expression of *MCFD2* and its protein product; according to our current mRNA results, we performed a protein study interrogating MCFD2 protein expression on invasive BC tissues. Further insights and comments are given in the following chapter, as we examined the altered expression of *MCFD2* on the protein level.

3.4.2.2. *CCT8*

The second LVI candidate obtained from the LVI-NT100 cluster was the *CCT8* gene. This gene encodes the *theta* subunit of the CCT chaperonin and may be involved in the assembly and transport of the de novo proteins. The altered expression of *CCT8* in the LVI$^{+/-}$ had no associations with clinical factors, molecular phenotypes, or clusters, except its positive relationship with tumours of high grade. In the larger cohort of the METABRIC, the majority of *CCT8*$^{+}$ patients were associated with aggressive subtypes of PAM50 and with the high-grade IntClusts.

The altered expression of *CCT8* could be one of the passenger events at the mRNA level that accompany LVI in subclasses of invasive BC patients. However,

the protein expression of CCT8 could reveal better insights into its role in LVI regulations. An earlier cancer study investigated the effects of the altered expression of *CCT8*/CCT8 in the high-grade HCC, and it specified *CCT8* with oncogenic and proliferative properties [203]. Another study stated that the overexpression of *CCT8* in glioma could promote proliferation and invasion, and the siRNA-silencing of *CCT8* reduced the invasion and proliferation potentials of glioma associated with cytoskeletal dysfunction [204]. A recent study stated that the overexpression of the *CCT8* protein was positively observed in tissues of oesophageal squamous cell carcinoma (OeSCC) patients with LNM^+, and *CCT8* could regulate the migration and invasion properties of the OSCC through interactions with the cytoskeletal proteins of α-actin and β-tubulin [205]. Collectively, the altered expression of *CCT8* and its protein is not well studied in invasive BC patients. Our current results, in addition to the conclusions of the referred *CCT8* studies, nominated *CCT8* as an aberrantly expressed molecule that is associated with high-grade, migratory, and invasive tumours. These findings emphasise the need to study the protein expression of *CCT8* on invasive BC tissues. Additional insights on the effects of the *CCT8* protein in BC progression and LVI are stated in the next chapter.

3.4.3. **LVI-related transcripts from the LVI-NT283 cluster**

The methodology used to obtain the LVI-NT283 cluster of LVI-related DEGs was progressive. In this approach, we aimed to exhaustively analyse the whole transcriptome of the limma analysis output considering the log2FC, AvE, and the status of perfect probe binding, as stated previously. The resulting DEGs in this cluster were divided into two upregulated sets of LVI-associated transcripts. One was associated with the LVI^+ patients, while the second was for the LVI^- patients. In the following subsections, we will briefly report on the clinicopathological impacts for a few targets from both sets, focusing on their LVI underlying mechanisms and prognostic relevance in the invasive BC patients.

3.4.3.1. *LVI-related transcripts from the UP-13/LVI^+ signature*

3.4.3.1.1. **SEC14L1**

Among the LVI^+ patients in the $LVI^{+/-}$ cohort, the sum rank of the aberrant *SEC14L1* expression showed the second-highest gene-upregulation mean rank. The chi-square goodness of fit test determined that there were positive associations between *SEC14L1* overexpression and the LVI definite occurrences in invasive BC patients. The *SEC14L1* gene is located within a discrete region of 17q25 that frequently shows copy number alterations in BC [206]. A proximate locus, the 17q23 locus is well documented in BC, and the key feature of this amplified locus

is its oncogenic activity in BC both in vitro and in vivo [207]. The altered expression of *SEC14L1* was suggested in a 12-gene signature of candidate predictors of progression in prostate cancer (PrC) patients. The protein expression of *SEC14L1* was detected by immunostaining, and high histological scores of the *SEC14L1* protein were correlated with advanced PrC stage, Gleason scores, and unfavourable outcomes [208]. Similarly, our results showed that the aberrant gene expression of *SEC14L1* was associated with high tumour grade, advanced LN stage, worse NPI indices, and poor BCSS in both cohorts in our study. In addition, the association with LVI was totally abolished when the LVI^{-}/LN^{+} patients were included, suggesting that the early detection of the upregulated *SEC14L1* could be a useful tool to predict the invasion of lymphatics through the lymphatic vessel routes.

Prior to our study, *SEC14L1* had not been suggested as an LVI-related biomarker or investigated at the protein level in invasive BC patients. Recently, we successfully published our work on *SEC14L1* in [191]. Currently, our group is designing and optimising models of functional studies to expand our knowledge regarding the mechanistic actions and the prognostic power of *SEC14L1* in invasive BC. In the next chapter, we will provide more details on the effects of *SEC14L1* protein expression on LVI and cancer progression in invasive BC patients.

3.4.3.1.2. PPM1G

The results of the altered expression of *PPM1G* exhibited one of the best statistical associations with LVI positivity in invasive BC patients in this study. Relatively, the overall results for the altered expression of *PPM1G* appeared better than the results of *SEC14L1*. Both transcripts were significantly upregulated in the LVI^{+} patients of the $LVI^{+/-}$ cohort, but not in the LVI^{-}/LN^{+} subset of patients. Additionally, both transcripts were upregulated and associated with invasive BC patients of the aggressive phenotypes, luminal-B, to some extent, HER2-enriched, and basal. The justification for favouring *SEC14L1* over *PPM1G* to procced with protein studies in the next chapter was due to the higher impact of $SEC14L1^{+}$ on worsening BCSS in the $LVI^{+/-}$ patients, specifically. In fact, the invasive BC patients with *SEC14L1* overexpression in the $LVI^{+/-}$ cohort scored the shortest period of survival in comparison to the other LVI-related transcripts in the suggested UP-13/LVI+ signature, including *PPM1G*.

Regarding *SEC14L1*, little is known about how *PPM1G* could influence LVI and invasive BC progression. From the perspective of novelty, this fact may be of vital importance to the current findings in this study. However, it would be difficult to suggest a specific mode of action for *PPM1G* without functional research on a representative cohort of invasive BC patients or cell lines. At the protein level,

PPM1G is also known as *PP2CG*, and it is a member of the PP2C family of serine/threonine protein phosphatases. Several members from the PP2C family of proteins were previously reported to be upregulated in invasive BC.

An integrative transcriptome and proteome study identified that protein phosphatase 1F (*PPM1F*) could regulate cellular motility through the regulation of phosphorylation levels of MAP kinase (*MAPK*) and glycogen synthase kinase-3 (*GSK3*), which in turn affects other transcriptional factors of cellular proliferation. The knockdown of the *PPM1F* gene in the MDA-MB-231 BC cell line reduced their migration speed and proliferation [209]. Additionally, it was demonstrated that the minimisation of protein phosphatase 1F phosphorylation activity could also alter the microtubule dynamics in metastatic cancer cells through less phosphorylation of stathmin: a protein that prevents the assembly of microtubules independently from *MAPK* and *GSK3* [209]. A later study indicated that the highest association for the overexpression of stathmin protein was immunohistochemically detected in the cytoplasm of the triple-negative BC tissues, and the *STMN1* gene was overly expressed and associated with the cell lines of the basal subtype [210]. According to our current results for *PPM1G*, and in light of the conclusions from the aforementioned citations, we hypothesise that there is an interactive axis between *PPM1G* and *STMN1* and their proteins, which could be associated with LVI enhancement and BC aggressiveness in specific molecular subtypes in invasive BC. To understand the relationships between *PPM1G* and LVI in invasive BC, other proteins should be included in a separate study in the future. These factors could be the hormonal receptors ER, PR, EGFR, ki67 as a general biomarker for proliferation, cytokeratins such as CK5/6, cadherins such as E- and N- cadherins, and stem cell biomarkers such as aldehyde dehydrogenase (ALDH), CD24, and CD44.

3.4.3.1.3. **P4HA2**

The altered expression of *P4HA2* was unique in this study. The *P4HA2* transcript was the only molecule that showed a conservative association with LVI in both the $LVI^{+/-}$ and METABRIC cohorts. In addition, it displayed strong association with worse outcomes in the two cohorts and independent prognosis features. The altered expression of *P4HA2* in the $LVI^{+/-}$ patients was preferentially upregulated in the luminal-B and HER2-enriched intrinsic subtypes, and in IntClust-5 that was defined with ERBB2 amplification as a molecular feature.

The protein-encoding gene of *P4HA2* is harboured in the 5q31.1 chromosomal region, and it was found to be ubiquitously expressed in normal human lung and kidney human tissues, and in the stomach and intestine during human foetal

development [211]. The encoded protein prolyl 4-hydroxylase is an oxidoreductase that catalyses the post-translational formation of 4-hydroxyproline, and it is essential for the proper three-dimensional folding of the de novo procollagen chains. An earlier study utilised cDNA microarray analysis to identify clusters of DEGs that are associated with the amplification of ERBB2 in luminal subtypes of human BC [212]. The study named *P4HA2*, with fibronectin-1 (*FN1*) and P-cadherin (*CDH3*), as one of the DEGs that was associated with cell-matrix interactions in the poor-prognosis luminal phenotypes of human BC. Another study recently stated that the *P4HA2* gene was associated with expression of *COL1A1*, *COL3A1*, and *COL4A1* during BC development, and the elevation in the *P4HA2* gene transcription was also associated with poor outcomes, as our results illustrate. Furthermore, the short hairpin RNA (shRNA)-silencing of *P4HA2* hindered BC growth and metastasis in immunodeficient SCID mice [213].

The enriched expression of *P4HA2* was also discovered by other researchers in a subset of BC stem cells of triple-negative that exhibited $ALDH^{+}/CD24^{-}/CD44^{+}$ profiles [214]. Additional studies stated that the *P4HA2* gene could be one of the genes associated with aggressive clones of cells in primary triple-negative BC. In the latter study, *P4HA2* expression was upregulated in successful tumour engraftments in the patient-derived xenograft model, and the silenced expression of *P4HA2* in the MDA-MB-231 BC cell line led to a significant inhibition of cancer cell invasion [215]. In the $LVI^{+/-}$ cohort of this study, 85.7% of the $ER^{+/-}/HER2^{+}$ were LVI^{+}. Similarly, 60% of basal subtype tumours were also LVI^{+}. The altered expression of *P4HA2* may be a key player in LVI mechanisms in phenotype-dependent manners. Targeting *P4HA2* in the high-risk class of HER2-enriched and basal patients is warranted and requires further attention in future studies.

3.4.3.2. *LVI-related transcripts from the UP-22/LVI^{-} signature*

3.4.3.2.1. **CXCL12**

The altered expression of *CXCL12* in invasive BC patients was studied in depth. However, the direct contribution of *CXCL12* in developing LVI is still not clear. The protein-encoding gene of *CXCL12* is located at the chromosomal region of 10q11.21, and it was identified as being ubiquitously expressed in normal human endometrium and spleen tissues [216]. The encoded protein is the stromal cell-derived-α (SDF-1), a chemokine member of the intercrine family of secreted proteins. In this study, we identified *CXCL12* as one of the DEGs that was positively correlated with LVI absence, low tumour grade, negative LNM, favourable NPI, and better outcome. The identification of *CXCL12* as an LVI-inhibitor in the suggested

LVI^- molecular signature of this study reinforces the importance of studying TME components and their regulators of LVI mechanisms in invasive BC patients.

In a transcriptomic study on 115 invasive BC versus their normal-adjacent tissues, the associations between *CXCL12* and its receptors, *CXCR4* and *CXCR7*, in promoting BC invasion and metastasis was revealed. The authors stated that the expression of *CXCL12* in primary tumours showed no differential expression levels when compared against the normal-adjacent tissues by RT-qPCR, and did not exhibit any relevant effects on patients' survival. Alternatively, high *CXCL12* expression was identified in LN^+ tumours and showed significant differences against the low-expressed *CXCL12* in primary tumours [217]. Our results showed that the majority of patients in the normal-like subtype were labelled with high *CXCL12* expression profiles in both cohorts. As above, the minimum number of patients with $CXCL12^+$ expression profiles was associated with the aggressive intrinsic subtypes of invasive BC in all studied patients. In another study, the detection of mRNA and protein expressions of *CXCL12* and *CXCR4* in the MDA-MB-231 invasive BC cell line were performed by qRT-PCR and WB. It was found that the *CXCL12* mRNA and protein were weakly expressed, while *CXCR4* was strongly expressed in the MDA-MB-231 cell line. When the MDA-MB-231 cells were stably transfected with *CXCL12*, the biological behaviour changed in these cell and promoted apoptosis process, but they kept their invasion potentials [218]. A different study identified positive IHC staining in normal tissues and the total absence of IHC staining in tumour legions in paired tissue specimens of invasive BC. Patients with $CXCL12^+$ expression profiles showed better survival rates. However, the negative expression of *CXCL12* was significantly correlated with LNM^+ [219].

In this study, we have provided further evidence that emphasised the good prognostic indications of the upregulated *CXCL12* in primary BC. Nonetheless, it is crucial to target the expression status of the *CXCL12* receptor, *CXCR4*, as an axis promoter that could develop LNM in invasive BC patients.

3.4.3.2.2. LPAR1

Our results indicated that the overexpression of *LPAR1* in invasive BC patients is strongly correlated with LVI absence. The transcript of *LPAR1* was strongly associated with the LVI^- patients in the $LVI^{+/-}$ patients and in the $LVI^-/LN^{+/-}$ subset of patients in the METABRIC cohort. The high expression of *LPAR1* was also correlated with the low tumour grade and favourable NPI categories in both cohorts. However, there was a significant difference between the $LVI^{+/-}$ and the METABRIC patients in their altered expression of *LPAR1* against the LN status. Two-thirds of stage-I patients were $LPAR1^+$ in the $LVI^{+/-}$ cohort. The best results for a longer

BCSS among all of the transcripts in the UP-22/LVI^{-} signature was for the patients of the *LPAR1*$^{+}$ expression profiles in the LVI$^{+/-}$ and METABRIC cohorts, and the downregulated levels of *LPAR1* were an independent sign to predict the high-risk subclass of invasive BC patients. The protein-coding gene of *LPAR1* is located at the 9q31.3 chromosomal location, and it is ubiquitously expressed in human brain tissue. The lysophosphatidic acid (LPA) receptor-1 is the translated protein of *LPAR1*. The LPA is a lysophospholipid that could be transported in the bloodstream, and it activates myriad BPs through LPA receptor-1: a member from the unique family of G-protein-coupled receptors called endothelial differentiation gene-2 (EDG2) receptors.

Earlier, the *LPAR1* receptor, or EDG2, was reported in ovarian cancer with increasing apoptosis rates, and it countered the stimulated effects by other members of the EDG family of receptors that may promote ovarian cancer cell proliferation and survival [220]. The following research states that the EDG2 and its ligand LPA could promote BC motility and metastatic potential specifically in the motile phenotype of the metastatic MDA-MB-435 and MDA-MB-231 BC cell lines [221]. The effects of EDG2 on cancer cell progression could be contextualised by several cellular and biochemical factors. Different studies have investigated the associations between LPA signalling pathways though *LPAR1* and *LPAR3* on the reactive oxygen species (ROS)-induced cellular motility in mouse fibroblast 3T3 cells. The overexpression of *LPAR1*, but not *LPAR3*, significantly supressed motility potentials of the 3T3 cells [222]. The LPA signalling through *LPAR2* was recently identified as being associated with the progression of osteosarcoma cells [223]. Nonetheless, there was an emphasis on *LPAR1* to serve different biological functions in several pathological scenarios [224].

Our results in this study may favour the upregulation of *LPAR1* in invasive BC patients; however, the biological behaviour of *LPAR1* in certain subclasses of invasive BC patients could be different. By scrutinising IntClust-2, we discovered that the majority of individuals in this cluster, a total of 62.5%, exhibited the *LPAR1*$^{+}$ expression profile. The patients of IntClust-2 were associated with a high genomic instability and the worst prognosis of all luminal tumours in the METABRIC cohort, and they comprised 50% of patients of the intrinsic subtype of luminal-B [39]. Generally, we can conclude that the molecular regulations for the positive expression of *LPAR1* could attenuate tumour progression and LVI and predict good prognosis for the vast majority of invasive BC patients, as our results illustrate. However, there could be a unique class of invasive BC patients for whom *LPAR1* is

of no distinct clinicopathological use. In fact, they may represent a high-risk subset of patients that will require further attention in future studies.

3.4.3.2.3. **FBLN1**

The fibulin-1 encoding gene is located at the 22q13.31 cytogenetic band. The RNA sequencing of 27 different human tissues showed that the gall bladder, the placenta, and the urinary bladder ranked as the top three normal human tissues with upregulated levels of *FBLN1* [216]. In this study, we identified *FBLN1* as one of the overexpressed genes of good prognosis that may inhibit LVI progression in the $LVI^{+/-}$ cohort. In addition, the majority of individuals with upregulated *FBLN1* were associated with the IntClusts of low tumour grade and good survival, IntClust-3, IntClust-4, and IntClust-8. The associations between *FBLN1*, LVI^{+}, and LN^{+} was not significant in the METABRIC cohort, as our results show. The aberrant expression of *FBLN1* in the $LVI^{+/-}$ patients was validated on the protein level, as our HPLC/MS results show in Chapter 5 of this study. We provide a comprehensive discussion on the clinicopathological impact for *FBLN1* and its protein, and also their altered expressions against a large number of tumour progressions and LVI-related biomarkers in Chapter 5, section 5.4.6.

3.5. Conclusion

In conclusion, the present study has identified that the initiation or inhibition of LVI mechanisms in invasive BC patients is strongly associated with specific transcriptional regulations of specific genes. In addition, the biological associations between LVI and the different molecular phenotypes of invasive BC can be governed with an LVI gene signature. The HER2-enriched invasive BC patients worth of better attention as they showed a positive association with PPMG1, SEC14L1, P4HA2, and CTSD the top four upregulated genes in LVI^{+} cases. The development of molecular methods that aid in enriching predictive biomarker panels is essential for better clinical intervention. The inclusion of newly discovered LVI driver molecules in future research could provide a better understanding of LVI mechanisms, and offer novel therapeutic options for invasive BC patients. In this part of the study, we attempted to deliver a molecular gene signature for LVI. The results presented here confirm that LVI mechanisms are highly sophisticated and comprise several biophysical and biochemical regulators. Applying clustering analysis to the proposed LVI-related protein-encoding genes could be a promising approach to discover novel prognostic and therapeutic targets in the future.

Summary of the top differentially expressed genes in LVI-supervised cohort of invasive breast cancer patients; The green-labelled genes are the qualified genes for the immunohistochemistry analysis								
Gene ID	Full name	Differentially expresed gene cluster	Statistical analysis	CNA data	CNA gain ratio	Adjusted p-value (LIMMA)	p-value (Jonckheere-Trepstra)	p-value (Mann-Whitney U-test)
ANXA5	Annexin A5	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.013987948	NA	0.008
APCDD1	Protein APCDD1	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.042116707	NA	0.005
C17orf81	Elongator acetyltransferase complex protein 5	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.035196174	NA	0.002
C1S	Complement C1s subcomponent	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.00154156	NA	0.022
CCT8	T-complex protein 1 subunit theta	LVI-NT100 (Top DEG+CNA-gain)	LIMMA, Jonckheere-Trepstra	Yes	2.7363	0.001287521	0.0385	NA
CNRIP1	CB1 cannabinoid receptor-interacting protein 1	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.029223368	NA	0.02
CTSD	Cathepsin D	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.014493894	NA	0.009
CXCL12	Stromal cell-derived factor 1	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.010081724	NA	0.001
ECM2	Extracellular matrix protein 2	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.015582381	NA	0.004
FBLN1	Fibulin-1	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.00150583	NA	0.005
FEZ1	Fasciculation and elongation protein zeta-1	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.026504355	NA	0.016
IKZF3	Zinc finger protein Aiolos	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.04427484	NA	0.036
INKA1	PAK4-inhibitor INKA1	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.043818375	NA	0.011
ITPRIP	Inositol 1,4,5-trisphosphate receptor-interacting protein	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.024242739	NA	0.016
ITSN1	Intersectin-1	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.010561393	NA	0.021
LPAR1	Lysophosphatidic acid receptor 1	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.00041222	NA	0.0002
MCFD2	Multiple coagulation factor deficiency protein 2	LVI-NT100 (Top DEG+CNA-gain)	LIMMA, Jonckheere-Trepstra	Yes	4.5604	0.001459333	0.0209	NA
NBR1	Next to BRCA1 gene 1 protein	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.022352683	NA	0.018
P4HA2	Prolyl 4-hydroxylase subunit alpha-2	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.007640178	NA	0.007
PDLIM7	PDZ and LIM domain protein 7	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.045265592	NA	0.032
PILRA	Paired immunoglobulin-like type 2 receptor alpha	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.029465389	NA	0.031
PPM1G	Protein phosphatase 1G	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.012683088	NA	0.0001
PRRX1	Paired mesoderm homeobox protein 1	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.002234422	NA	0.03
RNASE1	Ribonuclease pancreatic	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.023053485	NA	0.032
SEC14L1	SEC14-like protein 1;SEC14L1	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.027693888	NA	0.005
SIGLEC1	Sialoadhesin	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.022610171	NA	0.039
SLC3A2	4F2 cell-surface antigen heavy chain	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.03946014	NA	0.032
SLC6A9	Sodium- and chloride-dependent glycine transporter 1	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.011467043	NA	0.047
STAT3	Signal transducer and activator of transcription 3	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.003022918	NA	0.035
THAP4	THAP domain-containing protein 4	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.04489407	NA	0.016
TMEM98	Transmembrane protein 98	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.042904932	NA	0.012
TNC	Tenascin	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.032016003	NA	0.028
TNFRSF19	Tumor necrosis factor receptor superfamily member 19	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.009729556	NA	0.016
TSC1	Hamartin	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.006977967	NA	0.036
TWIST2	Twist-related protein 2	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.039096979	NA	0.047
WISP2	WNT1-inducible-signaling pathway protein 2	LVI-NT283 (UP-22/LVI$^-$)	LIMMA, Mann-Whitney U-test	No	NA	0.03260996	NA	0.019
ZDHHC9	Palmitoyltransferase ZDHHC9	LVI-NT283 (UP-13/LVI$^+$)	LIMMA, Mann-Whitney U-test	No	NA	0.017928918	NA	0.031

Chapter 4: IHC Assessment of MCFD2, CCT8, SNAP23, and SEC14L1 as Potential Regulators for LVI in Invasive BC

4.1. Introduction

Invasive BC is one of the most fatal diseases to threaten women's lives worldwide. Although early detection of BC may help to provide better clinical interventions, there are still several BC-associated factors that could deteriorate the desirable therapeutic responses of BC patients. Decades ago, the assessment of invasive BC was performed morphologically without the ability to detect the key molecular drivers that dictate adverse effects on invasive BC patients' survival. The molecular plateau of invasive BC was revealed by utilising modern genomic, transcriptomic, and proteomic technologies such as DNA and RNA sequencing, immunoassays, and high-throughput chromatographic techniques backed up with spectrophotometric detectors. This breakthrough in experimental design and availability of tools in molecular pathology developed novel hallmarks for several cancers, including BC [7].

The identification of LVI driver proteins with independent and prognostic features could help to classify invasive BC patients into high- or low-risk subclasses. Tackling the regulatory proteins of LVI in invasive BC using synthetic drugs could be an efficient therapeutic option for many patients with aggressive BC phenotypes. Genomic and transcriptomic analyses would provide suggestions for specific enriched genes/mRNAs that could dictate LVI mechanisms; however, the functional players are the translated proteins. The functions of a specific protein may be altered due to genetic mutations, mRNA splicing, PTM, PPI, and aberrant cellular localisation. As demonstrated in the introduction, IHC was used to identify several protein targets in invasive BC patients. However, the exact mechanisms are yet to be determined. Considering the primary tumour phenotypes, the intratumoural heterogeneity, the contribution of the stromal components of TME, and the architecture of the surrounding stromata are examples of how defining the underlying mechanisms of LVI is a challenging task.

The exact molecular activators for invasion and metastasis were enigmatic for several years. The invasion-metastasis mechanisms involved loss of cell-cell and cell-ECM attachments such as E-cadherin, irregularity in cell structure and shape, upregulation of migration proteins related to embryogenic changes and inflammatory responses such as N-cadherin, altered metabolism, and aberrant gene expression in myriad intracellular signalling networks such as the EMT regulatory pathways [225]. The journey of the metastatic BC cell starts with a local invasion of the surrounding tissue followed by successful docking and intravasation of an adjacent blood or lymphatic vessel [226]. Extravasation of the in-transit BC cell and forming small nodules of BC cells into the functional tissues of a distant

organ is called micro-metastasis. The colonisation of distant tissues by the growing micro-metastatic cells leads to the formation of macroscopic tumours. From this presentation, the discrete targeting of the LVI molecular drivers as an early step in the invasion-metastasis cascade could be critical.

In this study, a large number of aberrantly expressed genes associated with LVI were identified. This study provided a large panel of prognostic and therapeutic candidates associated with LVI. Therefore, it was important to prioritise a selection of identifiers for protein-encoding genes to be included in the IHC study. Here, the implications of a specific cluster of LVI-related proteins will be introduced, analysed, and discussed. These proteins included *MCFD2* and *CCT8* from the LVI-NT100 cluster, as they were the top LVI-DEGs with CNA gain ratios >2.5 and have not been studied before in invasive BC patients. Another list of proteins for which the encoding genes showed CNA gain ratios >1.0 were also studied by IHC, and for which the results were published by our group individually, included *RAPH1* [190], *ARHGAP18* [193], *KLHL7* [192], and *MED7* [43]. As the valid option for initial target selection was limited to *MCFD2* and *CCT8* in this study, we followed a different approach utilising specific hallmark sets of gene set enrichment analysis (GSEA). We identified *SNAP23* as an *MCFD2* and *CCT8* co-enriched molecule. *SNAP23* was also a significant DEG in the cluster of LVI-NT1158, but it did not appear in the LVI-NT100. The protein expression of *SEC14L1* was also assessed by IHC as one of the top LVI-DEGs from the cluster of LVI-NT283.

This study aimed to examine selected LVI-associated DEGs at the protein level by applying IHC procedures to a large cohort of invasive BC patients of NPS. As we mentioned in the second chapter, the selection criteria for tumour specimens for the DGE experiment were identified by H&E and IHC staining. Therefore, the study perused the key molecular drivers for the LVI, but not those associated with BVI. As the results of the previous chapter stated, *MCFD2*, *CCT8*, and *SEC14L1* were the promising LVI nominees according to the transcriptomic analyses. In the next subsections, the interrogated proteins in this chapter and their molecular functions, biological processes, cellular compartments, and related pathways are introduced.

4.1.1. **Multiple coagulation factor deficiency protein 2**

Multiple coagulation factor deficiency protein 2 (MCFD2) is an endoplasmic reticulum (EnR) protein (16.390Da) encoded by the *MCFD2* gene located at chromosomal locus 2p21. Rearrangement of the genetic architecture at this locus is reported to be associated with risk of thyroid adenomas and renal cell carcinoma [227, 228]. MCFD2 is also reported as a novel tumour biomarker for testicular cancer stem cells in seminoma [202]. The potent stem cell self-renewal capacity of

MCFD2 in the adult central nervous system (CNS) and human embryonic stem cells has also been suggested [229].

The MFs for MCFD2, identified by the GO database (GO:0005509), is limited to calcium ion binding [230]. MCFD2 is a transporter protein that facilitates specific protein transportation mechanisms. The BPs for MCFD2 are: **1**) the addition of COPII and adaptor proteins to the EnR membranes to form a vesicle coat during the formation of transport vesicles (GO:0048208); **2**) direct movement of the newly synthesised proteins from the EnR to Golgi vesicle-mediated transport (GO:0006888); and **3**) the directed motion of proteins into a cell, out of a cell, within a cell, or between cells (GO:0015031) [185]. The MCFD2 protein was positively detected intracellularly by immunofluorescence confocal microscopy in the EnR-Golgi intermediate compartment (ERGIC) of the fibroblast-like cell line, COS-1 [231]. The biochemical pathways in which MCFD2 was documented to be enriched are: **1**) cargo concentration in the EnR; **2**) COPII-mediated vesicle transport; **3**) anterograde transport from EnR to Golgi; **4**) transport to the Golgi and subsequent modification; **5**) asparagine N-linked glycosylation; **6**) membrane trafficking; **7**) vesicle-mediated transport; **8**) PTM; and **9**) metabolism of proteins.

The contribution of MCFD2 in regulating the invasion-metastasis cascade in invasive BC is unclear. Published studies that relate MCFD2 to cancer biology in PubMed are scarce. The first study that stated MCFD2 as a novel germ cell marker in testicular seminoma was published in 2007 [202]. Since then, no studies have been conducted. Recently, MCFD2 was suggested as a vital promoter for regional lymph node metastasis in OSCC [200]. Another study states that the deletion of 24 neighbouring genes at 2p16.3-p21 could be the cause of Lynch syndrome in a 52-year-old male patient with prostatic adenocarcinoma [201]. Since *MCFD2* was one of the top DEGs with a high CNA gain ratio between LVI^+ and LVI^- patients, according to our results in Chapter 3, this chapter additionally interrogated the prognostic and biological significance of MCFD2 protein expression in invasive BC tissues using the large and well-characterised BC series of NPS. To the best of our knowledge, this study is the first attempt to interrogate the altered protein expression of MCFD2 and LVI in invasive BC patients.

4.1.2. Chaperonin containing T-complex protein 1 (TCP1) subunit theta

Chaperonin containing T-complex protein 1 (CCT8) is a subunit of the ring-shaped hetero-oligomeric molecular chaperone, CCT, which is also known as the eukaryotic chaperonin TRiC. The CCT subunits 1 to 8 create a central ring-like pocket and catalyse the folding of specific proteins [232]. There are no specific studies that suggest direct or indirect mechanisms that associate CCT8 with LVI in

invasive BC. Instead, the aberrant expression of CCT subunits has been reported in many pathological conditions. The substitution of histidine for arginine at position 147 of CCT5 resulted in a missense mutation that could cause autosomal recessive mutilating sensory neuropathy [233]. Another study identified CCT3 as the protein-folding subunit that regulates STAT3 protein folding within eukaryotic cells [234]. The cancer-linked p53 protein was also identified as a substrate for CCT1 and CCT5 through its interaction with other chaperones such as the 70 and 90 kDa heat shock proteins, HSP70 and HSP90 [235, 236]. The misfolding tumorigenesis could be promoted by mutations in the CCT-binding sites of the von Hippel-Lindau (VHL) disease tumour suppressor [237]. Few studies have explained the association between the altered expression of CCT8 and the progression of several cancers in the last five years [205, 238, 239]. However, none of these studies identified LVI-regulating roles for CCT8 in invasive BC. CCT8 was also identified to physiologically interact with the cytoplasmic dynein intermediate chain 2, an intracellular protein that participates in the retrograde motility of vesicles and organelles along microtubules [240]. Interestingly, the altered expression of the heat shock protein 90 α, HSP90-α (*HSP90AA1*), and the cytoplasmic dynein light chain 2A (*DYNLRB1*) were identified as LVI^{+}-associated differentially expressed proteins/genes in our RNA-microarray/MS study in the following chapter.

A broader MF was identified for CCT8 in comparison to MCFD2. ATPase activity (GO:0042623), ATP binding (GO:0005524), cadherin binding (GO:0045296), protein binding involved in protein folding (GO:0044183), and unfolded protein binding (GO:0051082) are the GO-MFs for CCT8. There are several GO-BPs for CCT8 in human cells, and some of these are: positive regulation of protein localisation to Cajal body (GO:1904871), de novo protein folding (GO:0006458), chaperone-mediated protein folding (GO:0061077), pore complex assembly (GO:0046931), toxin transport (GO:1901998), and protein stabilisation (GO:0050821). According to the Reactome pathway knowledge base, CCT8 is involved in the following pathways: **1**) folding of actin by CCT/TRiC; **2**) BBSome-mediated cargo-targeting to cilium; **3**) formation of tubulin-folding intermediates by CCT/TRiC; **4**) pre-folding-mediated transfer of substrate to CCT/TRiC; **5**) cooperation of pre-folding and TRiC/CCT in actin and tubulin folding; **6**) association of TRiC/CCT with target proteins during biosynthesis; **7**) cooperation of PDCL (PhLP1) and TRiC/CCT in G-protein beta folding; **8**) cargo trafficking to the periciliary membrane; **9**) chaperonin-mediated protein folding; **10**) protein folding; **11**) cilium assembly; **12**) organelle biogenesis and maintenance; **13**) neutrophil degranulation; **14**) innate immune system; and **15**) metabolism of proteins [183]. This study provides the first report of a promising association between LVI

mechanisms and the aberrant expression of the CCT8 protein in a large cohort of invasive BC patients.

4.1.3. **Synaptosomal-associated protein 23**

In the early stages of this study, the identification of LVI-DEGs was based on their significant correlation with histologically defined LVI status and CNA gain ratios. This method of assessment was limited to the DEGs in the LVI-NT100 cluster. Beyond *MCFD2* and *CCT8* from the latter cluster, we also selected and examined other LVI candidates for further study [43, 190, 192, 193]. The slow retrieval process of the complete transcriptomic data of the METABRIC cohort influenced us to identify LVI-related genes using an experimental approach. We aimed to scrutinise the gene sets from the molecular signatures database (MSigDB) [241, 242] in order to identify further LVI candidates. We interrogated the enriched gene expression for *MCFD2* and *CCT8* in 68 different GSEA hallmark sets. *MCFD2* appeared in five sets, and *CCT8* in four sets. The *SNAP23* protein-encoding gene was co-expressed with MCFD2 in: **1**) alveolar rhabdomyosarcoma, systematic name: M19541, provided by a mesenchymal tumours study [243]; and **2**) GO intracellular vesicle, systematic name: M17371. Furthermore, *SNAP23* was co-expressed with *CCT8* in one GSEA hallmark set, GO cell projection, systematic name: M17656. We also identified that *SNAP23* was co-expressed with the *CCT8*-correlated genes, *HSP90AA1* and *DYNLRB1*, in the last set. Both *MCFD2* and *CCT8* were upregulated in an individual GSEA set, systematic name: M1103, provided by the BC apoptosis by doxorubicin study [244].

The PWs for the SNAP23 protein were identified using the Reactome-Kb. These pathways are: **1**) translocation of *SLC2A4* (GLUT4) to the plasma membrane; **2**) clathrin-derived vesicle budding; **3**) trans-Golgi network vesicle budding; **4**) EnR-phagosome pathway; **5**) antigen processing-cross presentation; **6**) class-I major histocompatibility complex (MHC)-mediated antigen processing and presentation; **7**) neutrophil degranulation; **8**) membrane trafficking; **9**) vesicle-mediated transport; **10**) adaptive immune system; **11**) innate immune system; and **12**) the immune system. The molecular characterisation and mechanisms of the SNAP23 protein in invasive BC cancer progression is still ambiguous. The number of studies that investigated the prognostic value of SNAP23 as one of the soluble N-ethylmaleimide-sensitive factor activating protein receptor (SNARE) proteins in invasive BC is few; therefore, this study aimed to define the clinical importance of SNAP23 as one of the recently reported and necessary molecules in the SNARE-mediated membrane trafficking and docking-fusion machinery in BC invasion.

4.1.4. **SEC14-like protein 1**

The upregulated expression of *SEC14L1* mRNA could be a significant driver for LVI, poor prognosis, and unfavourable outcomes in invasive BC patients; primarily in the $ER^-/HER2^+$ or $ER^-/PR^-/HER2^-$, as we illustrated previously in Chapter 3. The *SEC14L1* gene is located within a discrete region of 17q25 that frequently shows copy number alterations in BC [206]. A proximate locus, 17q23, is well documented in BC, and the key feature of this amplified locus is its oncogenic activity in BC both in vitro and in vivo [207]. The IHC examination of SEC14L1 was carried out on TMA sections derived from the $LVI^{+/-}$ cohort and a large cohort recently constructed from invasive BC specimens derived from NUH patients. However, we presented the IHC data from the $LVI^{+/-}$ cohort only, as the clinical data of the new TMA cohort is still under validation.

The cytosolic factor *Saccharomyces cerevisiae*-like 1, or SEC14-like lipid binding-1, was reported as a negative regulator for signal transduction (GO:0009968). A protein with such MF could activate or inhibit myriad cellular events. According to the UniProt knowledge base, SEC14-like protein 1 is an MF regulator (GO:0098772) and a protein that binds to the retinoic acid-inducible gene 1 protein, DDX58, previously RIG-1 (GO:0039552). The BPs in which SEC14-like protein 1 is involved are: **1**) choline transport (GO:0015871); **2**) the innate immune response (GO:0045087); and **3**) negative regulation of the DDX58/RIG-1 signalling pathway (GO:0039536). Interestingly, there were no specific annotated pathways for *SEC14L1* in the Reactome knowledge base or PANTHER database. *SEC14L1* was one of the top DEGs that was strongly associated with LVI^+ patients, according to the results in Chapter 3. Here, we aimed to question the prognostic and biological significance of SEC14L1 protein expression in invasive BC tissues using the large and well-characterised BC series of NPS. The investigation of the prognostic impact for SEC14L1 in invasive BC patients was recently published by our group [191].

The aim of the present study was to evaluate the LVI relationship, prognostic significance, protein-protein interactions, and the altered protein expression of MCFD2, CCT8, SNAP23, and SEC14L1 in invasive BC patients. The interrogated clinicopathological prognostic variables included LVI, tumour grade, tumour size, LN involvement, NPI groups, molecular phenotypes, and patient outcome. The interrogated protein-protein network was based on the IHC results for previously assessed proteins in the same studied cohort. Our interpretation of the prognostic and molecular impact of MCFD2, CCT8, SNAP23, and SEC14L1 altered protein expression may reveal new insights into the underlying pathways of LVI in invasive BC.

4.2. Materials and methods

4.2.1. Study cohorts

4.2.1.1. *Nottingham Primary Series of invasive BC (NPS)*

4.2.1.1.1. Construction of the TMA blocks from the NPS cohort

The non-selective primary operable invasive BC specimens from this cohort of patients were used in the construction of the 13 TMA FFPE blocks that were sectioned and immuno-stained in this study. The retrieved FFPE blocks were sectioned and stained with H&E to assess the optimum blocks with sufficient invasive legions for proper sampling. The representative tumour areas on each slide for each patient were marked alongside their corresponding spots on the FFPE blocks. The marked tumours from the donor blocks were collected using a 0.6mm probing needle, and the recipient blocks were designed to hold a maximum of 150 cases. The TMA maps were designed using Microsoft Excel (2013) software to reflect the actual distribution of the TMA cores on the slides. Each slide was labelled with an identifier from PS1–PS13, and the rows were numbered from Y1–Y17 as we skipped one row after each five loadings, and the columns were numbered from X1– X11. The X1 column was used to orient our navigation during the process of histological scoring, and it contained TMA cores of liver or kidney tissue, which were also used as controls. The constructed TMA from this cohort was IHC stained with anti-MCFD2, anti-CCT8, and anti-SNAP23 primary Abs.

4.2.1.1.2. Clinicopathological and histological profiles of the NPS cohort

The detailed clinicopathological profiles of the study samples included patients' age, tumour size, tumour grade-I, -II, and -III, $LN^{+/-}$ status, LN stage-I, -II, and -III, NPI groups, the different LVI statuses (LVI^{-}/LN^{-}, LVI^{-}/LN^{+}, and LVI^{+}), tumour histological types, ER, PR, HER2, and triple-negative status of hormonal receptors expression. The BCSS, DFS, and DMFS were recorded for up to 25 years and included information on local and regional recurrences. The median for BCSS was 177 months (with a range of 1–308 months) and 114.50 months (with a range of 1–310 months) for DFS. This cohort showed that 65.8% of patients were over the age of 50 years, 48.0% showed a tumour size >20mm, 19.9% had tumour grade-I, and 34.2% exhibited tumour grade-II, 63.6% were LN^{-}, whereas 26.2% illustrated $LN^{stage\text{-}II}$, 7.8% were $LN^{stage\text{-}III}$, and 51.4% and 15.4% showed moderate and poor NPI, respectively. The LVI statuses were: **1**) 44.8% as LVI^{-}/LN^{-}; **2**) 18.0% as LVI^{-}/LN^{+}; and **3**) 37.2% as LVI^{+}. The IDC was the dominant histological type in 80.3% of patients. The hormonal receptor expression profiles were 74.5% as ER^{+}, 59.3% as PR^{+}, 12.3% as $HER2^{+}$ (IHC/CISH), 15.7% as triple-negative, 25.3% as

basal, and 11.0% as triple-negative and basal. Cohort features are mentioned in chapter two (Table 2.4).

4.2.1.2. *LVI-supervised cohort ($LVI^{+/-}$)*

This cohort comprised only the $LVI^{+/-}$ samples that were used in the DGE in chapter three. Fewer TMA slides were constructed individually for this subset of the METABRIC patients that were originally derived from the NPS. The total number of patients in this cohort was 170, as the number of patients with LVI^{+} status was 90 and LVI^{-}/LN^{-} was 80. However, the number of patients could be different from the number stated above owing to the absence of tumour cores in the sectioned TMA. The final number of TMA cores with scorable tumour staining was 168. The constructed TMA from this cohort was IHC stained with anti-SEC14L1 primary Ab only. The BCSS, DFS, and DMFS were recorded for up to 25 years and included information on local and regional recurrences. The median for BCSS was 158.50 months (with a range of 5–248 months) and 140.50 months (with a range of 5–186 months) for DFS. Cohort features are mentioned in chapter two (Table 2.2).

4.2.2. **Tissue handling and sectioning**

All details regarding the handling and processing of tissues and the FFPE blocks full-face and TMA sectioning are detailed in Chapter 2 (section 2).

4.2.3. **Purchased antibodies and their optimised conditions for IHC and WB**

In this study, the use of primary Abs was for the semi-quantitative detection of protein expression in the invasive BC TMA cores, and for descriptive characterisation of protein expression on full-face sections applying IHC procedures. Also, specific primary Abs were used to qualitatively detect the Ab-Ag binding specificity in the WB methods.

The first Ab that was purchased and used was the anti-MCFD2 primary Ab produced in mouse (Abcam, UK, ab57843). The optimised dilution factor for the anti-MCFD2 primary Ab was 1:250 for IHC. However, the anti-MCFD2 primary Ab detected no protein bands in WB. This result was validated externally by reviewing the results of other manufacturers (Sigma-Aldrich, UK, HPA034593; and Merck, UK, MAB4324). In order to assess the specific binding capacity of the anti-MCFD2 primary Ab, we purchased the MCFD2 WB recombinant protein (Abnova, UK, H00090411-P01) to perform IPB protocol on two consecutive tissue sections from the same FFPE block. The optimised concentration of the recombinant Ag solution required us to totally neutralise the binding capacity of the anti-MCFD2 primary Ab, and was 0.25µg/ml.

The second antibody was the anti-CCT8 primary Ab produced in mouse (Sigma-Aldrich, UK, HPA021051). The optimised dilution factors for the anti-CCT8 primary Ab were 1:100 for IHC and 1:250 for WB. The third Ab was the anti-SNAP23 primary Ab produced in rabbit (Sigma-Aldrich, UK, HPA001214). The optimised dilution factors for the anti-SNAP23 primary Ab were 1:150 for IHC and 1:500 for WB. Finally, for the anti-SEC14L1 primary Ab produced in rabbit (Sigma-Aldrich, UK, HPA028703), the optimised dilution factors were 1:25 for IHC and 1:500 for WB. Extended details on the methodologies for the validation of specific Ab-Ag binding by WB and IPB are detailed in (sections 2.4. and 2.5.).

4.2.4. IHC, digital imaging, and scoring of stained sections

In this study, the detailed procedures of IHC, digital imaging, and H-scoring are referred to in Chapter 2. Briefly, all full-face sections for protein expression characterisation and TMA sections were loaded onto charged 25 × 75mm adhesion slides. The slides were labelled with an IHC marker pen. The labelling included the date of the experiment, protein identifiers, user name, dilution factor of applied antibody, antigen retrieval method, and buffer. The slides were deparaffinised on a Leica hot plate (Leica HI1220 flattening table, product code: 14042321474, Leica Biosystems, UK) face up at 60°C for 10 minutes. The slides were allowed to cool, then they were inserted into the xylene chamber of the Leica AXL. The automated dewaxing programme was activated and run as detailed in the second chapter.

After complete removal of paraffin wax from all tissue slides, the HIER method, citrate buffer, and pH 6.00 were designated as optimum conditions for the antigen exposure. A plastic vessel was filled with 500ml of citrate buffer solution, pH 6.00, and the dewaxed slides were submerged. The vessel was closed carefully with a plastic lid, with no metal handles. The slide-containing vessel was placed inside a laboratory microwave and operated at maximum power for 20 minutes. Protective glasses were worn and the vessel was removed from the microwave with a heat-resistant pair of gloves. The vessel lid was removed in the sink and cool water was applied into the vessel, displacing the citrate buffer and cooling the slides for five minutes. All slides were assembled onto Shandon cover-plates™ (catalogue number: 72110017, Fisher Scientific, part of ThermoFisher Scientific, UK) and inserted into the Shandon Sequenza immunostaining slide rack (Product number: 36105, Ted Pella, USA). Quality control checks were applied to ensure proper staining.

The IHC procedure was performed using Novolink™ Max Polymer Detection System (product code: RE7280-K, Leica, Biosystems, UK), which is suitable for antigen detection by using mouse IgG and rabbit IgG. The first step was to block

the endogenous peroxidases using 100µl of peroxidase blocking reagent (product number: RE7101, Leica, Biosystems, UK) onto each slide for five minutes. Two TBS washes were carried out for five minutes each. Then, non-specific reacting proteins were neutralised by applying of 100µL of the protein block reagent (product number: RE7102, Leica, Biosystems, UK), followed by two TBS washings for five minutes.

All primary antibodies were diluted in a specific volume of Bond™ primary antibody diluent (catalogue number: AR9352, Leica Biosystems, UK) to reach the optimum working dilution: 1:250 for anti-MCFD2 primary Ab, 1:100 for anti-CCT8 primary A, 1:500 for anti-SNAP23 primary Ab, and 1:25 for anti-SEC14L1 primary Ab. All primary antibodies were applied at 100µl and incubated with tissues for 60 minutes at room temperature. Two TBS washes were carried out for five minutes each. Post-primary antibody mix (product number: RE7111, Leica Biosystems, UK) was applied at 100µl to all slides from the previous step and incubated for 30 minutes. This mixture contained rabbit anti-mouse IgG <10 µg/ml in 10% (v/v) animal serum in TBS (0.09%) with ProClin™ 950 as a preservative. Two TBS washes were carried out for five minutes each. Then, 100µl of the Novolink™ polymer (product number: RE7112, Leica Biosystems, UK) was added to each slide, followed by two TBS washes for five minutes each. This polymer binds to rabbit IgGs and detects post-primary and rabbit-hosted primary antibodies. The chemical composition of this reagent is anti-rabbit poly-HRP-IgG <25µg/ml containing 10% (v/v) animal serum in TBS (0.09 %), with ProClin™ 950 as a preservative. The DAB working solution was prepared by mixing DAB chromogen (product number: RE7105, Leica Biosystems, UK) with Novolink™ DAB substrate buffer polymer (product number: RE7143, Leica Biosystems, UK) at 1:20. The DAB working solution was applied to all slides at 100µl for five minutes. Two TBS washes were carried out for five minutes each. Then, all tissue slides were counterstained with 100µl of haematoxylin for six minutes. The slides were gently disassembled from the Shandon cover-plates in a cold-water bath and the tissues were dehydrated using the Leica AXL activating dehydrating programme. Then, all slides were mounted in DPX medium and sealed with a transparent cover glass, or coverslip. The stained slides were archived in physical and digital formats.

High-resolution digital images of the slides were generated with the Nanozoomer Digital Pathology scanner (NDP C9600 series; Hamamatsu Photonics, Welwyn Garden City, Hertfordshire, UK). Scanning at 20-times magnification was applied to all slides and NDP.view2 software (Hamamatsu Photonics Europe GmbH, Systems Division, UK) was used to facilitate the scoring of the TMA cores using a

high-resolution screen (24 inches, 1920 x 1080 pixels). The protein expression of MCFD2, CCT8, SNAP23, and SEC14L1 was semi-quantitatively scored, as explained in Chapter 2.

4.2.5. Statistical analysis

Statistical analyses were conducted using SPSS statistical software v24.0 (IBM, Armonk, NY, USA). A chi-square test was used in the analysis of the categorical MCFD2, CCT8, and SNAP23 protein expression levels in relation to clinical variables, including the different categories of LVI and several proteins. The protein expression was assessed for ER, PR, HER2, CD3, CD8, CD20, CD68, FOXP3, CD24, CD44, AR, GATA3, Ki-67, E-cadherin, N-cadherin, p53, and TWIST2. KM survival curves for BCSS, DMFS, and DFS were plotted, and the log-rank test was conducted. The BCSS in this study was determined as the time from the day of surgery to the day of death caused by BC. Cox proportional hazards regression model, 95% CI, was applied for the multivariate survival assessment of clinicopathological factors, including the MCFD2, CCT8, SNAP23, and SEC14L1 protein expression. The median protein expression value was set as a cut-off point for MCFD2, CCT8, SNAP23, and SEC14L1 in all statistical analyses. For all tests, a two-tailed p-value of <0.05 was considered significant.

4.3. Results

4.3.1. Evaluation of the results of WB and IPB

The anti-MCFD2 primary antibody did not show any protein bands, but the protein bands for β-actin were visualised on the same transferring membrane. The same result was produced three times with different dilution factors. Our results were validated externally using different anti-MCFD2 primary antibodies from different manufacturers. Therefore, the use of the IPB technique was suggested especially after the optimisation invasive BC tissue sections stained positively with the anti-MCFD2 primary antibody in this study. The comparative examination of two invasive BC TMA consecutive sections revealed that the TMA slide of which the primary anti-MCFD2 primary antibody was totally blocked illustrated total immunostaining negativity. On the other hand, the TMA cores on the parallel slide showed different intensities of immunostaining (Figure 4.1). The evaluation of WB images for the anti-CCT8, anti-SNAP23, and anti-SEC14L1 primary antibodies supported a specific binding affinity to a single protein band for each cell lysate around the predicted molecular masses of 59 kDa for the CCT8 protein (Figure 4.2), 23 kDa for the SNAP23 protein (Figure 4.3), and 77 kDa for the SEC14L1 protein (Figure 4.4).

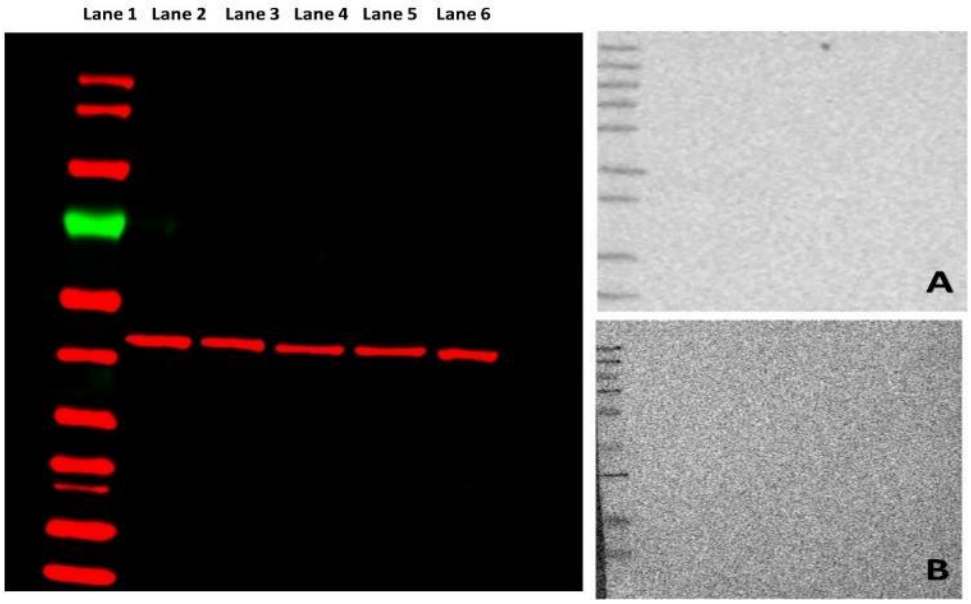


Figure 4. 1: **Left-sided image:** binding specificity of the anti-MCFD2 antibody, Abcam, UK, ab57843. WB membrane showed no antigen-binding for anti-MCFD2 antibody (green bands) around ~16 KDa. The positive protein expression of β-actin, as endogenous control, was detected around ~42 KDa (red bands). Protein ladder (180 to 10 KDa), lane 1, MCF7, lane 2, SKBR3, lane 3, HeLa, lane 4, MDA-MB-231, lane 5, MDA-MB-436, lane 6.
Right-sided images A and B: binding specificity of the anti-MCFD2 antibody in western blotting manufactured by (A) Sigma-Aldrich, HPA034593, and (B) Merck, MAB4324. Neither MCFD2 nor β-actin protein bands were detected on both WB membranes.

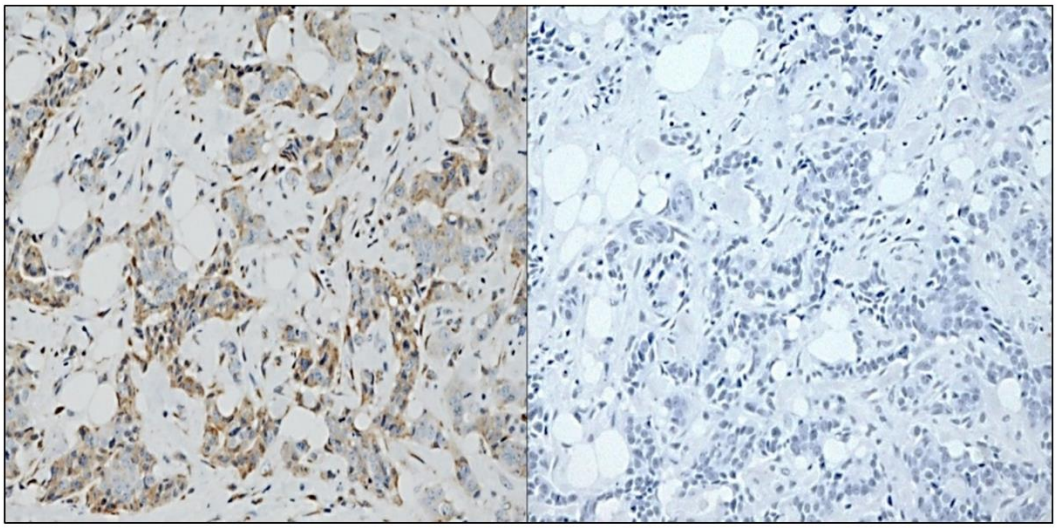


Figure 4. 2: immunising peptide blocking protocol on two consecutive invasive BC tissue sections from the same FFPE block. **Right-sided image:** positive IHC staining with the non-blocked anti-MCFD2 antibody, Abcam, UK, ab57843. **Left-sided image:** Negative IHC staining of anti-MCFD2 antibody, Abcam, UK, ab57843, blocked with 0.25µg/ml MCFD2 WB recombinant protein (Abnova, UK, H00090411-P01).

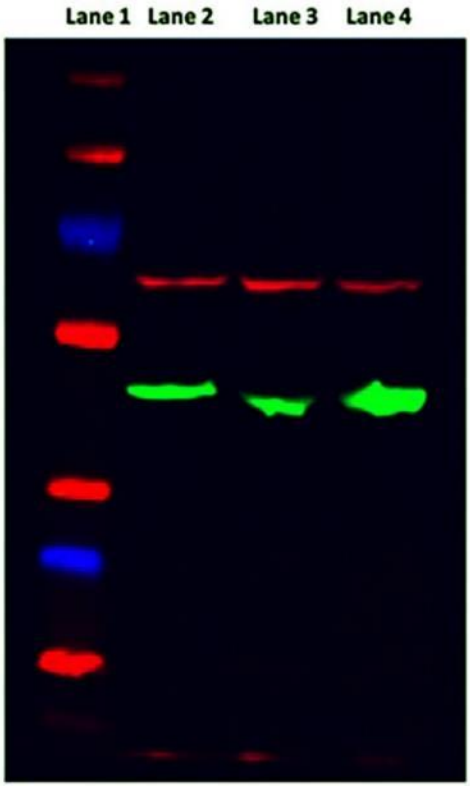


Figure 4. 3: Binding specificity of the anti-CCT8 antibody, Sigma-Aldrich, UK, HPA021051.
WB membrane showed specific antigen-binding of for anti-CCT8 antibody (red bands) around ~59 KDa. The positive protein expression of β-actin, as endogenous control, was detected around ~42 KDa (green bands). Protein ladder (180 to 10 KDa), lane 1, MCF7, lane 2, SKBR3, lane 3, MDA-MB-231, lane 4.

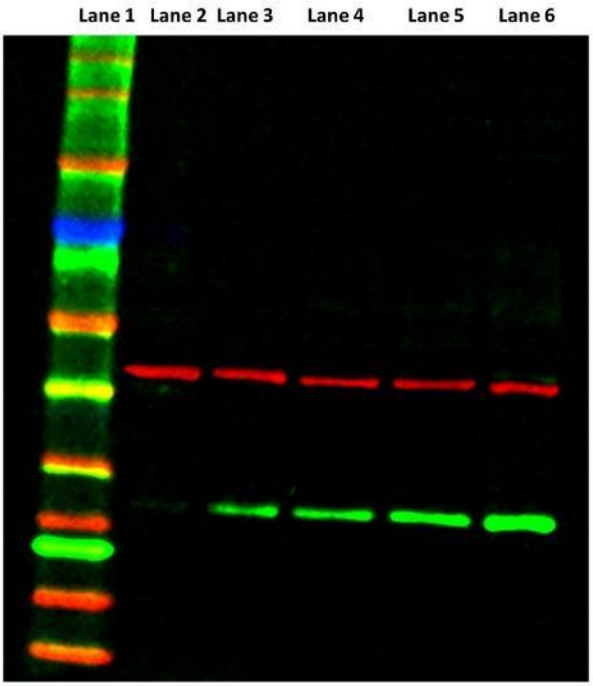


Figure 4. 4: Binding specificity of the anti-SNAP23 antibody, Sigma-Aldrich, UK, HPA001214.
WB membrane showed specific antigen-binding of for anti-SNAP23 antibody (green bands) around ~23 KDa. The positive protein expression of β-actin, as endogenous control, was detected around ~42 KDa (red bands). Protein ladder (180 to 10 KDa), lane 1, MCF7, lane 2, SKBR3, lane 3, HeLa, lane 4, MDA-MB-231, lane 5, MDA-MB-436, lane 6.

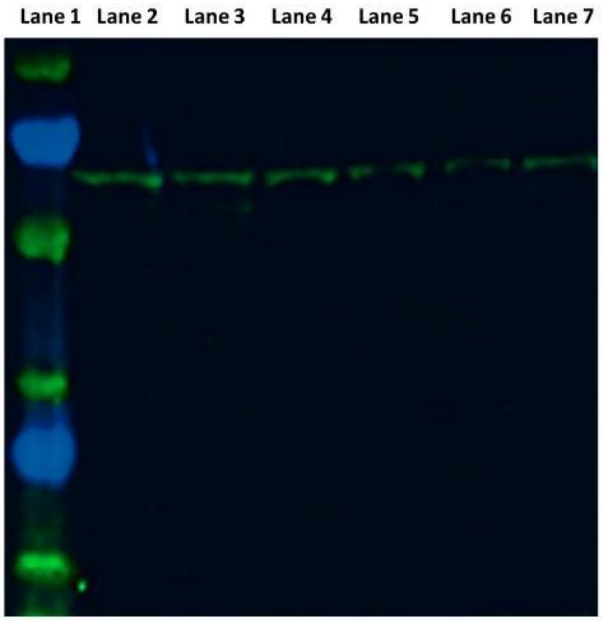


Figure 4. 5: Binding specificity of the anti-SEC14L1 antibody, Sigma-Aldrich, UK, HPA028703.
WB membrane showed specific antigen-binding of for anti-SNAP23 antibody (green bands) around ~77 KDa. Protein ladder (180 to 10 KDa), lane 1, MCF7, lane 2, HeLa, lane 3, MDA-MB-231, lane 4, Jurkat, lane 5, HEK293, lane 6, MDA-MB-468, lane 7. The detection of positive protein expression of β-actin, as endogenous control, was detected around ~42 KDa on a separate WB membrane for the same cell lysates used.

4.3.2. **Characterisation of protein expression patterns on full-face tissue sections**

Full-face sections of an invasive BC case showed a decrease in IHC staining intensity for the MCFD2 protein from normal to invasive, from A to C (Figure 4.6). The CCT8 protein was homogeneously expressed in the examined full-face tissues with positive stromal staining as illustrated in (Figure 4.7). The protein expression patterns for SNAP23 is shown in (Figure 4.8). Lastly, SEC14L1 protein expression patterns are described in (Figure 4.9).

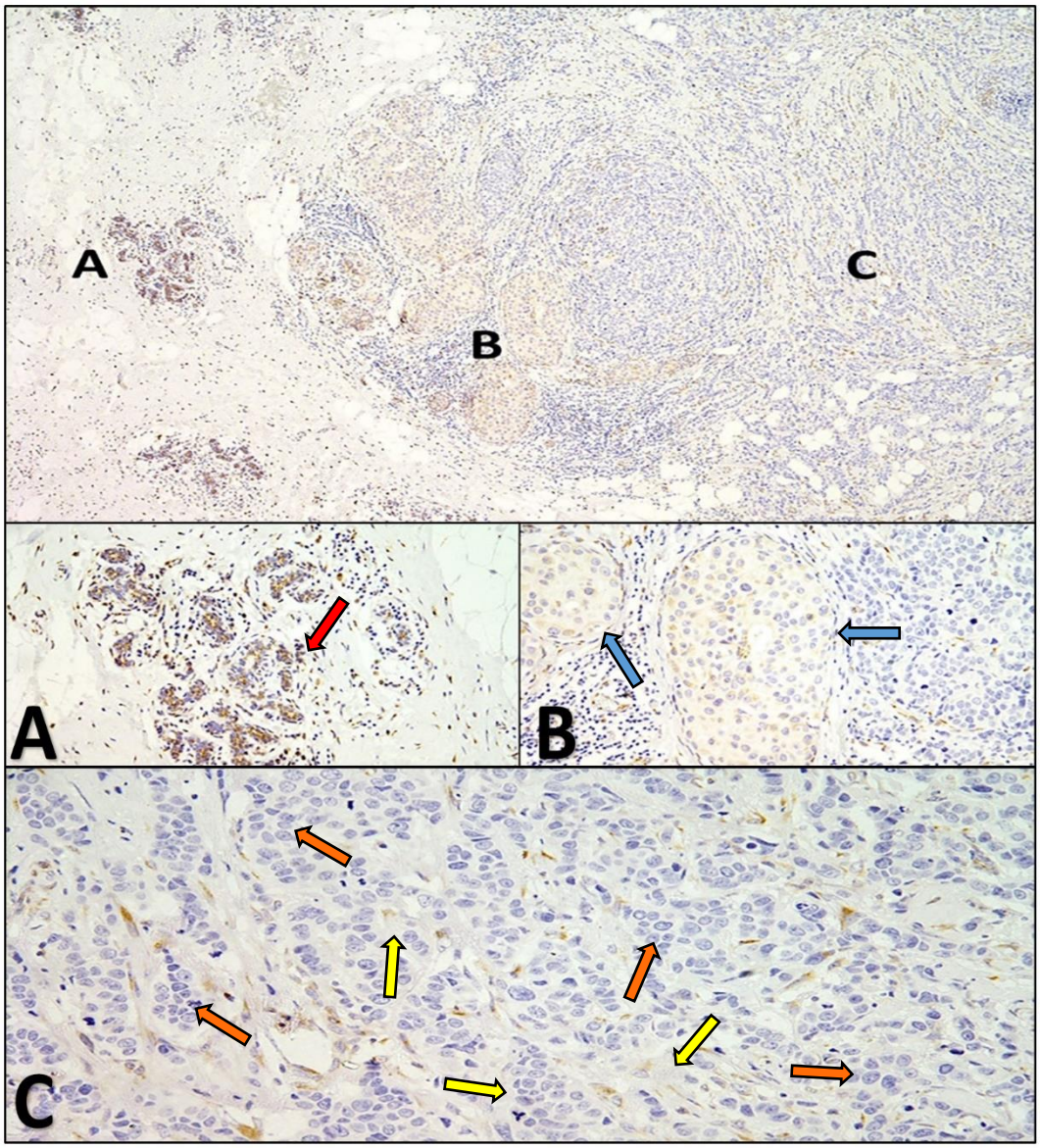


Figure 4. 6: Protein expression characterisation of MCFD2 in invasive BC. A normal TDLU adjacent to a malignant legion, red arrow, showed strong expression of MCFD2 in the luminal epithelia (Figure 4.6.A). A DCIS component exhibited a significant decrease in MCFD staining, blue arrow, (Figure 4.6.B). Total absence of MCFD2 staining positivity is associated with invasiveness features, yellow arrows (Figure 4.6.C). The MCFD2 protein was only localised to the cytoplasm and no nuclear or stromal staining were observed, orange arrows. The expression of MCFD2 showed homogeneous characteristics. Images A, B were captured at 10-times magnification. Image C was captured at 20-times magnification

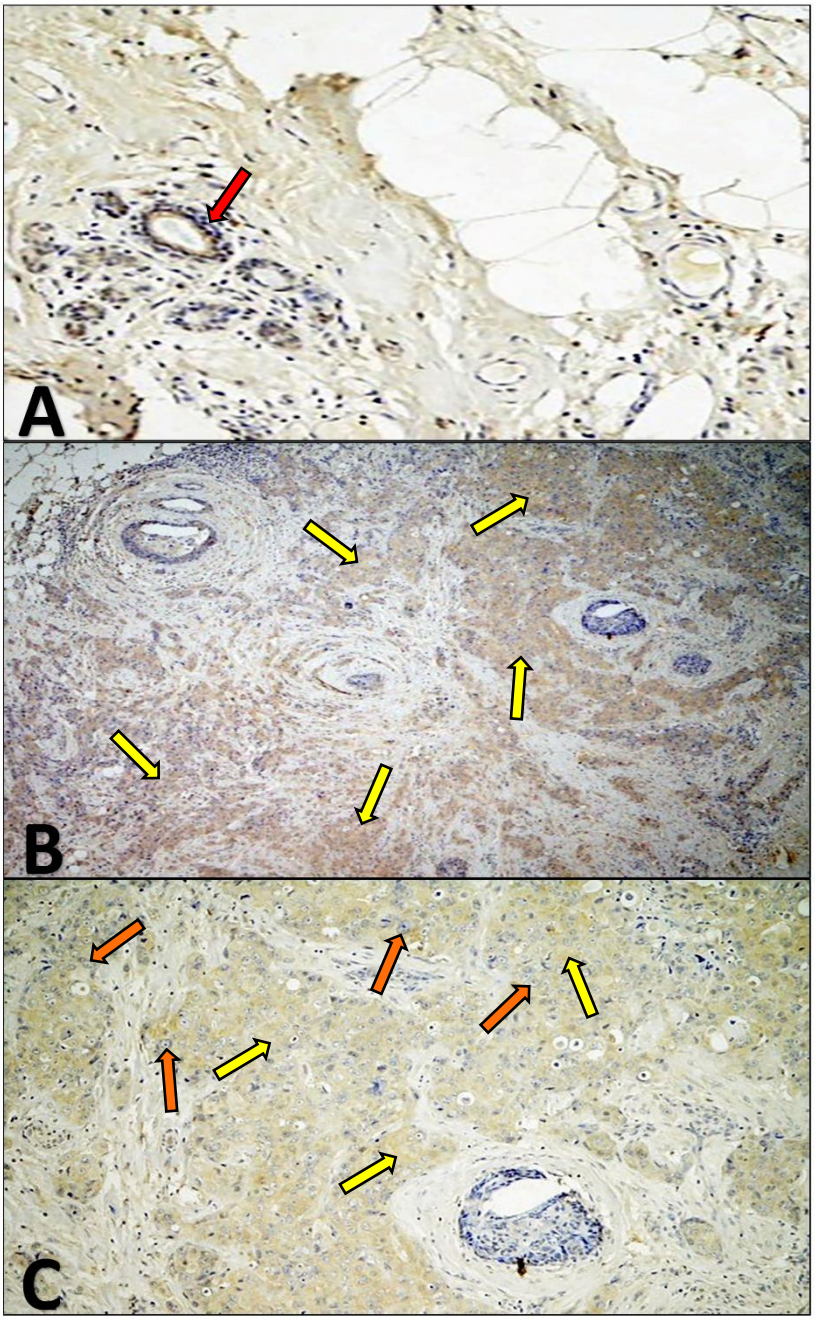


Figure 4. 7: Protein expression characterisation of CCT8 in invasive BC. The luminal epithelial cells of the TDLU, red arrow, showed strong nuclear staining (Figure 4.7.A). Nuclear staining was completely absent in all invasive BC lesions, orange arrows, that showed moderate to strong cytoplasmic protein expression of CCT8, yellow arrows (Figures 4.7.B and 4.7.C). Images A and C were capture at 20-times magnification, while image B was captured at 10-times magnification.

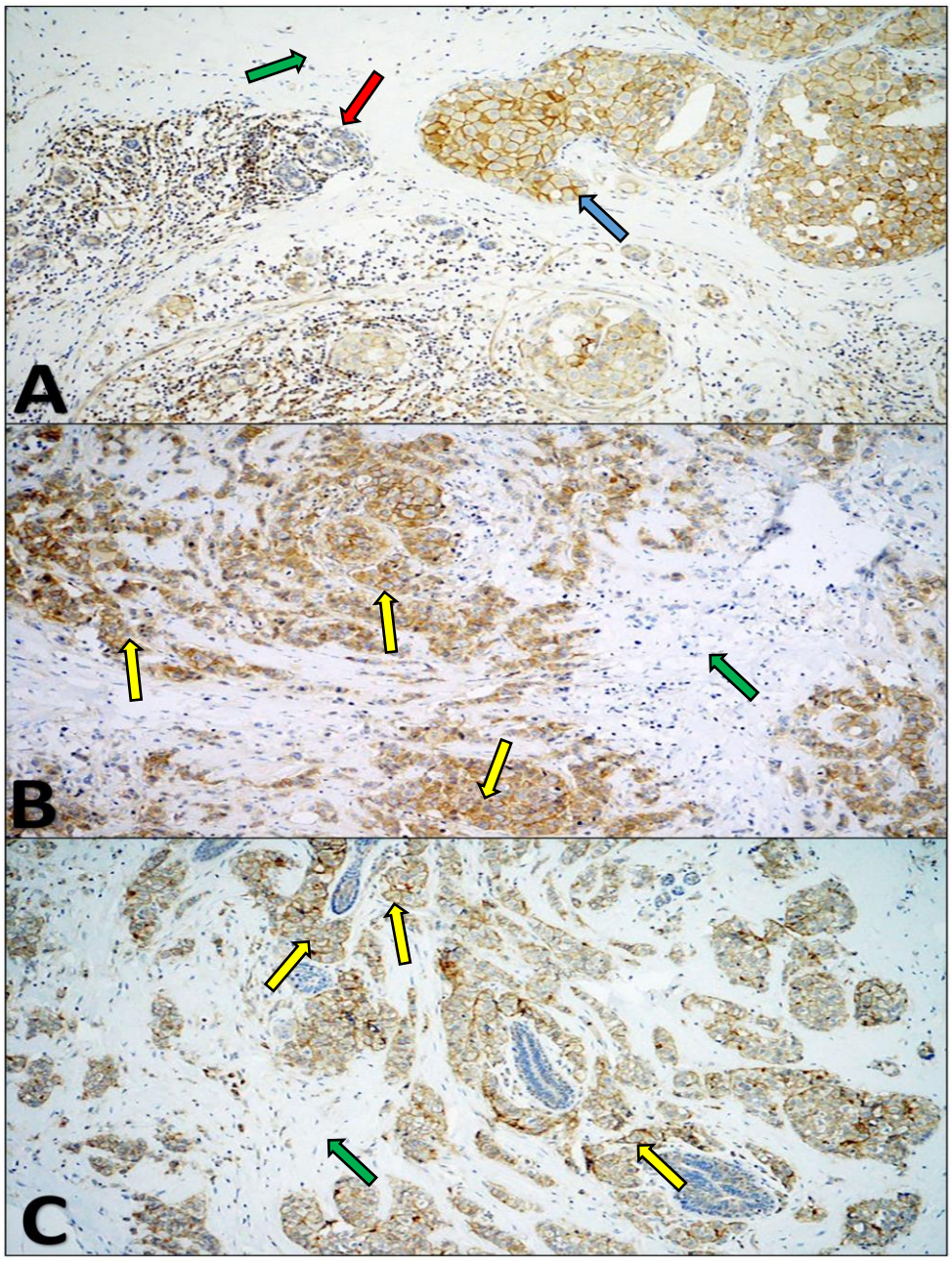


Figure 4. 8: Protein expression characterisation of SNAP23 in invasive BC. The expression of SNAP23 protein was absent or very weak in the luminal epithelial cells of TDLU, red arrow, (Figure 4.8.A). Protein expression of SNAP23 was localised to the cytoplasm and the cellular membranes of the malignant compartment, blue arrow. (Figure 4.8.A). In the invasive BC cells (Figures 4.8.B and 4.8.C), SNAP23 protein expression was observed in the cytoplasm of invasive cells, primarily, but few cells also showed positive membranous staining, yellow arrows. The protein expression of SNAP23 was not detected in the tumour-surrounding stroma, green arrows. All images were captured at 10-times magnification.

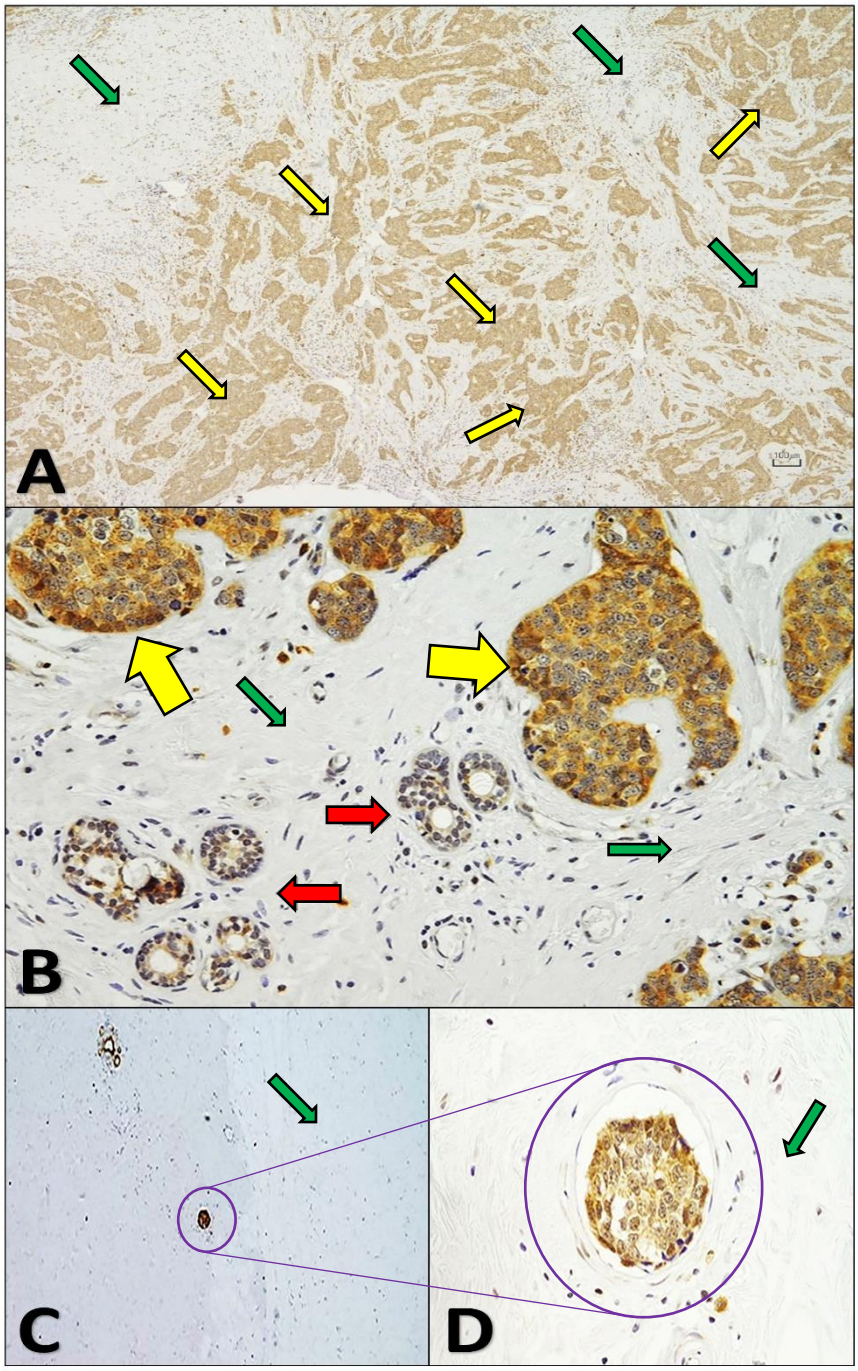


Figure 4. 9: Protein expression characterisation of SEC14L1 in invasive BC. The SEC14L1 protein showed strong cytoplasmic expression in the epithelial cells of the invasive lesions, yellow arrows, (Figure 4.8.A). The intensity of SEC14L1 staining was higher in the peripheral cells, thick yellow arrows. Weak SEC14L1 protein expression was detected in the adjacent normal TDLU, red arrows (Figure 4.8.B). A single observation of a possible LVI/BVI incidence (Figure 4.8.C) at 10X, and (Figure 4.8.D) at 40 X. No stromal staining was observed, green arrows. Assessment of full-face sections revealed a homogeneous expression pattern for all examined proteins, confirming the reliability of using TMA to detect its expression.

4.3.3. Clinicopathological relevance and protein interactions for the selected LVI candidates

In this study, we applied IHC methodologies to assess the altered protein expression of MCFD2, CCT8, SNAP23, and SEC14L1 for their prognostic importance, including the association with LVI, and their protein interactions against a selected cluster of proteins associated with BC progression. The examination of the immuno-stained TMA sections revealed a unique localised distribution for the proteins mentioned above among the different cellular compartments. The protein expressions of MCFD2 and SEC14L1 were cytoplasmic, (Figure 4.10) and (Figure 4.11). The protein expression of CCT8 was cytoplasmic, (Figure 4.12), and nuclear, (Figure 4.13). The protein expression of SNAP23 was cytoplasmic, (Figure 4.14), and membranous, (Figure 4.15). The designated cut-off points were at 55 H-score for cytoplasmic MCFD2, 10 H-score for cytoplasmic CCT8, 5 H-score for nuclear CCT8, 100 H-score for cytoplasmic SNAP23, 50 H-score for membranous SNAP23, and 70 H-score for cytoplasmic SEC14L1. The IHC staining intensities showed negative, weak, moderate, and strong staining for the four interrogated proteins.

Positive MCFD2 protein expression was observed in 49.0% of patients, while negative expression of MCFD2 was confirmed in 51.0%. Similarly, positive cytoplasmic CCT8 expression was seen in 49.5%, while the remaining patients (50.5%) showed negative cytoplasmic expression for CCT8. Positive nuclear CCT8 was observed in 46.8%, and negative nuclear expression of CCT8 was confirmed in 53.2%. The combined CCT8 was as follows: cytoplasmic^{+}/nuclear^{+} in 21.3%, cytoplasmic^{+}/nuclear^{-} in 27.8%, cytoplasmic^{-}/nuclear^{+} in 24.5%, and cytoplasmic^{-}/nuclear^{-} in 26.4%. The cytoplasmic expression of SNAP23 was confirmed in 33.7% of patients, while the remaining patients (66.3%) showed negative cytoplasmic expression of SNAP23. The membranous positive expression of SNAP23 was observed in 48.4% of patients, whereas 51.6% showed negative membranous expression of SNAP23. The combined expression profiles for SNAP23 were as follows: cytoplasmic^{+}/membranous^{+} in 27.6%, cytoplasmic^{+}/membranous^{-} in 6.1%, cytoplasmic^{-}/membranous^{+} in 20.7%, and cytoplasmic^{-}/membranous^{-} in 45.6%. Finally, the positive expression of SEC14L1 protein was observed in 40.5% of patients, while negative expression of SEC14L1 was confirmed in 59.5%. In the following subsections, current findings for these four molecular regulators of LVI are presented.

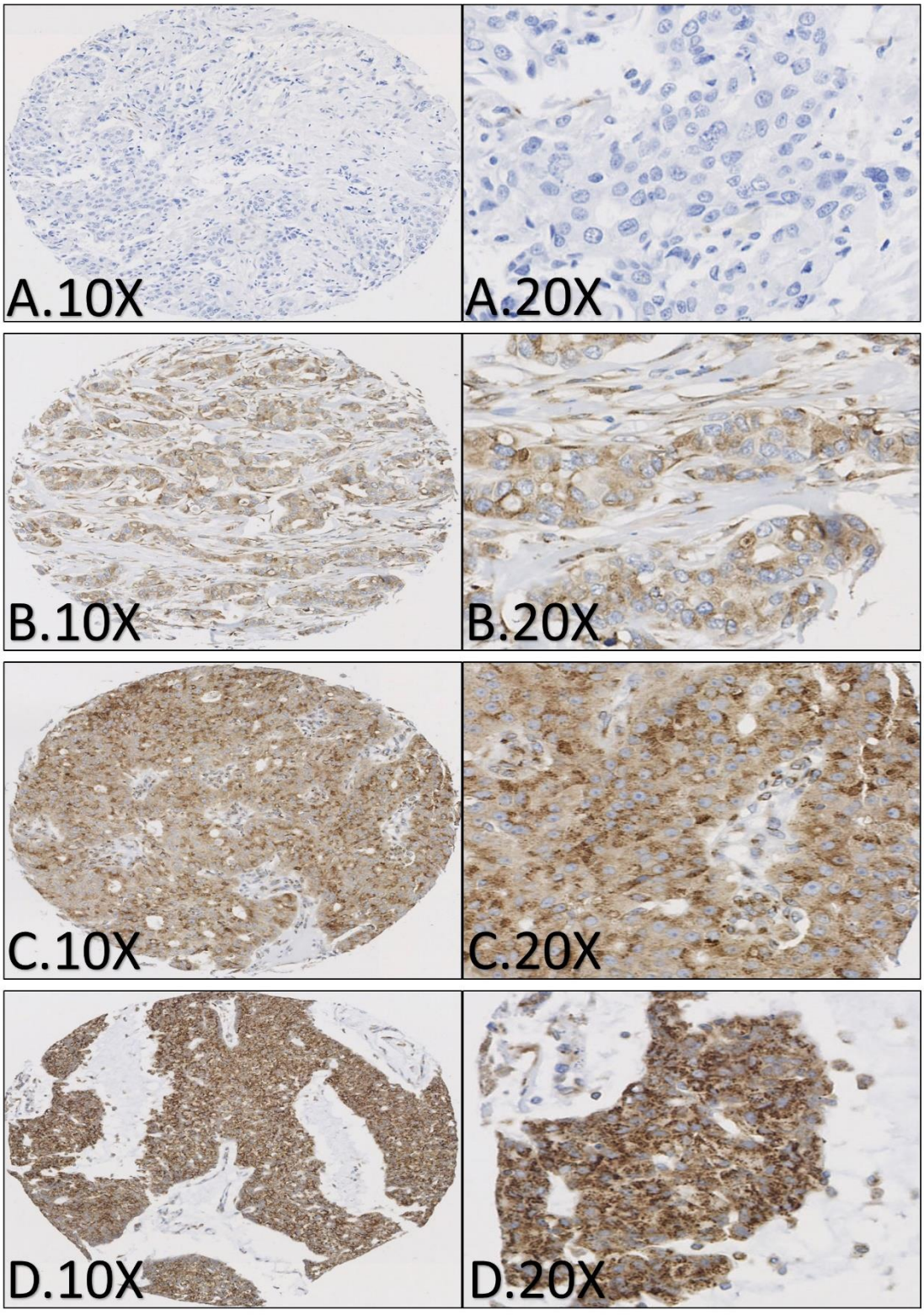


Figure 4. 10: Immunohistochemistry expression of MCFD2 in tissue microarray images.
Cytoplasm expression of MCFD2 in cancer cells was distributed as follows; **a**) No staining, **b**) weak staining, **c**) moderate staining, and **d**) strong staining. Images were captured at 10- and 20-times magnifications.

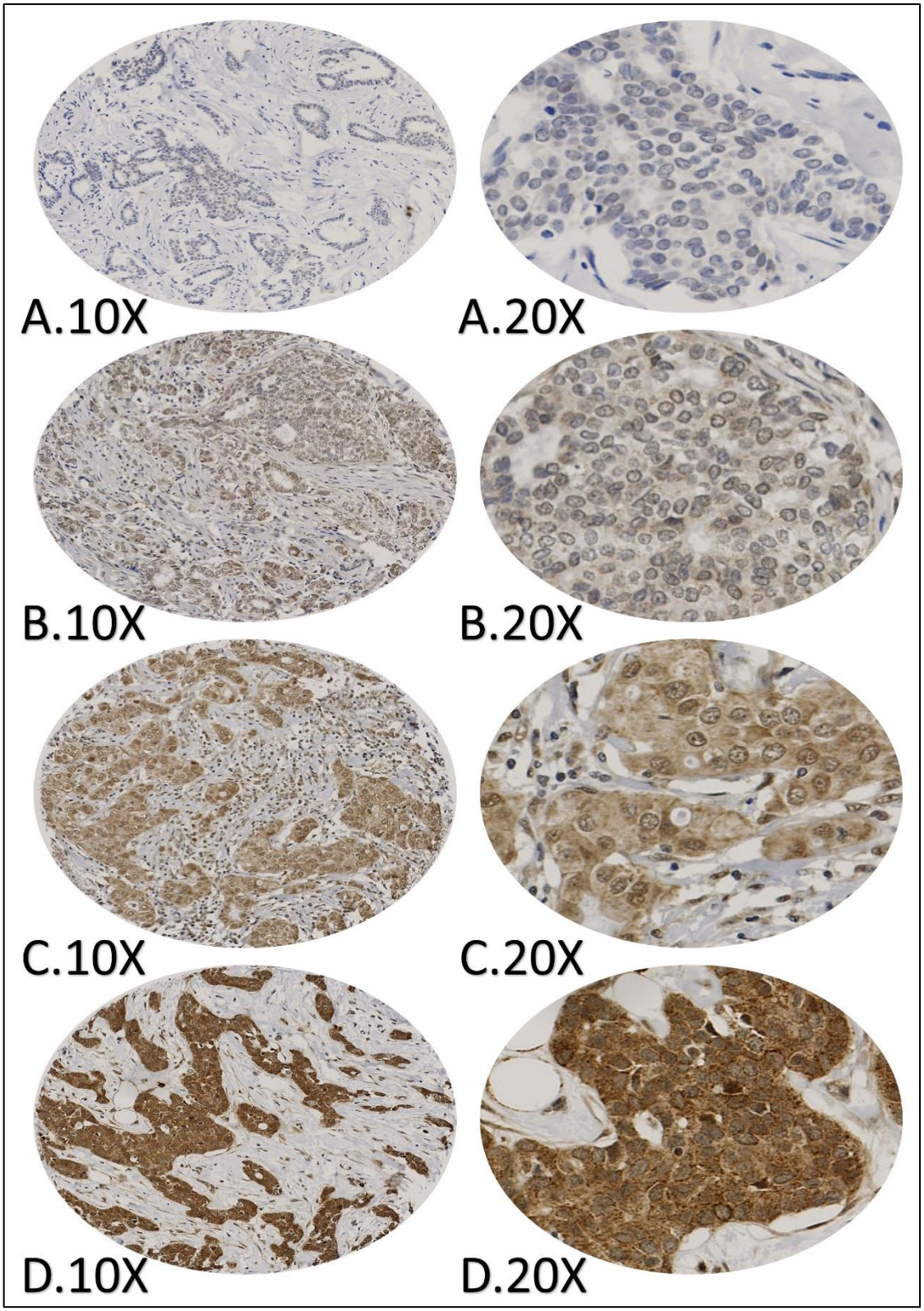


Figure 4. 11: Immunohistochemistry expression of SEC14L1 in tissue microarray images.
Cytoplasm expression of SEC14L1 in cancer cells was distributed as follows; a) No staining, b) weak staining, c) moderate staining, and d) strong staining. Images were captured at 10- and 20-times magnifications.

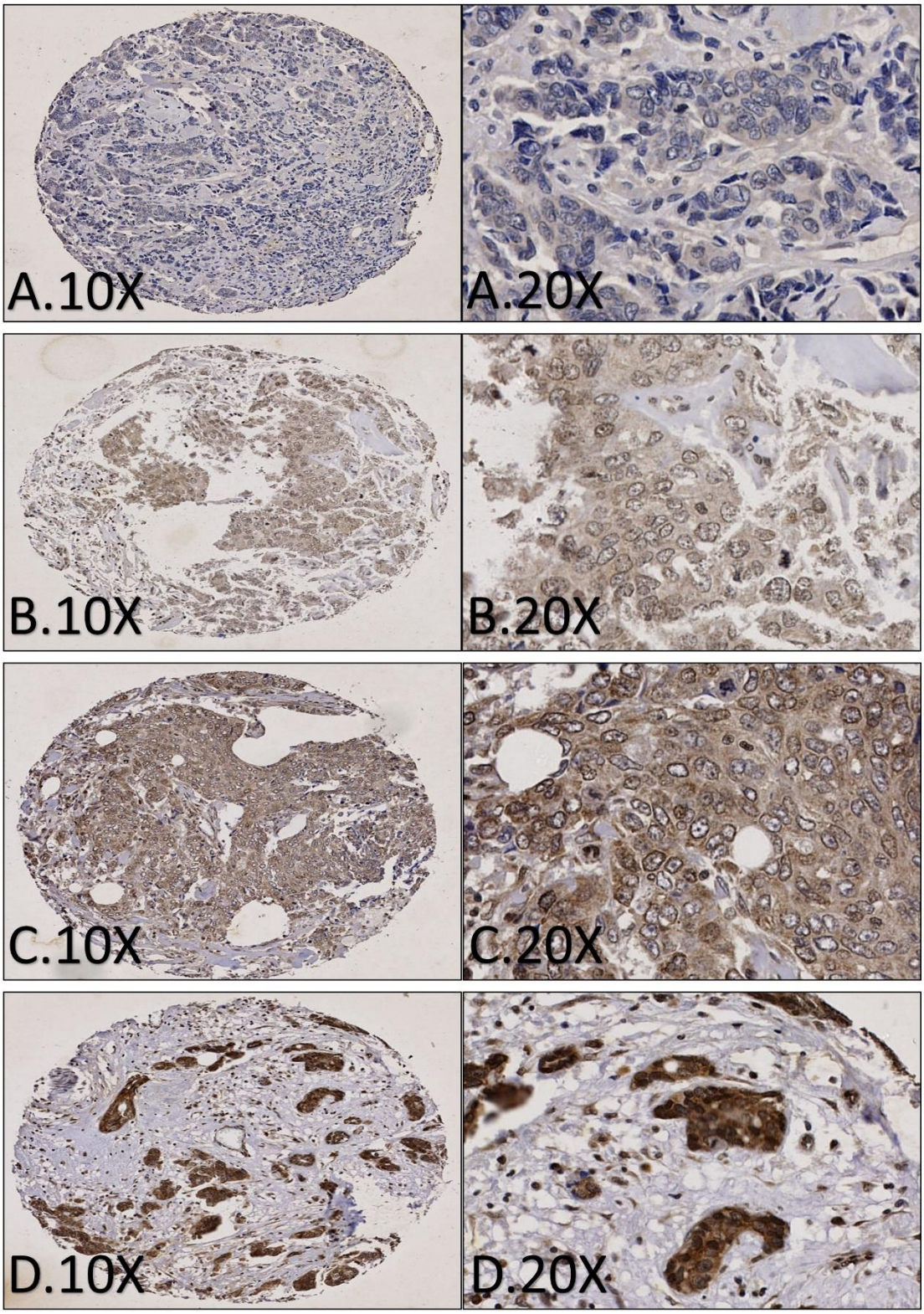


Figure 4. 12: Immunohistochemistry expression of cytoplasmic CCT8 in tissue microarray images. Cytoplasm expression of CCT8 in cancer cells was distributed as follows; a) No staining, b) weak staining, c) moderate staining, and d) strong staining. Images were captured at 10- and 20-times magnifications.

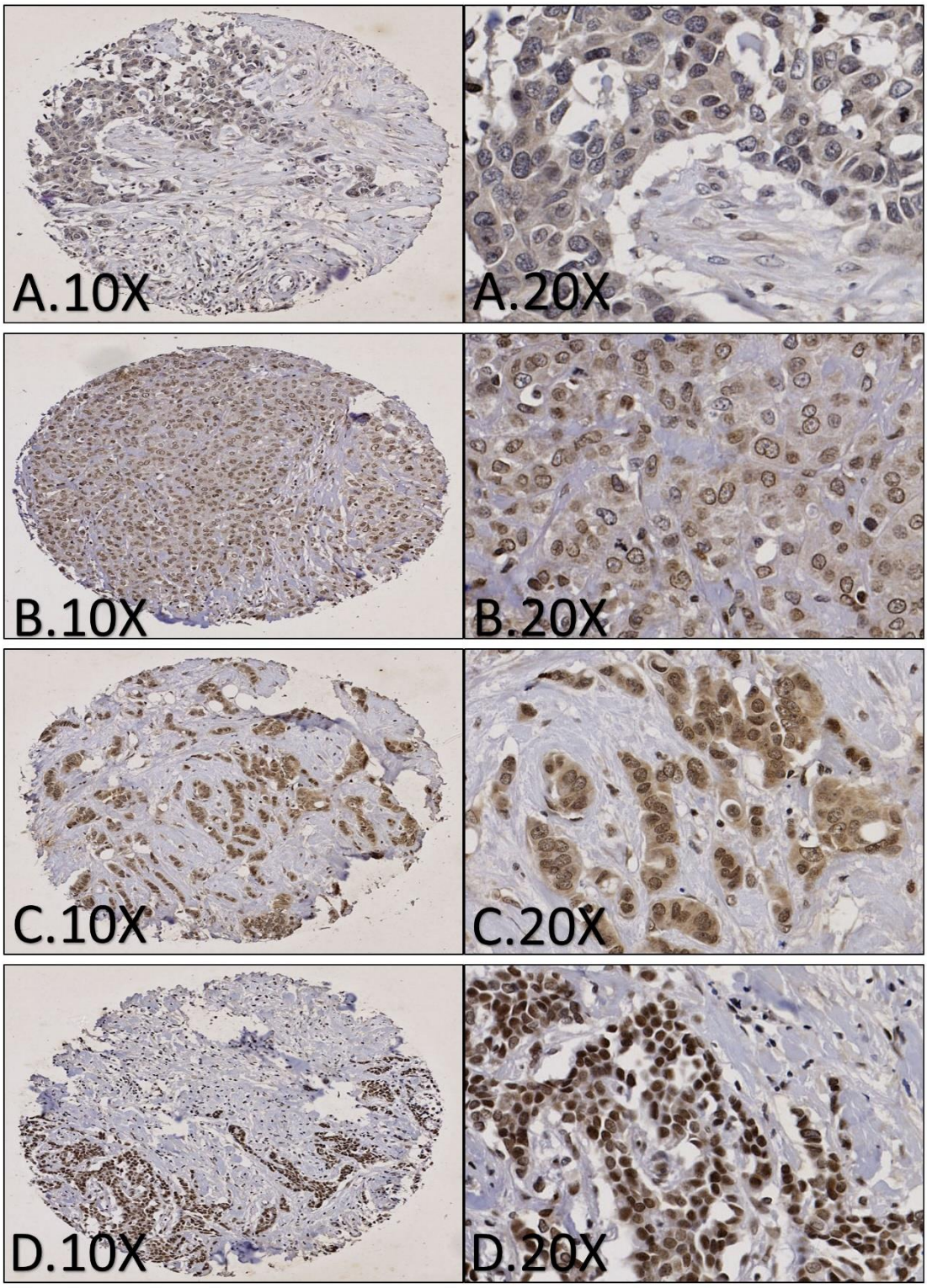


Figure 4. 13: Immunohistochemistry expression of nuclear CCT8 in tissue microarray images. Nuclear expression of CCT8 in cancer cells was distributed as follows; a) No staining, b) weak staining, c) moderate staining, and d) strong staining. Images were captured at 10- and 20-times magnifications.

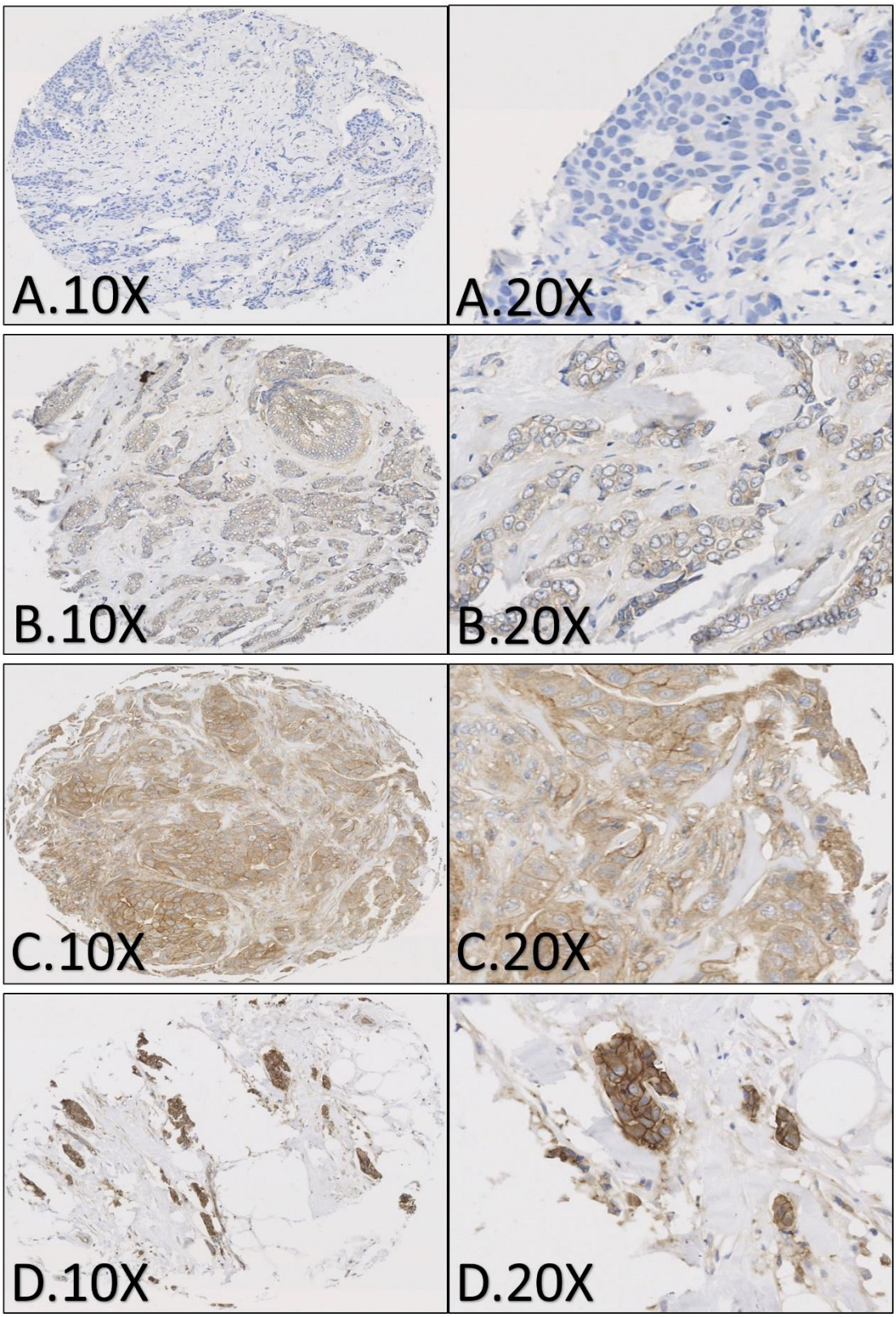


Figure 4. 14: Immunohistochemistry expression of cytoplasmic SNAP23 in tissue microarray images. Cytoplasm expression of SNAP23 in cancer cells was distributed as follows; a) No staining, b) weak staining, c) moderate staining, and d) strong staining. Images were captured at 10- and 20-times magnifications.

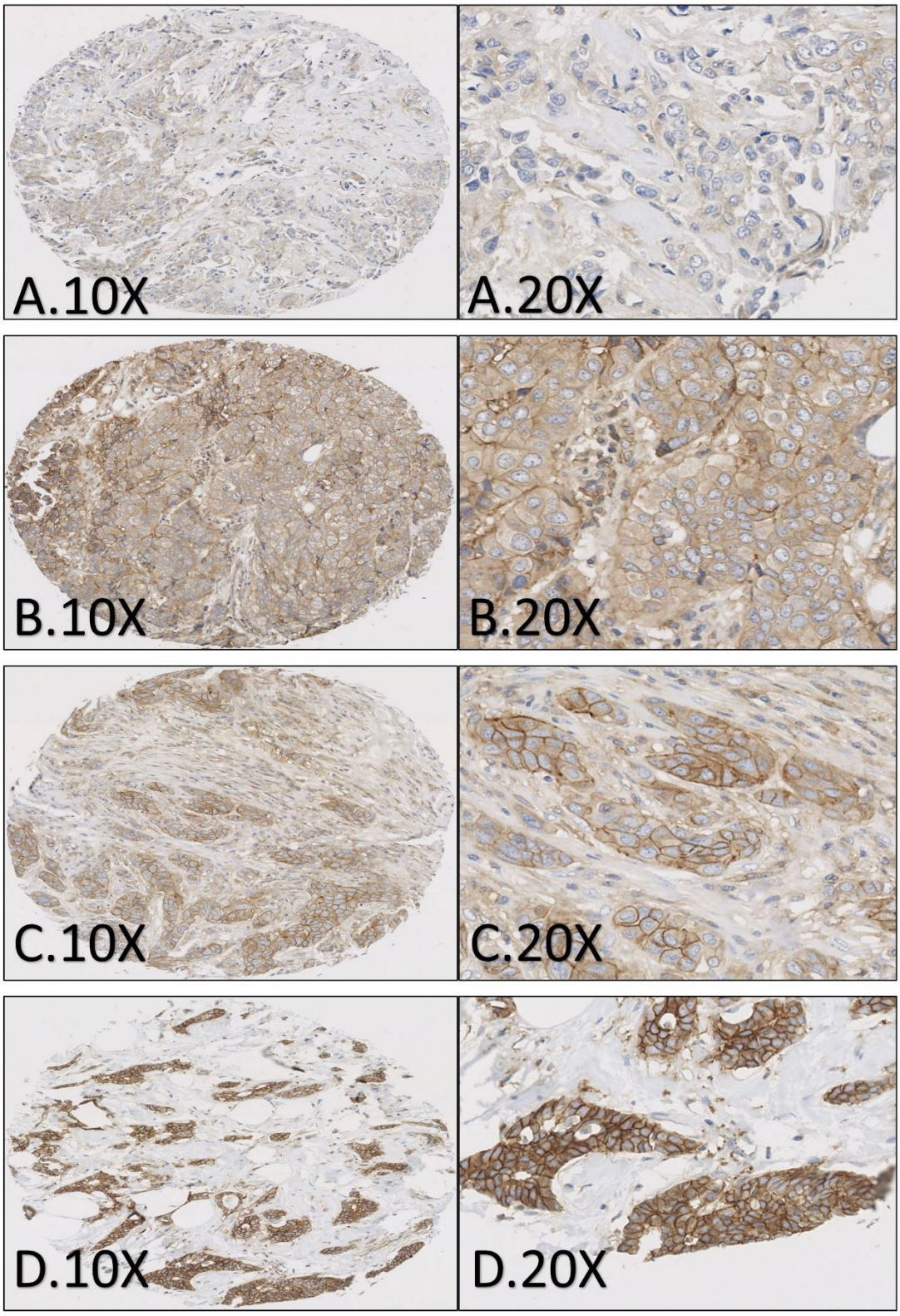


Figure 4. 15: Immunohistochemistry expression of membranous SNAP23 in tissue microarray images. Membranous expression of SNAP23 in cancer cells was distributed as follows; a) No staining, b) weak staining, c) moderate staining, and d) strong staining. Images were captured at 10- and 20-times magnifications.

4.3.3.1. *Clinicopathological impact for MCFD2 protein*

4.3.3.1.1. Cytoplasmic expression of MCFD2

Although the selection of *MCFD2* was based on its CNA gain ratios in LVI^+ patients, and it was also one of the highly ranked DEGs, as illustrated in the previous chapter, protein expression of MCFD2 showed strong features as a good prognostic biomarker for invasive BC patients. Upregulated protein expression of MCFD2 was strongly associated with the LVI^-/LN^- patients (p-value $\leq$0.003). The positive expression of MCFD2 was also correlated with small tumour size (p-value $\leq$0.021), tumour grade-I (p-value $\leq$5.7E-13), and good NPI index (p-value $\leq$5.7E-7). The absence of MCFD2 protein was related to the tumours of ER^- (p-value $\leq$4.0E-6), and $HER2^+$ (p-value $\leq$0.021) (Table 4.1). The MCFD2 protein was interrelated with the protein expression of CD8 (p-value $\leq$0.017), CD20 (p-value $\leq$0.005), CD68 (p-value $\leq$0.006), FOXP3 (p-value $\leq$0.003), androgen receptor (AR) (p-value $\leq$7.3E-11), GATA3 (p-value $\leq$0.0002), ki-67 (p-value $\leq$0.0002), E-cadherin (p-value $\leq$1.6E-7), N-cadherin (p-value $\leq$0.012), and p53 (p-value $\leq$0.003) (Table 4.2).

The KM survival analyses of BCSS, DMFS, and DFS showed favourable outcomes for patients with the upregulated expression of MCFD2 protein (p-value $\leq$0.002; p-value $\leq$0.023; and p-value $\leq$0.003, respectively) (Figure 4.16). However, a multivariate analysis illustrated that the upregulation in MCFD2 could show an independent predictive feature for BC recurrence (HR=1.461; p-value $\leq$0.031) (Table. 4.3).

4.3.3.2. *Clinicopathological impact for CCT8 protein*

4.3.3.2.1. Cytoplasmic expression of CCT8

Clinically, the cytoplasmic expression of CCT8 had no clinical impact (Table 4.1) and (Table 4.4). A univariate survival analysis showed no statistical relevance for cytoplasmic CCT8 in invasive BC (Figure 4.17). Cytoplasmic CCT8 showed strong positive association with CD24 (p-value $\leq$0.001), ki-67 (p-value $\leq$0.013), and p53 (p-value $\leq$0.0004) (Table 4.2).

4.3.3.2.2. Nuclear expression of CCT8

For MCFD2, the upregulated expression of nuclear CCT8 was an indicator for good prognosis. Positive nuclear expression of CCT8 was associated with the LVI^-/LN^- patients (p-value $\leq$0.010), small tumour size (p-value $\leq$0.00004), tumour grade-I (p-value $\leq$1.9E-28), negative LN involvement (p-value $\leq$0.036), and favourable NPI index (p-value $\leq$1.8E-16). The lack of CCT8 nuclear protein was related to the tumours of ER^- (p-value $\leq$5.3E-17) and $HER2^+$ (p-value $\leq$2.0E-6)

(Table 4.1). Positive nuclear expression of CCT8 protein showed strong negative association against ki-67, (p-value ≤9.8E-12), and N-cadherin (p-value ≤2.8E-12). The positive nuclear expression of CCT8 protein displayed strong positive correlations with AR (p-value ≤1.2E-17), GATA3, (p-value ≤1.4E-17), and E-cadherin, (p-value ≤0.00001) (Table 4.2).

The KM survival analyses of BCSS, DMFS, and DFS showed favourable outcomes for patients with upregulated expression of CCT8 protein (p-value ≤4.8E-8; p-value ≤5.0E-6; and p-value ≤0.00003, respectively) (Figure 4.18). The multivariate analysis revealed that the downregulated nuclear expression of CCT8 exhibited independent prognostic features that may classify a subset of invasive BC patients as a high-risk group: BCSS (HR=0.711; p-value ≤0.002), DMFS (HR=0.751; p-value ≤0.008), DFS (HR=0.783; p-value ≤0.014) (Table 4.5).

4.3.3.3. *Synaptosomal-associated protein 23*

4.3.3.3.1. Cytoplasmic expression of SNAP23

The altered protein expression of SNAP23 in the cytoplasm did not appear to be a promising prognostic indicator for invasive BC. However, the loss of cytoplasmic SNAP23 protein was correlated with poorer NPI (p-value ≤0.024) and with more patients with negative ER expression profiles (p-value ≤6.2E-9) (Table 4.1). Univariate and multivariate survival analyses showed no statistical relevance for cytoplasmic SNAP23 in invasive BC (Table 4.6 and Figure 4.19). Cytoplasmic expression of SNAP23 had a positive association with AR (p-value ≤0.0004), GATA3 (p-value ≤0.00003), and E-cadherin (p-value ≤0.00005). Cytoplasmic expression of SNAP23 showed negative association with N-cadherin (p-value ≤0.006). The remaining protein-protein interactions for cytoplasmic SNAP23 are shown in Table 4.2.

4.3.3.3.2. Membranous expression of SNAP23

Our results show that positive membranous protein expression of SNAP23 has a significant correlation with LVI^+ patients (p-value ≤0.020). The highest percentage of patients with positive expression of membranous SNAP23 was related to the patients of tumour grade-II (p-value ≤0.0002). This number was considerably reduced in the patients of tumour grade-III. The positive expression of membranous SNAP23 was statistically correlated with ER^+ and $HER2^+$ (p-value ≤0.2.2E-11 and p-value ≤0.002, respectively) (Table 4.1). The KM survival and the Cox regression analyses of BCSS, DMFS, and DFS for SNAP23 membranous expression were non-significant (Figure 4.20 and Table 4.7).

Regarding the protein-protein interactions for membranous SNAP23, no statistical interrelations were identified against the tested immune cells biomarkers. Strong positive association was identified between AR and membranous SNAP23 (p-value ≤0.00006), GATA3 (p-value ≤0.003), and E-cadherin (p-value ≤2.0E-6). The downregulated protein expression of membranous SNAP23 was inversely associated with p53 (p-value ≤0.048) (Table 4.2).

4.3.3.4. *SEC14-like protein 1*

4.3.3.4.1. **Cytoplasmic expression of SEC14L1**

Our results show that the cytoplasmic overexpression of SEC14L1 was strongly associated with the LVI^{+} patients (p-value ≤0.0003). The cytoplasmic protein expression of SEC14L1 was preferentially upregulated in triple-negative patients (p-value ≤0.016). However, we could not identify any further clinical correlation for the altered protein expression of SEC14L1 in this cohort (Table 4.8). At the designated cut-off point, a trend toward significance was calculated for the positive cytoplasmic expression of SEC14L1 with shorter BCSS (p-value ≤0.090) (Figure 4.21). Cytoplasmic SEC14L1 expression was negatively associated with AR (p-value ≤0.040), but positively correlated with E-cadherin (p-value ≤0.037) (Table 4.9). The results of the Cox regression model did not show predictive features for SEC14L1 using the same statistical conditions (Table 4.10).

Table 4. 1: Clinicopathological significance of the LVI-correlated proteins in the patients of Nottingham series of primary invasive breast carcinoma. Significant values are in bold.

	MCFD2 (cytoplasmic) at 55 H-score			CCT8 (cytoplasmic) at 10 H-score			CCT8 (nuclear) at 5 H-score			SNAP23 (cytoplasmic) at 100 H-score			SNAP23 (membranous) at 50 H-score		
	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)
Lymphovascular invasion (LVI) / lymph node status (LN), (H&E/IHC)															
LVI^-/LN^-	245 (47.0)	276 (53.0)		244 (50.2)	242 (49.8)		237 (48.5)	252 (51.5)		328 (66.1)	168 (33.9)		279 (56.5)	215 (43.5)	
LVI^-/LN^+	108 (47.6)	119 (52.4)	11.469 (**0.003**)	109 (49.8)	110 (50.2)	0.144 (0931)	115 (52.5)	104 (47.5)	9.202 (**0.010**)	145 (64.2)	81 (35.8)	1.427 (0.490)	110 (49.1)	114 (50.9)	7.857 (**0.020**)
$LVI^+/LN^{+/-}$	294 (56.9)	223 (43.1)		245 (51.1)	234 (48.9)		282 (58.1)	203 (41.9)		345 (68.5)	159 (31.5)		241 (48.0)	261 (52.0)	
Age (years)															
≤ 50	252 (54.2)	213 (45.8)	2.952 (0.086)	210 (48.3)	225 (51.7)	1.303 (0.254)	244 (56.0)	192 (44.0)	2.044 (0.153)	301 (67.2)	147 (32.8)	0.223 (0.637)	238 (53.1)	210 (46.9)	0.639 (0.424)
> 50	402 (49.2)	415 (50.8)		394 (51.7)	368 (48.3)		398 (51.7)	372 (48.3)		523 (65.9)	271 (34.1)		400 (50.8)	388 (49.2)	
Tumour size															
≤ 2.0cm	285 (47.7)	312 (52.3)	5.330 (**0.021**)	267 (48.6)	282 (51.4)	1.561 (0.211)	259 (46.7)	296 (53.3)	16.925 (**0.00004**)	370 (63.9)	209 (36.1)	3.460 (0.063)	290 (50.3)	286 (49.7)	0.717 (0.397)
> 2.0cm	366 (54.2)	309 (45.8)		334 (52.3)	305 (47.7)		376 (58.6)	266 (41.4)		450 (68.9)	203 (31.1)		343 (52.8)	307 (47.2)	
Tumour Grade															
Grade-I	69 (35.4)	126 (64.6)		105 (54.1)	89 (45.9)		55 (28.2)	140 (71.8)		117 (62.9)	69 (37.1)		90 (48.9)	94 (51.1)	
Grade-II	184 (43.2)	242 (56.8)	56.376 (**5.7E-13**)	204 (52.3)	186 (47.7)	2.652 (0.266)	166 (41.7)	232 (58.3)	127.634 (**1.9E-28**)	264 (63.8)	150 (36.2)	4.621 (0.099)	182 (44.2)	230 (55.8)	17.454 (**0.0002**)
Grade-III	396 (61.2)	251 (38.8)		290 (48.3)	310 (51.7)		412 (68.7)	188 (31.3)		435 (69.3)	193 (30.7)		358 (57.2)	268 (42.8)	
Nodal involvement, (NPI staging system)															
Stage-I	393 (50.1)	391 (49.9)		374 (51.8)	348 (48.2)		365 (50.2)	362 (49.8)		487 (65.5)	257 (34.5)		390 (52.7)	350 (47.3)	
Stage-II	197 (51.7)	184 (48.3)	1.671 (0.434)	178 (48.5)	189 (51.5)	1.111 (0.574)	215 (58.4)	153 (41.6)	6.669 (**0.036**)	258 (66.7)	129 (33.3)	1.916 (0.384)	193 (50.1)	192 (49.9)	0.963 (0.618)
Stage-III	59 (56.7)	45 (43.3)		47 (49.5)	48 (50.5)		53 (54.1)	45 (45.9)		71 (72.4)	27 (27.6)		48 (49.0)	50 (51.0)	
Nottingham Prognostic Index (NPI)															
good	151 (40.2)	225 (59.8)		195 (55.1)	159 (44.9)		130 (36.2)	229 (63.8)		220 (62.7)	131 (37.3)		168 (48.1)	181 (51.9)	
moderate	371 (54.2)	313 (45.8)	28.772 (**5.7E-7**)	304 (48.0)	329 (52.0)	4.529 (0.104)	360 (56.6)	276 (43.4)	72.550 (**1.8E-16**)	447 (66.3)	227 (33.7)	7.422 (**0.024**)	351 (52.4)	319 (47.6)	2.841 (0.242)
poor	129 (60.8)	83 (39.2)		102 (50.7)	99 (49.3)		145 (71.8)	57 (28.2)		153 (73.9)	54 (26.1)		114 (55.1)	93 (44.9)	

	MCFD2 (cytoplasmic) at 55 H-score			CCT8 (cytoplasmic) at 10 H-score			CCT8 (nuclear) at 5 H-score			SNAP23 (cytoplasmic) at 100 H-score			SNAP23 (membranous) at 50 H-score		
	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)
Tumour histological types															
Invasive ductal/NST	547 (51.8)	510 (48.2)		494 (49.8)	498 (50.2)		554 (55.7)	441 (44.3)		689 (66.6)	346 (33.4)		529 (51.4)	501 (48.6)	
Lobular	55 (47.0)	62 (53.0)		64 (61.0)	41 (39.0)		41 (37.3)	69 (62.7)		77 (67.5)	37 (32.5)		64 (56.1)	50 (43.9)	
Medullary-like	19 (70.4)	8 (29.6)	11.065 (**0.026**)	7 (29.2)	17 (70.8)	10.410 (**0.034**)	20 (83.3)	4 (16.7)	38.562 (**8.6E-8**)	19 (76.0)	6 (24.0)	3.007 (0.557)	14 (58.3)	10 (41.7)	5.176 (0.270)
Tubular/Mucinous	6 (75.0)	2 (25.0)		5 (71.4)	2 (28.6)		1 (14.3)	6 (85.7)		4 (80.0)	1 (20.0)		4 (80.0)	1 (20.0)	
Miscellaneous	24 (38.1)	39 (61.9)		31 (51.7)	29 (48.3)		19 (31.1)	42 (68.9)		31 (58.5)	22 (41.5)		22 (41.5)	31 (58.5)	
Estrogen receptor status (IHC)															
Negative	204 (62.0)	125 (38.0)	21.276 (**4.0E-6**)	148 (49.2)	153 (50.8)	0.271 (0.603)	225 (74.3)	78 (25.7)	70.211 (**5.3E-17**)	254 (79.6)	65 (20.4)	33.757 (**6.2E-9**)	215 (67.6)	103 (32.4)	44.789 (**2.2E-11**)
Positive	445 (47.2)	497 (52.8)		449 (50.9)	433 (49.1)		413 (46.5)	476 (53.5)		564 (61.8)	349 (38.2)		416 (45.8)	492 (54.2)	
Progesterone receptor status (IHC)															
Negative	303 (59.4)	207 (40.6)	21.830 (**3.0E-6**)	246 (52.5)	223 (47.5)	2.266 (0.132)	313 (66.3)	159 (33.7)	49.198 (**2.3E-12**)	371 (74.5)	127 (25.5)	22.247 (**2.0E-6**)	295 (59.5)	201 (40.5)	21.642 (**3.0E-6**)
Positive	329 (45.9)	388 (54.1)		324 (47.9)	352 (52.1)		309 (45.4)	372 (54.6)		424 (61.4)	266 (38.6)		314 (45.8)	372 (54.2)	
ERBB2 (IHC/-ISH)															
Negative	536 (50.3)	529 (49.7)	5.289 (**0.021**)	491 (49.3)	504 (50.7)	0.580 (0.446)	508 (50.6)	496 (49.4)	22.905 (**2.0E-6**)	684 (66.5)	344 (33.5)	0.016 (0.900)	546 (53.4)	476 (46.6)	9.216 (**0.002**)
Positive	95 (60.1)	63 (39.9)		78 (52.7)	70 (47.3)		106 (71.6)	42 (28.4)		103 (66.0)	53 (34.0)		63 (40.4)	93 (59.6)	
Triple negative (IHC)															
Negative	488 (47.9)	531 (52.1)	20.377 (**6.0E-6**)	487 (51.0)	467 (49.0)	1.340 (0.247)	467 (48.6)	494 (51.4)	47.265 (**6.2E-12**)	616 (62.6)	368 (37.4)	31.092 (**2.5E-8**)	449 (45.9)	530 (54.1)	62.250 (**3.0E-15**)
Positive	144 (64.6)	79 (35.4)		96 (46.6)	110 (53.4)		155 (74.9)	52 (25.1)		181 (82.3)	39 (17.7)		165 (75.3)	54 (24.7)	

Table 4. 2: Protein-protein interactions for the LVI-correlated proteins in the Nottingham series of primary invasive breast carcinoma, significant values are in bold

	MCFD2 (cytoplasmic) at 55 H-score			CCT8 (cytoplasmic) at 10 H-score			CCT8 (nuclear) at 5 H-score			SNAP23 (cytoplasmic) at 100 H-score			SNAP23 (membranous) at 50 H-score		
	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)	- n (%)	+ n (%)	χ^2 (*p*-value≤)
T-cell surface glycoprotein Cluster of differentiation 3, (CD3): present or absent															
Negative	4 (26.7)	11 (73.3)	3.871 (0.051)	7 (46.7)	8 (53.3)	0.020 (0.890)	3 (20.0)	12 (80.)	7.677 (**0.006**)	7 (46.7)	8 (53.3)	2.840 (0.092)	5 (33.3)	10 (66.7)	1.989 (0.158)
Positive	517 (52.1)	476 (47.9)		454 (48.5)	482 (51.5)		525 (55.9)	415 (44.1)		648 (67.3)	314 (32.7)		493 (51.7)	461 (48.3)	
T-cell surface glycoprotein Cluster of differentiation 8, (CD8): present or absent															
Negative	85 (61.6)	53 (38.4)	5.735 (**0.017**)	77 (59.2)	53 (40.8)	5.984 (**0.014**)	66 (50.0)	66 (50.0)	2.279 (0.131)	80 (60.2)	53 (39.8)	2.769 (0.096)	67 (50.4)	66 (49.6)	0.141 (0.708)
Positive	403 (50.6)	394 (49.4)		360 (47.6)	396 (52.4)		432 (57.1)	325 (42.9)		526 (67.5)	253 (32.5)		403 (52.1)	370 (47.9)	
B-lymphocyte antigen, CD20: positive ≥ 5 cells															
Negative	346 (49.0)	360 (51.0)	8.065 (**0.005**)	329 (48.9)	344 (51.1)	0.032 (0.857)	342 (50.5)	335 (49.5)	19.510 (**0.00001**)	450 (65.6)	236 (34.4)	1.933 (0.164)	339 (49.7)	343 (50.3)	3.358 (0.067)
Positive	183 (58.7)	129 (41.3)		138 (48.3)	148 (51.7)		190 (66.0)	98 (34.0)		209 (70.1)	89 (29.9)		166 (56.1)	130 (43.9)	
Macrosialin, CD68: positive ≥ 17 cells															
Negative	140 (45.6)	167 (54.4)	7.416 (**0.006**)	146 (51.0)	140 (49.0)	1.616 (0.204)	132 (45.7)	157 (54.3)	21.843 (**3.0E-6**)	195 (66.1)	100 (33.9)	0.013 (0.911)	150 (51.4)	142 (48.6)	0.147 (0.701)
Positive	340 (55.1)	277 (44.9)		269 (46.5)	310 (53.5)		361 (62.3)	218 (37.7)		395 (65.7)	206 (34.3)		299 (50.0)	299 (50.0)	
Forkhead box protein P3, (FOXP3): positive ≥ 3 cells															
Negative	155 (45.5)	186 (54.5)	8.882 (**0.003**)	164 (51.4)	155 (48.6)	1.141 (0.285)	159 (49.4)	163 (50.6)	8.439 (**0.004**)	216 (64.3)	120 (35.7)	1.683 (0.194)	162 (48.6)	171 (51.4)	1.430 (0.232)
Positive	365 (55.4)	294 (44.6)		294 (47.7)	322 (52.3)		367 (59.3)	252 (40.7)		433 (68.4)	200 (31.6)		332 (52.7)	298 (47.3)	
Signal transducer CD24:															
Negative	69 (49.3)	71 (50.7)	0.746 (0.388)	84 (61.8)	52 (38.2)	10.401 (**0.001**)	74 (54.0)	63 (46.0)	0.246 (0.620)	82 (59.9)	55 (40.1)	4.291 (**0.036**)	61 (44.5)	76 (55.5)	2.470 (0.116)
Positive	320 (53.3)	280 (46.7)		261 (46.4)	302 (53.6)		319 (56.4)	247 (43.6)		400 (69.1)	179 (30.9)		300 (52.0)	277 (48.0)	
CD44 antigen: positive ≥ 50 H-score															
Negative	187 (55.0)	153 (45.0)	0.267 (0.606)	166 (52.4)	151 (47.6)	1.996 (0.158)	188 (59.1)	130 (40.9)	2.062 (0.151)	223 (67.6)	107 (32.4)	1.767 (0.184)	168 (51.1)	161 (48.9)	0.030 (0.863)
Positive	153 (52.9)	136 (47.1)		128 (46.5)	147 (53.5)		147 (53.3)	129 (46.7)		178 (62.5)	107 (37.5)		147 (51.8)	137 (48.2)	

	MCFD2 (cytoplasmic) at 55 H-score			CCT8 (cytoplasmic) at 10 H-score			CCT8 (nuclear) at 5 H-score			SNAP23 (cytoplasmic) at 100 H-score			SNAP23 (membranous) at 50 H-score		
	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)	- n (%)	+ n (%)	χ2 (*p*-value≤)
Androgen receptor:															
Negative	274 (63.4)	158 (36.6)	42.430 (**7.3E-11**)	204 (50.9)	197 (49.1)	0.177 (0.674)	282 (70.1)	120 (29.9)	73.160 (**1.2E-17**)	315 (75.2)	104 (24.8)	17.073 (**0.0004**)	246 (59.3)	169 (40.7)	15.975 (**0.00006**)
Positive	305 (43.5)	396 (56.5)		319 (49.5)	325 (50.5)		280 (43.1)	370 (56.9)		429 (63.2)	250 (36.8)		317 (46.8)	360 (53.2)	
Trans-acting T-cell-specific transcription factor, (GATA3): positive ≥ 60 H-score															
Negative	318 (56.5)	245 (43.5)	13.432 (**0.0002**)	252 (47.5)	278 (52.5)	5.017 (**0.025**)	352 (66.3)	179 (33.7)	72.876 (**1.4E-17**)	390 (70.3)	165 (29.7)	17.220 (**0.00003**)	309 (55.9)	244 (44.1)	9.077 (**0.003**)
Positive	73 (40.8)	106 (59.2)		96 (57.5)	71 (42.5)		49 (29.0)	120 (71.0)		92 (53.2)	81 (46.8)		74 (42.8)	99 (57.2)	
Proliferation marker protein Ki-67: positive ≥ 10 Percentage, full-face															
Negative	154 (43.1)	203 (56.9)	13.946 (**0.0002**)	189 (55.6)	151 (44.4)	6.189 (**0.013**)	134 (38.8)	211 (61.2)	46.371 (**9.8E-12**)	232 (66.1)	119 (33.9)	0.004 (0.951)	167 (48.0)	181 (52.0)	2.601 (0.107)
Positive	355 (55.5)	285 (44.5)		278 (47.1)	312 (52.9)		365 (61.9)	225 (38.1)		413 (66.3)	210 (33.7)		331 (53.4)	289 (46.6)	
Cadherin-1, (E-cadherin): positive ≥ 110 H-score															
Negative	394 (57.9)	287 (42.1)	27.460 (**1.6E-7**)	330 (52.4)	300 (47.6)	3.901 (**0.048**)	376 (59.4)	257 (40.6)	18.985 (**0.00001**)	477 (72.2)	184 (27.8)	16.602 (**0.00005**)	379 (57.5)	280 (42.5)	22.630 (**2.0E-6**)
Positive	222 (42.6)	299 (57.4)		228 (46.4)	263 (53.6)		230 (46.4)	266 (53.6)		308 (60.9)	198 (39.1)		218 (43.4)	284 (56.6)	
Cadherin-2, (N-cadherin): positive ≥ 100 H-score															
Negative	113 (43.5)	147 (56.5)	6.315 (**0.012**)	104 (42.8)	139 (57.2)	4.849 (**0.028**)	89 (36.3)	156 (63.7)	48.826 (**2.8E-12**)	152 (61.0)	97 (39.0)	7.635 (**0.006**)	124 (50.2)	123 (49.8)	0.645 (0.422)
Positive	345 (52.7)	310 (47.3)		317 (51.1)	303 (48.9)		389 (62.5)	233 (37.5)		453 (70.7)	188 (29.3)		340 (53.2)	299 (46.8)	
Cellular tumor antigen p53, (p53):															
Negative	421 (48.7)	444 (51.3)	8.583 (**0.003**)	436 (53.3)	382 (46.7)	12.425 (**0.0004**)	413 (50.2)	410 (49.8)	12.273 (**0.0005**)	556 (66.3)	282 (33.7)	1.094 (0.296)	415 (49.8)	418 (50.2)	3.907 (**0.048**)
Positive	205 (57.9)	149 (42.1)		132 (41.6)	185 (58.4)		198 (61.7)	123 (38.3)		237 (69.5)	104 (30.5)		191 (56.2)	149 (43.8)	
Twist-related protein 2: positive ≥ 150 H-score															
Negative	181 (47.4)	201 (52.6)	2.337 (0.126)	177 (50.1)	176 (49.9)	1.698 (0.193)	153 (42.6)	206 (57.4)	32.767 (**1.0E-8**)	222 (62.0)	136 (38.0)	3.409 (0.065)	185 (51.8)	172 (48.2)	0.867 (0.352)
Positive	232 (52.7)	208 (47.3)		189 (45.4)	227 (54.6)		264 (63.2)	154 (36.8)		295 (68.3)	137 (31.7)		208 (48.5)	221 (51.5)	

Table 4. 3: Multivariate Cox regression analyses for prediction of breast cancer-specific survival, distant metastasis-free survival, and disease-free survival, altered protein expression of MCFD2.

Breast cancer specific survival (BCSS), n= 1245◊

Cut-off point 55 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
MCFD2 protein (cytoplasmic)	1.177	0.964	1.438	0.110
Tumour grade	1.663	1.390	1.989	0.000
Nodal staging*	1.902	1.648	2.194	0.000
Tumour size	1.431	1.157	1.771	0.001
Estrogen expression**	1.052	0.833	1.328	0.670
Her2 expression***	1.673	1.296	2.160	0.000

Distant metastasis free survival (DMFS), n= 1244 ◊

Cut-off point 55 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
MCFD2 protein (cytoplasmic)	1.065	0.879	1.291	0.519
Tumour grade	1.515	1.287	1.783	0.000
Nodal staging*	1.856	1.614	2.134	0.000
Tumour size	1.431	1.166	1.757	0.001
Estrogen expression**	1.201	0.950	1.519	0.126
Her2 expression***	1.502	1.165	1.937	0.002

Disease free survival (DFS), n= 1247◊

Cut-off point 55 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
MCFD2 protein (cytoplasmic)	1.220	1.018	1.461	**0.031**
Tumour grade	1.301	1.121	1.511	0.001
Nodal staging*	1.631	1.427	1.864	0.000
Tumour size	1.321	1.091	1.599	0.004
Estrogen expression**	1.075	0.862	1.341	0.521
Her2 expression***	1.483	1.164	1.889	0.001

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
◊ Sum of patients does not equal to 1901 because of missing cases among different categories.

Table 4. 4: Multivariate Cox regression analyses for prediction of breast cancer-specific survival, distant metastasis-free survival, and disease-free survival, altered protein expression of cytoplasmic CCT8.

Breast cancer specific survival (BCSS), n= 1168◊

Cut-off point 10 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
CCT8 protein (cytoplasmic)	1.073	0.875	1.315	0.499
Tumour grade	1.686	1.402	2.026	0.000
Nodal staging*	1.909	1.638	2.224	0.000
Tumour size	1.572	1.250	1.976	0.000
Estrogen expression**	1.104	0.863	1.412	0.433
Her2 expression***	1.658	1.272	2.162	0.000

Distant metastasis free survival (DMFS), n= 1168◊

Cut-off point 10 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
CCT8 protein (cytoplasmic)	1.026	0.844	1.249	0.794
Tumour grade	1.537	1.302	1.814	0.000
Nodal staging*	1.840	1.588	2.131	0.000
Tumour size	1.526	1.230	1.893	0.000
Estrogen expression**	1.320	1.031	1.690	0.028
Her2 expression***	1.522	1.171	1.978	0.002

Disease free survival (DFS), n= 1165◊

Cut-off point 10 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
CCT8 protein (cytoplasmic)	0.924	0.770	1.110	0.401
Tumour grade	1.340	1.152	1.559	0.000
Nodal staging*	1.626	1.414	1.870	0.000
Tumour size	1.339	1.099	1.632	0.004
Estrogen expression**	1.198	0.949	1.512	0.128
Her2 expression***	1.525	1.189	1.957	0.001

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
◊ Sum of patients does not equal to 1901 because of missing cases among different categories.

Table 4. 5: Multivariate Cox regression analyses for prediction of breast cancer-specific survival, distant metastasis-free survival, and disease-free survival, altered protein expression of nuclear CCT8.

Breast cancer specific survival (BCSS), n= 1168◊				
Cut-off point 5 H-score	**Hazard ratio (95% CI)**	**Lower**	**Upper**	***p*-value**
CCT8 protein (nuclear)	0.711	0.570	0.885	**0.002**
Tumour grade	1.594	1.325	1.918	0.000
Nodal staging*	1.932	1.660	2.248	0.000
Tumour size	1.552	1.238	1.947	0.000
Estrogen expression**	1.193	0.931	1.529	0.163
Her2 expression***	1.636	1.255	2.134	0.000
Distant metastasis free survival (DMFS), n= 1168◊				
Cut-off point 5 H-score	**Hazard ratio (95% CI)**	**Lower**	**Upper**	***p*-value**
CCT8 protein (nuclear)	0.751	0.609	0.927	**0.008**
Tumour grade	1.474	1.247	1.743	0.000
Nodal staging*	1.867	1.613	2.162	0.000
Tumour size	1.506	1.216	1.865	0.000
Estrogen expression**	1.393	1.087	1.787	0.009
Her2 expression***	1.481	1.139	1.925	0.003
Disease free survival (DFS), n= 1165◊				
Cut-off point 5 H-score	**Hazard ratio (95% CI)**	**Lower**	**Upper**	***p*-value**
CCT8 protein (nuclear)	0.783	0.644	0.951	**0.014**
Tumour grade	1.288	1.106	1.500	0.001
Nodal staging*	1.642	1.429	1.887	0.000
Tumour size	1.324	1.088	1.611	0.005
Estrogen expression**	1.255	0.993	1.587	0.058
Her2 expression***	1.492	1.163	1.913	0.002

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
◊ Sum of patients does not equal to 1901 because of missing cases among different categories.

Table 4. 6: Multivariate Cox regression analyses for prediction of breast cancer-specific survival, distant metastasis-free survival, and disease-free survival, altered protein expression of cytoplasmic SNAP23.

Breast cancer specific survival (BCSS), n= 1205◊

Cut-off point 100 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
SNAP23 protein (cytoplasmic)	0.951	0.767	1.179	0.646
Tumour grade	1.684	1.405	2.019	0.000
Nodal staging*	1.892	1.632	2.194	0.000
Tumour size	1.520	1.220	1.892	0.000
Estrogen expression**	1.091	0.857	1.388	0.481
Her2 expression***	1.679	1.297	2.174	0.000

Distant metastasis free survival (DMFS), n= 1205◊

Cut-off point 100 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
SNAP23 protein (cytoplasmic)	0.964	0.785	1.184	0.729
Tumour grade	1.509	1.280	1.779	0.000
Nodal staging*	1.820	1.577	2.100	0.000
Tumour size	1.531	1.243	1.887	0.000
Estrogen expression**	1.274	1.000	1.623	0.050
Her2 expression***	1.487	1.149	1.926	0.003

Disease free survival (DFS), n= 1208◊

Cut-off point 100 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
SNAP23 protein (cytoplasmic)	0.945	0.779	1.146	0.564
Tumour grade	1.327	1.142	1.543	0.000
Nodal staging*	1.594	1.391	1.826	0.000
Tumour size	1.389	1.145	1.684	0.001
Estrogen expression**	1.175	0.936	1.477	0.165
Her2 expression***	1.487	1.165	1.900	0.001

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
◊ Sum of patients does not equal to 1901 because of missing cases among different categories.

Table 4. 7: Multivariate Cox regression analyses for prediction of breast cancer-specific survival, distant metastasis-free survival, and disease-free survival, altered protein expression of membranous SNAP23.

Breast cancer specific survival (BCSS), n= 1205◊

Cut-off point 50 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
SNAP23 protein (membranous)	1.047	0.853	1.286	0.660
Tumour grade	1.665	1.388	1.997	0.000
Nodal staging*	1.905	1.643	2.208	0.000
Tumour size	1.526	1.225	1.901	0.000
Estrogen expression**	1.074	0.842	1.370	0.567
Her2 expression***	1.673	1.287	2.175	0.000

Distant metastasis free survival (DMFS), n= 1205◊

Cut-off point 50 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
SNAP23 protein (membranous)	1.035	0.850	1.260	0.734
Tumour grade	1.494	1.266	1.762	0.000
Nodal staging*	1.828	1.584	2.110	0.000
Tumour size	1.536	1.246	1.894	0.000
Estrogen expression**	1.264	0.991	1.613	0.059
Her2 expression***	1.487	1.145	1.932	0.003

Disease free survival (DFS), n= 1208◊

Cut-off point 50 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
SNAP23 protein (membranous)	0.969	0.805	1.166	0.738
Tumour grade	1.317	1.132	1.532	0.000
Nodal staging*	1.596	1.393	1.830	0.000
Tumour size	1.398	1.152	1.697	0.001
Estrogen expression**	1.177	0.936	1.480	0.163
Her2 expression***	1.505	1.174	1.929	0.001

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
◊ Sum of patients does not equal to 1901 because of missing cases among different categories.

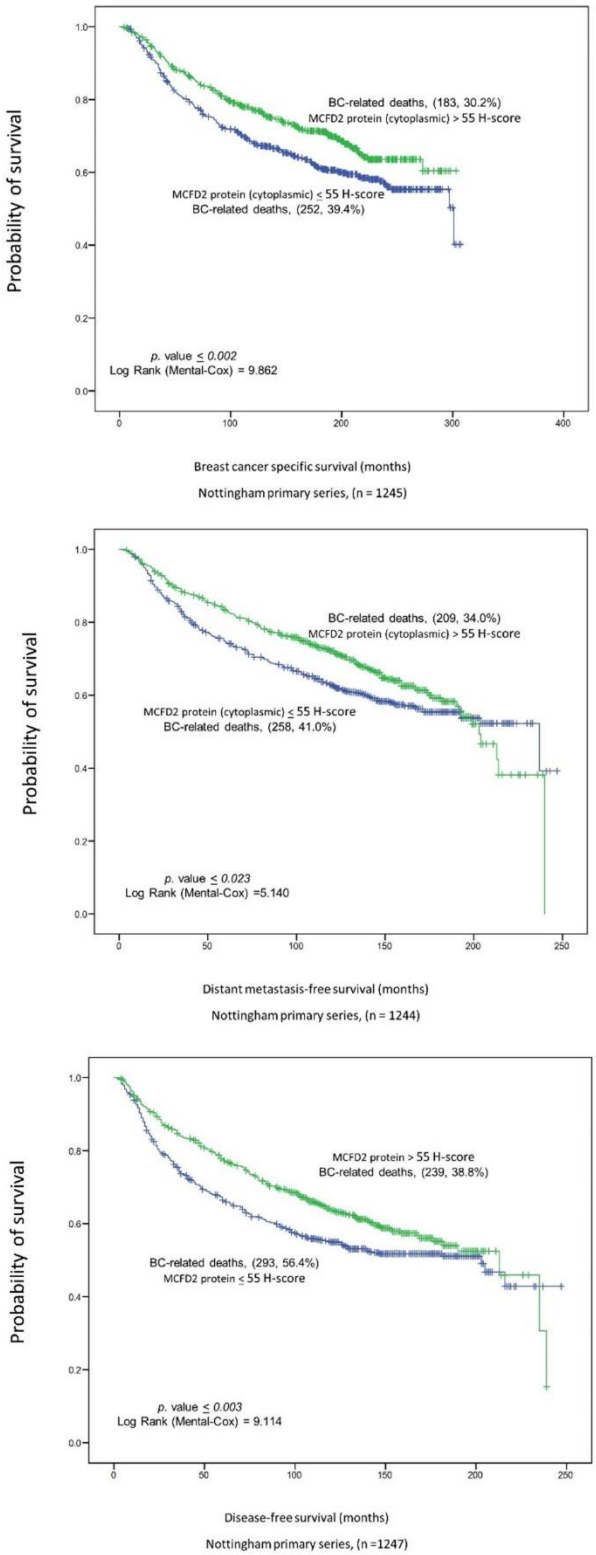


Figure 4. 16: Associations between the altered expression of MCFD2 protein and breast cancer-specific survival, distant metastasis-free survival, and disease-free survival in the Nottingham series of primary invasive breast carcinoma

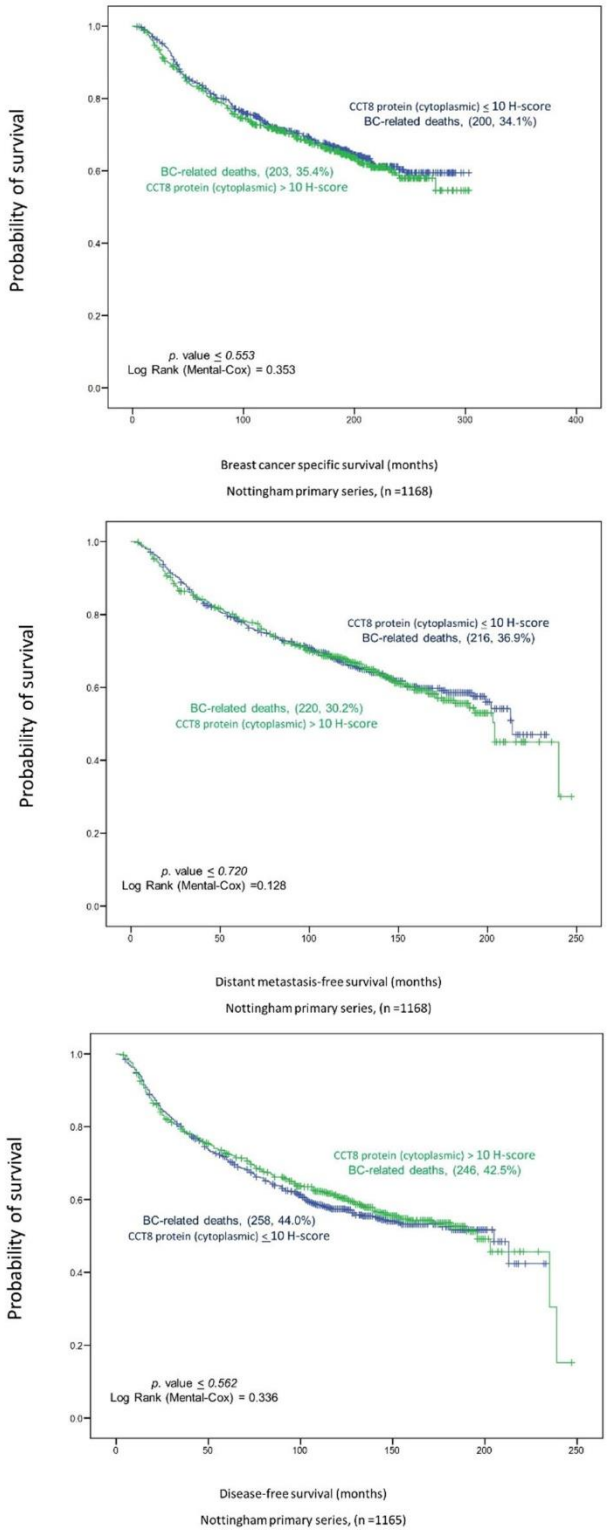


Figure 4. 17: Associations between the altered expression of cytoplasmic CCT8 protein and breast cancer-specific survival, distant metastasis-free survival, and disease-free survival in the Nottingham series of primary invasive breast carcinoma.

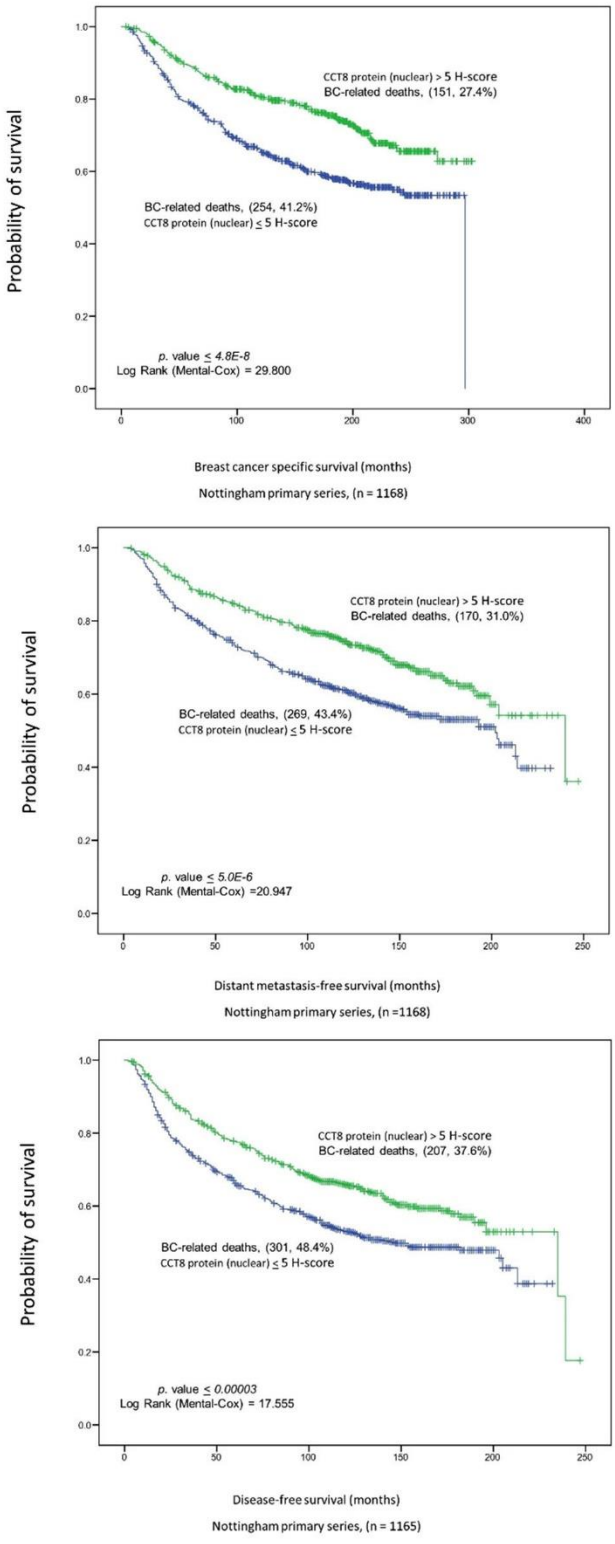


Figure 4. 18: Associations between the altered expression of nuclear CCT8 protein and breast cancer-specific survival, distant metastasis-free survival, and disease-free survival in the Nottingham series of primary invasive breast carcinoma.

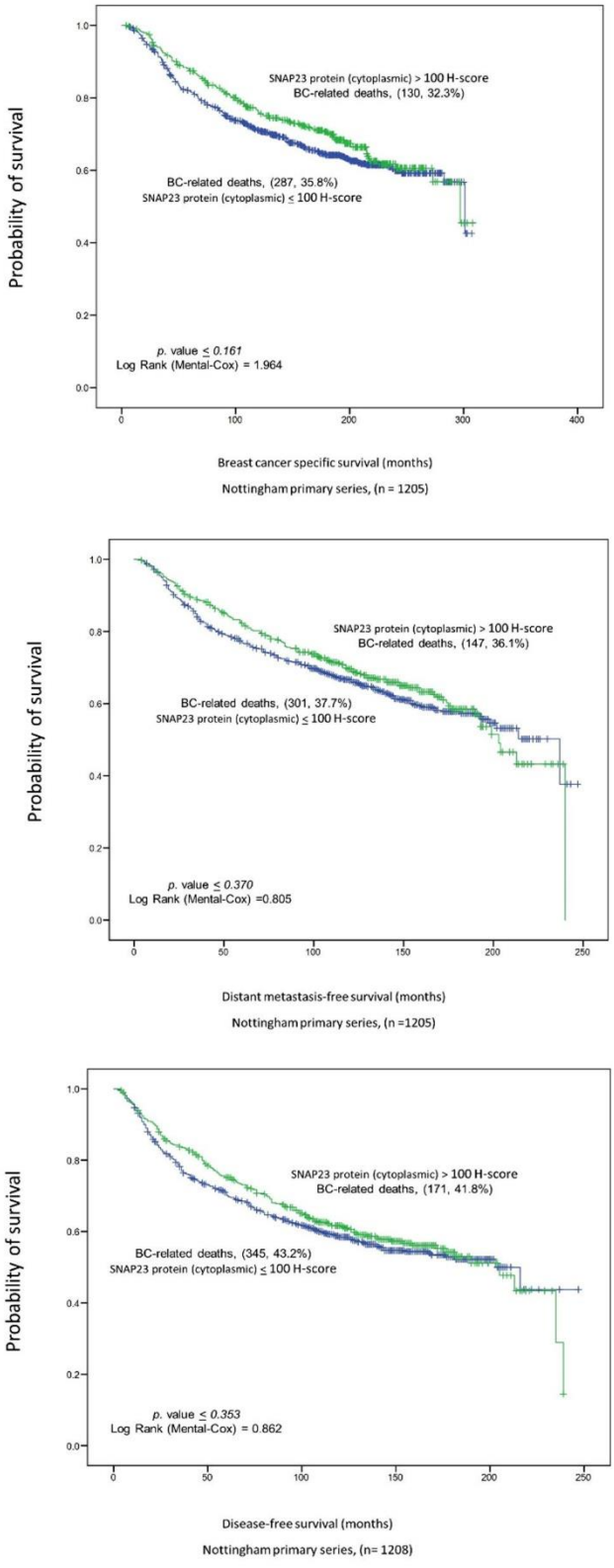


Figure 4. 19: Associations between the altered expression of cytoplasmic SNAP23 protein and breast cancer-specific survival, distant metastasis-free survival, and disease-free survival in the Nottingham series of primary invasive breast carcinoma.

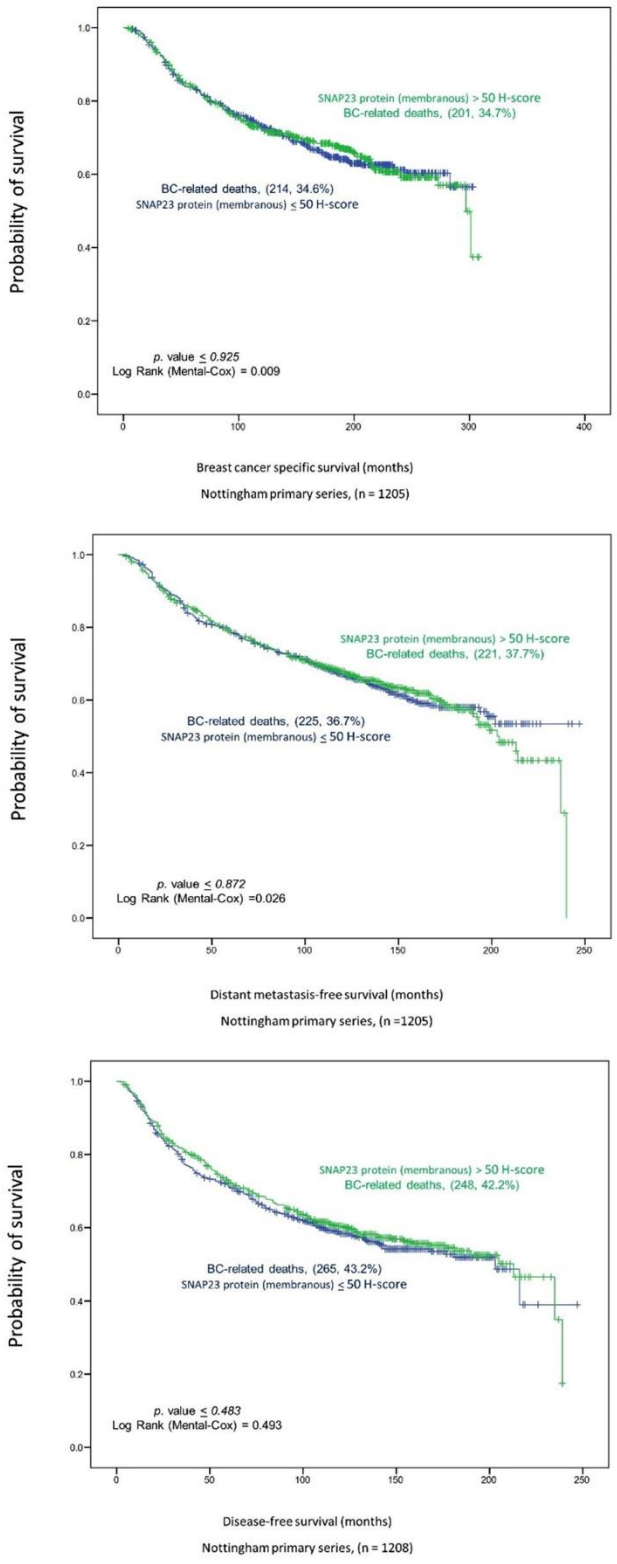


Figure 4. 20: Associations between the altered expression of membranous SNAP23 protein and breast cancer-specific survival, distant metastasis-free survival, and disease-free survival in the Nottingham series of primary invasive breast carcinoma.

Table 4. 8: Clinicopathological significance of SEC14L1protein in the patients of LVI-supervised subset of primary invasive breast carcinoma. Significant values are in bold.

SEC14L1 (cytoplasmic) 70 H-score	- n (%)	+ n (%)	χ2 (*p*-value)
Lymphovascular invasion (LVI) / lymph node status (LN), (H&E/IHC)			
LVI^-/LN^-	37 (74.0)	13 (26.0)	
LVI^-/LN^+	26 (76.5)	8 (23.5)	16.134 (**0.0003**)
$LVI^+/LN^{+/-}$	37 (44.6)	46 (55.4)	
Age (years)			
≤ 50	30 (55.6)	24 (44.4)	0.621 (0.431)
> 50	70 (61.9)	43 (38.1)	
Tumour size			
≤ 2.0cm	43 (64.2)	24 (35.8)	0.861 (0.354)
> 2.0cm	57 (57.0)	43 (43.0)	
Tumour Grade			
Grade-I	13 (72.2)	5 (27.8)	
Grade-II	39 (62.9)	23 (37.1)	2.180 (0.336)
Grade-III	48 (55.2)	39 (44.8)	
Nodal involvement, (NPI staging system)			
Stage-I	53 (63.9)	30 (36.1)	
Stage-II	35 (54.7)	29 (45.3)	1.264 (0.531)
Stage-III	12 (60.0)	8 (40.0)	
Nottingham Prognostic Index (NPI)			
good	29 (70.7)	12 (29.3)	
moderate	45 (55.6)	36 (44.4)	2.723 (0.256)
poor	26 (57.8)	19 (42.2)	
Tumour histological types			
Invasive ductal/NST	85 (56.7)	65 (43.3)	
Lobular	11 (100.0)	0 (0.0)	
Medullary-like	2 (66.7)	1 (33.3)	8.130 (**0.043**)
Tubular/Mucinous	0 (0.0)	0 (0.0)	
Miscellaneous	2 (66.7)	1 (33.3)	
Estrogen receptor status (IHC)			
Negative	30 (66.7)	15 (33.3)	1.302 (0.254)
Positive	70 (56.9)	53 (43.1)	
Progesterone receptor status (IHC)			
Negative	38 (60.3)	25 (39.7)	0.002 (0.968)
Positive	57 (60.6)	37 (39.4)	
ERBB2 (IHC/-ISH)			
Negative	82 (62.1)	50 (37.9)	0.624 (0.430)
Positive	14 (53.8)	12 (46.2)	

Table 4. 9: Protein-protein interactions of SEC14L1 protein in the LVI-supervised subset of primary invasive breast carcinoma, significant values are in bold.

SEC14L1 (cytoplasmic) At 70 H-score	- n (%)	+ n (%)	χ2 (*p*-value≤)
T-cell surface glycoprotein Cluster of differentiation 3, (CD3): present or absent			
Negative	1 (33.3)	2 (66.7)	0.930 (0.335)
Positive	84 (60.9)	54 (39.1)	
T-cell surface glycoprotein Cluster of differentiation 8, (CD8): present or absent			
Negative	8 (72.7)	3 (27.3)	0.980 (0.322)
Positive	70 (57.4)	52 (42.6)	
B-lymphocyte antigen, CD20: positive ≥ 5 cells			
Negative	60 (61.9)	37 (38.1)	1.026 (0.311)
Positive	21 (52.5)	19 (47.5)	
Macrosialin, CD68: positive ≥ 17 cells0			
Negative	27 (62.8)	16 (37.2)	0.400 (0.527)
Positive	49 (57.0)	37 (43.0)	
Forkhead box protein P3, (FOXP3): positive ≥ 3 cells			
Negative	26 (60.5)	17 (39.5)	0.028 (0.866)
Positive	56 (58.9)	39 (41.1)	
Signal transducer CD24:			
Negative	8 (57.1)	6 (42.9)	0.018 (0.894)
Positive	49 (59.0)	34 (41.0)	
CD44 antigen: positive ≥ 50 H-score			
Negative	24 (54.5)	20 (45.5)	0.713 (0.398)
Positive	23 (63.9)	13 (36.1)	
Androgen receptor:			
Negative	30 (51.7)	28 (48.3)	4.238 (**0.040**)
Positive	63 (68.5)	29 (31.5)	
Trans-acting T-cell-specific transcription factor, (GATA3): positive ≥ 60 H-score			
Negative	44 (55.0)	36 (45.0)	0.355 (0.552)
Positive	5 (45.5)	6 (54.5)	
Proliferation marker protein Ki-67: positive ≥ 10 Percentage, full-face			
Negative	31 (73.8)	11 (26.2)	2.923 (0.087)
Positive	55 (58.5)	39 (41.5)	
Cadherin-1, (E-cadherin): positive ≥ 110 H-score			
Negative	48 (70.6)	20 (29.4)	4.345 (**0.037**)
Positive	50 (54.3)	42 (45.7)	
Cadherin-2, (N-cadherin): positive ≥ 100 H-score			
Negative	15 (46.9)	17 (53.1)	2.684 (0.101)
Positive	62 (63.3)	36 (36.7)	
Cellular tumor antigen p53, (p53):			
Negative	68 (62.4)	41 (37.6)	0.811 (0.368)
Positive	28 (54.9)	23 (45.1)	
Twist-related protein 2: positive ≥ 150 H-score			

Table 4. 10: Multivariate Cox regression analyses for prediction of breast cancer-specific survival, distant metastasis-free survival, and disease-free survival, altered protein expression of cytoplasmic SEC14L1.

Breast cancer specific survival (BCSS), n= 153◊

Cut-off point 70 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
SEC14L1 protein (cytoplasmic)	1.223	0.699	2.137	0.481
Tumour grade	1.792	0.989	3.250	0.055
Nodal staging*	1.546	1.018	2.349	0.041
Tumour size	1.862	0.955	3.630	0.068
Estrogen expression**	1.073	0.582	1.977	0.822
Her2 expression***	3.678	1.995	6.780	0.000

Distant metastasis free survival (DMFS), n= 153◊

Cut-off point = 70 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
SEC14L1 protein (cytoplasmic)	1.081	0.628	1.860	0.779
Tumour grade	1.575	0.942	2.632	0.083
Nodal staging*	1.432	0.960	2.137	0.079
Tumour size	2.130	1.145	3.963	0.017
Estrogen expression**	1.174	0.637	2.164	0.606
Her2 expression***	2.721	1.481	4.996	0.001

Disease free survival (DFS), n= 153◊

Cut-off point = 70 H-score	Hazard ratio (95% CI)	Lower	Upper	*p*-value
SEC14L1 protein (cytoplasmic)	1.117	0.675	1.849	0.665
Tumour grade	1.230	0.792	1.912	0.357
Nodal staging*	1.270	0.869	1.855	0.217
Tumour size	1.467	0.858	2.511	0.162
Estrogen expression**	1.165	0.638	2.128	0.618
Her2 expression***	2.746	1.532	4.924	0.001

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
◊ Sum of patients does not equal to 170 because of missing cases among different categories.

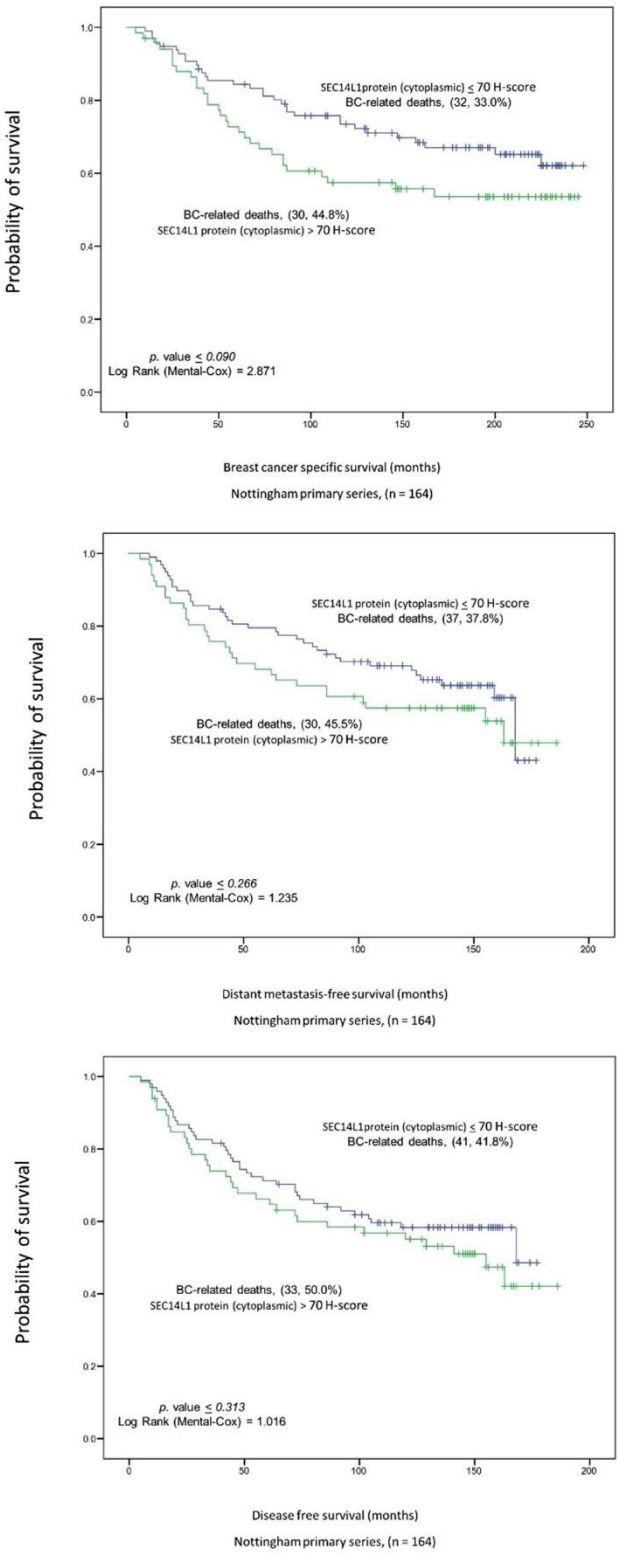


Figure 4. 21: Associations between the altered expression of cytoplasmic SEC14L1 protein and breast cancer-specific survival, distant metastasis-free survival, and disease-free survival in the Nottingham series of primary invasive breast carcinoma.

4.4. Discussion

Despite the extensive investigation of various biomarkers in invasive BC, the differential gene/protein expression associated with LVI and the roles of MCFD2, CCT8, SNAP23, and SEC14L1 in invasive BC remain poorly characterised. DEG analysis is a robust starting point to identify aberrant gene expression in breast tumours [245]. Accurate characterisation of the differentiation strata for the histological specimens, [$LVI^{+}/LN^{+/-}$] contrasted with [LVI^{-}/LN^{-}], was crucial to obtain informative transcriptomic data. Ideal modelling of validation methods, the sampling settings, and cohort construction, and well-designed experimental techniques have lent further strength to our hypothesis of LVI exploration. Well-known variations between transcription and translation profiles [202] may be verified by IHC as validation for protein expression. In this study, we reported novel findings for MCFD2, CCT8, SNAP23, and SEC14L1 altered protein expression in invasive BC patients with LVI data. However, different protein localisation and molecular subtypes showed different observations, in keeping with the heterogeneity of invasive BC, not only in morphology but also in the evolutionary molecular changes through its primary, metastatic, and recurrent phases [246].

4.4.1. Multiple coagulation factor deficiency protein 2

The MCFD2 protein was previously known as stem cell-derived neural stem/progenitor cell supporting factor (SDNSF). Multiple studies have reported pathogenic effects for the MCFD2-Lectin mannose-binding 1 (LMAN1) protein complex and their roles as cargo receptor proteins in combined factor-V and factor-VIII deficiency (F5F8D) in blood coagulation disorders [247-252]. MCFD2 dimerises with LMAN1 forming a complex specific cargo receptor for the EnR-to-Golgi transport of selected proteins [231], and this protein-transporting complex plays a critical role in the secretion of coagulation factors. Active transport mechanisms of specific proteins in BC from EnR to plasma membrane via Golgi apparatus were reported in BC in addition to various tumours. It was suggested that the aberrant trafficking of essential proteins for BC invasiveness and metastasis such as E-cadherin and MMP14 is regulated through a Golgi-mediated secretory pathway [253]. Other researchers focused on the effects of protein-protein interactions as a frequent biochemical reaction that took place on the newly synthesised proteins. Aberrant interference between MCFD2 and LMAN1 was suggested to promote the initiation, growth, and

metastasis of CRC [254]. Deficiency of the MCFD2-LMAN1 protein complex in the microsatellite unstable (MSI-H) CRC cell lines caused impaired secretion of the anti-angiogenic and growth-inhibiting protein α-1-antitrypsin (A1AT), indirectly promoting angiogenesis and vascularity of tumours [254].

The disruption of the MCFD2-LMAN1 complex may be thought to create a pro-vasculogenic environment with the potential of metastases. Our results show that the low MCFD2 protein level is remarkably associated with LVI^+ and high tumour grade invasive BC patients. Also, we identified that MCFD2 loss is statistically correlated with large tumour size and worse NPI. These clinicopathological factors are known to be correlated with LVI and hence their correlation of MCFD2 could be significant to invasive BC progression [255]. The low expression of MCFD2 marked the $ER^-/HER2^+$ and triple-negative patients as two aggressive and highly proliferative molecular types characterised with the worst survival profiles among invasive BC patients. According to our findings, the downregulated MCFD2 protein was associated with shorter BCSS, DMFS, and DFS. From previous research, we are aware of the triangular association between LVI, LN metastasis, and lymphatic vessel density (LVD) [256]. Higher risk of LNM is associated with LVI in the peripheral tumour areas with higher LVD. If MCFD2 plays a role in the transport of anti-vascularisation molecules, we suggest that the downregulated expression of MCFD2 protein could be useful as an early high-risk indicator for LVI, LNM, and recurrence of invasive BC, especially for patients with LN stage-I. The opposing impact for MCFD2 expression on the progression or inhibition of LVI in the studied cohort could support our claim and enhance the hypothesised LVI predictive impact for MCFD2 altered protein expression in invasive BC patients, regardless of whether the LVI^- patients belonged to positive or negative LN subgroups. Further morphometric studies may assess MCFD2 protein in various defined tumour compartments and associate relative expression with aggressive features [257].

As mentioned in the results section, regarding the altered gene expression of *MCFD2*, different tumour histological and molecular contexts seem to govern MCFD2 biology. On the one hand, HER2-enriched patients were associated with downregulated *MCFD2* gene transcription. On the other hand, the majority of basal intrinsic subtype patients exhibited elevated levels of *MCFD2*. Similarly, the HER2-enriched IntClust5,

IntClust6, and IntClust9 were the top clusters of patients with *MCFD2* loss, whereas IntClust10 showed a higher percentage of patients with upregulated gene expression of *MCFD2*. IntClust10 comprised patients with an aggressive phenotype of higher grade, larger tumour size, and poorer outcome, especially in the first five years after surgery [39]. Deeper investigation of what seemed a contradictory result revealed that the majority of these patients (38.2%) were diagnosed with basal-like BC. The latter molecular subtype is highly associated with poor prognosis and has no effective targeted therapy. Our findings support a key role for MCFD2 expression in cancer stem cell potency in basal-like invasive BC, which is supported in other studies [202, 229, 258]. It should be noted that the gene signature of any METABRIC integrating cluster may not be consistent with the protein expression level of a specified gene for the corresponding cluster of patients. Moreover, the vast variety of other clinical and histological factors of invasive BC may alter this molecular regulation, leading to conflicting findings. The molecular regulation at the mRNA level is not necessarily reflected in protein functionality. That being said, tackling positive expression of MCFD2 as a possible cancer stem cell maintainer of pluripotency in the basal-like molecular subtype of invasive BC is warranted [202] and could be associated with therapy resistance and shorter DFS, as our findings showed that the upregulation of MCFD2 was identified as an independent prognostic factor for BC recurrence.

At the protein level, tumours lacking MCFD2 are associated with $HER2^+$ and triple-negative patients. This may suggest strong transcription and translation of *MCFD2* regulated by the high expression of HER2; all patients' specimens in the METABRIC study were collected before the general availability of trastuzumab [39]. Our results for the MCFD2-interacting proteins revealed a strong positive interrelation with the AR. A recent IHC-based study revealed that the positive expression of AR was associated with luminal BCs, while the lowest expression was correlated with the basal phenotype [259]. Our results also indicate a positive interrelation between MCFD2 and the trans-acting T-cell-specific transcription factor, GATA-3 (*GATA3*), which is essential for proper ER function [260] and may act as a key suppressor for metastasis in triple-negative BC when activated [261]. MCFD2 appears to have a specific intrinsic biological role in ER^+ tumours.

Uncontrolled mitosis is a hallmark of invasive BC. The proliferation marker protein, Ki67, was considered a prognostic biomarker for invasive BC [262, 263]. In a large cohort study, the expression of Ki67 protein and its prognostic value in invasive BC patients was assessed [264]. The authors stated that Ki67 positive expression is generally associated with adverse prognostic factors. We presented the strong relationship between the $MCFD2^-$and $Ki67^+$ patients with supposedly highly proliferative tumours. Functional studies on MCFD2 and Ki67 expression are warranted. The lack of adhesive forces between the luminal monolayer of epithelial cells in breast ducts is a feature of invasion-related biological mechanisms. The disruption of apical-basal orientation of these cells is a signature of EMT [265]. In our cohort, positivity in MCFD2 and high E-cadherin expression provided a positive outcome. The loss of E-cadherin is well studied and is associated with higher propensities of invasion and metastasis, and intermediates the activation of other molecules that are related to invasion pathways in various malignancies [266-268]. The significant association with the positive N-cadherin expression when MCFD2 is downregulated confirmed good prognosis characteristics of MCFD2 positive expression. Tumours with negative expression of MCFD2 should be specifically scrutinised with early invasive activities and vigorous activation of EMT pathways in the primary malignant lesions. Hence, we propose a unique link between MCFD2 and the regulation of EMT pathways. Further genetic and protein interaction data for MCFD2 is illustrated on page 413, Appendix C.

In summary, our study reveals that the loss of MCFD2 protein expression in invasive epithelium of BC is highly correlated with definite LVI. Invasive BC patients who positively expressed MCFD2 demonstrated a higher survival rate, less risk of distant metastasis, and lower recurrence events than their counterparts with negative MCFD2 expression. Our study also showed that the altered MCFD2 protein expression could be differentially interpreted between ER^+, $ER^-/HER2^+$, and triple-negative patients. The protein interactions of MCFD2 with other proteins suggested that MCFD2 could be a novel key player in the stem cell, proliferation, and EMT pathways in invasive BC. Our results suggest a pivotal multi-factorial role of MCFD2 in invasive BC biology, and potentiates further scientific and clinical approaches in prognostic biomarkers and therapeutics.

4.4.2. T-complex protein 1 subunit theta

The CCT8 protein is a component of the chaperonin-containing T-complex, a molecular chaperone complex consisting of seven other CCT proteins, CCT1–CCT7. CCT8 assists the folding of proteins upon ATP hydrolysis [269]. The CCT8 complex was also suggested to facilitate the regulation of telomere maintenance through folding of the telomerase Cajal body protein-1, TCAB1, an RNA chaperone that is responsible for telomerase trafficking [270]. As part of the CCT complex, other studies state that active CCT8 could play a critical role in BBSome assembly, a complex involved in the primary biogenesis of cilium, and regulating vesicle transportation to the cilia [271]. The oligomeric CCT chaperone was demonstrated in other research as an essential mediator for the folding of members of the eukaryotic cytoskeletal protein families, such as actin and tubulin [232]. Recently, a novel function for the CCT4 monomer was discovered. It was stated that CCT4 is indirectly involved in the microtubule minus end-directed transport via the motor protein, dynein [272]. The interaction between CCT4 and dynein was identified as promoting the formation of cellular protrusions and extensive cellular membrane ruffling, and this function was also found to be exclusive to the CCT4 monomer. Another study identified CCT8 as interacting with the C-terminus of the newly synthesised dynein intermediate chain-2 (IC-2) by co-immunoprecipitation in the A375 melanoma cell line [273]. These studies suggest that the biological functions of monomeric CCT8 differ from its oligomeric form. Interestingly, dynein light chain roadblock-type-1 protein, DYNLRB1, was identified as one of the differentially expressed proteins when we scrutinised the LVI-related differentially expressed proteins using MS techniques. Further details are mentioned in the following chapter. Here, we identified CCT8 as one of the differentially expressed proteins that showed high association with LVI and invasive BC prognosis. Our results indicate a significant difference in the prognostic impact between the cytoplasmic and nuclear subcellular localisation of the CCT8 protein in invasive BC patients.

The cytoplasmic expression of CCT8 did not show a reliable prognostic relevance or any association with BC molecular phenotypes; however, the positive cytoplasmic expression of CCT8 was strongly and positively correlated with the protein expression of CD24 and p53, but negatively associated with GATA3. The signal transducer CD24 is a glycosylated

mucin-like glycosyl-phosphatidylinositol-anchored cell surface protein. The strong expression of CD24, when CD44 is negative, was identified as a poor prognostic biomarker in early invasive BC [274]. Positive expression of CD24 was also associated with unfavourable outcomes in other cancers [275-279]. The knocked-down CD24 in the MCF7 cell line resulted in a significant reduction in proliferation and invasion capacity [280]. In contrast, the triple-negative aggressive phenotype of $CD24^{-}/CD44^{+}$ cells was claimed to illustrate higher migration and invasion potential [281]. This may suggest that the positive cytoplasmic CCT8 may indicate a high-risk subpopulation of invasive BC patients in which tumours evolve cancer stem cell-like properties, or intrinsic features of the basal-like subtype. Further investigations are required as the current experimental methods were not sufficient to precisely define the altered cellular localisation of cytoplasmic CCT8 in invasive BC patients.

The nuclear CCT8 expression, on the other hand, showed high prognostic impact in invasive BC patients. The strong nuclear expression of CCT8 was inversely correlated with LVI^{+}, large tumour size, high tumour grade, nodal involvement, and poor NPI. Patients who expressed high nuclear levels of CCT8 exhibited better BCSS, DMFS, and DFS. In addition, the downregulated nuclear expression of CCT8 was identified as a protein of independent predictive power for worse survival, DM, and recurrence. Lack of nuclear CCT8 was positively correlated with the aggressive molecular phenotypes such as $ER^{-}/HER2^{+}$ and triple-negative. The positive protein-protein interactions between AR, GATA3, E-cadherin, and nuclear CCT8 suggested a preferential upregulation and nuclear localisation of CCT8 in the cohesive ER^{+} tumours. To our knowledge, this study is the first attempt to provide strong evidence of the prognostic value of positive nuclear CCT8 expression in invasive BC patients.

Interest in the underlying mechanisms of CCT8 in cancer progression is ongoing. Previously, CCT8 expression was stated as modulating metastasis in HCC cell lines through CCT8/c-Jun/EMT signalling [239]. The key of this finding was produced by the interrogation of the role of endoplasmin (HSP90B1), previously known as GRP94, in HCC progression. The depletion of HSP90B1 was associated with decreased CCT8 expression, proliferation, and migration abilities of the HepJ5 HCC cell lines. The transient overexpression of CCT8 in HSP90B1-knocked-down cells restored the migration potential. The elevation of CCT8 production was accompanied

by an overexpression of p-JNK, JNK, c-Jun, fibronectin, and N-cadherin, but a reduction in E-cadherin [239]. Our results show similarities regarding the expression status of the cadherin proteins and the ki67 proliferation marker. In this cohort, cross-comparison between the distribution of the $Ki67^{+}$ expression profiles among the categorical subgroups of cytoplasmic and nuclear CCT8 expression revealed a positive association with CCT8 in the cytoplasmic compartment, but not in nuclear subcellular localisation. The latter expression status of CCT8 was strongly and inversely related to ki67. The percentage of patients in the positive cytoplasmic CCT8 expression category and E-$cadherin^{-}$ and N-$cadherin^{+}$ was relatively higher than those with CCT8 nuclear subcellular localisation.

A recent IHC study identified that overexpression of CCT8 was associated with definite LNM and poor survival in 128 patients with OeSCC [205]. A following functional assay revealed that the α-actin and β-tubulin-regulated migration and invasion of OeSCC cells were endorsed by cytosolic CCT8 [205]. Another study on patients of B-cell non-Hodgkin's lymphoma showed that cytosolic CCT8 was highly expressed in proliferating germinal centre cells compared with dormant cells [238]. An association between high cytosolic expression of CCT8 and cancer proliferation was also identified by a previous IHC study in 102 HCC patients [203]. The expression of CCT8 was also investigated in glioma tissues applying IHC methods in a further study [204]. The examination of the FFPE immuno-stained tissue sections showed that CCT8 was predominantly expressed in the cytoplasmic compartment. High CCT8 expression was associated with poor outcome in glioma patients, and the silencing of *CCT8* decreased proliferation and invasion abilities in the U251MG human glioma cell line [204]. Further genetic and protein interaction data for CCT8 is illustrated on page 414, Appendix C.

From these studies, and from our results, we propose that the molecular regulation of CCT8 in invasive BC is differentiated according to the subcellular localisation of the CCT8 protein. Our study showed that high CCT8 protein nuclear expression might be related to good outcomes in invasive BC, and that nuclear CCT8 expression supresses the LVI mechanisms by suspending other invasion-related physiological hallmarks such as EMT and rapid proliferation. The positive nuclear CCT8 could be described as having tumour-suppressing qualities. We also suggest a pro-oncogenic role for the CCT8 conditioned by its cytoplasmic localisation. The

studies cited above did not identify nuclear expression for CCT8 in different malignancies, and the unfavourable prognostic conclusions for high CCT8 were based on its cytosolic detection. Therefore, the physiological mechanisms for the CCT8 protein in the different cellular compartments of invasive BC cells are worthy of further assessment in the future.

4.4.3. Synaptosomal-associated protein 23

Synaptosomal-associated protein-23 is a 23 KDa protein that is essential for the docking and fusion of protein transporting granules in vesicle-mediated pathways. SNAP23 belongs to the SNARE family of proteins. The SNARE complex family of proteins is one largely studied cluster that mediates membrane docking and fusion in eukaryotic cells [41]. SNARE proteins are classified into two anchoring protein vesicles (v-SNARE) and target (t-SNARE), which together form the SNARE complex protein in the membranes of transporting vesicles and terminal targets, respectively [42]. Hundreds of studies have investigated the pathological roles of several SNARE proteins such as vesicle-associated membrane proteins (VAMPs), syntaxins (STXs), and synaptosomal-associated proteins (SNAPs) in different pathological conditions [43–45].

The elevated biological activity in invasive BC, specifically, the greater demand on protein production, indicates potential clinical importance of the secretory vesicles in exocytosis pathways. Efficient vesicular trafficking of intracellular protein production may be crucial for BC invasion and metastasis. The excretion of certain proteins and cytokines intracellularly or into extracellular spaces regulates the dynamic signalling between the invasive BC cells and many components of the TME [282, 283]. Thus, vesicular transport pathways require specific protein-trafficking mechanisms that organise the trafficking of the protein transport vesicles through the different cellular compartments until they reach their targeted destinations [284]. Protein transport vesicles are molecular ampules loaded with large molecular weight proteins. Translocation of the wrapped cargos into their corresponding functional sites in vesicular transport mechanisms, internally or externally, pass through four fundamental molecular steps: **1**) the formation of clathrin-coated vesicles; **2**) budding and detachment of the newly formed coated vesicle [285]; **3**) attachment of motor proteins to the protein-transporting vesicles and coordination of their directions along microtubules of the cytoskeleton [286, 287]; and **4**) docking and fusion onto the target plasma membrane [288]. The machinery for the docking

and fusion of cargo transport vesicles is highly sophisticated and comprises myriad biochemical interactions between various proteins on the ERGIC. The vesicular-tubular structure of ERGIC hosts a vast number of protein-related reactions, and it begins by shipping the newly produced protein from the EnR towards the Golgi apparatus in an anterograde fashion [289]. Further post-transcriptional modifications take place in the Golgi apparatus, which receives the synthesised protein from EnR, and the discharge of this protein or its intracellular utilisation is decided through exocytosis or retrograde transport to the EnR. The determining factors of which pathway should occur are the retrieval signal luminal proteins in the Golgi apparatus cavities [290].

In this study, the examination of the stained TMA sections showed that SNAP23 sub-localised in the cytosol and cellular membrane, and it was not strongly expressed in around half of the samples. Membranous expression of SNAP23 was the dominant localisation for this protein. Very few samples showed absent membranous expression accompanied with positive cytoplasmic localisation of SNAP23. The overexpression of membranous SNAP23 was positively associated with LVI incidence, $HER2^{+}$ tumours, and tumour grade-II. The high cytoplasmic expression of SNAP23 did not appear to associate LVI in the studied cohort. The aberrant subcellular localisation of the SNAP23 protein was stated to elicit various biological processes in several cancer studies [291]. A recent OC study linked the high expression of SNAP23 protein with poor survival. Silencing of SNAP23 expression weakened the migration and invasion potentials of OC in vitro, and it was suggested that SNAP23 is a key regulator for apoptosis inhibition and aberrant metabolism in OC [292]. A further study presented conflicting results for upregulated SNAP23 in cervical cancer (CvC). The IHC staining of SNAP23 showed low-intensity SNAP23 expression in the CvC tissues in comparison to adjacent normal tissues. Downregulated SNAP23 was inversely associated with unfavourable outcomes, and it was suggested that overexpression of SNAP23 may induce G2/M cell-cycle arrest [293]. The regulation of invasive BC invasion pathways by SNAP23 could be multifactorial.

Here, our results show that the aberrant localisation of the SNAP23 protein has different implications for LVI. The successful invasion mechanism by BC cells into their surrounding tissues requires highly sophisticated and regulated physical and biochemical interactions. As a t-

SNARE, SNAP23 could be involved with other proteins to initiate a unique intracellular trafficking machinery for specific invasion-related proteins in the invasive phenotype of BC [294]. Previously, a protein co-localisation study was performed on the MDA-MB-231 invasive BC cell line to identify the key driver molecules involved in the transportation and secretion of MMP14 [295]. The protein complex of SNAP23-Syntaxin4-VAMP7 was proposed as an efficient transportation apparatus for MMP14 to invadopodia. First, the v-SNARE vesicle-associated membrane protein-7 (VAMP7) could mediate the vesicular protein transportation of MMP14 to its targeted destination intracellularly during the early stages of invadopodia formation. Second, SNAP23, situated in the cellular membrane, forms a complex with the dephosphorylated syntaxin-4, a cellular membrane t-SNARE protein that facilitates docking of transport vesicles. The latter complex is localised to the site of invadopodia formation. Finally, VAMP7 binds to the SNAP23-sytaxin-4 complex, and the membrane of the MMP14-containing vesicle is fused with the cellular membrane, enabling MMP14 to be extracellularly released and degrade the surrounding ECM [295].

The active interaction between SNAP23 and syntaxin-4 could be interrupted by syntaxin-binding protein-3 (Munc-18c), which specifically binds to the N-terminal of syntaxin-4, leading to a conformational change in the syntaxin-4 structure, inhibiting the association with SNAP23 [296]. What decides the optimum complex partners could be set by the type of v-SNARE and the contents of trafficked cargo [297]. The induction of Munc-18c protein may exchange the SNAP23-syntaxin-4 complex with SNAP25-syntaxin-4, SNAP23-syntaxin-2, or SNAP23-syntaxin-3 complexes [298]. In a different approach, an immune-evading role for SNAP23 was identified in melanoma [299]. Although our results did not pinpoint a specific correlation between SNAP23 and the IHC-stained biomarkers of immune cells, there was a trend toward a significant association between the overexpression of membranous SNAP23 and the low expression of the B-lymphocyte antigen, CD20. This may suggest that membranous SNAP23 may negatively influence the immune response in invasive BC patients. Further genetic and protein interaction data for SNAP23 is illustrated on page 415, Appendix C.

Taken together, SNAP23 is a critical membrane-trafficking protein that is associated with pathogenesis. The membranous localisation of SNAP23 protein mirrors the biological relevance for this molecule; however,

being part of several vesicle-transportation complexes would make the interpretation of its association with LVI extremely challenging, especially with multiple reports stating contradictory prognostic relevance for SNAP23 in different cancers. Scrutinising the trafficking machinery of proteolytic enzymes and their association with SNAP23 expression in invasive BC may enhance our current understanding of LVI in invasive BC, and it could define a novel LVI driver and therapeutic target in the future.

4.4.4. SEC14-like protein 1

In the present research, we aimed to decipher the molecular mechanisms underlying LVI in invasive BC. High-throughput molecular techniques coupled with bioinformatics and strict definition of LVI status were utilised to identify genes associated with LVI that can potentially be used as therapeutic targets. Following internal and external validation, SEC14L1 was identified as one of the top DEGs associated with LVI and also with BC prognosis. Further examination of SEC14L1 at protein expression level confirmed its association with LVI and indicated its independent prognostic value in invasive BC.

Our investigation of the clinicopathological impact of the altered protein expression of SEC14L1 protein in the invasive BC unveiled strong positive association with definite LVI incidence. The optimal cut-off point for SEC14L1 was set with X-tile software [300] in the published version [191]. In this study, we intended to keep the consistency of the selection of cut-off points at the median for SEC14L1 as we did for MCFD2, CCT8, and SNAP23. Although the interrogated clinical factors did not show statistical significance with the median-categorised SEC14L1 as they illustrated with the X-tiles in the published version, the association between the overly expressed SEC14L1 and LVI^+ patients was conserved in both. This SEC14L1 protein expression showed a significant relationship with luminal BC and high histological grade. Further examination of the 46 patients with $SEC14L1^+/LVI^+$ profiles revealed that 67.4% were of high tumour grade, and 76.1% were ER^+ (data not shown). We previously reported that LVI was an independent predictor for BCSS in ER^+ patients [52]. The discussed MCFD2, CCT8, and SNAP23 proteins were positively associated with the ER^+ patients, suggesting EnR-operated signalling that regulates LVI mechanisms. Moreover, our results show that SEC14L1 was positively associated with cytoplasmic CCT8 and membranous SNAP23, two proteins that might be associated with poor outcome and progressive LVI.

Previous studies suggested that SEC14L1 might play an important role in trafficking proteins to the cell membrane [301, 302]. SEC14L1 belongs to the SEC14 cytosolic factor family, which play a role in the intracellular transport system [303] and innate immunity [304]. SEC14L1 overexpression may influence intracellular signalling pathways and bridge the BC cell to several components of the TME to promote tumour growth, LVI, and metastasis.

A recent investigation of PC showed a prognostic 12-gene signature, including SEC14L1, associated with criteria of aggressive clinical outcome, which was also confirmed by RT-PCR and IHC assays [208]. SEC14L1 cytoplasmic protein overexpression was frequently found in the transmembrane protease, serine 2 (TMPRSS2)/Ets-related gene (ERG) fusion-positive PrC, and the prognostic power of SEC14L1 depends strongly on this fusion status in PrC [305]. The TMPRSS2/ERG fusion has been revealed to be associated with several molecular alterations, including deletion of the phosphatase and tensin homolog gene (*PTEN*) [306]. Several previous studies have suggested that *PTEN* loss is strongly correlated with transcriptional instability and is associated with poor outcome of $HER2^+$ BC [307, 308]. The latter statement may suggest further attention to LVI regulations in invasive BC cancer patients of the $SEC14L1^+/PTEN^-$ protein expression profiles, especially in the $ER^+/HER2^{+/-}$ phenotypes. TMPRSS2/ERG fusion was also thought to be correlated with AR activity [309], transcriptional stability [310], and stem cell maintenance [311] in cancer. Our results show that the majority of AR^+ patients demonstrated low expression of SEC14L1, suggesting a good prognostic feature for the $SEC14L1^-/AR^+$ subgroup of invasive BC patients. These relationships between SEC14L1 and intracellular transportation and oncogenic properties associate PTEN loss as essential for cell proliferation, adhesion, invasion, and migration. Thus, SEC14L1 may play a pivotal role in the regulation of cell growth/tissue development and cell adhesion, underlying LVI and metastasis in invasive BC. In human HCC, the SEC14 domain-containing protein, clavesin-1 (*CLVS1*), has been identified as a potential prognostic biomarker that is mainly sub-localised in the cytoplasm of malignant cells [312]. Further genetic and protein interaction data for SEC14L1 is illustrated on page 416, Appendix C.

This study revealed that the upregulated expression of SEC14L1 protein is a significant prognostic factor for invasive BC. Overexpression of

SEC14L1 protein may play an important role in the development of LVI, invasive BC progression, and metastasis through several intracellular mechanisms. Further functional assessment of SEC14L1 to determine its mechanistic roles and therapeutic potential in invasive BC is warranted.

4.5. Conclusion

Lymphovascular invasion is a key event in the invasion-metastasis cascade in invasive BC. In this study, we aimed to identify the driver proteins that are significantly associated and differentially expressed in LVI pathways. The present study identified four proteins that showed significant association with LVI. MCFD2 and CCT8 were strongly associated with the absence of LVI and favourable outcomes. SNAP23 and SEC14L1 were associated with poor prognosis and definite LVI occurrence. Clinically, our study provided novel findings and useful prognostic indications. However, it appeared that the underlying mechanisms of LVI are contextualised with myriad biophysical and biochemical factors.

The molecular subtypes of invasive BC could be a crucial factor that can alter the biological functions of the discussed LVI-associated proteins. For example, MCFD2 showed positive outcomes, but its upregulated expression in triple-negative BC may indicate the presence of a subpopulation of indolent CSCs, which may increase chemo-resistance and disease recurrence. In the conclusion of the third chapter of this study, we stated that $HER2^+$ patients exhibited a positive relationship with the top upregulated genes associated with LVI^+ incidences. Here, $HER2^+$ patients showed an opposite association with the cytoplasmic MCFD2 and nuclear CCT8 expressions. The sub-cellular localisation was another vital element that could dictate different LVI scenarios in invasive BC. In this study, we identified several promising LVI candidates that were associated with various pathways. The pathways related vesicular trafficking of protein cargos to the cellular membrane, and the cargo receptor for the EnR-to-Golgi transport of selected proteins could be the key pathways that are related to LVI progression. Further functional studies should be conducted to consider the different molecular phenotypes of invasive BC, protein sub-cellular localisation, and PPI due to the sophisticated mechanisms underlying LVI.

LVI associations, cellular localisations, and relationship with hormonal receptors for MCFD2, CCT8, SNAP23, and SEC14L1

Gene ID	MCFD2	CCT8	SNAP23	SEC14L1
Protein full name	Multiple coagulation factor deficiency protein 2	T-complex protein 1 subunit theta	Synaptosomal-associated protein 23	SEC14-like protein 1;SEC14L1
Differentially expresed gene cluster	LVI-NT100 (Top DEG+CNA-gain)	LVI-NT100 (Top DEG+CNA-gain)	LVI-NT1154 (Gene set enrichment analysis)	LVI-NT283 (UP-13/LVI^+)
Statistical analysis	LIMMA, Jonckheere-Trepstra	LIMMA, Jonckheere-Trepstra	LIMMA	LIMMA, Mann-Whitney U-test
Adjusted p-value (LIMMA, Differential gene expression	0.00145	0.00128	0.0015	0.02769
p-value (Jonckheere-Trepstra, Copy number	0.0209	0.0385	NA	NA
Antibody specificity (Western blotting, or Immunising peptide blocking)	Validated with IPB	Validated with WB	Validated with WB	Validated with WB
Cellular localisation (IHC)	Cytoplasmic	Cytoplasmic & Nuclear	Cytoplasmic & Membranous	Cytoplasmic
Protein type	Endoplasmic reticulum	Chaperone; Motility/Polarity/Chemotaxis	Vesicle	Transport
Association with LVI	Upregulated in LVI^-	Cytoplasmic: No association Nuclear: Upregulated in LVI^-	Cytoplasmic: No association Membranous: Upregulated in LVI^+	Upregulated in LVI^+
p-value (Chi-square test)	0.003	Cytoplasmic: 0.931 Nuclear: 0.010	Cytoplasmic: 0.490 Membranous: 0.020	0.0003
Association with Hormonal receptor	Downregulated in $HER2^+$, ER^-, and TNBC patients	Cytoplasmic: No association Nuclear: Downregulated in HER2+, ER-, and TNBC patients	Cytoplasmic: Downregulated in ER^- and TNBC patients Membranous: Upregulated in $HER2^+$ and ER^- patients	No association
p-value (Chi-square test)	$HER2^+$ = 0.021 ER^- = 4.0E-6 TNBC = 6.0E-6	Cytoplasmic: No significance Nuclear: $HER2^+$ = 2.0E-6 ER^- = 5.3E-17 TNBC = 6.0E-12	Cytoplasmic: ER^- = 6.2E-9 TNBC = 2.5E-8 Membranous: $HER2^+$ = 0.002 ER^- = 2.2E-11	No significance

Chapter 5: LVI Molecular Signatures in Invasive BC: RNA microarray and HPLC/MS Study

5.1. Introduction

Invasive BC could be a fatal disease if not discovered early and treated. Cancer mortality statistics for 2014, published in 2016 by Cancer Research UK (CRUK), state that the mortality rate of invasive BC per 100,000 females in the UK is 34.6; among the top 20 registered cases of cancer, invasive BC was ranked in second position. Invasive BC represents 14.8% of all malignant neoplasms diagnosed in female patients in the UK. In 2014, there were 11,400 deaths caused by BC, which is equivalent to 31 deaths per day (CRUK, 2016).

The enhancement of the treatment and management of invasive BC could provide alternative prognostic and therapeutic options for subsets of invasive BC patients. However, the phenotypical and molecular heterogeneous natures of invasive BC may challenge this approach. Promising breakthroughs in discovering novel BC gene/protein signatures have expanded in the last few years. A recent study identified a gene signature of 22 expressed genes that may predict a worse survival in advanced HNSCC patients after chemotherapy, and underscored a subset of genes among the suggested 22-gene signature that is responsible for driving the chemotherapy resistance mechanisms in a high-risk group of HNSCC patients [313]. Similarly, a later study identified an ethnicity-related radiosensitive molecular signature that would improve the prediction of the overall survival (OS) in post-radiotherapy invasive BC patients [314]. Moreover, the critical involvement of the tumour microenvironment components in cancer progression was demonstrated through the identification of a gene signature consisting of 15 differentially expressed genes that may play key roles in macrophage-tumour signalling and are associated with poor survival in invasive BC patients [315]. From the previous examples, we emphasise that the differential gene and protein expressions in any cancer are not only governed by the malignancy type or the organ affected; in fact, the forces that could drive BC progression, including LVI mechanisms, may encompass several clinical, molecular, cellular, and biochemical aberrant interactions.

As BC develops as a multi-step disease, myriad key genes and proteins are associated with tumour progression and maliciously impact the BC patient's survival. The progressive environment for invasive BC comprises heterotypic and collaborative junctions of multiple pathways under specific biological contexts [316]. Hanahan and Weinberg conceptualised the basis for the hallmarks of cancer in their renowned work in 2011, and they summarised the pathogenesis for, perhaps, all cancers [7]. Their illustration of cancer's biological capabilities included several biological pathways such as proliferative signalling, angiogenic, resisting

cell death, altered metabolism, inflammatory, metastasis pathways, and vascular invasion. From the perspective of the modern molecular pathology, it is critical to link the morphological changes occurring in the observed malignant tissues with their unobserved drivers at the molecular levels. LVI is a morphological and independent parameter of BC prognosis [52]. The early detection of tumour emboli within a lymphatic lumen, especially in primary invasive BC patients with LN^- status, predicts an early metastatic potential and unfavourable outcome [78]. In addition, other morphological features of invasive BC were reported to be exponentially associated with LVI, such as tumour grade and tumour size [52, 99]. The distinctive molecular mechanisms of LVI in invasive BC are still ambiguous.

In the context of BC prognosis, molecular pathologists may combine histological methodologies with high-tech molecular biology procedures and computational integrative analyses to enhance the build-up of a molecular prognostication for invasive BC patients upon diagnosis [317-321]. Proper prognosis of BC patients could be misguided because of the heterogeneous nature and the substantial diversity of genotypic and phenotypic characteristics of invasive BC. Recently, rapid advancements in molecular research technologies have allowed a better understanding of the unique molecular patterns of invasive BC. Whole genome sequencing, bio-molecular and computational spectroscopic analyses, and microarray genetic testing have been utilised to discover new genome and proteome signatures in BC [164]. Biological knowledge bases for molecular signatures and gene expression data are invaluable resources for online data mining as well [322-326].

In this study, we aimed to discover a cluster of novel mRNA/proteins that are associated with robustly defined LVI patients. The present study is based on tracking the LVI-DEGs that could be successfully translated into proteins by comparing the results of RNA-microarray DGE analysis against the results of HPLC/MS under similar experimental conditions. This work is an initial approach for the molecular classification of invasive BC based on LVI status. This study could improve BC clinical management based on a proposed LVI molecular signature for each BC intrinsic subtype. In addition, we aimed to investigate the transcriptomic and proteomic architectures that operate and regulate LVI mechanisms by applying bioinformatics, mRNA microarray, and HPLC/MS methods. In the future, we would like to convey this understanding of LVI molecular biology into translational clinical research and design systematic functional studies accordingly to confirm our construction of novel prognostic biomarker clusters for invasive BC patients in the future.

5.2. Materials and methods

5.2.1. **Study cohorts**

5.2.1.1. *METABRIC cohort*

The construction of this cohort was explained in detail in chapter two.

5.2.1.2. *LVI-supervised cohort*

The construction of this cohort was explained in detail in chapter two.

5.2.1.3. *The luminal-A subtype cohort*

The construction of this cohort was explained in detail in chapter two. The clinical data of the luminal-A cohort is shown in Table 2.5.

5.2.2. **DGE analyses**

Extensive details of the DEG experiment are presented in chapter three, (section 3.2.2.).

5.2.3. **Total protein extraction and quantification**

5.2.3.1. *Tissue harvesting off FFPE tissue sections*

The specimens in this study were derived from a cohort of luminal-A BC patients; specific details on this cohort are mentioned in section 2.1.5. Corresponding H&E stained sections, 4µm, for each specimen were available. Tissue harvesting with a surgical scalpel was applied to four to six tissue sections, 8µm thickness each, of the same FFPE block. The collected areas with high tumour burden were marked previously on the H&E stained sections and tracked. Harvested tissues were accordingly collected in 1.5ml collection tubes (Qproteome FFPE Tissue Kit, catalogue number: 37623, Qiagen, UK). Tissues were harvested from the slides while they were still wet after microtome cutting. If the tissues dried on the slide, the amount of collected tissues decreased dramatically.

5.2.3.2. *Dewaxing of the harvested tissues*

Sample preparation and protein extraction were carried out as per the Qproteome FFPE Handbook 01/2011 (Qiagen) with the moderate alterations arising from extensive optimisation. Deparaffinisation was performed by adding 500µl of n-heptane (99%) (product code: 10664912, Fisher Scientific, UK) as an inorganic solvent. The tubes were sealed tightly and vortexed thoroughly for 10 seconds. The mixture was incubated for 60 minutes at room temperature. Aliquots of 25µl of MeOH were added to each tube and the tubes were re-sealed and vortexed as per the previous step. The dewaxed and dehydrated tissues were micro-centrifuged at 9000 × g for two minutes and, consequently, precipitated as semi-solid white

pellets. The supernatant was removed from each tube very carefully without disrupting the pellets at the bottom.

5.2.3.3. *Total protein extraction using Qproteome FFPE Tissue Kit*

The protein extraction buffer (PEB), provided with the Kit, was supplemented with 6% of β-mercaptoethanol as a denaturing agent that reduced the protein's disulphide bonds. 100µl of supplemented PEB was added to Eppendorf tubes containing the protein pellets. All tubes were mixed vigorously for one minute, then properly sealed with tube-sealing clips. The tubes were incubated for 20 minutes at 100°C on a heating block. On another heating block, which had a shaking function, all tubes were incubated for 120 minutes at 80°C with agitation at 750 rpm. Then, all tubes were placed in the fridge at 4°C for one minute. The sealing clips were removed carefully and all tubes were placed into a refrigerated centrifuge with an angled rotor that held reaction vials of 2.2–1.5ml and centrifuged for 15 minutes at 14,000 × g. Supernatants were transferred carefully into new and labelled 1.5ml collection tubes. 10µl aliquots were taken from each 100µl protein-containing supernatant to estimate the total protein concentration in each sample quantitatively before tryptic digestion and protein precipitation.

5.2.3.4. *Bradford protein assay*

Please refer to Chapter 2 (section 2.10.4.1) for complete details.

5.2.4. **Preparation of protein sample for tryptic digestion**

Please refer to Chapter 2 (section 2.11) for complete details.

5.2.4.1. *Tryptic digestion of protein extracts for mass spectroscopy*

Please refer to Chapter 2 (section 2.12.1) for complete details.

5.2.5. **Reverse-phase high-performance liquid chromatography electrospray ionisation tandem mass spectrometry (RP-HPLC-ESI-MS/MS) in SWATH and IDA modes**

Please refer to Chapter 2 (section 2.12.2) for complete details.

5.2.5.1. *Construction of peptides library, spectral normalisation, and fold change analysis*

Please refer to Chapter 2 (section 2.12.3) for complete details.

5.2.6. **Statistical analyses**

In this study, the integrated statistical analyses were performed using the online SCIEX OneOmics™ software for SWATH Acquisition on a SCIEX TripleTOF®

system. The expression fold changes between the digested proteins/peptides from the experimental set of LVI^+ against the control set of LVI^-/LN^-specimens from SWATH® acquisition data were comprehensively normalised and computed by the Assembler application of the OneOmics™ cloud processing software (SCIEX, Warrington, Cheshire, UK). The resulting statistics of the sample grouping proteins, peptides, and transitions, normalisation bar plots, protein fold changes, top BP, and the enriched GO were saved and browsed using the Protein Expression Workflow application (SCIEX, Warrington, Cheshire, UK), available online at Illumina's BaseSpace website. A list of DEPs was generated, considering the re-nomenclating proteins' identifiers to gene-identifiers format for proper alignment against their expressing genes.

SPSS statistical software v24.0 (IBM, Armonk, NY, USA) was used in the characterisations of the studied cohorts, which were determined by utilising the descriptive analytical commands. The Pearson correlation test was used to evaluate the gene expression correlations and the directions of their expression between: 1) the identified overlapping DEGs/DEPs (n=56) from the RNA-microarray/HPLC-MS experiments; and 2) assessment of the gene expression correlations between the LVI-related transcripts against a cluster of selected transcripts (n=61) related to tumour invasion, migration, EMT, and proliferation. The chi-square test was used to assess the distribution of the $LVI^{+/-}$ patients and the METABRIC cohort among the different expression levels of the DEG/DEP candidates. The rest of the chi-square statistics considered for the patients of the latter cohorts were: age at diagnosis, tumour grade, LN involvement, NPI, PAM50, and the IntClusts of METABRIC. KM survival curves for BCSS were plotted and the log-rank test was conducted. The BCSS in this study was determined as the time from the day of surgery to the day of death caused by breast cancer. For univariate and multivariate survival assessment, clinicopathological factors, including the mRNA identifiers for the identified DEPs by HPLC/MS, were assessed using the Cox proportional hazards regression model, 95% CI. The median mRNA expression value was set as a cut-off point in these mRNA-based statistical analyses.

5.2.7. Ethical approval and authentication of material transferring

This study was approved by Nottingham Research Ethics Committee 2 under the title 'Development of a molecular genetic classification of breast cancer', and in compliance with current ethical and legal guidelines of the UK. All experiments in this study were conducted in accordance with the 1975 Helsinki declaration and its later amendments or comparable ethical standards. A material transfer agreement (MTA) was made between NHSB, based at NUH, the NHS trust with principal offices

at City Campus, the David Evans Medical Research Centre, Hucknall Road, Nottingham, NG5 1PB, and the NTU, John Van Geest Cancer Centre, with principal offices at 50 Shakespeare Street, Nottingham, NG1 4FQ. All biological materials were transferred from NHSB to the proteomic laboratory at the JvGCRC, School of Science and Technology, NTU, Clifton Lane, Nottingham, NG11 8NS, and solely used for the purpose of this study, within the research group of Professor Graham Pockley and under the supervision of Dr Amanda Miles.

5.3. Results

5.3.1. **Features of the studied cohorts**

5.3.1.1. *METABRIC cohort*

The detailed features for the METABRIC cohort are stated in chapter three, (section 3.2.1.2).

5.3.1.2. *LVI-supervised sub-cohort of the METABRIC*

The detailed features for the $LVI^{+/-}$ cohort are stated in chapter three, (section 3.2.1.1).

5.3.1.3. *LVI-supervised and luminal-A cohort for mass spectroscopy*

Increasing the sample size for the HPLC/MS experiment, and avoiding particular molecular subtypes of BC with genomic instability and higher proliferative molecular profiles, prioritised luminal-A for LVI study. Similar to the $LVI^{+/-}$ (n=170) statistics, LVI^{+} and LVI^{-} cases for the MS study were practically equalised: (41, 51.3%) and (39, 48.8%), respectively. In parallel to the mRNA cohorts, moderate NPI was calculated for 51.3% of the total number of patients, though tumour grade-I was seen in only 15 patients, while tumour grade-III was seen in 31 (38.8%). Positive nodal involvement was recorded in only 20 patients, yielding 21 LVI^{+}/LN^{-} patients, and 39 LVI^{-}/LN^{-} patients, in this cohort. Histologically, IDC and ILC were confirmed in 65 (81.3%) and 13 (16.3%), correspondingly.

5.3.2. **Protein quantitation and differentially expressed proteome**

Regression analysis revealed reliable goodness of fit for standard curve regression line R^2=0.995 (Figure 5.X). Calculation of total protein concentrations was based on 10µl aliquots of total protein extracts. Spectrophotometric absorbance at 595nm of all aliquots resulted in protein concentrations ranging between 1.5–2.6 mg/ml, with a median/mean of 2.30/2.19 for LVI^{+} specimens, and 1.7–3.2 mg/ml, with a median of 2.40/2.39 for LVI^{-}/LN^{-} specimens. Optimum protein concentrations allowed us to proceed with the tryptic digestion procedure. Combining the best-performing peptide libraries from the LVI^{+} and LVI^{-}/LN^{-}

subgroups, we obtained a list of 1,179 proteins, 4,737 peptides, and 28,394 transitions. The top five GO processes were: 1) 271 proteins enriched for small molecule metabolic process; 2) 248 proteins enriched for gene expression; 3) 190 proteins enriched for cellular protein metabolic process; 4) 169 proteins enriched for viral process; and 5) 137 proteins enriched for innate immune response.

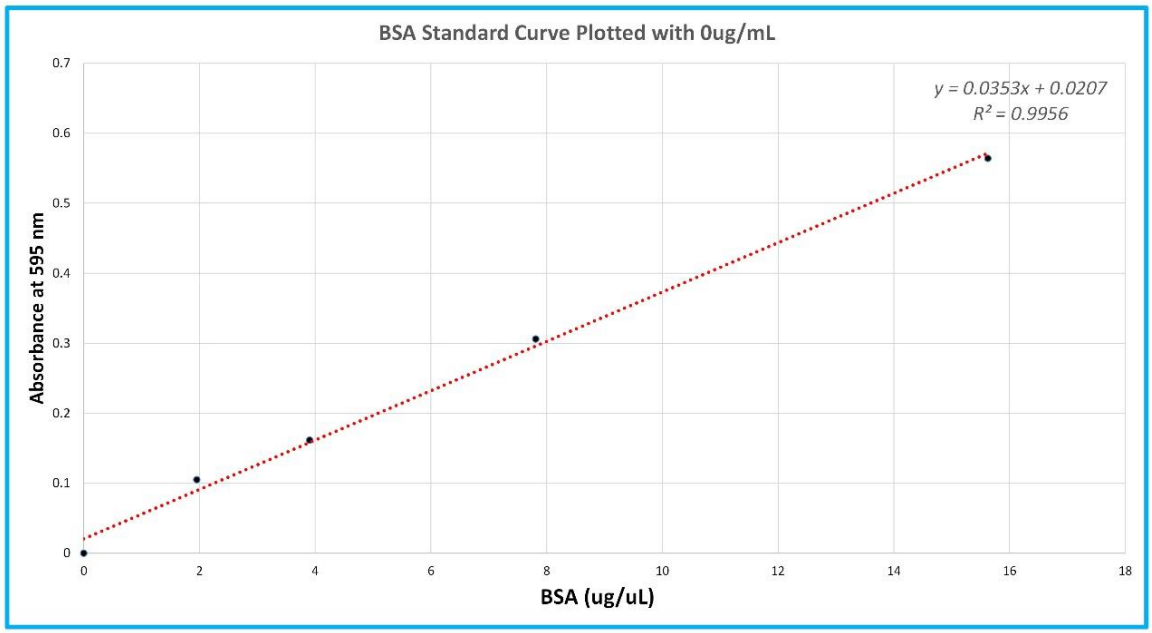


Figure 5. 1: Bovine serum albumin standard curve

The alignment of IDA spectral data to SWATH data and the quantitative analyses of validated and unique peptide revealed a list of DEPs (n=370) that could be related to the differential LVI status. As the entry names for the identified list of DEPs were imported as a UniProtKB entry, we renamed them into gene nomenclature format, then lined them up against the DEGs of the LVI-NT1154 cluster that were verified from the microarray experiment previously (see Chapter 3, section 3.3.1) Both lists of proteins and protein-coding genes were aligned against each other, and an overlying cluster of LVI differentially expressed molecules (LVI-DEM, n=56) was obtained. The flow of genetic information of the identified LVI-DEM could be translated into proteins in the specimens of this study. This promising and distinctive cluster comprised the following protein-encoding genes: *ACTB*, *ACTG1*, *ANP32A*, *ARHGDIA*, *ARPC5*, *CCT8*, *COL6A2*, *COMT*, *DBI*, *DDX17*, *DYNLRB1*, *EIF4E*, *ERH*, *FBLN1*, *H2AFY*, *HMGB2*, *HNRNPA2B1*, *HNRNPH3*, *HNRNPK*, *HSP90AA1*, *HSPA4*, *ILF3*, *KCTD12*, *LGALS3*, *LMNA*, *OLA1*, *PGAM1*, *PPIA*, *PRRC1*, *PSMA6*, *PTGES3*, *PTMA*, *RAC1*, *RBBP4*, *RPL10A*, *RPL7*, *RPL9*, *RPLP1*, *RPS11*, *RPS18*, *RPS2*, *RPS28*, *RPS4X*, *RPS7*, *RPSA*, *SAR1A*, *SEC61B*, *SNRPE*, *SRSF3*, *SUB1*, *TPI1*, *TUBB6*, *UFM1*, *UPF1*, *VDAC2*, and *YWHAE* (Table 5.1)

Table 5. 1: The list of the LVI-differentially expressed and overlapped genes/proteins produced by RNA microarray and HPLC-MS protocols, (n=56), colour code: Green= low rank, Red= high rank.

LVI-DEM Identifiers	LVI Status	N	Mean Rank	Sum of Ranks	p-value (u-test)
ACTB	Negative	80	83.1125	6649	0.550979529
	Positive	90	87.62222222	7886	
ACTG1	Negative	80	82.375	6590	0.435103584
	Positive	90	88.27777778	7945	
ANP32A	Negative	80	85.3625	6829	0.972604848
	Positive	90	85.62222222	7706	
ARHGDIA	Negative	80	74.075	5926	**0.004324532**
	Positive	90	95.65555556	8609	
ARPC5	Negative	80	80.4125	6433	0.203858618
	Positive	90	90.02222222	8102	
CCT8	Negative	80	79.575	6366	0.138925139
	Positive	90	90.76666667	8169	
COL6A2	Negative	80	86.7	6936	0.764400308
	Positive	90	84.43333333	7599	
COMT	Negative	80	85.5625	6845	0.987545701
	Positive	90	85.44444444	7690	
DBI	Negative	80	86.6125	6929	0.781124597
	Positive	90	84.51111111	7606	
DDX17	Negative	80	85.9125	6873	0.917943574
	Positive	90	85.13333333	7662	
DYNLRB1	Negative	80	76.03125	6082.5	**0.018035462**
	Positive	90	93.91666667	8452.5	
EIF4E	Negative	80	80.6125	6449	0.222205874
	Positive	90	89.84444444	8086	
ERH	Negative	80	86.85	6948	0.735988079
	Positive	90	84.3	7587	
FBLN1	Negative	80	96.86875	7749.5	**0.004519521**
	Positive	90	75.39444444	6785.5	
H2AFY	Negative	80	85.36875	6829.5	0.97384961
	Positive	90	85.61666667	7705.5	
HMGB2	Negative	80	80.6	6448	0.221025624
	Positive	90	89.85555556	8087	
HNRNPA2B1	Negative	80	85.25	6820	0.950213165
	Positive	90	85.72222222	7715	
HNRNPH3	Negative	80	85.0375	6803	0.908039055
	Positive	90	85.91111111	7732	
HNRNPK	Negative	80	88.175	7054	0.504071116
	Positive	90	83.12222222	7481	
HSP90AA1	Negative	80	70.025	5602	**0.000111099**
	Positive	90	99.25555556	8933	
HSPA4	Negative	80	80.49375	6439.5	0.211174063
	Positive	90	89.95	8095.5	
ILF3	Negative	80	85.025	6802	0.905565167
	Positive	90	85.92222222	7733	
KCTD12	Negative	80	92.8125	7425	0.067798712
	Positive	90	79	7110	

LGALS3	Negative	80	80.9	6472	0.250606045
	Positive	90	89.58888889	8063	
LMNA	Negative	80	85.8	6864	0.940272902
	Positive	90	85.23333333	7671	
OLA1	Negative	80	86.5875	6927	0.78592187
	Positive	90	84.53333333	7608	
PGAM1	Negative	80	85.2	6816	0.940272902
	Positive	90	85.76666667	7719	
PPIA	Negative	80	77.55	6204	**0.047081988**
	Positive	90	92.56666667	8331	
PRRC1	Negative	80	81.6125	6529	0.331584191
	Positive	90	88.95555556	8006	
PSMA6	Negative	80	80.975	6478	0.258414339
	Positive	90	89.52222222	8057	
PTGES3	Negative	80	79.875	6390	0.160057346
	Positive	90	90.5	8145	
PTMA	Negative	80	80.65	6452	0.225773687
	Positive	90	89.81111111	8083	
RAC1	Negative	80	82.2	6576	0.409828475
	Positive	90	88.43333333	7959	
RBBP4	Negative	80	87.8875	7031	0.550979529
	Positive	90	83.37777778	7504	
RPL10A	Negative	80	93.5	7480	**0.045711269**
	Positive	90	78.38888889	7055	
RPL7	Negative	80	84.6	6768	0.822149963
	Positive	90	86.3	7767	
RPL9	Negative	80	88.775	7102	0.41338482
	Positive	90	82.58888889	7433	
RPLP1	Negative	80	85.8625	6869	0.927860788
	Positive	90	85.17777778	7666	
RPS11	Negative	80	89.825	7186	0.280055071
	Positive	90	81.65555556	7349	
RPS18	Negative	80	93.85	7508	**0.037027166**
	Positive	90	78.07777778	7027	
RPS2	Negative	80	89.225	7138	0.352194087
	Positive	90	82.18888889	7397	
RPS28	Negative	80	90.3125	7225	0.229382207
	Positive	90	81.22222222	7310	
RPS4X	Negative	80	90.0125	7201	0.259731942
	Positive	90	81.48888889	7334	
RPS7	Negative	80	85.475	6838	0.995018117
	Positive	90	85.52222222	7697	
RPSA	Negative	80	92.1125	7369	0.098633872
	Positive	90	79.62222222	7166	
SAR1A	Negative	80	86.8125	6945	0.743058778
	Positive	90	84.33333333	7590	
SEC61B	Negative	80	89.5875	7167	0.30731199
	Positive	90	81.86666667	7368	

SNRPE	Negative	80	85.35	6828	0.970115496
	Positive	90	85.63333333	7707	
SRSF3	Negative	80	92.9125	7433	0.064123662
	Positive	90	78.91111111	7102	
SUB1	Negative	80	92.0625	7365	0.101208018
	Positive	90	79.66666667	7170	
TPI1	Negative	80	79.9625	6397	0.166657186
	Positive	90	90.42222222	8138	
TUBB6	Negative	80	89.05	7124	0.375275385
	Positive	90	82.34444444	7411	
UFM1	Negative	80	86.35	6908	0.831878625
	Positive	90	84.74444444	7627	
UPF1	Negative	80	81.9375	6555	0.373596637
	Positive	90	88.66666667	7980	
VDAC2	Negative	80	76.825	6146	**0.030262683**
	Positive	90	93.21111111	8389	
YWHAE	Negative	80	87.5	7000	0.61741843
	Positive	90	83.72222222	7535	

5.3.2.1. *Biological drivers for LVI and their distribution of expression among the LVI-supervised patients*

Reliability of the association between LVI and the discovered LVI-DEM, mentioned in the previous section, was statistically assessed. First, the difference of means for the overlapping LVI-DEM was interrogated against LVI categories with the non-parametric test of Mann-Whitney (u-test), as each of the LVI-DEM signifies a dependent variable against independent subgroups of LVI. Differential upregulation of: 1) heat shock protein HSP90-α, ***HSP90AA1*** (p-value $\leq$0.0001); 2) Rho-GDP dissociation inhibitor-1, ***ARHGDIA*** (p-value $\leq$0.004); 3) voltage-dependent anion-selective channel protein-2, ***VDAC2*** (p-value $\leq$0.030); 4) dynein light chain roadblock-type-1, ***DYNLRB1*** (p-value $\leq$0.018); and 5) peptidyl-prolyl cis-trans isomerase-A, ***PPIA*** (p-value $\leq$0.047) were found in patients with definite LVI profiles. On the other hand, differential upregulation of: 1) Fibulin-1, ***FBLN1*** (p-value $\leq$0.005); 2) 40S ribosomal protein S18, ***RPS18*** (p-value $\leq$0.037); and, 3) 60S ribosomal protein L10a, ***RPL10A*** (p-value $\leq$0.046) were found in patients with LVI^{-}/LN^{-}. Utilising the accurate categorisation of LVI status in the $LVI^{+/-}$ cohort of patients, non-parametric statistics on LVI-DEM revealed significant associations with promising candidates as LVI drivers (n=8). The label for these eight LVI-DEMs is henceforth T8. The T8 drivers of LVI were stratified into two sets: LVI^{+} set (n=5) labelled in red; and LVI^{-}/LN^{-} set (n=3) labelled in green (Table 5.2).

To approve the discriminability of the LVI subgroups is essentially dependent on the expression level of the established list of T8; we categorised the transcriptomic data for each LVI-DEM of the T8 list around their statistical median cut-off points. The segregation of the transcriptomic data between the precisely defined LVI categories is critical to discriminate any coincidental statistical association with LVI, especially for patients around the marginal zones of mRNA cut-off points. Interestingly, the positive associations between LVI and the expression levels of *HSP90AA1* and *VDAC2* were increased: (p-value $\leq$1.7E-5) and (p-value $\leq$0.006), individually. *ARHGDIA* kept a trend toward statistical significance (p-value $\leq$0.065), and its positive expression was associated with 51 patients (56.7%) with LVI definite profiles. Variables of *PPIA* and *DYNLRB1* showed no statistical associations with LVI upon categorisation. On the other hand, *FBLN1* was the only molecule that reserved, almost, the same statistical value in both the u-test (p-value $\leq$0.005) and chi-square statistics (p-value $\leq$0.006). *RPS18* and *RPL10A* lost their statistical significance with LVI (Table 5.2). The predicted functional association networks between the LVI-differentially expressed T8 protiens/genes that are identified with RNA microarray and HPLC/MS experiments are illustrated on page 417, Appendix D.

Table 5. 2: Statistical associations of the T8 list of LVI-DEM in the LVI-supervised cohort of patients, (n=170)

T8 identifiers	LVI status	N	Mean Rank	Sum of Ranks	p-value (u-test)	p-value (chi-square)
HSP90AA1	Negative	80	70.025	5602	0.0001	1.7E-5
	Positive	90	99.25555556	8933		
ARHGDIA	Negative	80	74.075	5926	0.004	0.065
	Positive	90	95.65555556	8609		
DYNLRB1	Negative	80	76.03125	6082.5	0.018	0.357
	Positive	90	93.91666667	8452.5		
VDAC2	Negative	80	76.825	6146	0.030	0.006
	Positive	90	93.21111111	8389		
PPIA	Negative	80	77.55	6204	0.047	0.124
	Positive	90	92.56666667	8331		
FBLN1	Negative	80	96.86875	7749.5	0.005	0.006
	Positive	90	75.39444444	6785.5		
RPS18	Negative	80	93.85	7508	0.037	0.219
	Positive	90	78.07777778	7027		
RPL10A	Negative	80	93.5	7480	0.046	0.124
	Positive	90	78.38888889	7055		

5.3.3. Prognostic importance and molecular interactions of T8 in invasive BC patients

All tables and figures are placed at the end of this section.

5.3.3.1. *HSP90AA1*

5.3.3.1.1. Clinicopathological impact of *HSP90AA1*

Upregulated expression of *HSP90AA1* was associated with definite LVI (p-value ≤1.7E-5; 59, 65.6%), and also low expression of *HSP90AA1* was associated with LVI^- patients (54, 67.5%) in the $LVI^{+/-}$ cohort. Higher levels of *HSP90AA1* in the same cohort were associated with unfavourable prognostic profiles as: larger tumour size (p-value ≤5.0E-5; 48, 68.6%), higher tumour grade (p-value ≤2.9E-5; 55, 62.5%); while 23 (88.5%) patients with tumour grade-I had low expression of *HSP90AA1*, positive nodal metastasis, combined stage-II/stage-III (p-value ≤0.001; 43, 66.2%), and poor NPI (p-value ≤0.004) (Table 5.3). In the METABRIC cohort, the *HSP90AA1*$^-$ gene expression profile was associated with 66.0% of patients with LVI^- (p-value ≤0.009). However, an identical number of LVI^+ patients in the LVI^+/LVI^- and LN^{mixed} sub-cohort was split equally between both *HSP90AA1* gene expression positive and negative categories. The remaining clinicopathological statistics were concordant with their equivalents from the $LVI^{+/-}$ cohort: larger tumour size (p-value ≤0.017; 588, 53.4%), higher tumour grade (p-value ≤6.4E-10; 554, 58.2%), positive axillary LN metastasis, stage-II and stage-III (p-value ≤0.010; 56, 53.9%), and poor NPI (p-value ≤0.009; 111, 55.8%) (Table 5.4).

According to KM estimates of the BCSS for patients of the $LVI^{+/-}$ cohort, the expression of *HSP90AA1* high level was correlated with BC-related death (p-value ≤0.023; 31, 36.90%). The results of the BCSS estimation in the METABRIC cohort confirmed the poor prognosis indications of *HSP90AA1* (p-value ≤0.00005; 294, 29.30%). The worse outcome in BC patients was associated with upregulated expression of *HSP90AA1* (p-value ≤0.0001), disregarding the molecular subtypes (Figure 5.2).

The Cox regression model of multivariate analysis involving categorical *HSP90AA1* levels of expression was applied to the larger METRABRIC cohort, exclusively, against clinicopathological parameters found to be significant in the univariate analysis of BC-specific survival. Higher levels of *HSP90AA1* had an HR of 1.22, and this suggests that upregulated *HSP90AA1* gene expression is an independent predictor of BC progression. Prognostic independence of *HSP90AA1* is backed up with an external pooled cohort of BC patients, with an HR of 1.37 at 95% CI (Table 5.15).

5.3.3.1.2. Distribution of *HSP90AA1* expression and molecular classification of invasive BC

Intrinsic molecular subtyping by PAM50 showed a positive association between *HSP90AA1* expression and patients with luminal-B, HER2-enriched, and basal biological intrinsic subtypes (p-value <0.0026). In contrast, the lowest expression level of *HSP90AA1* was associated with patients of the normal-like and luminal-A intrinsic subtypes. The results of the PAM50 classifier were on the same trend for the large cohort of METABRIC (p-value ≤1.85E-21). Genetic modification in invasive BC is diverse and requires meticulous molecular classification that considers germline variants, such as CNVs and single nucleotide polymorphism (SNPs), and somatic aberrations represented by CNAs. Therefore, we analysed the relationship between *HSP90AA1* and each of the ten IntClusts. The statistical difference of *HSP90AA1* expression stratified the METABRIC's ten integrating clusters into two notable sects. Upregulated expression of *HSP90AA1* was common in IntClust-1, IntClust-5, and IntClust-9 for the $LVI^{+/-}$ cohort (p-value ≤0.003) and also the METABRIC cohort (p-value ≤1.57E-21), with a preferential upregulation for IntClust-10 in the latter cohort. Expression profiles for *HSP90AA1* of the remaining clustered patients were generally low (Table 5.5).

5.3.3.1.3. Expression correlations between *HSP90AA1* altered expression and a cluster of transcripts for matrix metalloproteinases

The co-expression associations between MMP encoding genes (n=20) and *HSP90AA1* were assessed in the $LVI^{+/-}$ cohort of patients. Positive expression was identified between *HSP90AA1* and *MMP10* (r=0.151, p-value ≤0.049) in particular. Two opposing relationships of gene expression for *HSP90AA1* were found to be associated with *MMP2* (r=-0.268, p-value ≤0.0004) and *MMP28* (r=-0.168, p-value ≤0.029). No other gene expression associations were identified for *HSP90AA1* in this cohort (Table 5.9).

Nine expression correlations were identified between MMP encoding genes and *HSP90AA1* in the METABRIC cohort. Positive expression correlations for *HSP90AA1* were interrelated with *MMP1* (r=0.073, p-value ≤0.001); *MMP10* (r=0.100, p-value ≤8.27E-06); *MMP13* (r=0.162, p-value ≤3.7E-13); and *MMP19* (r=0.075, p-value ≤0.001). In contrast, the negative expression relationships were linked to *MMP2* (r=-0.071, p-value ≤0.002); *MMP15* (r=-0.171, p-value ≤1.7E-14); *MMP21* (r=-0.151, p-value ≤1.4E-11); *MMP24* (r=-0.053, p-value ≤0.018); and *MMP25* (r=-0.051, p-value ≤0.024) (Table 5.10).

5.3.3.1.4. Expression correlations between *HSP90AA1* altered expression and a group of transcripts for a cluster of differentiation of lymphocytes

A positive correlation of expression was found between *HSP90AA1* and *CD80* in the LVI$^{+/-}$ cohort (r=0.167, p-value ≤0.030). No inverted expression relationships were identified for *HSP90AA1* against other lymphocyte CDs in the same groups of patients (Table 5.11).

The positive correlations of *HSP90AA1* expression with lymphocytes' antigens was expanded in the METABRIC cohort. The expression of *HSP90AA1* was positively correlated with *CD80* (r=0.120, p-value ≤9.0E-08); *CD45* (r=0.089, p-value ≤0.0001); and *CD269* (r=0.061, p-value ≤0.007). The negative correlations of expression between *HSP90AA1* and other CDs were *CD3G* (r=-0.047, p-value ≤0.036); *CD3E* (r=-0.182, p-value ≤3.5E-16); *CD3Z* (r=-0.005, p-value ≤0.007); *CD4* (r=-0.164, p-value ≤2.1E-13); *CD28* (r=-0.076, p-value ≤0.001); *CD19* (r=-0.089, p-value ≤0.007); *CD22* (r=-0.210, p-value ≤3.8E-21); *CD40* (r=-0.080, p-value ≤0.0004); *CD267* (r=-0.127, p-value ≤1.3E-08); *CD268* (r=-0.141, p-value ≤3.1E-10) (Table 5.12).

5.3.3.1.5. Expression correlations between *HSP90AA1* altered expression and a cluster of transcripts for invasion-related biomarkers

The expression of *HSP90AA1* in the LVI$^{+/-}$ cohort of patients showed specific interactions with the expression of seven invasion-related transcripts. *HSP90AA1* was expressed positively against the expression of *CTNNB1* (r=0.169, p-value ≤0.028); *MKI67* (r=0.204, p-value ≤0.0.008); *CDH1* (r=0.214, p-value ≤0.005); and *CDH2* (r=0.230, p-value ≤0.003). The negative correlations of *HSP90AA1* expression were against *CXCL12* (r=-0.195, p-value ≤0.011); *THY1* (r=-0.230, p-value ≤0.003); and *CAV1* (r=-0.165, p-value ≤0.031) (Table 5.13).

Larger expression interactions were identified for *HSP90AA1* in the METABRIC cohort in comparison to the LVI controlled cohort. The positively correlated expressions against *HSP90AA1* were *CTNNB1* (r=0.328, p-value ≤5.2E-51); *FN1* (r=0.050, p-value ≤0.025); *COL3A1* (r=0.150, p-value ≤1.7E-11); *SNAI2* (r=0.150, p-value ≤1.9E-11); *CDH1* (r=0.204, p-value ≤5.7E-20); *VIM* (r=0.259, p-value ≤1.1E-31); *CXCL12* (r=0.060, p-value ≤0.007); *MET* (r=0.087, p-value ≤0.0001); *PDGFC* (r=0.277, p-value ≤4.2E-36); *FAP* (r=0.164, p-value ≤2.5E-13); *CD24* (r=0.121, p-value ≤6.3E-08); *CD44* (r=0.133, p-value ≤2.7E-09); *S100A4* (r=0.137, p-value ≤1.0E-09); and *CAV1* (r=0.161, p-value ≤5.3E-13). On the other hand, only one negative expression correlation was found between *HSP90AA1* and *THY1* in the METABRIC cohort (r=-0.214, p-value ≤6.8e-22) (Table 5.14).

5.3.3.2. *VDAC2*

5.3.3.2.1. Clinicopathological impact of *VDAC2*

An increase in *VDAC2* expression was positively associated with LVI^{+} patients of the $LVI^{+/-}$ cohort (p-value ≤0.006; 54, 60.0%). Furthermore, this association was backed up with the positive association between the sum of positive axillary LNM patients and their upregulated profiles of *VDAC2* (p-value ≤0.018; 40, 61.5%), as of *HSP90AA1*. Levels of *VDAC2* were highly expressed in larger tumour cells of the same patients, statistically (p-value ≤0.005). In addition, patients with upregulated *VDAC2* were profiled with poorer NPI (p-value ≤0.021). (Table 5.3).

The gene expression of *VDAC2* against LVI was not significant among the patients with LVI data in the METABRIC cohort (p-value ≤0.109). Higher expression of *VDAC2* was progressively associated with the number of positive LNs of the axilla (p-value ≤0.007; [stage-I: 72, 42.1%], [stage-II: 331, 43.0%], [stage-III: 490, 51.5%]). Similar to patients of the $LVI^{+/-}$ cohort, larger tumour cells and poor NPI profile were correlated with highly expressed *VDAC2* (p-value ≤0.005) and (p-value ≤0.001). In particular, attenuated *VDAC2* expression was associated with grade-I, well-differentiated, and slow-growing tumour cells (p-value ≤0.001; 99, 57.9%) (Table 5.4).

Estimation of BCSS for patients with the different levels of *VDAC2* expression revealed a strong association between BC mortality and elevated expression of this gene. For $LVI^{+/-}$, METABRIC with LVI data, and the pooled external cohorts of invasive BC (n=5,352), high *VDAC2* estimated larger numbers of deceased patients: (p-value ≤0.014), (p-value ≤0.009), and (p-value ≤0.0001), respectively (Figure 5.3) . In the Cox regression model and against the independent prognostic parameters of BC, *VDAC2* expression appeared to be a dependently expressed molecule with an HR of 1.09 (Table 5.15).

5.3.3.2.2. Distribution of *VDAC2* expression and molecular classification of invasive BC

PAM50 classification for the patients of the $LVI^{+/-}$ cohort appointed upregulation in *VDAC2* expression to be associated with patients of the HER2-enriched and basal intrinsic subtypes. On the other hand, the lower expression of *VDAC2* was associated with luminal-A subtype patients (p-value ≤0.014). The highest percentage of patients with high *VDAC2* levels in the larger cohort, with no LVI strict conditions for sample selection, was linked to the luminal-B subtype, followed by basal (p-value ≤7.8E-18). Patients with normal-like and luminal-A intrinsic subtypes accounted for the majority of patients with low-expression

profiles of *VDAC2*. No association between the integrated genomic clusters of the METABRIC and *VDAC2* was verified in the LVI$^{+/-}$ cohort. However, upregulation of *VDAC2* was observed in approximately two-thirds of IntClust-1, IntClust-5, and IntClust-10 in the larger cohort. In addition, the minimum level of *VDAC2* expression was associated with IntClust-7 (p-value ≤2.5E-14) (Table 5.5).

5.3.3.2.3. Expression correlations between *VDAC2* altered expression and a cluster of transcripts for matrix metalloproteinases

Positive expression correlations were found between *VDAC2* and *MMP1* (*r*=0.193, p-value ≤0.012); *MMP12* (*r*=0.214, p-value ≤0.005); and *MMP24* (*r*=0.155, p-value ≤0.043) in the LVI$^{+/-}$ cohort. *MMP2* (*r*=-0.238, p-value ≤0.002) and *MMP27* (*r*=-0.165, p-value ≤0.031) were negatively associated with the expression of *VDAC2* (Table 5.9).

In the METABRIC cohort, a total of 12 MMP encoding genes were statistically associated with the expression of *VDAC2*. Positive correlations of *VDAC2* expression were with *MMP1* (*r*=0.096, p-value ≤0.00002); *MMP12* (*r*=0.136, p-value ≤1.3E-09); *MMP14* (*r*=0.055, p-value ≤0.015); *MMP16* (*r*=0.082, p-value ≤0.0002). *MMP20* (*r*=0.048, p-value ≤0.034); and *MMP24* (*r*=0.047, p-value ≤0.037). Negative expression correlations for *VDAC2* were related to *MMP2* (*r*=-0.142, p-value ≤1.99E-10); *MMP11* (*r*=-0.085, p-value ≤0.0001); *MMP15* (*r*=-0.068, p-value ≤0.002); *MMP21* (*r*=-0.063, p-value ≤0.005); *MMP25* (*r*=-0.079, p-value ≤0.0004); and *MMP28* (*r*=-0.081, p-value ≤0.0003) (Table 5.10).

5.3.3.2.4. Expression correlations between *VDAC2* altered expression and a group of transcripts for cluster of differentiation of lymphocytes

No positive correlations for *VDAC2* expression were found in the LVI$^{+/-}$ cohort against any tested marker of lymphocytes. All of the relationships for *VDAC2* expression were in the negative direction. The expression of *VDAC2* was negatively associated with the expression of *CD3E* (*r*=-0.220, p-value ≤0.004); *CD4* (*r*=-0.176, p-value ≤0.022); *CD8A* (*r*=-0.161, p-value ≤0.036); *CD28* (*r*=-0.157, p-value ≤0.04); *CD86* (*r*=-0.165, p-value ≤0.031); *CD22* (*r*=-0.205, p-value ≤0.007); *CD52* (*r*=-0.154, p-value ≤0.045); and *CD267* (*r*=-0.202, p-value ≤0.008) (Table 5.11).

The positive correlations of *VDAC2* expression were exclusive with *CD80* (*r*=0.058, p-value ≤0.01) in the METABRIC cohort. The remaining expression correlations were opposing the direction of *VDAC2* expression in this cohort. *VDAC2* expression was negatively associated with *CD2* (*r*=-0.044, p-value ≤0.049); *CD3G* (*r*=-0.101, p-value ≤6.1E-06); *CD3D* (*r*=-0.045, p-value ≤0.046); *CD3E* (*r*=-

0.178, p-value ≤1.4E-15); *CD3Z* (*r*=-0.083, p-value ≤0.0002); *CD4* (*r*=-0.099, p-value ≤9.8E-06); *CD8A* (*r*=-0.054, p-value ≤0.016); *CD*28 (*r*=-0.129, p-value ≤8.0E-09); *CD86* (*r*=-0.066, p-value ≤0.004); *CD19* (*r*=-0.108, p-value ≤1.4E-06); *CD22* (*r*=-0.151, p-value ≤1.3E-11); *CD40* (*r*=-0.072, p-value ≤0.001); *CD267* (*r*=-0.147, p-value ≤4.6E-11); and *CD268* (*r*=-0.052, p-value ≤0.020) (Table 5.12).

5.3.3.2.5. Expression correlations between *VDAC2* altered expression and a cluster of transcripts for invasion-related biomarkers

The positive expression associations of *VDAC2* were limited to the expression of *MKI67* (*r*=0.297, p-value ≤0.0001) and *CDH2* (*r*=0.200, p-value ≤0.009) for the $LVI^{+/-}$ patients. The negative expression correlations of *VDAC2* were against the expression of *CXCL12* (*r*=-0.204, p-value ≤0.008); *THY1* (*r*=-0.238, p-value ≤0.002); and *CD267* (*r*=-0.158, p-value ≤0.039) (Table 5.13).

In the METABRIC cohort, the identified positive correlations of *VDAC2* expression were against the expression of *CTNNB1* (*r*=0.184, p-value ≤1.7E-16); *ETS1* (*r*=0.174, p-value ≤0.022); *MKI67* (*r*=0.174, p-value ≤7.1E-15); *CDH1* (*r*=0.073, p-value ≤0.001); *CDH2* (*r*=0.070, p-value ≤0.002); *VIM* (*r*=0.125, p-value ≤2.5E-08); *MET* (*r*=0.143, p-value ≤1.5E-10); *PDGFC* (*r*=0.120, p-value ≤8.5E-08); *CD44* (*r*=0.065, p-value ≤0.004); and *S100A4* (*r*=0.073, p-value ≤0.001). The inverted expression correlations for *VDAC2* were against the expression of *CXCL12* (*r*=-0.085, p-value ≤0.0001); *FAP* (*r*=-0.056, p-value ≤0.012); and *THY1* (*r*=-0.204, p-value ≤4.6E-20) (Table 5.14).

5.3.3.3. *ARHGDIA*

5.3.3.3.1. Clinicopathological impact of *ARHGDIA*

As a continuous variable, *ARHGDIA* showed strong association with histological LVI (p-value ≤0.004). However, this statistical correlation was reduced then lost on categorisation of LVI in both $LVI^{+/-}$ (p-value ≤0.065) and the larger METABRIC (p-value ≤0.105) cohorts. As it is not associated with LVI, no associations were specified for alterations in *ARHGDIA* expression versus LNM (p-value ≤0.164) and (p-value ≤0.477) in the latter cohorts, accordingly. Marginal association with tumour grade was found in the $LVI^{+/-}$ patients (p-value ≤0.044), in comparison to the very strong association found in the larger METABRIC cohort (p-value ≤0.000001). Both cohorts showed similar correlation with poor NPI (p-value ≤0.028) and (p-value ≤0.021), consequently. Smaller tumour size was associated with low *ARHGDIA* expression in the larger cohort only (p-value ≤0.008) (Table 5.3) for $LVI^{+/-}$ cohort, and (Table 5.4) for the METABRIC's.

Analysis of BCSS against *ARHGDIA* expression in the LVI$^{+/-}$ cohort revealed a trend toward significance (p-value ≤0.071). Nonetheless, *ARHGDIA* upregulation was associated with a greater number of BC-related deaths in the METABRIC cohort (p-value ≤0.001). The external and pooled cohort of ER^{+} invasive BC patients (n=3,840) showed a significant correlation of higher *ARHGDIA* with mortality (p-value ≤0.036) (Figure 5.4). The results of proportional hazards regression analyses showed independent prognostic implication for *ARHGDIA* when it was positively expressed in the METABRIC cohort of patients (p-value ≤0.019, HR =1.24). The independent prognostic feature for ARHGDIA was also detected in the external and pooled cohort of ER^{+} invasive BC patients (HR =1.13) (Table 5.15).

5.3.3.3.2. Distribution of *ARHGDIA* expression and molecular classification of invasive BC

The expression of *ARHGDIA* in the LVI$^{+/-}$ cohort was not statistically associated with any molecular intrinsic subtypes according to PAM50 (p-value ≤0.066) or even with integrating clusters of METABRIC (p-value ≤0.067). On the contrary, cohort expansion affected the molecular stratification of patients based on *ARHGDIA* expression levels. Patients with HER2-enriched subtype (p-value ≤5.0E-10; 148, 61.7%) showed positive values for *ARHGDIA*. The low expression level of *ARHGDIA* was associated with the luminal-A subtype. Moreover, genetic regulation of *ARHGDIA* was statistically correlated with IntClust (p-value ≤6.5E-18). Typically, patients with higher expression of *ARHGDIA* were associated with IntClust-1, IntClust-5, and IntClust-9, whereas patients with lower expression of *ARHGDIA* were associated with IntClust-3 and IntClust-8 (Table 5.5).

5.3.3.3.3. Expression correlations between *ARHGDIA* altered expression and a cluster of transcripts for matrix metalloproteinases

Two positive and two negative expression relationships were found between *ARHGDIA* and MMP encoding genes in the LVI$^{+/-}$ patients. Positive associations for *ARHGDIA* expression were connected with *MMP9* (r=0.355, p-value ≤2.1E-06) and *MMP12* (r=0.199, p-value ≤0.009), whereas the negative associations were linked to *MMP3* (r=-0.170, p-value ≤0.027) and *MMP27* (r=-0.175, p-value ≤0.022) (Table 5.9).

Broader correlations with MMP encoding genes were identified for *ARHGDIA* in the METABRIC cohort. Positive correlations were identified between *ARHGDIA* and *MMP1* (r=0.119, p-value ≤1.18E-07); *MMP2* (r=0.055, p-value ≤0.015); *MMP9* (r=0.233, p-value ≤6.65E-26); *MMP11* (r=0.181, p-value ≤4.4E-16); *MMP12* (r=0.076, p-value ≤0.001); *MMP15* (r=0.157, p-value ≤2.4E-12); *MMP16*

(*r*=0.083, p-value ≤0.0002); *MMP24* (*r*=0.049, p-value ≤0.029); and *MMP25* (*r*=0.048, p-value ≤0.034). The negative associations were associated with *MMP3* (*r*=-0.047, p-value ≤0.037); *MMP19* (*r*=-0.062, p-value ≤0.005); *MMP27* (*r*=-0.097, p-value ≤0.00001); and *MMP28* (*r*=-0.122, p-value ≤5.2E-08) (Table 5.10).

5.3.3.3.4. Expression correlations between *ARHGDIA* altered expression and a group of transcripts for cluster of differentiation of lymphocytes

The expression correlation between *ARHGDIA* and *CD4* showed a positive direction in the LVI$^{+/-}$ cohort (*r*=0.178, p-value ≤0.020). The rest of the correlations were in the negative direction of expression. *ARHGDIA* was negatively expressed against *CD3G* (*r*=-0.170, p-value ≤0.027); *CD3D* (*r*=-0.174, p-value ≤0.023); *CD8A* (*r*=-0.135, p-value ≤0.080); *CD52* (*r*=-0.151, p-value ≤0.049); *CD267* (*r*=-0.178, p-value ≤0.020); and *CD269* (*r*=-0.201, p-value ≤0.009) (Table 5.11).

In the METABRIC cohort, *ARHGDIA* was positively expressed against *CD3E* (*r*=0.114, p-value ≤3.8E-07); *CD4* (*r*=0.198, p-value ≤7.0E-19); *CD80* (*r*=0.086, p-value ≤0.0001); *CD22* (*r*=0.101, p-value ≤7.3E-06); *CD40* (*r*=0.044, p-value ≤0.049); and *CD268* (*r*=0.102, p-value ≤5.6E-06). The negative correlations for the expression of *ARHGDIA* against the other lymphocyte antigens were *CD8B* (*r*=-0.058, p-value ≤0.010); *CD45* (*r*=-0.050, p-value ≤0.025); *CD52* (*r*=-0.057, p-value ≤0.011); and *CD269* (*r*=-0.072, p-value ≤0.001) (Table 5.12).

5.3.3.3.5. Expression correlations between *ARHGDIA* altered expression and a cluster of transcripts for invasion-related biomarkers

Positive expression of *ARHGDIA* was statistically correlated to *FN1* (*r*=0.194, p-value ≤0.011); *MKI67* (*r*=0.318, p-value ≤0.00002); and *CDH1* (*r*=0.196, p-value ≤0.011) in the LVI$^{+/-}$ cohort. The negative associations of expression for *ARHGDIA* were against the expression of *CXCL12* (*r*=-0.194, p-value ≤0.011); *TGFB1* (*r*=-0.152, p-value ≤0.049); and *CAV1* (*r*=-0.227, p-value ≤0.003) (Table 5.13).

Additional expression relationships were identified for *ARHGDIA* in the METABRIC cohort. Positive associations of *ARHGDIA*'s expression were against the expression of *SDC1* (*r*=0.110, p-value ≤9.9E-07); *FN1* (*r*=0.078, p-value ≤0.001); *MKI67* (*r*=0.160, p-value ≤8.1E-13); *CDH2* (*r*=0.105, p-value ≤2.6E-06); and *THY1* (*r*=0.251, p-value ≤7.9E-30). The negative expression associations for *ARHGDIA* were as opposed to the expression of *CTNNB1* (*r*=-0.209, p-value ≤5.5E-21); *VIM* (*r*=-0.065, p-value ≤0.004); *CXCL12* (*r*=-0.123, p-value ≤3.8E-08); *PDGFC* (*r*=-0.145, p-value ≤8.2E-11); *FAP* (*r*=-0.059, p-value ≤0.009); *CD44* (*r*=-0.054, p-value ≤0.015); and *CAV1* (*r*=-0.168, p-value ≤4.7E-14) (Table 5.14).

5.3.3.4. *DYNLRB1*

5.3.3.4.1. Clinicopathological impact of *DYNLRB1*

DYNLRB1 was one of the differentially expressed genes in microarray analyses and also one of the detected upregulated proteins in LVI^{+} patients according to the HPLC-MS experiment. Non-parametric statistics displayed significant impact for higher levels of *DYNLRB1* in patients with definite LVI status (p-value ≤0.018). Nonetheless, no clinicopathological power was perceived for *DYNLRB1* in both categorised LVI$^{+/-}$ and LVI patients of the METABRIC cohorts (Tables 5.3 and 5.4). Additionally, the KM curves of the BCSS estimates showed no impacts for *DYNLRB1* on mortality risk in both cohorts (p-value ≤0.958) and (p-value ≤0.517), in the same order (Figure 5.5) *DYNLRB1* did not exhibit prognostic independence features in the Cox regression model (p-value ≤0.454) (Table 5.15).

5.3.3.4.2. Distribution of *DYNLRB1* expression and molecular classification of invasive BC

The upregulation of *DYNLRB1* was most likely associated with the luminal-B intrinsic subtypes in both cohorts, with (p-value ≤2.0E-5) and (p-value ≤0.009), accordingly. More than 60% of IntClust-1, IntClust-2, IntClust-6, and IntClust-9 in the METABRIC cohort exhibited a relevant upregulation of *DYNLRB1* (p-value ≤2.5E-20). On the other hand, IntClust-1, IntClust-2, and IntClust-7 showed high levels of *DYNLRB1* in the LVI+/- patients (Table 5.5).

5.3.3.4.3. Expression correlations between *DYNLRB1* altered expression and a cluster of transcripts for matrix metalloproteinases

A negative and single association was found between *DYNLRB1* and *MMP2* (r=-0.250, p-value ≤0.001) in the LVI$^{+/-}$ cohort of patients. No positive correlations were identified in this cohort (Table 5.9). On the other hand, a wider spectrum of statistical correlations for the expression of *DYNLRB1* against MMP encoding genes was found in the METABRIC cohort. The positive direction of *DYNLRB1* expression was linked to *MMP1* (r=0.065, p-value ≤0.004); *MMP11* (r=0.051, p-value ≤0.024); *MMP12* (r=0.47, p-value ≤0.037); *MMP16* (r=0.097, p-value ≤0.00001); and *MMP24* (r=0.075, p-value ≤0.001). Negative expression correlations for *DYNLRB1* were associated with seven MMP encoding genes: *MMP2* (r=-0.145, p-value ≤8.7E-11); *MMP3* (r=-0.119, p-value ≤11.05E-07); *MMP9* (r=-0.053, p-value ≤0.018); *MMP15* (r=-0.085, p-value ≤0.0001); *MMP20* (r=-0.056, p-value ≤0.013); *MMP21* (r=-0.085, p-value ≤0.0001); and *MMP28* (r=-0.076, p-value ≤0.001) (Table 5.10).

5.3.3.4.4. Expression correlations between *DYNLRB1* altered expression and a group of transcripts for cluster of differentiation of lymphocytes

All expression associations for *DYNLRB1* against the expression of all inspected CDs were negative in the LVI$^{+/-}$ cohort. The direction of expression for *DYNLRB1* was against the expression of *CD2* (*r*=-0.293, p-value ≤0.0001); *CD3G* (*r*=-0.350, p-value ≤2.9E-06); *CD3D* (*r*=-0.333, p-value ≤9.0E-06); *CD3E* (*r*=-0.296, p-value ≤0.0001); *CD3Z* (*r*=-0.319, p-value ≤0.00002); *CD4* (*r*=-0.156, p-value ≤0.042); *CD8A* (*r*=-0.352, p-value ≤2.5E-06); *CD28* (*r*=-0.196, p-value ≤0.01); *CD86* (*r*=-0.307, p-value ≤0.00005); *CD19* (*r*=-0.258, p-value ≤0.001); *CD45* (*r*=-0.199, p-value ≤0.009); *CD52* (*r*=-0.332, p-value ≤9.8E-06); *CD40* (*r*=-0.250, p-value ≤0.001); *CD267* (*r*=-0.204, p-value ≤0.008); *CD268* (*r*=-0.215, p-value ≤0.005); and *CD269* (*r*=-0.238, p-value ≤0.002) (Table 5.11).

The negative statistical trend for the expression associations between *DYNLRB1* and CDs persisted in the METABRIC cohort. *DYNLRB1* was negatively correlated with the expression of *CD2* (*r*=-0.137, p-value ≤8.3E-10); *CD3G* (*r*=-0.210, p-value ≤3.2E-21); *CD3D* (*r*=-0.141, p-value ≤3.2E-10); *CD3E* (*r*=-0.190, p-value ≤1.6E-17); *CD3Z* (*r*=-0.173, p-value ≤8.7E-15); *CD4* (*r*=-0.083, p-value ≤0.0002); *CD8A* (*r*=-0.158, p-value ≤1.6E-12); *CD8B* (*r*=-0.106, p-value ≤2.1E-06); *CD28* (*r*=-0.155, p-value ≤4.0E-12); *CD86* (*r*=-0.171, p-value ≤2.0E-14); *CD19* (*r*=-0.162, p-value ≤3.6E-13); *CD45* (*r*=-0.110, p-value ≤9.9E-07); *CD52* (*r*=-0.130, p-value ≤7.2E-09); *CD40* (*r*=-0.192, p-value ≤6.8E-18); *CD267* (*r*=-0.160, p-value ≤8.1E-13); *CD268* (*r*=-0.051, p-value ≤0.023); and *CD269* (*r*=-0.102, p-value ≤5.9E-06) (Table 5.12).

5.3.3.4.5. Expression correlations between *DYNLRB1* altered expression and a cluster of transcripts for invasion-related biomarkers

DYNLRB1 positive expression correlations were specific to *CTNNB1* (*r*=0.192, p-value ≤0.012); and *CDH1* (*r*=0.230, p-value ≤0.003) in the LVI$^{+/-}$ cohort. The negative expression of *DYNLRB1* was associated with the expression of *THY1* (*r*=-0.151, p-value ≤0.049); *CD44* (*r*=-0.197, p-value ≤0.010); and *CAV1* (*r*=-0.166, p-value ≤0.031) (Table 5.13).

The trend of expression relationships for *DYNLRB1* in the METABRIC cohort were in accordance with the expression of *ETS1* (*r*=0.212, p-value ≤1.4E-21); *CDH1* (*r*=0.156, p-value ≤2.9E-12); *MET* (*r*=0.094, p-value ≤0.00003); and *PDGFC* (*r*=0.151, p-value ≤1.3E-11). The opposite expression correlations for *DYNLRB1* were against the expression of *SNAI2* (*r*=-0.088, p-value ≤0.0008); *CXCL12* (*r*=-0.144, p-value ≤1.3E-10); *FAP* (*r*=-0.074, p-value ≤0.001); *THY1* (*r*=-0.096, p-

value ≤0.00001); *CD24* (*r*=-0.089, p-value ≤0.00007); *CD44* (*r*=-0.075, p-value ≤0.001); and *CAV1* (*r*=-0.093, p-value ≤0.00003) (Table 5.14).

5.3.3.5. *FBLN1*

5.3.3.5.1. Clinicopathological impact of *FBLN1*

Strong favourable associations were specified for upregulated *FBLN1* in invasive BC patients. For *HSP90AA1*, statistical correlations between continuous and categorised *FBLN1* expression data and LVI status were consistent in both the LVI$^{+/-}$ and METABRIC cohorts. Under LVI$^{+/-}$ conditions, upregulation of *FBLN1* accompanied patients with LVI negative and no LNM (p-value ≤0.006; 49, 61.3%). Additionally, patients in the category of definite LVI were commonly linked to low levels of *FBLN1* in the same cohort. Lack of genetic expression of *FBLN1* was strongly correlated with higher tumour grades and vice versa (p-value ≤0.00017). Positive LN was also observed in patients with lower levels of *FBLN1* (p-value ≤0.028). Consequently, downregulation of *FBLN1* classified patients with poor NPI (p-value ≤8.0E-5) (Table 5.6).

On the other hand, the association with LVI in the METABRIC cohort was debilitated. A trend toward significance was calculated for the association between higher *FBLN1* and negative LVI in this cohort (p-value ≤0.09). However, in the LVI^{+} patients, the percentage of patients with *FBLN1*$^{+/-}$ expressions was relatively equal. Younger patients exhibited upregulation in *FBLN1* in comparison to older ones (p-value ≤0.001). Smaller tumour cells were observed more often in patients with upregulated genetic expression of *FBLN1* (p-value ≤0.001). Elevated levels of *FBLN1* gene expression was found in grade-I tumours (p-value ≤4.9E-14). Altered expression of *FBLN1* had no associations with LNM in the METABRIC cohort. All nodal involvements stages showed similar expression profiles for each *FBLN1* category (p-value ≤0.299), in contrast to the results of the LVI$^{+/-}$ patients. Finally, patients with upregulated *FBLN1* were allocated among the NPI poor prognostic group (p-value ≤1.0E-6) (Table 5.7).

KM estimates for BCSS disclosed *FBLN1* upregulation with better survival. In the LVI$^{+/-}$ cohort, highly expressed *FBLN1* was associated with a favourable outcome (p-value ≤0.002), as well as in the METABRIC (p-value ≤0.002). The longer survival for patients with upregulated *FBLN1* was confirmed statistically to an external large cohort of invasive BC (n=5,352) (p-value ≤0.0001) with HR of 0.75 at 95% CI (Figure 5.6). *FBLN1* was not interpreted as a prognostic biomarker with independent characteristics against independent prognostic parameters of BC (p-value ≤0.102) (Table 5.15).

5.3.3.5.2. Distribution of *FBLN1* expression and molecular classification of invasive BC

The PAM50 molecular classifier showed a strong bond between highly expressed *FBLN1* and the normal-like and luminal-A intrinsic subtypes. In contrast, low genetic regulation of *FBLN1* stamped patients with a basal subtype (p-value ≤8.2E-7). Distribution of the altered expression of *FBLN1* among the genomic-driven IntClusts was extensive. Frequently, patients in IntClust-3, IntClust-4, and IntClust-8 showed the highest levels of *FBLN1*, whereas IntClust-1, IntClust-5, IntClust-6, IntClust-9, and IntClust-10 comprised patients with low *FBLN1* expression profiles (p-value ≤2.0E-5). Cohort expansion did not change the associations of PAM50 as this was in the $LVI^{+/-}$ patients (p-value ≤3.1E-59). In the METABRIC cohort, *FBLN1* expression seemed to be upregulated in patients of IntClust-3 and IntClust-4 only, but not IntClust-8. Downregulation of *FBLN1* was associated with IntClust-1, IntClust-2, IntClust-9, and IntClust-10, but not IntClust-5 and IntClust-6 (p-value ≤1.36E-55) (Table 5.8).

5.3.3.5.3. Expression correlations between *FBLN1* altered expression and a cluster of transcripts for matrix metalloproteinases

Many expression correlations were identified between *FBLN1* and several MMP encoding genes in the $LVI^{+/-}$ cohort. *FBLN1* expression was positively associated with *MMP2* (r=0.684, p-value ≤8.7E-25); *MMP3* (r=0.372, p-value ≤5.9E-07); *MMP10* (r=0.177, p-value ≤0.021); *MMP11* (r=0.172, p-value ≤0.025); *MMP13* (r=0.346, p-value ≤3.9E-06); *MMP14* (r=0.182, p-value ≤0.018); *MMP27* (r=0.381, p-value ≤2.9E-07); and *MMP28* (r=0.548, p-value ≤1.0E-14). *FBLN1* negative direction of expression was associated with *MMP1* (r=-0.172, p-value ≤0.025); *MMP12* (r=-0.194, p-value ≤0.011); *MMP15* (r=-0.235, p-value ≤0.002); and *MMP25* (r=-0.155, p-value ≤0.044) (Table 5.9).

More expression associations were pointed out with *MMPs* for *FBLN1* in the METABRIC cohort of patients. Expression of *FBLN1* was positively correlated with *MMP2* (r=0.667, p-value ≤2.2E-255); *MMP3* (r=0.470, p-value ≤1.5E-109); *MMP10* (r=0.154, p-value ≤6.2E-12); *MMP*11 (r=0.255, p-value ≤1.0E-30); *MMP13* (r=0.297, p-value ≤1.5E-41); *MMP14* (r=0.112, p-value ≤6.1E-07); *MMP19* (r=0.166, p-value ≤1.2E-13); *MMP27* (r=0.241, p-value ≤1.8E-27); and *MMP28* (r=0.466, p-value ≤1.5E-107). The negative direction of *FBLN1* expression was linked to *MMP12* (r=-0.116, p-value ≤2.1E-07); *MMP15* (r=-0.080, p-value ≤0.00003); *MMP16* (r=-0.065, p-value ≤0.004); *MMP20* (r=-0.101, p-value

≤6.1E-06); *MMP24* (*r*=-0.204, p-value ≤4.8E-20); *MMP25* (*r*=-0.105, p-value ≤2.7E-06); and *MMP26* (*r*=-0.068, p-value ≤0.002) (Table 5.10).

5.3.3.5.4. Expression correlations between *FBLN1* altered expression and a group of transcripts for a cluster of differentiation of lymphocytes

In the LVI$^{+/-}$ patients, the expression of *FBLN1* was inversely correlated to the expression of *CD80* (*r*=-0.259, p-value ≤0.001) and *CD20* (*r*=-0.166, p-value ≤0.030). No additional expression relationships between *FBLN1* and other CDs were identified in this cohort (Table 5.11).

In the METABRIC cohort, many positive expression interrelations were pointed out between *FBLN1* and different CDs. The expression of *FBLN1* was accordingly correlated to the expression of *CD2* (*r*=0.146, p-value ≤6.0E-11); *CD3G* (*r*=0.080, p-value ≤0.0003); *CD3D* (*r*=0.143, p-value ≤1.8E-10); *CD3E* (*r*=0.072, p-value ≤0.001); *CD3Z* (*r*=0.136, p-value ≤1.1E-09); *CD4* (*r*=0.079, p-value ≤0.0004); *CD8A* (*r*=0.132, p-value ≤4.1E-09); *CD28* (*r*=0.092, p-value ≤0.00003); *CD86* (*r*=0.077, p-value ≤0.001); *CD19* (*r*=0.045, p-value ≤0.044); *CD45* (*r*=0.071, p-value ≤0.001); *CD52* (*r*=0.172, p-value ≤1.1E-14); *CD40* (*r*=0.142, p-value ≤2.0E-10); *CD267* (*r*=0.142, p-value ≤2.4E-10); and *CD269* (*r*=0.121, p-value ≤6.5E-08). The negative expression of *FBLN1* was correlated to the expression of *CD80* (*r*=-0.162, p-value ≤4.4E-13); *CD20* (*r*=-0.101, p-value ≤6.7E-06); and *CD268* (*r*=-0.130, p-value ≤7.3E-09) (Table 5.12).

5.3.3.5.5. Expression correlations between *FBLN1* altered expression and a cluster of transcripts for invasion-related biomarkers

Various gene expression interrelations were identified for *FBLN1* in the LVI$^{+/-}$ cohort. The positive associations of gene expression for *FBLN1* were correlated to the expression of *SDC1* (*r*=0.338, p-value ≤6.4E-06); *FN1* (*r*=0.174, p-value ≤0.023); *COL3A1* (*r*=0.675, p-value ≤5.8E-24); *SNAI2* (*r*=0.526, p-value ≤1.8E-13); *ETS1* (*r*=0.190, p-value ≤0.013); *VIM* (*r*=0.502, p-value ≤3.1E-12); *CXCL12* (*r*=0.768, p-value ≤2.6E-34); *PDGFC* (*r*=0.414, p-value ≤2.0E-08); *FAP* (*r*=0.601, p-value ≤4.3E-18); *THY1* (*r*=0.595, p-value ≤1.3E-17); *S100A4* (*r*=0.192, p-value ≤0.012); and *CD19* (*r*=0.652, p-value ≤6.3E-22). Two relationships of negative gene expression were pointed out for *FBLN1* in LVI$^{+/-}$ patients versus the genes of *MKI67* (*r*=-0.333, p-value ≤9.1E-06) and *CDH1* (*r*=-0.289, p-value ≤0.0001) (Table 5.13).

Several gene expression interactions for *FBLN1* were also found in the METABRIC cohort of patients. Positive gene expression correlations of *FBLN1* were related to the expression of *SDC1* (*r*=0.345, p-value ≤2.2E-56); *CTNNB1* (*r*=0.145,

p-value ≤8.7E-11); *FN1* (*r*=0.248, p-value ≤3.1E-29); *COL3A1* (*r*=0.696, p-value ≤3.0E-287); *SNAI2* (*r*=0.516, p-value ≤5.5E-135); *ETS1* (*r*=0.179, p-value ≤9.8E-16); *CDH2* (*r*=0.124, p-value ≤3.5E-08); *VIM* (*r*=0.502, p-value ≤5.0E-127); *CXCL12* (*r*=0.676, p-value ≤1.7E-264); *PDGFC* (*r*=0.317, p-value ≤5.8E-24); *FAP* (*r*=0.648, p-value ≤4.4E-236); *THY1* (*r*=0.580, p-value ≤4.7E-178); *CD24* (*r*=0.085, p-value ≤0.0001); *CD44* (*r*=0.055, p-value <0.014); *S100A4* (*r*=0.367, p-value <3.4E-64); and *CAV1* (*r*=0.546, p-value ≤1.9E-154). The same negative gene expression correlations that were assigned to *FBLN1* in the $LVI^{+/-}$ patients were also identified in the METABRIC cohort. *FBLN1* expression was negatively associated with *MKI67* (*r*=-0.332, p-value ≤4.5E-52); and *CDH1* (*r*=-0.197, p-value ≤9.0E-19) (Table 5.14).

Table 5. 3: Clinicopathological significance of the T8 upregulated genes associated LVI positivity in the LVI-supervised cohort, (n=170). Significant values are in bold.

	HSP90AA1			*ARHGDIA*			*DYNLRB1*			*VDAC2*			*PPIA*		
Cohort: LVI$^{+/-}$	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Lymphovascular invasion (LVI), (H&E/IHC)															
Negative	54 (67.5)	26 (32.5)	18.51 (**1.7E-5**)	46 (57.5)	34 (42.5)	3.40 (0.065)	43 (53.8)	37 (46.3)	0.850 (0.357)	49 (61.3)	31 (38.8)	7.65 (**0.006**)	45 (56.3)	35 (43.8)	2.36 (0.124)
Definite	31 (34.4)	59 (65.6)		39 (43.3)	51 (56.7)		42 (46.7)	48 (53.3)		36 (40.0)	54 (60.0)		40 (44.4)	50 (55.6)	
Age (years)															
≤ 50	29 (43.9)	37 (56.1)	1.585 (0.208)	38 (57.6)	28 (42.4)	2.48 (0.116)	34 (51.5)	32 (48.5)	0.10 (0.753)	29 (43.9)	37 (56.1)	1.585 (0.208)	27 (40.9)	39 (59.1)	3.57 (0.059)
> 50	56 (53.6)	48 (46.2)		47 (45.2)	57 (54.8)		51 (49.0)	53 (51.0)		56 (53.8)	48 (46.2)		58 (55.8)	46 (44.2)	
Tumour size															
≤ 2.0cm	63 (63.0)	37 (37.0)	16.42 (**5.0E-5**)	56 (56.0)	44 (44.0)	3.50 (0.061)	54 (54.0)	46 (46.0)	1.554 (0.213)	59 (59.0)	41 (41.0)	7.87 (**0.005**)	56 (56.0)	44 (44.0)	3.50 (0.061)
> 2.0cm	22 (31.4)	48 (68.6)		29 (41.4)	41 (58.6)		31 (44.3)	39 (55.7)		26 (37.1)	44 (62.5)		29 (41.4)	41 (58.6)	
Tumour Grade															
Grade-I	23 (88.5)	3 (11.5)		17 (65.4)	9 (34.6)		15 (57.5)	11 (42.3)		14 (53.8)	12 (46.2)		21 (80.8)	5 (19.2)	
Grade-II	28 (50.9)	27 (49.1)	20.90 (**2.9E-5**)	31 (56.4)	24 (43.6)	6.26 (**0.044**)	24 (43.6)	31 (56.4)	1.55 (0.462)	33 (60.0)	22 (40.0)	4.58 (0.102)	25 (45.5)	30 (54.5)	11.43 (**0.003**)
Grade-III	33 (37.5)	55 (62.5)		36 (40.6)	52 (59.1)		45 (51.1)	43 (48.9)		37 (42.0)	51 (58.0)		39 (44.3)	49 (55.7)	
Nodal involvement, (NPI staging system)															
Stage-I	63 (60.0)	42 (40.0)		58 (55.2)	47 (44.8)		53 (50.5)	52 (49.5)		60 (57.1)	45 (42.9)		58 (55.2)	47 (44.8)	
Stage-II	16 (32.7)	33 (67.3)	11.10 (**0.004**)	19 (38.8)	30 (61.2)	3.62 (0.164)	25 (51.0)	24 (49.0)	0.280 (0.869)	19 (38.8)	30 (61.2)	5.61 (0.060)	24 (49.0)	25 (51.0)	7.42 (**0.024**)
Stage-III	6 (37.5)	10 (62.5)		8 (50.0)	8 (50.0)		7 (43.8)	9 (56.3)		6 (37.5)	10 (62.5)		3 (18.8)	13 (81.3)	
Nottingham Prognostic Index (NPI)															
good	47 (68.1)	22 (31.9)		43 (62.3)	26 (37.7)		35 (50.7)	34 (49.3)		43 (62.3)	26 (37.7)		41 (59.4)	28 (40.6)	
moderate	34 (38.6)	54 (61.4)	15.53 (**0.004**)	36 (40.9)	52 (59.1)	7.17 (**0.028**)	44 (50.0)	44 (50.0)	0.091 (0.995)	38 (43.2)	50 (56.8)	7.75 (**0.021**)	42 (47.7)	46 (52.3)	8.86 (**0.012**)
poor	4 (30.8)	9 (69.2)		6 (46.2)	7 (53.8)		6 (46.2)	7 (53.8)		4 (30.8)	9 (69.2)		2 (15.4)	11 (84.6)	

Table 5. 4: Clinicopathological significance of the T8 upregulated genes associated LVI positivity in the METABRIC cohort, (n=1980). Significant values are in bold.

	HSP90AA1			***ARHGDIA***			***DYNLRB1***			***VDAC2***			***PPIA***		
Cohort: METABRIC	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Lymphovascular invasion (LVI), (H&E/IHC)															
Negative	66 (66.0)	34 (34.0)	6.861 (**0.009**)	95 (57.9)	69 (42.1)	2.62 (0.105)	66 (40.2)	98 (59.8)	1.50 (0.221)	67 (40.9)	97 (59.1)	2.57 (0.109)	39 (23.8)	125 (76.2)	0.136 (0.713)
Positive	98 (50.0)	98 (50.0)		64 (48.5)	68 (51.5)		44 (33.3)	88 (66.7)		42 (31.8)	90 (68.2)		29 (22.0)	103 (78.0)	
Age (years)															
≤ 50	194 (45.2)	235 (54.8)	3.28 (0.070)	237 (55.2)	192 (44.8)	2.43 (0.119)	245 (57.1)	184 (42.9)	12.87 (**0.0003**)	210 (49.0)	219 (51.0)	3.30 (0.069)	219 (51.0)	210 (49.0)	4.303 (**0.038**)
> 50	778 (50.2)	773 (49.8)		791 (51.0)	760 (49.0)		734 (47.3)	817 (52.7)		836 (53.9)	715 (46.1)		879 (56.7)	672 (43.3)	
Tumour size															
≤ 2.0cm	447 (52.1)	411 (47.9)	5.75 (**0.017**)	476 (55.5)	382 (44.5)	7.011 (**0.008**)	442 (51.5)	416 (48.5)	2.14 (0.144)	484 (56.4)	374 (43.6)	7.73 (**0.005**)	489 (57.0)	369 (43.0)	1.36 (0.244)
> 2.0cm	514 (46.6)	588 (53.4)		545 (49.5)	557 (50.5)		531 (48.2)	571 (51.8)		552 (50.1)	550 (49.9)		599 (54.4)	503 (45.6)	
Tumour Grade															
Grade-I	109 (63.7)	62 (36.3)		115 (67.3)	56 (32.7)		86 (50.3)	85 (49.7)		99 (57.9)	72 (42.1)		103 (60.2)	68 (39.8)	
Grade-II	416 (54.0)	354 (46.0)	42.33 (**6.4E-10**)	417 (54.2)	353 (45.8)	27.010 (**1.0E-6**)	355 (46.1)	415 (53.9)	5.00 (0.082)	439 (57.0)	331 (43.0)	14.23 (**0.001**)	450 (58.4)	41.6 (464)	10.96 (**0.004**)
Grade-III	398 (41.8)	554 (58.2)		447 (47.0)	505 (53.0)		490 (51.5)	462 (48.5)		462 (48.5)	490 (51.5)		488 (51.3)	464 (48.7)	
Nodal involvement, (NPI staging system)*															
Stage-I	537 (51.9)	498 (48.1)		528 (51.0)	507 (49.0)		523 (50.5)	512 (49.5)		581 (56.1)	454 (43.9)		600 (58.0)	435 (42.0)	
Stage-II	276 (44.4)	346 (55.6)	8.78 (**0.012**)	335 (53.9)	287 (46.1)	1.48 (0.477)	308 (49.5)	314 (50.5)	2.39 (0.303)	312 (50.2)	310 (49.8)	9.96 (**0.007**)	340 (54.7)	282 (45.3)	7.55 (**0.023**)
Stage-III	156 (49.4)	160 (50.6)		160 (50.6)	156 (49.4)		144 (45.6)	172 (54.4)		150 (47.5)	166 (52.5)		156 (49.4)	160 (50.6)	
Nottingham Prognostic Index (NPI)															
good	386 (56.8)	294 (43.2)		375 (55.1)	305 (44.9)		330 (48.5)	350 (51.5)		392 (57.6)	288 (42.4)		421 (61.9)	259 (38.1)	
moderate	498 (45.2)	603 (54.8)	24.47 (**5.0E-6**)	565 (51.3)	536 (48.7)	7.72 (**0.021**)	556 (50.5)	545 (49.5)	1.30 (0.521)	566 (51.4)	535 (48.6)	13.144 (**0.001**)	585 (53.1)	516 (46.9)	20.73 (**3.0E-5**)
poor	88 (44.2)	111 (55.8)		88 (44.2)	111 (952)		93 (46.7)	106 (53.3)		88 (44.2)	111 (55.8)		92 (46.2)	107 (53.8)	

Table 5. 5: Distribution of the LVI-supervised and METABRIC patients on the PAM50 and IntClusts classification systems against gene expression profiles of the T8 upregulated genes associated LVI positivity.

	HSP90AA1			***ARHGDIA***			***DYNLRB1***			***VDAC2***			***PPIA***		
Cohort: $LVI^{+/-}$	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50															
Normal-like	7 (58.3)	5 (41.7)		5 (41.7)	7 (58.3)		8 (66.7)	4 (33.3)		7 (58.3)	5 (41.7)		9 (75.0)	3 (25.0)	
Luminal-A	44 (68.8)	20 (31.3)		41 (64.1)	23 (35.9)		31 (48.4)	33 (51.6)		41 (64.1)	23 (35.9)		39 (60.9)	25 (39.1)	
Luminal-B	11 (24.4)	34 (75.6)	21.40 (**0.0026**)	17 (37.8)	28 (62.2)	8.80 (0.066)	10 (22.2)	35 (77.8)	26.74 (**2.0E-5**)	21 (46.7)	24 (53.3)	12.48 (**0.014**)	18 (40.0)	27 (60.0)	11.15 (**0.025**)
Her2-enriched	6 (42.9)	8 (57.1)		7 (50.0)	7 (50.0)		9 (64.3)	5 (35.7)		3 (21.4)	11 (78.6)		4 (28.6)	10 (71.4)	
Basal	17 (48.6)	18 (51.4)		15 (42.9)	20 (57.1)		27 (77.1)	8 (22.9)		13 (37.1)	22 (62.9)		15 (42.9)	20 (57.1)	
Integrated classification clusters of METABRIC															
cluster-1	2 (10.5)	17 (89.5)		4 (21.1)	15 (78.9)		5 (26.3)	14 (73.7)		12 (63.2)	7 (36.8)		7 (36.8)	12 (63.2)	
cluster -2	4 (66.7)	2 (33.3)		3 (50.0)	3 (50.0)		2 (33.3)	4 (66.7)		3 (50.0)	3 (50.0)		4 (66.7)	2 (33.3)	
cluster -3	20 (66.7)	10 (33.3)		16 (53.3)	14 (46.7)		18 (60.0)	12 (40.0)		19 (63.3)	11 (36.7)		17 (56.7)	13 (43.3)	
cluster -4	15 (60.0)	10 (40.0)		14 (56.0)	11 (44.0)		15 (60.0)	10 (40.0)		12 (48.0)	13 (52.0)		17 (68.0)	8 (32.0)	
cluster -5	8 (40.0)	12 (60.0)		10 (50.0)	10 (50.0)		9 (45.0)	11 (55.0)		7 (35.0)	13 (65.0)		5 (25.0)	15 (75.0)	
cluster -6	2 (66.7)	1 (33.3)	24.66 (**0.003**)	2 (66.7)	1 (33.3)	15.60 (0.067)	2 (66.7)	1 (33.3)	14.22 (0.115)	1 (33.3)	2 (66.7)	13.10 (0.158)	2 (66.7)	1 (33.3)	22.20 (**0.008**)
cluster -7	10 (66.7)	5 (33.3)		8 (53.3)	7 (46.7)		4 (26.7)	11 (73.3)		10 (66.7)	5 (33.3)		12 (80.0)	3 (20.0)	
cluster -8	11 (57.9)	8 (42.1)		15 (78.9)	4 (21.1)		9 (47.4)	10 (52.6)		11 (57.9)	8 (42.1)		11 (57.9)	8 (42.1)	
cluster -9	1 (12.5)	7 (87.5)		2 (25.0)	6 (75.0)		4 (50.0)	4 (50.0)		3 (37.5)	5 (62.5)		2 (25.0)	6 (75.0)	
cluster -10	12 (48.0)	13 (52.0)		11 (44.0)	15 (56.0)		17 (68.0)	8 (32.0)		7 (28.0)	18 (72.0)		8 (32.0)	17 (68.0)	

Continue on the next page

	HSP90AA1			*ARHGDIA*			*DYNLRB1*			*VDAC2*			*PPIA*		
Cohort: METABRIC	-	+	$\chi 2$ (*p*-value≤)	-	+	$\chi 2$ (*p*-value≤)	-	+	$\chi 2$ (*p*-value≤)	-	+	$\chi 2$ (*p*-value≤)	-	+	$\chi 2$ (*p*-value≤)
	n (%)	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**		**n (%)**	**n (%)**	
Prediction analysis of microarray 50, PAM50															
Normal-like	136 (66.3)	69 (33.7)		115 (56.1)	90 (43.9)		129 (62.9)	76 (37.1)		137 (66.8)	68 (33.2)		141 (68.8)	64 (31.2)	
Luminal-A	422 (58.8)	296 (41.2)		436 (60.7)	282 (39.3)		361 (50.3)	357 (49.7)		446 (62.1)	272 (37.9)		458 (63.8)	260 (36.2)	
Luminal-B	165 (33.8)	323 (66.2)	103.40 (**1.85E-21**)	226 (46.3)	262 (53.7)	49.325 (**5.0E-10**)	130 (26.6)	358 (73.4)	168.83 (**0.009**)	193 (39.5)	295 (60.5)	86.36 (**7.8E-18**)	209 (42.8)	279 (57.2)	71.45 (**1.1E-14**)
Her2-enriched	105 (43.8)	135 (56.3)		92 (38.3)	148 (61.7)		133 (55.4)	107 (44.6)		126 (52.5)	114 (47.5)		126 (52.5)	47.5 (165)	
Basal	144 (43.8)	185 (56.2)		159 (48.3)	170 (51.7)		226 (68.7)	103 (31.3)		144 (43.5)	185 (56.2)		164 (49.8)	165 (50.2)	
Integrated classification clusters of METABRIC															
cluster-1	38 (27.3)	101 (72.2)		29 (20.9)	110 (79.1)		37 (26.6)	102 (73.4)		52 (37.4)	87 (62.6)		55 (39.6)	84 (60.4)	
cluster -2	31 (43.1)	41 (56.9)		38 (52.8)	34 (47.2)		22 (30.6)	50 (69.4)		30 (41.7)	42 (58.3)		37 (51.4)	35 (48.6)	
cluster -3	170 (58.6)	120 (41.4)		188 (64.8)	102 (35.2)		144 (49.7)	146 (50.3)		172 (59.3)	118 (40.7)		169 (58.3)	121 (41.7)	
cluster -4	224 (65.3)	119 (34.7)		187 (54.5)	156 (45.5)		215 (62.7)	128 (37.3)		213 (62.1)	130 (37.9)		216 (63.0)	127 (37.0)	
cluster -5	76 (40.0)	114 (60.0)	119.66 (**1.57E-21**)	79 (41.6)	111 (58.4)	101.88 (**6.5E-18**)	97 (51.1)	93 (48.9)	113.80 (**2.5E-20**)	74 (38.9)	116 (61.1)	84.06 (**2.51E-14**)	90 (47.4)	100 (52.6)	47.79 (**2.8E-7**)
cluster -6	35 (41.2)	50 (58.8)		46 (54.1)	39 (45.9)		26 (30.6)	59 (69.4)		45 (52.9)	40 (47.1)		45 (52.9)	40 (47.1)	
cluster -7	112 (58.9)	78 (41.1)		94 (49.5)	96 (50.5)		84 (44.2)	106 ()		120 (63.2)	70 (36.8)		115 (60.5)	75 (39.5)	
cluster -8	155 (51.8)	144 (48.2)		183 (61.2)	116 (38.8)		151 (50.5)	148 (49.5)		178 (59.5)	121 (40.5)		190 (63.5)	109 (36.5)	
cluster -9	54 (37.0)	92 (63.0)		59 (40.4)	87 (59.6)		53 (36.3)	93 (63.7)		78 (53.4)	68 (46.6)		78 (53.4)	68 (46.6)	
cluster -10	77 (34.1)	149 (65.9)		125 (55.3)	101 (44.7)		150 (66.4)	76 (33.6)		84 (37.2)	142 (62.8)		103 (45.6)	123 (54.4)	

Table 5. 6: Clinicopathological significance of the T8 upregulated genes associated LVI negativity in the LVI-supervised cohort, (n=170). Significant values are in bold.

	FBLN1			*RPS18*			*RPL10A*		
Cohort: LVI+-	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Lymphovascular invasion (LVI), (H&E/IHC)									
Negative	31 (38.8)	49 (61.3)	7.65 (**0.006**)	36 (45.0)	44 (55.0)	1.51 (0.219)	35 (43.8)	45 (56.3)	2.36 (0.124)
Positive	54 (60.0)	36 (40.0)		49 (54.4)	41 (45.6)		50 (55.6)	40 (44.4)	
Age (years)									
≤ 50	32 (48.5)	34 (51.5)	0.099 (0.753)	36 (54.5)	30 (45.5)	0.892 (0.345)	32 (48.5)	34 (51.5)	0.10 (0.753)
> 50	53 (51.0)	51 (49.0)		49 (47.1)	55 (52.9)		53 (51.0)	51 (49.0)	
Tumour size									
≤ 2.0cm	44 (44.0)	56 (56.0)	3.50 (0.060)	49 (49.0)	51 (51.0)	0.097 (0.755)	48 (48.0)	52 (52.0)	0.389 (0.533)
> 2.0cm	41 (58.6)	29 (41.4)		36 (51.4)	34 (48.6)		37 (52.9)	33 (47.1)	
Tumour Grade									
Grade-I	6 (23.1)	20 (76.9)		11 (42.3)	15 (57.7)		10 (38.5)	16 (61.5)	
Grade-II	22 (40.0)	33 (60.0)	17.42 (**0.00017**)	27 (49.1)	28 (50.9)	1.04 (0.595)	22 (40.0)	33 (60.0)	7.26 (**0.027**)
Grade-III	57 (64.8)	31 (35.2)		47 (53.4)	41 (46.6)		53 (60.2)	35 (39.8)	
Nodal involvement, (NPI staging system)*									
Stage-I	45 (42.9)	60 (57.1)		45 (42.9)	60 (57.1)		46 (43.8)	59 (56.2)	
Stage-II	28 (57.1)	21 (42.9)	7.14 (**0.028**)	33 (67.3)	16 (32.7)	8.291 (**0.016**)	32 (65.3)	17 (34.7)	6.45 (**0.040**)
Stage-III	12 (75.0)	4 (25.0)		7 (43.8)	9 (56.3)		7 (43.8)	9 (56.3)	
Nottingham Prognostic Index (NPI)									
good	21 (30.4)	48 (69.6)		31 (44.9)	38 (55.1)		28 (40.6)	41 (59.4)	
moderate	54 (61.4)	34 (38.6)	18.89 (**8.0E-5**)	48 (54.5)	40 (45.5)	1.51 (0.469)	51 (58.0)	37 (42.0)	4.75 (0.093)
poor	10 (76.9)	3 (23.1)		6 (46.2)	7 (53.8)		6 (46.2)	7 (53.8)	

Table 5. 7: Clinicopathological significance of the T8 upregulated genes associated LVI negativity in the METABRIC cohort, (n=1980). Significant values are in bold.

	FBLN1			*RPS18*			*RPL10A*		
Cohort: METABRIC	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)	-	+	χ^2 (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Lymphovascular invasion (LVI), (H&E/IHC)									
Negative	61 (37.2)	103 (62.8)	2.877 (0.090)	53 (32.3)	111 (67.7)	2.34 (0.126)	20 (12.2)	144 (87.8)	1.20 (0.273)
Positive	62 (47.0)	70 (53.0)		54 (40.9)	78 (59.1)		22 (16.7)	110 (83.3)	
Age (years)									
≤ 50	168 (39.2)	261 (60.8)	10.80 (**0.001**)	171 (39.9)	258 (60.1)	24.12 (**9.04E-7**)	177 (41.3)	252 (58.7)	7.68 (**0.006**)
> 50	746 (48.1)	805 (51.9)		826 (53.3)	725 (46.7)		757 (48.8)	794 (51.2)	
Tumour size									
≤ 2.0cm	357 (41.6)	501 (58.4)	11.96 (**0.001**)	436 (50.8)	422 (49.2)	0.078 (0.780)	382 (44.5)	476 (55.5)	4.04 (**0.044**)
> 2.0cm	545 (49.5)	557 (50.5)		553 (50.2)	549 (49.8)		541 (49.1)	561 (50.9)	
Tumour Grade									
Grade-I	54 (31.6)	117 (68.4)	61.28 (**4.9E-14**)	77 (45.0)	94 (55.0)	2.06 (0.356)	60 (35.1)	111 (64.9)	13.44 (**0.001**)
Grade-II	297 (38.6)	473 (61.4)		393 (51.0)	377 (49.0)		352 (45.7)	418 (54.3)	
Grade-III	522 (54.8)	430 (45.2)		480 (50.4)	472 (49.6)		475 (49.9)	477 (50.1)	
Nodal involvement, (NPI staging system)*									
Stage-I	462 (44.6)	573 (55.4)	2.415 (0.299)	513 (49.6)	522 (50.4)	0.54 (0.763)	501 (48.4)	534 (51.6)	2.46 (0.292)
Stage-II	302 (48.6)	320 (51.4)		317 (51.0)	305 (49.0)		278 (44.7)	344 (55.3)	
Stage-III	147 (46.5)	169 (53.5)		163 (51.6)	153 (48.4)		154 (48.7)	162 (51.3)	
Nottingham Prognostic Index (NPI)									
good	259 (38.1)	421 (61.9)	27.29 (**1.0E-6**)	335 (49.3)	345 (50.7)	0.810 (0.667)	319 (46.9)	361 (53.1)	2.34 (0.310)
moderate	557 (50.6)	544 (49.4)		557 (50.6)	544 (49.4)		511 (46.4)	590 (53.6)	
poor	98 (49.2)	101 (50.8)		105 (52.8)	94 (47.2)		104 (52.3)	95 (47.7)	

Table 5. 8: Distribution of the LVI-supervised and METABRIC patients on the PAM50 and IntClusts classification systems against gene expression profiles of the T8 upregulated genes associated LVI negativity.

	FBLN1			*RPS18*			*RPL10A*		
Cohort: LVI$^{+/-}$	-	+	$\chi 2$ (*p*-value≤)	-	+	$\chi 2$ (*p*-value≤)	-	+	$\chi 2$ (*p*-value≤)
	n (%)	n (%)		n (%)	n (%)		n (%)	n (%)	
Prediction analysis of microarray 50, PAM50									
Normal-like	1 (8.3)	11 (91.7)		5 (41.7)	7 (58.3)		4 (33.3)	8 (66.7)	
Luminal-A	19 (29.7)	45 (70.3)		30 (46.9)	34 (53.1)		27 (42.2)	37 (57.8)	
Luminal-B	32 (71.1)	13 (28.9)	33.80 (**8.2E-7**)	25 (55.6)	20 (44.4)	1.45 (0.835)	28 (62.2)	17 (37.8)	8.13 (0.087)
Her2-enriched	11 (78.6)	3 (21.4)		8 (57.1)	6 (42.9)		5 (35.7)	9 (64.3)	
Basal	22 (62.9)	13 (37.1)		17 (48.6)	18 (51.4)		21 (60.0)	14 (40.0)	
Integrated classification clusters of METABRIC									
cluster-1	13 (68.4)	6 (31.6)		13 (68.4)	6 (31.6)		13 (68.4)	6 (31.6)	
cluster -2	3 (50.0)	3 (50.0)		1 (16.7)	5 (83.3)		3 (50.0)	3 (50.0)	
cluster -3	4 (13.3)	26 (86.7)		15 (50.)	15 (50.0)		12 (40.0)	18 (60.0)	
cluster -4	8 (32.0)	17 (68.0)		13 (52.0)	12 (48.0)		12 (48.0)	13 (52.0)	
cluster -5	15 (75.0)	5 (25.0)	37.43 (**2.0E-5**)	9 (45.0)	11 (55.0)	14.23 (0.114)	9 (45.0)	11 (55.0)	9.40 (0.401)
cluster -6	2 (66.7)	1 (33.3)		3 (100.0)	0 (0.0)		3 (100.0)	0 (0.0)	
cluster -7	8 (53.3)	7 (46.7)		10 (66.7)	5 (33.3)		7 (46.7)	8 (53.3)	
cluster -8	7 (36.8)	12 (63.2)		6 (31.6)	13 (68.4)		7 (36.8)	12 (63.2)	
cluster -9	6 (75.0)	2 (25.0)		5 (62.5)	3 (37.5)		4 (50.0)	4 (50.0)	
cluster -10	19 (76.0)	6 (24.0)		10 (40.0)	15 (60.0)		15 (60.0)	10 (40.0)	

Continue on the next page

Cohort: METABRIC	*FBLN1* - n (%)	*FBLN1* + n (%)	χ^2 (*p*-value≤)	*RPS18* - n (%)	*RPS18* + n (%)	χ^2 (*p*-value≤)	*RPL10A* - n (%)	*RPL10A* + n (%)	χ^2 (*p*-value≤)
Prediction analysis of microarray 50, PAM50									
Normal-like	30 (14.6)	175 (85.4)		95 (46.3)	110 (53.7)		77 (37.6)	128 (62.4)	
Luminal-A	226 (31.5)	492 (68.5)		354 (49.3)	364 (50.7)		319 (44.4)	399 (55.6)	
Luminal-B	344 (70.5)	144 (29.5)	103.40 (**3.1E-59**)	264 (54.1)	224 (45.9)	15.06 (**0.005**)	219 (44.9)	269 (55.1)	34.89 (**4.9E-7**)
Her2-enriched	130 (54.2)	110 (45.8)		139 (57.9)	101 (42.1)		149 (62.1)	91 (37.9)	
Basal	184 (55.9)	145 (44.1)		145 (44.1)	184 (55.9)		170 (51.7)	159 (48.3)	
Integrated classification clusters of METABRIC									
cluster-1	100 (71.9)	39 (28.1)		76 (54.7)	63 (45.3)		67 (48.2)	72 (51.8)	
cluster -2	45 (62.5)	27 (37.5)		39 (54.2)	33 (45.8)		35 (48.6)	37 (51.4)	
cluster -3	63 (21.7)	227 (78.3)		129 (44.5)	161 (55.5)		96 (33.1)	194 (66.9)	
cluster -4	80 (23.3)	263 (76.7)		201 (58.6)	142 (41.1)		153 (44.6)	190 (55.4)	
cluster -5	91 (47.9)	99 (52.1)	282.47 (**1.36E-55**)	105 (55.3)	85 (44.7)	39.91 (**8.0E-6**)	99 (52.1)	91 (47.9)	33.65 (**0.0001**)
cluster -6	47 (55.3)	38 (44.7)		54 (63.5)	31 (36.5)		43 (50.6)	42 (49.4)	
cluster -7	83 (43.7)	107 (56.3)		89 (46.8)	101 (53.2)		102 (53.7)	88 (46.3)	
cluster -8	141 (47.2)	158 (52.8)		141 (47.2)	158 (52.8)		150 (50.2)	149 (49.8)	
cluster -9	102 (69.9)	44 (30.1)		78 (53.4)	68 (46.6)		79 (54.1)	67 (45.9)	
cluster -10	162 (71.1)	64 (28.3)		85 (37.6)	141 (62.4)		110 (48.7)	116 (51.3)	

Table 5. 9: Gene expression correlations between the T8 genes and a cluster of MMP-encoding genes in the LVI-supervised patients, (n=170).

Protein-encoding gene	Protein name	HSP90AA1	ARHGDIA	VDAC2	DYNLRB1	PPIA	FBLN1	RPS18	RPL10A
		Labelling of correlation coefficient values , (Red = negative correlation); (Green = positive correlation); (White = no correlation)							
		P value < 0.050 is in bold							
MMP1	Interstitial collagenase	0.122	0.118	.193*	-0.127	.237**	-.172*	-0.029	-0.097
		0.114	0.125	**0.012**	0.098	**0.002**	**0.025**	0.709	0.208
MMP2	72 kDa type IV collagenase	-.268**	-0.042	-.238**	-.250**	-.263**	.684**	-0.132	-0.027
		0.0004	0.588	**0.002**	**0.001**	**0.001**	**8.7E-25**	0.086	0.727
MMP3	Stromelysin-1	-0.140	-.170*	-0.077	-0.140	-0.029	.372**	0.150	0.149
		0.069	**0.027**	0.320	0.069	0.704	**5.9E-07**	0.051	0.052
MMP9	Matrix metalloproteinase-9	-0.005	.355**	0.087	-0.012	-0.011	-0.041	-.307**	-.314**
		0.946	**2.1E-06**	0.262	0.881	0.888	0.597	**0.00004**	**0.0001**
MMP10	Stromelysin-2	0.151*	0.013	0.035	0.026	-0.021	.177*	-0.104	-0.032
		0.049	0.865	0.649	0.734	0.781	**0.021**	0.177	0.677
MMP11	Stromelysin-3	0.071	0.104	-0.053	0.020	0.064	.172*	-0.136	-.220**
		0.357	0.176	0.493	0.796	0.405	**0.025**	0.077	**0.004**
MMP12	Macrophage metalloelastase	0.009	.199**	.214**	-0.101	0.108	-.194*	-0.085	-0.112
		0.909	**0.009**	**0.005**	0.190	0.160	**0.011**	0.270	0.147
MMP13	Collagenase 3	0.076	-0.037	-0.012	0.013	0.127	.346**	-0.113	0.014
		0.324	0.635	0.881	0.868	0.098	**3.9E-06**	0.142	0.855
MMP14	Matrix metalloproteinase-14	-0.039	0.133	0.035	-0.098	-0.005	.182*	-.184*	-.180*
		0.611	0.083	0.653	0.204	0.953	**0.018**	**0.016**	**0.019**
MMP15	Matrix metalloproteinase-15	-0.085	0.146	-0.063	-0.096	-0.043	-.235**	-.251**	-.297**
		0.271	0.058	0.417	0.213	0.576	**0.002**	**0.001**	**0.0001**
MMP16	Matrix metalloproteinase-16	0.134	0.108	0.086	0.085	0.030	-0.142	-0.121	-.236**
		0.081	0.160	0.267	0.273	0.701	0.065	0.116	**0.002**
MMP17	Matrix metalloproteinase-17	0.008	0.009	0.020	-0.022	0.034	-0.081	-0.029	-0.045
		0.920	0.910	0.799	0.773	0.658	0.291	0.703	0.561
MMP19	Matrix metalloproteinase-19	-0.075	0.065	0.087	-0.036	0.006	-0.078	-0.089	-0.054
		0.330	0.399	0.261	0.639	0.941	0.312	0.251	0.487
MMP20	Matrix metalloproteinase-20	0.074	0.049	0.147	-0.087	0.102	-0.148	0.078	0.004
		0.336	0.527	0.055	0.260	0.184	0.055	0.313	0.961
MMP21	Matrix metalloproteinase-21	0.078	-0.037	-0.038	-0.117	0.043	-0.007	-0.006	0.025
		0.310	0.634	0.619	0.129	0.575	0.924	0.934	0.751
MMP24	Matrix metalloproteinase-24	0.043	0.082	.155*	0.071	-0.046	-0.106	-0.080	-0.046
		0.580	0.287	**0.043**	0.355	0.556	0.168	0.297	0.553
MMP25	Matrix metalloproteinase-25	-0.072	0.029	-0.025	0.096	-0.009	-.155*	0.007	-0.057
		0.348	0.711	0.745	0.214	0.908	**0.044**	0.925	0.457
MMP26	Matrix metalloproteinase-26	0.133	0.001	-0.031	0.030	0.052	-0.147	0.016	.216**
		0.083	0.990	0.688	0.701	0.497	0.055	0.836	**0.005**
MMP27	Matrix metalloproteinase-27	-0.040	-.175*	-.165*	0.083	0.028	.381**	0.096	.208**
		0.607	**0.022**	**0.031**	0.282	0.713	**2.9E-07**	0.214	**0.006**
MMP28	Matrix metalloproteinase-28	-.168*	-0.126	-0.095	0.043	-0.014	.548**	-0.016	0.075
		0.029	0.103	0.216	0.573	0.856	**1.0E-14**	0.841	0.333

Table 5. 10: Gene expression correlations between the T8 genes and a cluster of MMP-encoding genes in the METABRIC patients, (n=1980).

Protein-encoding gene	Protein name	HSP90AA1	ARHGDIA	VDAC2	DYNLRB1	PPIA	FBLN1	RPS18	RPL10A
		Labelling of correlation coefficient values , (Red = negative correlation); (Green = positive correlation); (White = no correlation)							
		P value < 0.050 is in bold							
MMP1	Interstitial collagenase	.073**	.119**	.096**	.065**	.273**	-0.018	0.022	0.038
		0.001	**1.18E-07**	**0.00002**	**0.004**	**3.2E-35**	0.416	0.338	0.092
MMP2	72 kDa type IV collagenase	-.071**	.055*	-.142**	-.145**	-.098**	.667**	-.185**	.067**
		0.002	**0.015**	**1.99E-10**	**8.7E-11**	**0.00001**	**2.2E-255**	**1.0E-16**	**0.003**
MMP3	Stromelysin-1	-0.003	-.047*	-0.042	-.119**	0.016	.470**	0.035	.137**
		0.888	**0.037**	0.062	**11.05E-07**	0.464	**1.5E-109**	0.118	**8.4E-10**
MMP9	Matrix metalloproteinase-9	-0.026	.233**	0.023	-.053*	-0.012	0.018	-.122**	-.136**
		0.252	**6.65E-26**	0.303	**0.018**	0.598	0.425	**4.8E-08**	**1.2E-09**
MMP10	Stromelysin-2	.100**	-0.030	-0.010	-0.020	0.033	.154**	-0.031	.112**
		8.27E-06	0.183	0.672	0.383	0.138	**6.2E-12**	0.168	**6.5E-07**
MMP11	Stromelysin-3	0.004	.181**	-.085**	.051*	-.049*	.255**	-.171**	-.135**
		0.864	**4.4E-16**	**0.0001**	**0.024**	**0.029**	**1.0E-30**	**1.6E-14**	**1.8E-09**
MMP12	Macrophage metalloelastase	0.041	.076**	.136**	-.047*	.159**	-.116**	0.002	0.013
		0.068	**0.001**	**1.3E-09**	**0.037**	**1.1E-12**	**2.1E-07**	0.927	0.549
MMP13	Collagenase 3	.162**	0.023	0.044	0.042	.192**	.297**	-.101**	.150**
		3.7E-13	0.315	0.052	0.062	**5.9E-18**	**1.5E-41**	**7.1E-06**	**1.8E-11**
MMP14	Matrix metalloproteinase-14	0.021	0.032	.055*	0.028	.136**	.112**	0.002	.083**
		0.344	0.158	0.015	0.212	**1.1E-09**	**6.1E-07**	0.922	**0.0002**
MMP15	Matrix metalloproteinase-15	-.171**	.157**	-.068**	-.085**	-.145**	-.080**	-.150**	-.374**
		1.7E-14	**2.4E-12**	**0.002**	**0.0001**	**8.1E-11**	**0.00003**	**2.3E-11**	**6.2E-67**
MMP16	Matrix metalloproteinase-16	0.035	.083**	.082**	.097**	.146**	-.065**	.066**	.103**
		0.115	**0.0002**	**0.0002**	**0.00001**	**6.9E-11**	**0.004**	**0.003**	**4.4E-06**
MMP17	Matrix metalloproteinase-17	-0.021	0.003	-0.012	-0.024	0.013	-0.038	0.004	-0.024
		0.355	0.882	0.579	0.277	0.556	0.089	0.848	0.294
MMP19	Matrix metalloproteinase-19	.075**	-.062**	0.013	-0.016	.135**	.166**	-.081**	.222**
		0.001	**0.005**	0.554	0.470	**1.6E-09**	**1.2E-13**	**0.0003**	**1.4E-23**
MMP20	Matrix metalloproteinase-20	0.026	-0.009	.048*	-.056*	0.023	-.101**	0.043	0.009
		0.253	0.682	**0.034**	**0.013**	0.310	**6.1E-06**	0.053	0.692
MMP21	Matrix metalloproteinase-21	-.151**	-0.010	-.063**	-.085**	-.151**	-0.020	-.065**	-.204**
		1.4E-11	0.653	**0.005**	**0.0001**	**1.4E-11**	0.380	**0.004**	**5.7E-20**
MMP24	Matrix metalloproteinase-24	-.053*	.049*	.047*	.075**	-0.030	-.204**	-.046*	-.066**
		0.018	**0.029**	**0.037**	**0.001**	0.177	**4.8E-20**	**0.042**	**0.003**
MMP25	Matrix metalloproteinase-25	-.051*	.048*	-.079**	0.031	-0.038	-.105**	-0.031	-.111**
		0.024	**0.034**	**0.0004**	0.170	0.094	**2.7E-06**	0.173	**6.7E-07**
MMP26	Matrix metalloproteinase-26	0.030	-0.030	-0.003	-0.022	-0.011	-.068**	-0.028	0.014
		0.181	0.188	0.901	0.331	0.614	**0.002**	0.214	0.538
MMP27	Matrix metalloproteinase-27	0.033	-.097**	-0.028	0.017	.102**	.241**	.075**	.212**
		0.147	**0.00001**	0.218	0.451	**5.2E-06**	**1.8E-27**	**0.001**	**1.4E-21**
MMP28	Matrix metalloproteinase-28	0.023	-.122**	-.081**	-.076**	0.007	.466**	-.129**	.190**
		0.298	**5.2E-08**	**0.0003**	**0.001**	0.761	**1.5E-107**	**7.4E-09**	**1.7E-17**

Table 5. 11: Gene expression correlations between the T8 genes and a cluster of CD-encoding genes in the LVI-supervised patients, (n=170).

Protein-encoding gene	Protein name	HSP90AA1	ARHGDIA	VDAC2	DYNLRB1	PPIA	FBLN1	RPS18	RPL10A
		Labelling of correlation coefficient values , (Red = negative correlation); (Green = positive correlation); (White = no correlation)							
		P value < 0.050 is in bold							
CD2	T-cell surface antigen CD2	-0.002	-0.101	-0.106	-.293**	-0.076	-0.004	-0.008	0.075
		0.981	0.188	0.167	**0.0001**	0.327	0.963	0.913	0.331
CD3G	T-cell surface glycoprotein CD3 gamma chain	0.007	-.170*	-0.116	-.350**	-0.109	-0.087	-0.021	0.044
		0.923	**0.027**	0.132	**2.9E-06**	0.158	0.258	0.785	0.568
CD3D	T-cell surface glycoprotein CD3 delta chain	-0.035	-.174*	-0.128	-.333**	-0.045	-0.012	0.066	.160*
		0.647	**0.023**	0.097	**9.0E-06**	0.558	0.877	0.395	**0.038**
CD3E	T-cell surface glycoprotein CD3 epsilon chain	-0.040	-0.086	-.220**	-.296**	-.167*	0.016	-0.005	0.080
		0.606	0.263	**0.004**	**0.0001**	**0.030**	0.835	0.950	0.302
CD3Z, (CD247)	T-cell surface glycoprotein CD3 zeta chain	-0.061	-0.150	-0.127	-.319**	-0.116	0.011	0.005	0.094
		0.433	0.052	0.099	**0.00002**	0.132	0.887	0.948	0.222
CD4	T-cell surface glycoprotein CD4	-0.131	.178*	-.176*	-.156*	-0.118	0.004	-.294**	-.159*
		0.088	**0.020**	**0.022**	**0.042**	0.124	0.963	**0.0001**	**0.038**
CD8A	T-cell surface glycoprotein CD8 alpha chain	-0.033	-0.135	-.161*	-.352**	-0.125	0.002	0.000	0.070
		0.671	**0.080**	**0.036**	**2.5E-06**	0.104	0.977	0.997	0.361
CD8B	T-cell surface glycoprotein CD8 beta chain	0.134	-0.107	-0.028	-0.118	-0.009	-0.036	0.038	0.124
		0.081	0.165	0.720	0.126	0.910	0.640	0.626	0.108
CD28	T-cell-specific surface glycoprotein CD28	0.030	-0.089	-.157*	-.196*	-0.067	0.141	0.009	.180*
		0.695	0.248	**0.040**	**0.010**	0.388	0.068	0.905	**0.019**
CD80	T-lymphocyte activation antigen CD80	.167*	0.052	0.092	-0.143	0.037	-.259**	-.159*	-0.141
		0.030	0.504	0.232	0.064	0.629	**0.001**	**0.038**	0.066
CD86	T-lymphocyte activation antigen CD86	-0.056	-0.014	-.165*	-.307**	-0.078	-0.064	-.260**	-0.103
		0.470	0.856	**0.031**	**0.00005**	0.313	0.408	**0.001**	0.180
CD19	B-lymphocyte antigen CD19	-0.014	-0.136	-0.134	-.258**	-0.091	-0.067	-0.010	0.105
		0.853	0.077	0.083	**0.001**	0.236	0.387	0.902	0.171
CD20, (MS4A1)	B-lymphocyte antigen CD20	0.042	0.076	0.108	0.043	0.098	-.166*	-0.002	-0.050
		0.589	0.324	0.162	0.580	0.203	**0.030**	0.984	0.519
CD22	B-cell receptor CD22	-0.001	-0.001	-.205**	0.026	-0.072	-0.018	-0.119	0.064
		0.989	0.988	**0.007**	0.735	0.353	0.812	0.121	0.410
CD45, (PTPRC)	Receptor-type tyrosine-protein phosphatase C	0.090	-0.127	-0.091	-.199**	-0.040	-0.077	-.233**	-0.036
		0.244	0.099	0.236	**0.009**	0.600	0.319	**0.002**	0.645
CD52	CAMPATH-1 antigen	-0.016	-.151*	-.154*	-.332**	-0.049	0.093	-0.008	.162*
		0.839	**0.049**	**0.045**	**9.8E-06**	0.522	0.226	0.913	**0.034**
CD40, (TNFRSF5)	Tumour necrosis factor receptor superfamily member 5	-0.109	-0.021	-0.081	-.250**	-0.124	0.100	-.190*	-0.081
		0.156	0.790	0.293	**0.001**	0.106	0.196	**0.013**	0.292
CD267, (TNFRSF13B)	Tumour necrosis factor receptor superfamily member 13B	-0.050	-.178*	-.202**	-.204**	-0.120	0.124	0.093	.220**
		0.520	**0.020**	**0.008**	**0.008**	0.119	0.108	0.229	**0.004**
CD268, (TNFRSF13C)	Tumour necrosis factor receptor superfamily member 13C	0.068	-0.050	-0.055	-.215**	-0.017	-0.136	-0.007	0.059
		0.378	0.519	0.475	**0.005**	0.825	0.076	0.924	0.448
CD269, (TNFRSF17)	Tumour necrosis factor receptor superfamily member 17	-0.021	-.201**	-0.081	-.238**	-0.046	0.007	0.047	0.151
		0.783	**0.009**	0.295	**0.002**	0.552	0.923	0.546	0.050

Table 5. 12: Gene expression correlations between the T8 genes and a cluster of CD-encoding genes in the METABRIC patients, (n=1980).

Protein-encoding gene	Protein name	HSP90AA1	ARHGDIA	VDAC2	DYNLRB1	PPIA	FBLN1	RPS18	RPL10A
		Labelling of correlation coefficient values , (Red = negative correlation); (Green = positive correlation); (White = no correlation)							
		P value < 0.050 is in bold							
CD2	T-cell surface antigen CD2	0.011	0.041	-.044*	-.137**	0.006	.146**	0.029	.055*
		0.628	0.067	**0.049**	**8.3E-10**	0.789	**6.0E-11**	0.204	**0.015**
CD3G	T-cell surface glycoprotein CD3 gamma chain	-.047*	-0.027	-.101**	-.210**	-.062**	.080**	0.027	-0.009
		0.036	0.230	**6.1E-06**	**3.2E-21**	**0.006**	**0.0003**	0.234	0.695
CD3D	T-cell surface glycoprotein CD3 delta chain	-0.002	-0.018	-.045*	-.141**	.069**	.143**	.087**	.118**
		0.945	0.424	**0.046**	**3.2E-10**	**0.002**	**1.8E-10**	**0.0001**	**1.5E-07**
CD3E	T-cell surface glycoprotein CD3 epsilon chain	-.182**	.114**	-.178**	-.190**	-.155**	.072**	0.022	-.154**
		3.5E-16	**3.8E-07**	**1.4E-15**	**1.6E-17**	**3.6E-12**	**0.001**	0.329	**6.0E-12**
CD3Z, (*CD247*)	T-cell surface glycoprotein CD3 zeta chain	-.063**	0.024	-.083**	-.173**	-0.039	.136**	.049*	0.012
		0.005	0.295	**0.0002**	**8.7E-15**	0.080	**1.1E-09**	**0.028**	0.603
CD4	T-cell surface glycoprotein CD4	-.164**	.198**	-.099**	-.083**	-.067**	.079**	-.096**	-.135**
		2.1E-13	**7.0E-19**	**9.8E-06**	**0.0002**	**0.003**	**0.0004**	**0.00002**	**1.5E-09**
CD8A	T-cell surface glycoprotein CD8 alpha chain	0.010	-0.016	-.054*	-.158**	0.013	.132**	0.033	.062**
		0.660	0.466	**0.016**	**1.6E-12**	0.566	**4.1E-09**	0.143	**0.006**
CD8B	T-cell surface glycoprotein CD8 beta chain	0.036	-.058**	-0.022	-.106**	0.032	0.031	0.025	.092**
		0.114	**0.010**	0.323	**2.1E-06**	0.154	0.171	0.271	**0.00005**
CD28	T-cell-specific surface glycoprotein CD28	-.076**	-0.005	-.129**	-.155**	-.058**	.092**	-0.033	-0.021
		0.001	0.811	**8.0E-09**	**4.0E-12**	**0.010**	**0.00003**	0.142	0.341
CD80	T-lymphocyte activation antigen CD80	.120**	.086**	.058*	-0.025	.130**	-.162**	-.086**	-0.043
		9.0E-08	**0.0001**	**0.010**	0.269	**6.8E-09**	**4.4E-13**	**0.0001**	0.053
CD86	T-lymphocyte activation antigen CD86	-0.022	0.029	-.066**	-.171**	.045*	.077**	-.129**	-0.004
		0.324	0.202	**0.004**	**2.0E-14**	**0.047**	**0.001**	**7.7E-09**	0.866
CD19	B-lymphocyte antigen CD19	-.089**	0.043	-.108**	-.162**	-.079**	.045*	0.040	-0.026
		0.0001	0.055	**1.4E-06**	**3.6E-13**	**0.0004**	**0.044**	0.072	0.250
CD20, (*MS4A1*)	B-lymphocyte antigen CD20	-0.002	0.032	0.021	0.025	.111**	-.101**	.112**	.059**
		0.940	0.149	0.348	0.261	**8.1E-07**	**6.7E-06**	**5.7E-07**	**0.009**
CD22	B-cell receptor CD22	-.210**	.101**	-.151**	-0.036	-.135**	-0.010	-0.007	-.158**
		3.8E-21	**7.3E-06**	**1.3E-11**	0.113	**1.9E-09**	0.662	0.761	**1.5E-12**
CD45, (*PTPRC*)	Receptor-type tyrosine-protein phosphatase C	.089**	-.050*	0.015	-.110**	.105**	.071**	-.075**	.120**
		0.0001	**0.025**	0.503	**9.9E-07**	**2.9E-06**	**0.001**	**0.001**	**9.0E-08**
CD52	CAMPATH-1 antigen	0.043	-.057*	-0.021	-.130**	.119**	.172**	.054*	.217**
		0.055	**0.011**	0.350	**7.2E-09**	**1.2E-07**	**1.1E-14**	**0.016**	**1.7E-22**
CD40, (*TNFRSF5*)	Tumour necrosis factor receptor superfamily member 5	-.080**	.044*	-.072**	-.192**	-.056*	.142**	-.115**	-0.038
		0.0004	**0.049**	**0.001**	**6.8E-18**	**0.013**	**2.0E-10**	**2.8E-07**	0.091
CD267, (*TNFRSF13B*)	Tumour necrosis factor receptor superfamily member 13B	-.127**	0.005	-.147**	-.160**	-.105**	.142**	.060**	0.004
		1.3E-08	0.818	**4.6E-11**	**8.1E-13**	**2.6E-06**	**2.4E-10**	**0.008**	0.871
CD268, (*TNFRSF13C*)	Tumour necrosis factor receptor superfamily member 13C	-.141**	.102**	-.052*	-.051*	-0.013	-.130**	.113**	-.091**
		3.1E-10	**5.6E-06**	**0.020**	**0.023**	0.558	**7.3E-09**	**4.3E-07**	**0.00005**
CD269, (*TNFRSF17*)	Tumour necrosis factor receptor superfamily member 17	.061**	-.072**	-0.013	-.102**	.092**	.121**	.068**	.178**
		0.007	**0.001**	0.573	**5.9E-06**	**0.00004**	**6.5E-08**	**0.003**	**1.2E-15**

Table 5. 13: Gene expression correlations between the T8 genes and a cluster of tumour progression genes in the LVI-supervised patients, (n=170).

Protein-encoding gene	Protein name	HSP90AA1	ARHGDIA	DYNLRB1	VDAC2	PPIA	FBLN1	RPS18	RPL10A
		Labelling of correlation coefficient values , (Red = negative correlation); (Green = positive correlation); (White = no correlation)							
		P value < 0.050 is in bold							
SDC1	Syndecan-1	-0.058	-0.013	-0.081	-0.004	0.029	.338**	0.088	-0.036
		0.450	0.865	0.295	0.959	0.711	**6.4E-06**	0.251	0.638
CTNNB1	Catenin beta-1	.169*	-0.055	.192*	0.127	.236**	0.147	0.036	0.088
		0.028	0.475	**0.012**	0.100	**0.002**	0.056	0.643	0.252
FN1	Fibronectin	-0.035	.194*	0.085	0.001	.151*	.174*	-.244**	-.174*
		0.647	**0.011**	0.271	0.987	**0.049**	**0.023**	**0.001**	**0.024**
COL3A1	Collagen alpha-1(III) chain	-0.072	-0.029	-0.096	-0.121	-0.022	.675**	0.109	0.113
		0.350	0.712	0.213	0.117	0.775	**5.8E-24**	0.157	0.144
SNAI1	Zinc finger protein SNAI1	-0.046	-0.085	-0.053	-0.042	-0.091	-0.032	-0.055	-0.064
		0.555	0.272	0.491	0.586	0.239	0.676	0.474	0.410
SNAI2	Zinc finger protein SNAI2	-0.014	0.046	-0.073	-0.073	-0.067	.526**	-0.093	0.065
		0.854	0.553	0.345	0.342	0.383	**1.8E-13**	0.229	0.401
ETS1	Protein C-ets-1	-0.055	-0.057	-0.049	-0.019	.234**	.190*	.158*	.258**
		0.474	0.463	0.523	0.811	**0.002**	**0.013**	**0.040**	**0.001**
MKI67	Proliferation marker protein Ki-67	.204**	.318**	0.124	.297**	.181*	-.333**	-0.087	-.310**
		0.008	**0.00002**	0.107	**0.0001**	**0.018**	**9.1E-06**	0.257	**0.00003**
CDH1	Cadherin-1	.214**	.196*	.230**	0.013	0.097	-.289**	-.197*	-.188*
		0.005	**0.011**	**0.003**	0.867	0.208	**0.0001**	**0.010**	**0.014**
CDH2	Cadherin-2	.230**	0.139	0.003	.200**	0.150	0.112	-0.022	-0.070
		0.003	0.071	0.970	**0.009**	0.051	0.145	0.771	0.361
VIM	Vimentin	-0.027	0.037	-0.032	0.064	0.059	.502**	0.062	.152*
		0.725	0.628	0.679	0.405	0.448	**3.1E-12**	0.425	**0.048**
CXCL12	Stromal cell-derived factor 1	-.195*	-.194*	-0.137	-.204**	-.193*	.768**	0.091	.293**
		0.011	**0.011**	0.074	**0.008**	**0.012**	**2.6E-34**	0.237	**0.0001**
TGFB1	Transforming growth factor beta-1	-0.080	-.152*	-0.008	0.038	0.068	0.050	0.075	0.116
		0.302	**0.049**	0.920	0.621	0.380	0.518	0.328	0.132
MET	Hepatocyte growth factor receptor	0.007	0.101	-0.058	0.112	.158*	-0.078	0.001	-0.029
		0.926	0.190	0.456	0.145	**0.040**	0.311	0.987	0.712
PDGFC	Platelet-derived growth factor C	-0.047	-0.060	0.095	0.011	0.095	.414**	0.019	0.140
		0.540	0.435	0.217	0.884	0.217	**2.0E-08**	0.811	0.069
FAP	Prolyl endopeptidase FAP	0.025	-0.121	-0.109	-0.102	0.033	.601**	-0.044	0.078
		0.743	0.115	0.158	0.187	0.670	**4.3E-18**	0.572	0.312
THY1, (CD90)	Thy-1 membrane glycoprotein	-.230**	0.103	-.151*	-.238**	-.228**	.595**	-0.081	-0.114
		0.003	0.181	**0.049**	**0.002**	**0.003**	**1.3E-17**	0.294	0.139
CD24	Signal transducer CD24	0.105	-0.007	-0.086	0.050	0.035	-0.034	0.136	-0.037
		0.172	0.929	0.265	0.520	0.646	0.664	0.076	0.633
CD44	CD44 antigen	-0.051	-0.076	-.197**	-0.015	-.172*	0.028	-0.106	-.193*
		0.513	0.326	**0.010**	0.850	**0.025**	0.717	0.169	**0.012**
S100A4	Protein S100-A4	0.043	0.077	-0.052	0.066	0.107	.192*	-0.122	0.039
		0.575	0.321	0.500	0.391	0.166	**0.012**	0.112	0.618
CAV1	Caveolin-1	-.165*	-.227**	-.166*	-.158*	-.179*	.652**	0.132	.307**
		0.031	**0.003**	**0.031**	**0.039**	**0.020**	**6.3E-22**	0.087	**0.00004**

Table 5. 14: Gene expression correlations between the T8 genes and a cluster of tumour progression genes in the METABRIC patients, (n=1980).

Protein-encoding gene	Protein name	HSP90AA1	ARHGDIA	DYNLRB1	VDAC2	PPIA	FBLN1	RPS18	RPL10A
		Labelling of correlation coefficient values , (Red = negative correlation); (Green = positive correlation); (White = no correlation)							
		P value < 0.050 is in bold							
SDC1	Syndecan-1	-0.017	.110**	-0.022	-0.013	0.027	.345**	-0.007	-.065**
		0.452	**9.9E-07**	0.337	0.560	0.232	**2.2E-56**	0.753	**0.004**
CTNNB1	Catenin beta-1	.328**	-.209**	-0.029	.184**	.196**	.145**	-.070**	.366**
		5.2E-51	**5.5E-21**	0.204	**1.7E-16**	**1.7E-18**	**8.7E-11**	**0.002**	**1.0E-63**
FN1	Fibronectin	.050*	.078**	-0.012	0.017	.093**	.248**	-.246**	0.015
		0.025	**0.001**	0.587	0.462	**0.00003**	**3.1E-29**	**1.4E-28**	0.495
COL3A1	Collagen alpha-1(III) chain	.150**	-0.035	-0.008	-0.034	.092**	.696**	-0.019	.252**
		1.7E-11	0.119	0.710	0.127	**0.00004**	**3.0E-287**	0.401	**4.4E-30**
SNAI1	Zinc finger protein SNAI1	-0.036	0.036	-0.016	0.001	-.050*	0.002	-0.042	-.087**
		0.107	0.105	0.473	0.980	**0.026**	0.926	0.061	**0.0001**
SNAI2	Zinc finger protein SNAI2	.150**	-0.041	-.088**	0.024	.104**	.516**	-.055*	.280**
		1.9E-11	0.070	**0.00008**	0.286	**3.4E-06**	**5.5E-135**	**0.014**	**4.9E-37**
ETS1	Protein C-ets-1	-0.023	0.032	.212**	.052*	.368**	.179**	.343**	.274**
		0.305	0.150	**1.4E-21**	**0.022**	**2.1E-64**	**9.8E-16**	**9.4E-56**	**2.5E-35**
MKI67	Proliferation marker protein Ki-67	0.022	.160**	-0.035	.174**	-.055*	-.332**	0.039	-.281**
		0.333	**8.1E-13**	0.119	**7.1E-15**	**0.015**	**4.5E-52**	0.085	**2.8E-37**
CDH1	Cadherin-1	.204**	0.029	.156**	.073**	.134**	-.197**	-.105**	0.005
		5.7E-20	0.191	**2.9E-12**	**0.001**	**2.4E-09**	**9.0E-19**	**2.6E-06**	0.823
CDH2	Cadherin-2	0.044	.105**	-0.016	.070**	.055*	.124**	-0.043	0.027
		0.050	**2.6E-06**	0.469	**0.002**	**0.015**	**3.5E-08**	0.053	0.231
VIM	Vimentin	.259**	-.065**	-0.040	.125**	.182**	.502**	0.018	.406**
		1.1E-31	**0.004**	0.078	**2.5E-08**	**3.9E-16**	**5.0E-127**	0.427	**1.5E-79**
CXCL12	Stromal cell-derived factor 1	.060**	-.123**	-.144**	-.085**	-0.013	.676**	-0.035	.318**
		0.007	**3.8E-08**	**1.3E-10**	**0.0001**	0.576	**1.7E-264**	0.124	**7.3E-48**
TGFB1	Transforming growth factor beta-1	-0.034	-0.030	-0.017	0.007	-0.020	0.004	-0.025	-0.026
		0.129	0.178	0.442	0.762	0.368	0.866	0.270	0.249
MET	Hepatocyte growth factor receptor	.087**	0.028	.094**	.143**	.265**	0.019	.154**	.161**
		0.0001	0.206	**0.00003**	**1.5E-10**	**3.3E-33**	0.402	**5.0E-12**	**6.4E-13**
PDGFC	Platelet-derived growth factor C	.277**	-.145**	.151**	.120**	.372**	.317**	0.023	.478**
		4.2E-36	**8.2E-11**	**1.3E-11**	**8.5E-08**	**4.3E-66**	**1.9E-47**	0.307	**1.6E-113**
FAP	Prolyl endopeptidase FAP	.164**	-.059**	-.074**	-.056*	.087**	.648**	-.167**	.223**
		2.5E-13	**0.009**	**0.001**	**0.012**	**0.0001**	**4.4E-236**	**7.3E-14**	**1.2E-23**
THY1, (CD90)	Thy-1 membrane glycoprotein	-.214**	.251**	-.096**	-.204**	-.205**	.580**	-.126**	-.115**
		6.8E-22	**7.9E-30**	**0.00001**	**4.6E-20**	**2.8E-20**	**4.7E-178**	**1.6E-08**	**2.6E-07**
CD24	Signal transducer CD24	.121**	0.023	-.089**	0.031	0.037	.085**	.052*	-0.017
		6.3E-08	0.315	**0.00007**	0.172	0.101	**0.0001**	**0.020**	0.451
CD44	CD44 antigen	.133**	-.054*	-.075**	.065**	-0.002	.055*	0.013	.091**
		2.7E-09	**0.015**	**0.001**	**0.004**	0.943	**0.014**	0.576	**0.0001**
S100A4	Protein S100-A4	.137**	-0.009	-0.012	.073**	.228**	.367**	-.077**	.314**
		1.0E-09	0.693	0.597	**0.001**	**8.2E-25**	**3.4E-64**	**0.001**	**2.0E-46**
CAV1	Caveolin-1	.161**	-.168**	-.093**	0.001	.093**	.546**	0.023	.408**
		5.3E-13	**4.7E-14**	**0.00003**	0.952	**0.00003**	**1.9E-154**	0.301	**2.9E-80**

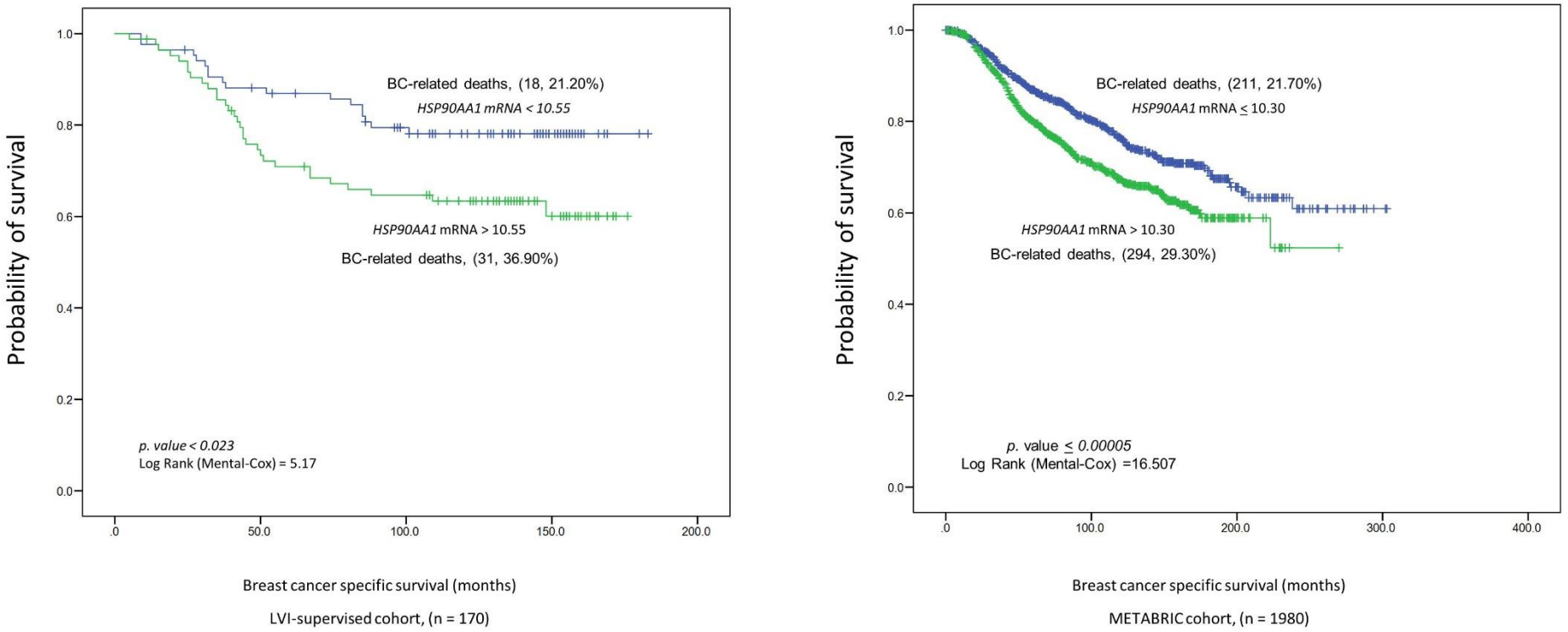


Figure 5. 2: Associations between the altered expression of *HSP90AA1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

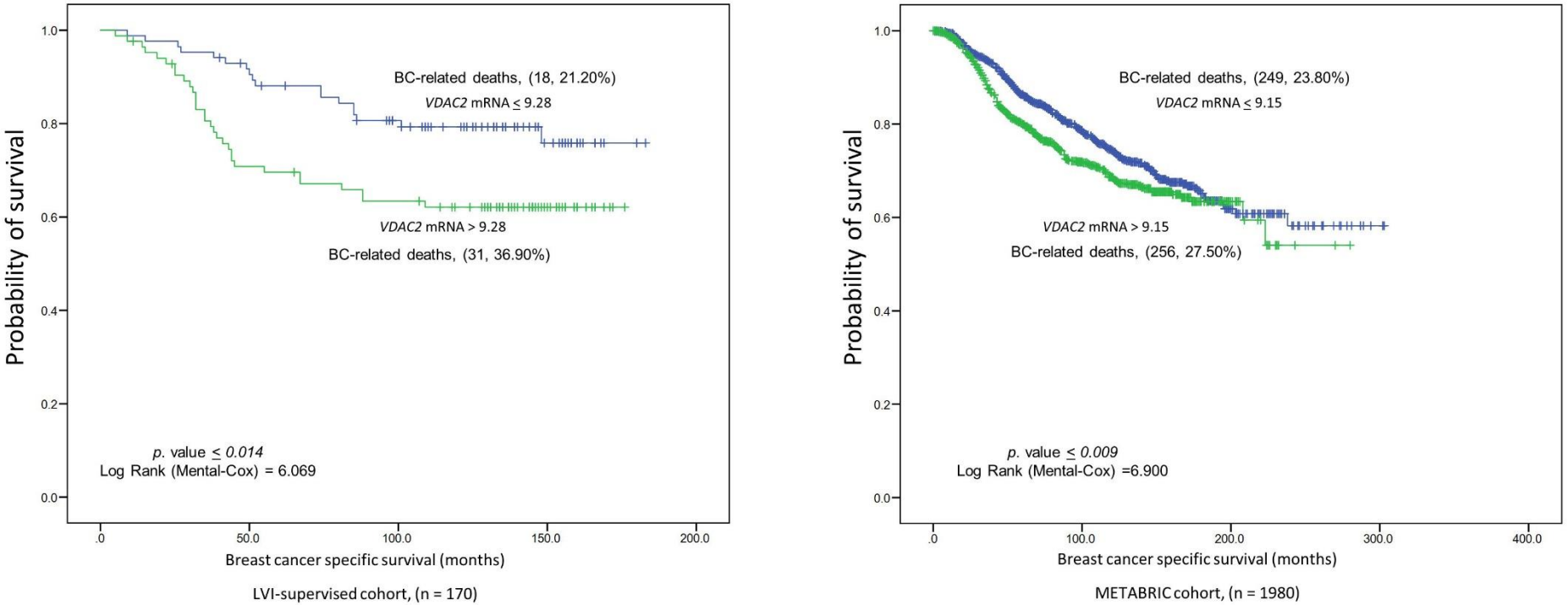


Figure 5. 3: Associations between the altered expression of *VDAC2* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

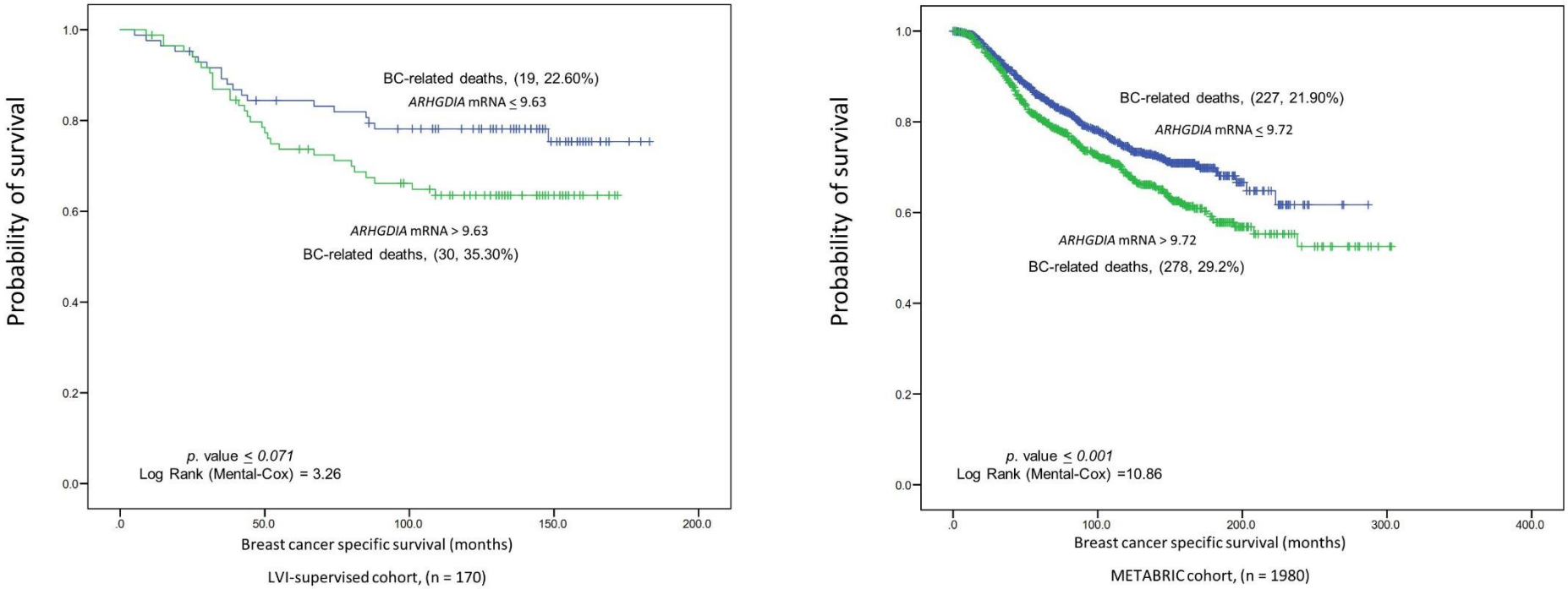


Figure 5. 4: Associations between the altered expression of *ARHGDIA* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

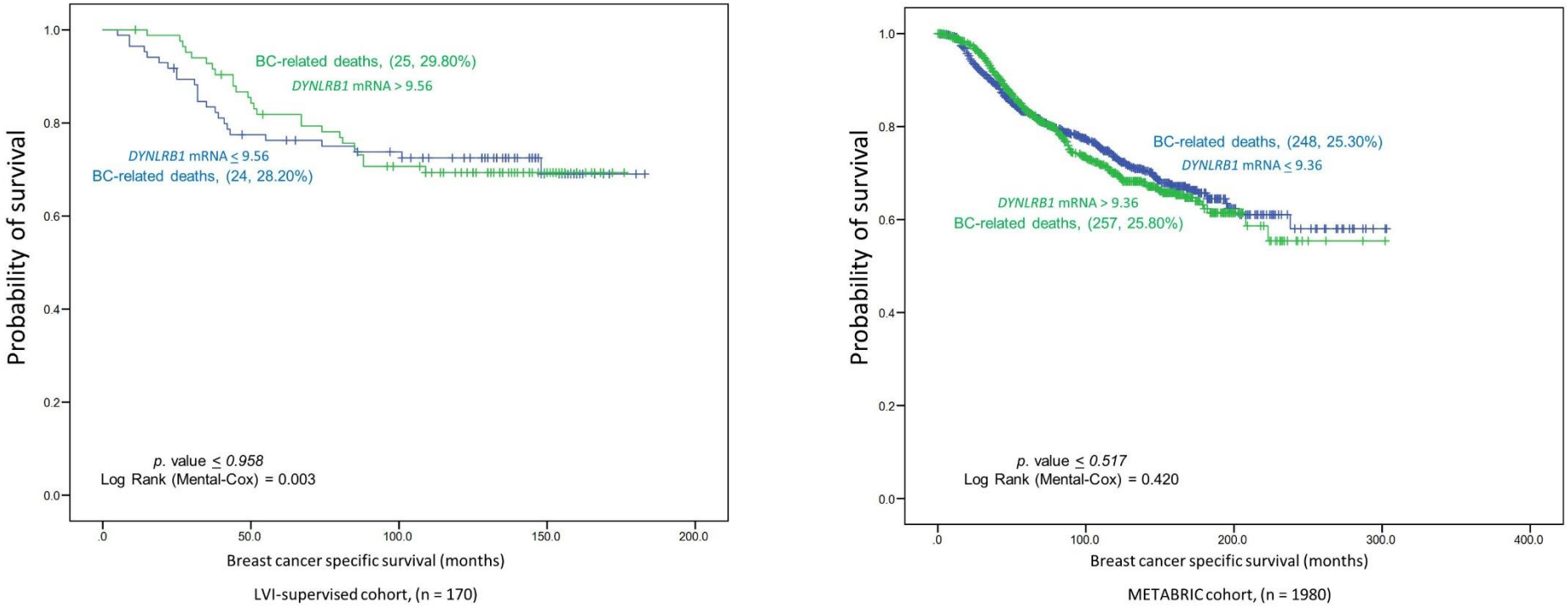


Figure 5. 5: Associations between the altered expression of *DYNLARB1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

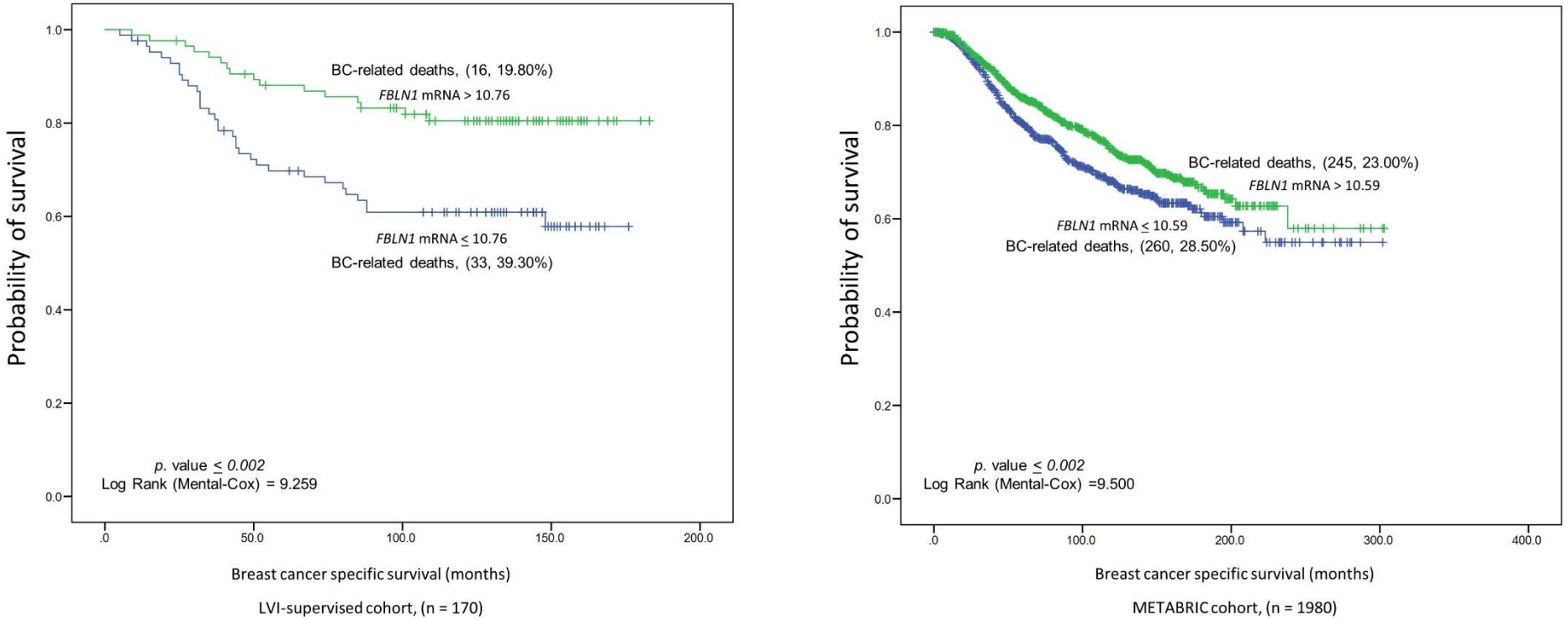


Figure 5. 6: Associations between the altered expression of *FBLN1* mRNA and breast cancer specific survival in the LVI-supervised and METABRIC cohorts

Table 5. 15: Multivariate Cox regression analyses for prediction of breast cancer-specific survival, altered gene expression of *HSP90AA1*, *VDAC2*, *ARHGDIA*, *DYNLRB1*, and *FBLN1*.

Breast cancer specific survival (BCSS), n= 1869[0]				
	Hazard ratio (95% CI)	**Lower**	**Upper**	***p*-value**
***HSP90AA1* mRNA**	1.222	1.016	1.469	**0.033**
Tumour grade	1.243	1.050	1.471	**0.012**
Nodal staging*	1.849	1.643	2.082	**2.4E-24**
Tumour size	1.544	1.267	1.883	**0.00002**
Estrogen expression**	.714	.577	.884	**0.002**
Her2 expression***	1.466	1.156	1.860	**0.002**
	Hazard ratio (95% CI)	**Lower**	**Upper**	***p*-value**
***VDAC2* mRNA**	1.093	.913	1.308	0.335
Tumour grade	1.267	1.071	1.499	**0.006**
Nodal staging*	1.841	1.635	2.073	**5.8E-24**
Tumour size	1.549	1.270	1.889	**0.00002**
Estrogen expression**	.720	.581	.891	**0.003**
Her2 expression***	1.499	1.183	1.900	**0.001**
	Hazard ratio (95% CI)	**Lower**	**Upper**	***p*-value**
ARHGDIA mRNA	1.241	1.035	1.487	**0.019**
Tumour grade	1.249	1.056	1.478	**0.009**
Nodal staging*	1.857	1.650	2.090	**9.3E-25**
Tumour size	1.529	1.253	1.865	**0.00003**
Estrogen expression**	0.715	0.578	0.885	**0.002**
Her2 expression***	1.479	1.167	1.875	**0.001**
	Hazard ratio (95% CI)	**Lower**	**Upper**	***p*-value**
***DYNLRB1* mRNA**	1.073	.892	1.290	0.454
Tumour grade	1.268	1.072	1.499	**0.006**
Nodal staging*	1.843	1.637	2.075	**4.1E-24**
Tumour size	1.555	1.275	1.895	**0.00001**
Estrogen expression**	.700	.562	.872	**0.001**
Her2 expression***	1.494	1.179	1.895	**0.001**
	Hazard ratio (95% CI)	**Lower**	**Upper**	***p*-value**
***FBLN1* mRNA**	.860	.718	1.031	.102
Tumour grade	1.249	1.055	1.479	**0.010**
Nodal staging*	1.854	1.647	2.086	**1.4E-24**
Tumour size	1.543	1.265	1.881	**0.00002**
Estrogen expression**	.720	.582	.891	**0.003**
Her2 expression***	1.501	1.184	1.902	**0.001**

*** NPI staging system; ** IHC; *** IHC/in situ hybridisation.**
[0] Sum of patients does not equal to 1980 because of missing cases among different categories.

5.4. Discussion

BC-related deaths in the UK have decreased by 32% since 1970. Regardless of the positive impression of BC mortality data in the UK, the identification of patients at higher risk than others is essential to improving prognosis accuracy, therapy effectiveness, and clinical management of BC patients. Therefore, deciphering biochemical networks of LVI could be employed to discover novel prognostic and therapeutic targets at the molecular level. Although several studies have investigated differential gene expression, few focus on LVI drivers in BC. Comprehensive and systematic analysis of the sophisticated LVI mechanisms is facilitated by applying high-throughput molecular technologies and bioinformatics tools, complemented by robust and validated clinicopathological profiles.

In this study, we identified 1,054 differentially expressed mRNAs between the LVI^+ and LVI^-/LN^- subgroups of invasive BC by applying RNA microarray. Under the same sample selection criteria, we identified 370 differentially expressed proteins by applying MS procedures between two cohorts of patients with the luminal-A invasive BC subtype with a meticulous definition of their LVI status. The latter cohorts were constructed in the same LVI-administrated manner as for the microarray experiment. Although the comparison was assessed on two different levels of molecular expression for two different cohorts of patients, a constant and overlapping cluster of 56 mRNAs/proteins was identified.

The upregulation of the T8 molecules was statistically correlated with BC patients with informative LVI profiles. These molecules were firmly assessed statistically to maximise the elimination of the effects of passenger events of the LVI-correlated biological mechanisms, and to specify the most accurate drivers/inhibitors for LVI. This study suggests *HSP90AA1* (HSP90-α) as a driver hub that hallmarks several LVI mechanisms under various biological contexts. Alternatively, *FBLN1* (Fibulin-1) was designated as a potent core that is associated with LVI inhibition when it is upregulated in particular pathways. The effects of the altered expression of the T8 mRNAs/proteins on the patients' clinicopathological parameters, the regulation of selected photolytic enzymes, the interactions with immune cells, and the associations with invasion-related molecules will be discussed. The discussion will focus on the unique gene expression trends for a gene between the $LVI^{+/-}$ cohorts and the METABRIC cohort.

5.4.1. ***HSP90AA1*: A molecular chaperone associated with LVI and poor prognosis in invasive BC**

5.4.1.1. *Overview*

The *HSP90AA1* gene (HGNC ID: 5253) is a protein-encoding gene located at 14q32.31 with 13 known exons. Heat shock protein 90 alpha family class A member 1 (HSP90-α, UniProtKB entry ID: P07900) is the full name of the encoded cytosolic protein by this gene (84.660 Da). Generally, heat shock proteins (HSPs) operate as molecular chaperones that assist protein folding, refolding of misfolded proteins or peptides, stabilisation of interacting or client proteins, and possibly degradation of irreparable proteins. Generally, the effective function of HSPs requires ATP-binding, the presence of co-chaperones, and client proteins [327]. HSPs are overly expressed as a cellular response to different stress activators. A few examples of HSPs promoters include acute or chronic inflammation, destructive activity of reactive oxygen species (ROS), microbial infections, and cancer, ultimately. Hence, HSPs are active under myriad signalling pathways, such as the HER2 signalling pathway, mitochondrial transport, positive regulation of the nitric oxide biosynthetic process, and VEGF receptor [186].

The biological general repository for interaction datasets (BioGRID, version 3.4) was used to explore genetic and protein interactions for *HSP90AA1*. A grid of biochemical connections shows 232 (interacting proteins, clients) with 566 edges (interactions) for *HSP90AA1*. Based on the GO KB, *HSP90AA1* is associated with 16 various biological processes and ten distinctive molecular functions [328, 329]. A literature search of the terms: [breast AND ("cancer" OR "carcinoma") AND (lymphovascular) AND (invasion) AND (HSP90AA1)] in the database resources of the National Centre for Biotechnology Information (NCBI) and the PubMed search engine [330] yielded no published studies at all. Although a massive number of published studies have discussed the associations between HSP90 proteins and cancer, the specific roles of *HSP90AA1*/HSP90-α in the LVI mechanisms are still elusive. Our aim in the following section is to discuss our current findings, highlighting the invasive mechanisms that involve the aberrant expression of *HSP90AA1* and LVI in invasive BC patients.

5.4.1.2. *Clinicopathological relevance of the HSP90AA1 altered gene expression*

Our current results pair the upregulation of *HSP90AA1* transcription with the unfavourable prognosis of invasive BC patients. This study provided statistical evidence on the linkage between the upregulation of HSP90AA1 and LVI occurrence, positive nodal involvement, large tumour size, high tumour grade, and poor NPI in two distinctive cohorts of invasive BC. The efficient and malicious roles of *HSP90AA1* may characterise patients with $ER^+/HER2^+$ profiles as the luminal-B and HER2-enriched molecular intrinsic subtypes, and IntClust-1, IntClust-5, and IntClust-9 in

particular. The luminal-B subtype was stated as having the highest association with LVI incidences [226], and IntClust-5 exhibited *ERBB2* amplification, high tumour grade, and extensive LNM [39].

Regulatory associations between kinase activity and kinase stability of the oncogenic tyrosine kinase *ERBB2* and the expression of HSP90-α were suggested [331-333]. A recent study suggested that patients of the HER2 subtype who highly expressed HSP90 protein and illustrated an elevated index for Ki67 expression represented an unequalled molecular class that should be investigated [334]. Luminal-B and HER2-enriched are highly proliferative and aggressive molecular subtypes of invasive BC that require neoadjuvant treatment with taxane-based chemotherapy in combination with trastuzumab and pertuzumab as a standard preoperative therapeutic regimen [335]. Therefore, we aimed to explore whether there is a possible association between the upregulations of *HSP90AA1* and *MKI67*, the general biomarker for proliferation and poor prognosis [336], in promoting LVI mechanisms. We pointed out a positive and strong gene expression correlation between *HSP90AA1* and MKI67 in the $LVI^{+/-}$ cohort in particular (r=0.204, p-value $\leq$0.008). This association was absent in the larger cohort of METABRIC (r=0.022, p-value $\leq$0.333). This suggests another signalling axis of proliferation and invasion involving HSP90-α and Ki67 proteins within LVI molecular contexts. $HER2^{+}/ER^{+}$ or $HER2^{+}/ER^{-}$ patients with LVI definite profiles and upregulated expression of HSP90-α/Ki67 indices may signify two novel molecular classes of invasive BC patients that are worth additional care – primarily, before surgery.

Our results demonstrated a significant increase in the percentage of $LVI^{+/-}$ patients with $HSP90AA1^{+}$/LN-stage-II. This increasing trend was also seen in the patients of the METABRIC cohort, but the percentage of patients was 12% lower than of those patients with LVI data. According to the NPI staging system, patients labelled as stage-II represent early LNM with three or fewer LN with metastatic tumours. The occurrence of invasive legions in resected BC tissues and regional LN of preoperative BC patients who received neoadjuvant chemotherapy may lessen the success of surgical intervention, especially in highly proliferative and aggressive types [337]. Therefore, *HSP90AA1* mRNA expression was recently suggested as a part of the pathologic complete response (PathCR) for BC patients, along with the universal proliferate biomarker ki67 (*MKI67*) [338]. Total RNA isolated from 93 BC patients was analysed by quantitative real-time PCR for expression status of *HSP90AA1*, *MKI67*, *HSP90AB1*, and cyclin-dependent kinase 1 (*CDK1*). All RNA specimens in the latter study were extracted from single-core needle biopsies collected before surgery and after administration of anthracycline and taxane.

HSP90AA1 and *HSP90AB1* transcripts were found to be downregulated in patients with PathCR. On the other hand, the researchers stated that there was no relevant difference in the *MKI67* and *CDK1* levels of expression among PathCR and no PathCR categories. The quantitative relationship between *HSP90AA1* and *CDK1* showed statistical significance and applicable diagnostic implications for PathCR patients. Low expression of *HSP90AA1* may tag patients with a satisfactory response to chemotherapy treatment, less the tendency to develop LVI, and less LN involvement. The last statement is highly comparable to our current findings. Low expression of *HSP90AA1* was associated with fewer nodal involvements and no LVI events in the $LVI^{+/-}$ and METABRIC cohorts.

The results of our study also suggested a preferential upregulation of *HSP90AA1* in a phenotype of aggressive $HER2^{-}$ tumours. In a gene expression study on a large pooled cohort of 4,010 invasive BC, upregulated expression of *HSP90AA1* and *HSP90AB1* defined a subgroup of BC patients with poor prognosis among patients with $HER2^{-}/ER^{+}$. In addition, the elevation of cytoplasmic HSP90-α and HSP90-β proposed a subpopulation of triple-negative BC patients with a high risk of recurrence [339]. High expression of these two transcripts with the amplification of the transcription factor of heat shock factor protein 1 (*HSF1*) may define a subpopulation of BC patients that are susceptible to developing more aggressive molecular types of BC. As our current results indicated a possible synergism between LVI and the upregulation of *HSP90AA1* in the luminal-B and $HER2^{-}$enriched intrinsic subtypes, we highly recommend a systematic investigation between the $HER2^{-}/ER^{+}$ and triple-negative classes of BC patients. This comparative study may reveal clearer insights into LVI regulatory mechanisms among different statuses of the hormonal receptors in invasive BC. Additionally, we may define a novel subclass of triple-negative BC patients based on their LVI status and HSP90 protein expression that need more attention as the option of hormonal therapy is not valid in their case.

5.4.1.3. *Pivotal gene expression associations between HSP90AA1 expression and matrix metalloproteinases*

Regardless of molecular subtypes, the results of our two cohorts of patients showed strong and progressive correlations between upregulated *HSP90AA1* mRNA and both higher tumour grade and larger tumour size. These associations with tumour grade and size led us to question whether the expression of *HSP90AA1*/HSP90-α may be influenced by or even dictate the expression of other client proteins in the tumour microenvironment (TME). In a recent study, the authors stated that there are associations between extracellular expression of the

HSP90-α protein (eHSP90-α) and the increase of cancer invasive and metastatic properties through molecular crosstalk with other proteins occurring intracellularly in diverse cancers [340]. In addition, biochemical activities for eHSP90-α with other client proteins deposited in the extracellular matrix were reported previously as increasing cancer cell motility, migration, and invasion abilities [341].

Matrix metalloproteinases (MMPs) are a decent representative family for secreted proteins with proteolytic activities among other pro-invasion cytokines [342]. MMPs belong to a larger superfamily of proteases known as the metzincin superfamily [343]. MMPs are calcium-dependent, zinc-containing endopeptidases, necessary for the degradation of extracellular matrix (ECM) proteins. Prognostic impacts on patients' survival and tumour invasion for some of the members of the metzincin superfamily were reported recently in melanoma [344] and BC [345]. MMPs are well known as key regulators for tissue remodelling, adhesion, migration, and tumour invasion when they are biologically activated in several malignancies, including BC [346-351]. Therefore, we searched for previous studies that showed a declared relationship between the expression of MMPs and *HSP90AA1* mRNA or its protein in invasive BC.

Adding to the aforementioned suggestions, MMPs were reported to be overexpressed in higher grades of invasive BC; in particular, MMP1, MMP9, MMP11, and MMP13 [352]. In the METABRIC cohort, the results of the gene expression correlations for *HSP90AA1* were positive against *MMP1*, *MMP10*, *MMP13*, and *MMP19*. Interstitial collagenase, *MMP1*, is one of the candidate markers for high-grade/LVI^{+} tumours [257]. Stromelysin-2, *MMP10*, increases triple-negative BC adhesion and degradation of the major components for basement membranes, laminin [353]. Collagenase-3, *MMP13*, was verified to be regulated through the binding of neuregulins to the ErbB/Her subfamily of receptor tyrosine kinases [351]. *MMP13* and *MMP19* were described as optimisers of BC progression signalling for efficient BC invasion and metastasis, and their expression may stimulate the protein expression of other MMPs [354].

In the LVI$^{+/-}$ patients, the only positive gene expression correlation that we identified for *HSP90AA1* against all of the tested MMP encoding genes was with *MMP10*, although it was at the threshold of statistical significance. Our results revealed an opposite gene expression correlation for *HSP90AA1* against *MMP2* (negative) and *MMP10* (positive) in both the LVI$^{+/-}$ and the METABRIC cohorts. We applied further correlation analyses that showed a positive expression fashion between *MMP2* and *MMP10* in the METABRIC cohort, but they did not show any statistical relevance in the LVI$^{+/-}$ samples (results not shown). Pair-wise statistical

methods for *MMP* co-expression against the clinicopathological parameters revealed a sort of regulation for the expression pattern between *MMP2* and *MMP10* in a previous exploratory study on primary luminal-A BC [355]. MMP2 was hypothesised to be activated through the extracellular HSP90-α (eHSP90-α), leading to an increase in invasion tumour migration capabilities of the triple-negative BC cell line, MDA-MB-231 [356]. Our results demonstrated a significant upregulation of *HSP90AA1* in the IntClust-10 patients of the METABRIC cohort (65.9%), which is a cluster of patients that consists majorly of the triple-negative BC molecular subtype. A basic comparison between the distributions of patients between IntClust-10 in the $LVI^{+/-}$ cohort against the METABRIC showed no difference in the number of patients between the up or down expressions of *HSP90AA1*. Based on these findings, we suggest that the associations between *HSP90AA1, MMP2,* and *MMP10* are multifactorial. *HSP90AA1 and MMP10* may collaborate synergistically to drive the invasion mechanisms of luminal-A tumours that upregulate key genes of the basal intrinsic subtype.

The gene expression of *HSP90AA1* in the $LVI^{+/-}$ patients showed exclusive and negative gene expression correlation against *MMP28*. The activities of proteolytic enzymes in the tumour epithelium are, typically, associated with the progression of different cancers. Nonetheless, *MMP28* was found to be downregulated in clinical stages 2–4 of colorectal cancer patients [357]. Downregulation in the expression of MMP28 at the protein level was shown to indicate a malignant transformation in normal tissues of the intestine, and the overexpression of MMP26 with the low levels of MMP19 and MMP28 may signify inflammatory bowel disease (IBD) and colon cancer [358]. Alternatively, high levels of MMP28 protein expression were suggested to be correlated with vascular invasion, LNM, and distant metastasis in gastric cancer (GC) patients [359]. The expression of MMP28 seems to be tissue dependent, and there are specific biological contexts that dictate the upregulation or downregulation of MMP28.

The effects of *MMP28* gene expression in invasive BC patients were stated in a proposed seven-gene prognostic signature (*AR*, *ESR2*, *GATA3*, *GBX2*, *KRT16*, *MMP28*, *WNT11*) for a subgroup of chemo-insensitive triple-negative BC patients [360]. The expressions of *WNT11* and *MMP28* were positively detected in the cancer stem cell-like, mesenchymal, and mesenchymal-like triple-negative BCs. The authors stated that triple-negative BCs that expressed the luminal-like genes, AR and GATA3, were relevantly different from triple-negative BCs expressing stem-cell biomarkers. The upregulation of *MMP28* is, potentially, exclusive for triple-negative BC, but not for the luminal-A class. Survival analysis, log-rank (Mantal-Cox), on

two pooled cohorts of ER^- and ER^+ tumours revealed shorter BCSS for the ER^-/*MMP28*$^+$ patients (p-value $\leq$0.0001). There was no significant difference in BCSS for patients with ER^+/*MMP28*$^+$ profiles (p-value $\leq$0.09) [188]. As the sample selection of our LVI/MS study was based exclusively on BC cases of the ER^+/$HER2^-$ subtype, we assumed that the sample selection may have eliminated the hostile effects of *MMP28* on the studied patients. The expression association between *HSP90AA1* and *MMP28* and the LVI mechanisms is a novel finding for ER^+ BC patients that requires better understanding. In addition, it could be a promising therapeutic target to be scrutinised for triple-negative BC patients in the future. The discrepancies in the conclusions of the previous studies and our results validate the need for further investigations on the biochemical reactivity between HSP90-α and MMP28 in different BC subtypes, especially for the triple-negative BC patient whose therapeutic options are limited. Although several MMPs are well known to be the cornerstone of the invasion-metastasis cascade of events, the logical concordance for their mRNA-to-protein expression can be arguable. On the one hand, the approach of transcriptomic studies can provide high-throughput data to analyse and identify novel LVI targets. On the other hand, studying these novel transcripts (e.g., *MMP28* VS. *HSP90AA1*) functionally on the protein level can explain their mode of biological actions. A limitation of mRNA-based data is that it neither reveal any PTMs on a translated protein nor any preferential cellular localisation for it. The current findings of this study suggest studying the LVI-regulations between the identified T8 list of mRNA/proteins and the different MMPs on both RNA and protein molecular levels.

5.4.1.4. *Key gene expression associations between HSP90AA1 and immune cells*

Distinctive results were obtained from the gene expression correlation analyses for *HSP90AA1* expression against the investigated CD encoding genes in the $LVI^{+/-}$ and the METABRIC cohorts. Multiple negative expression associations for several CD encoding genes were pointed out in the METABRIC cohort. On the other hand, a single and positive gene expression relationship was identified between *HSP90AA1* and the gene expression of T-lymphocyte activation antigen protein or CD80 (also known as B7-1). The expression of CD80 is limited on the cell surface in terms of the number of antigen-presenting cells (APCs) such as activated B-cells, dendritic cells, and macrophages [361]. CD80 interacts with either CD25 or CD152 (CTLA-4) in an essential co-stimulation mechanism for natural killer cells (NKs) and the tumour specific cytotoxic T lymphocytes (CTL) after the antigen presentation on major histocompatibility complex (MHC) class II. However, an early study stated

that the expression of CD80 was immunohistochemically positive on cancer cells and tumour-infiltrating lymphocytes (TILs) [362]. Protein expression of the stimulatory molecules of the immune system by the cancer cells is questionable.

The employing of immune-related proteins by tumour cells was reported previously as an immunosuppressing machinery to evade the host immune cells in the TME. For example, the abnormal expression of interleukin-6 (IL-6), a crucial cytokine during metastasis, may facilitate tumour progression and metastasis via Janus-activated kinase/signal transducer and activator of transcription 3 pathway [363]. As our results labelled *HSP90AA1* with strong associations with LVI and unfavourable survival, and as *HSP90AA1* is positively correlated with *CD80*, we highly recommend the application of dual-staining IHC to invasive BC tissue sections to determine a prognostic relevance for *HSP90AA1* and *CD80* on the protein level. In particular, when we noticed a negative gene expression correlation for *CD80* with *FBLN1*, one of the good prognosis genes in our study that seems to suppress LVI (section 12.4.6). Currently, no pathologists would use the expression of CD80 for prognostic purposes, but deciphering the molecular regulation of CD80 in BC invasion and its related invasion biomarkers may alter this trend. Focusing on factors of immune deficiencies in the TME may reveal a novel cross-talk in the CD80/HSP90-α axis that drives LVI in invasive BC.

5.4.1.5. *HSP90AA1 gene expression and tumour microenvironment and progression biomarker*

We assessed the gene expression of *HSP90AA1* against several genes related to the tumour aggressive behaviour. The number of significant and positive gene expression correlations are much larger in the METABRIC cohort than those in the $LVI^{+/-}$ patients. The top three gene expression associations of *HSP90AA1* for the METABRIC's patients were with *CTNNB1*, *PDGFC*, and *VIM*. In the $LVI^{+/-}$ patients, *HSP90AA1* gene expression was associated positively with only four genes, *MKI67*, *CDH1*, *CDH2*, and *CTNNB1*. Although the biological roles of *MKI67* and *CDH2* in aggressive tumours are renowned, they did not show any associations with the gene expression of *HSP90AA1* in patients of the METABRIC cohort. *CAV1* and *CXCL12* were negatively correlated with *HSP90AA1* in the $LVI^{+/-}$ patients; however, the same genes exhibited positive relationships with *HSP90AA1* in patients of the METABRIC cohort.

High levels of the proliferation marker protein Ki67 indicate shorter disease-free survival and an unfavourable outcome in early BC [364]. Ki67 stands out as an independent predictive biomarker for BC patients and their response to

neoadjuvant chemotherapy [336, 365]. The protein expression of HSP90 was synergistically inhabited with 17-Allylamino-17-demethoxy-geldanamycin (17-AAG) and enzalutamide, an antagonist for androgen in PrC. The co-inhibition of HSP90 with both of the aforementioned drugs in the LuCaP35CR xenografts significantly decreased the expression of Ki67 [366]. Interestingly, another IHC study indicated that higher expression levels of HSP90 and Ki67 were detected in the pterygium epithelium of the eye in comparison to normal conjunctiva [367]. Pterygium is usually not malignant; however, a previous study cautioned that higher vascularity in the reactive pterygium may indicate malignant changes [368]. We recommend a thorough investigation of the HSP90-α/Ki67 co-expression in early-stage BC patients with LVI^+ profiles.

In normal cells, HSP90-α plays an important role as a signal transduction maintainer for various pathways through interaction with different client proteins. Altered expression of HSP90-α may allow oncoproteins to sustain proteostatic signalling or even gain evolved functions in progressive tumours [369]. The potentiality for HSP90-α /*HSP90AA1* to be involved in trans-endothelial migration (TEM) is strongly suggested. The process of TEM comprises the cluster of proteins encoded by the *CTNNB1*, *CDH1*, and *CDH2* genes that interact biologically to successfully govern the TEM-invasion cascade.

The proto-oncogene c-Src, a non-receptor tyrosine kinase, modulates several downstream signalling pathways such as MAPK, STAT3, and PI3K [370]. The previous study stated that the upregulation in c-Src showed a significant impact on the progression of LVI, angiogenesis, and lymphangiogenesis in early-stage metastatic BC [371]. One of the hallmarks of early tumour invasiveness is the downregulation of cell-cell adhesion proteins. The discohesive and motile tumour cells may dock onto the surface of the endothelial cells of an adjacent lymphatic or blood vessel utilising N-cadherin-intermediated junctions. The N-cadherin from each cell type is attached to β-catenin at the cytoplasmic localisation. This protein bridging would trigger the Src kinase pathway, and the tight junctions of E-cadherin will be downregulated upon phosphorylation by the Src kinase activity. The lined and tightly connected cells of the lymphatic vessels may diverge, allowing TEM to occur [372]. One of the physical features for an observed invasive BC embolus in the lymphovascular cavity of a positive LVI specimen is the spheroid shape and the compact structure. The overexpression of *CDH1* and *CTNNB1* in inflammatory BC (iBC) has been reported previously to be an essential step in the formation of the lymphovascular emboli in vivo [373]. Theoretically, there is a possible role for

HSP90-α/*HSP90AA1* in expediting the TEM process, especially eHSP90-α. Further in vitro/in vivo investigations are warranted.

Inverse gene expression associations were found in the $LVI^{+/-}$ patients between *HSP90AA1* and *CAV1* and *CXCL12*, but not in the METABRIC patients. In fact, *CAV1* and *CXCL12* were positively correlated with the expression of *HSP90AA1* in the METABRIC patients. A meta-analysis study stated that the high protein expression of the stromal cell-derived factor-1, SDF-1 or CXCL12, is statistically relevant with better overall survival and $ER^{+}/HER2^{-}$ BC, giving CXCL12 good-prognosis properties [374]. This study and our results may indicate possible regulatory roles for *HSP90AA1* and BC progression in the luminal-A subtype. The loss of *CXCL12* expression in the *HSP90AA1*$^{+}$/luminal-A tumours may classify a subclass of high-risk BC patients. On the other hand, the upregulated expression of the CXCL12 protein was found to be higher in the stroma of luminal-B tumours, especially with the presence of pro-tumour macrophages [375]. This may explain the discrepancy in the gene co-expression associations of *HSP90AA1*/*CXCL12* between the $LVI^{+/-}$ patients, all composed of luminal-A tumours, and the METABRIC mixed cohort. Our finding may suggest a novel role for *HSP90AA1* in the regulation of signalling molecules in TME.

The caveolin-1 protein may function as a tumour suppressor in the early stages of cancer [376]. Being correlated positively with the gene expression of *HSP90AA1* in the METABRIC, the expression of caveolin-1 was found to enhance the motility and invasiveness properties of MDA-MB-231 cells. As our mRNA/protein study was based on a cohort of luminal-A patients, we propose an inhibitory role for *HSP90AA1* in the tumour-suppressing capabilities of *CAV1* in the early stages of luminal-A BC. This suggestion could be supported through in vitro work in the future. Studying the invasion-enhancing mechanisms of caveolin-1 in ER^{-} BC may lead to a better understanding of the controversial biology of caveolin-1.

5.4.1.6. *Inferences of HSP90AA1*

In this study, we identified that the high gene expression of *HSP90AA1* was related to definite LVI occurrence and poor survival in the $LVI^{+/-}$, METABRIC, and external pooled cohorts. Invasive BC patients with *HSP90AA1*$^{+}$ gene expression profiles represented a subclass of high-risk patients associated with LVI occurrences, larger tumours, high tumour grade, LNM, poor and NPI index in the studied cohorts. Our investigation revealed that the positive expression of *HSP90AA1* may draw the attention of translational cancer research to an $ER^{+}/HER2^{+}$ subclass of LVI^{+} invasive BC patients with high genomic instability. Furthermore, the altered expression of *HSP90AA1* showed independent prognostic quality in

invasive BC. At the molecular level, the high gene expression of *HSP90AA1* was positively correlated with the EMT and proliferation biomarkers in invasive BC, but it was marginally correlated with the expression of the MMP10 encoding gene specifically. Our results also showed that the altered expression of *HSP90AA1* may have weak interrelations in triggering the immune responses.

According to our results, we have provided statistical evidence to label *HSP90AA1* as a potential prognostic biomarker and LVI driver in invasive BC patients. The biological processes regulated by the *HSP90AA1* gene appear to be sophisticated in invasive BC. Various pathways that encompassed several co-expressed genes/proteins with *HSP90AA1*/HSP90-α were reported previously in different cancer studies. These co-expressed molecules may contribute to the passenger regulatory events of the LVI mechanism rather than initiating and leading the mechanisms of LVI in particular. To the extent of our knowledge, none of this provides specific conclusions as to how *HSP90AA1*/HSP90-α regulate LVI progression in the different invasive BC phenotypes. The altered expression of *HSP90AA1* could be a robust and independent prognostic and therapeutic target for the high-risk/LVI^{+} subgroup of invasive BC patients. We highly encourage further focused studies on the LVI-governing mechanisms by *HSP90AA1* in the future.

5.4.2. VDAC2: A mitochondrial transmembrane protein that may predict LVI, LN metastasis, and poor survival for invasive BC patients

5.4.2.1. *Overview*

The voltage-dependent anion-selective channel protein 2 (VDAC2) is a mitochondrial protein channel that regulates the trafficking of hydrophilic metabolites with low molecular weights, sized between 1–5 kDa, across the outer membrane of mitochondria [377]. Three isoforms of VDAC2 protein were identified previously. According to UniParc, an archive for protein sequences, isoform-1 contains 309 aa with a total mass of 33,372 Da, whereas isoform-2 exhibits a shorter sequence in length with 283 aa and a molecular weight of 30,412 Da. Isoform-3 was labelled as the canonical sequence of the VDAC2 protein, and it consists of 294 aa and weighs 31,567 Da (UniParc identifier: P45880-3). VDAC2 protein is one member of the VDAC pore-forming protein family, besides VDAC1 and VDAC3 [378]. At the genomic level, the VDAC2 encoding gene is located at the 10q22.2 cytogenetic map location [379, 380]. Genetic abnormalities linked to 95 genes on chromosome 10 were found to be involved in cancer development, of which 45 aberrant genes were reported in BC patients [381].

In this study, we identified VDAC2 as one of the differentially expressed mRNA/proteins in the $LVI^{+/-}$ patients, applying RNA microarray and MS procedures. We identified that the upregulation of *VDAC2* was associated with the occurrence of LVI, LNM, poor NPI, and shorter BCSS in invasive BC patients. Upregulation of *VDAC2* was statistically associated with BC that exhibits a negative expression of ER according to PAM50. Previous studies on several malignancies revealed a cluster of adjacent genes to *VDAC2* chromosomal locations that are associated with multiple mechanisms of invasion; for example, *ANK3* [382], *COL13A1* [383], and *MCU* [384, 385].

The outstanding question for this study is: what is the nature of direct/indirect contributions that the altered VDAC2 expression dictates throughout different steps of LVI development? To answer this inquiry, we will debate several aspects that may communicate VDAC2 with LVI molecular mechanisms through the altered functions of the mitochondria and tumour microenvironment.

5.4.2.2. *Clinicopathological relevance of the VDAC2 altered gene expression*

Remodelling of the cellular metabolism is a distinct biochemical feature that imprints the development of BC and many other malignancies [386]. The impaired metabolism of BC may also interrupt the organisation of biological signalling between cell-cell and cell-stroma in the tumour microenvironment. Aberrant metabolism in BC may also enhance drug-resistance to many treatment regimens [387], and multiple studies indicated several anomalous tumour metabolic pathways as the outstanding origin of adverse outcomes in cancer patients. Therefore, identifying novel prognostic biomarkers to targeting BC-altered metabolism is promising as it increases therapeutic options for patients and may suggest novel insights for invasion mechanisms.

In tumour metabolism, highly proliferative cancer cells rely on rapid aerobic glycolysis to produce their need for adenosine triphosphate (ATP) and produce higher levels of pyruvate and lactate. However, the amount of ATP production by this mechanism is far below the efficient yields of ATP by the mitochondrial oxidative phosphorylation under normoxic conditions; this is an energy-production mechanism termed the Warburg effect [388]. Favouring aerobic glycolysis by cancer cells is important for maintaining their stemness-related pathways' activity, a phenomenon that features tumour-initiating cells (TICs) [389]. Previously, *VDAC2* was characterised as one of the differentially expressed genes in a cohort of 31 BC patients with supraclavicular LNM [390]. Grouping the patients into poor and good outcome subgroups and applying cDNA microarray to analyse the differing gene expression patterns identified 321 DEGs. Interrogating the biological interaction

network databases showed that the c-Myc protein promotes the transcription of *VDAC2* and the proteasome subunit β type-5 (*PSMB5*) and that their transcription levels were higher in the group of patients with poor outcomes. Our results associated as unfavourable an outcome with upregulated *VDAC2* as for the rest of the identified genes of T8. We expanded our scope to look for more therapeutic candidates by examining the co-expression network around each of the T8 genes. We identified the strong association of co-expression between *VDAC2* and the proteasome subunit α -6 (*PSMA6*), a component of the proteolytic degradation complex that is associated with poor outcome in lung cancer. The knocking down of *PSMA6* expression led to cell cycle arrest and apoptosis in H460 and H1299 cancer cell lines, but not in normal lung cells [391].

Recently, 12 low-passage primary BC cell lines were tested for their glycolytic phenotypes. The researchers stated that a specific cell line with significant glucose uptake and lactate production was surgically isolated originally from an infiltrating ductal carcinoma that exhibits no expression of ER and PR receptors or EGFR amplification [392]. The remaining examined cells were derived and cultured from LNMs or pleural fluid effusions. The researchers indicated that the Warburg effect in some cells could be maintained through the activity of the c-Myc protein that regulates glycolysis in BC. Besides the regulation of *VDAC2* transcription, as suggested by [390], the c-Myc protein was identified recently as an enhancer for chemotherapy resistance in triple-negative BC through altered metabolism of the mitochondria [393]. In the same year of the latter study, high-risk BC patients were subjected to in silico analysis in order to identify biomarkers with prognostic potentials for metastatic relapse and treatment failure events. This study comprised luminal-A BC patients treated with tamoxifen, but no chemotherapy, and with prolonged follow-up data. The researchers identified the mitochondrial membrane protein VDAC2 among the discovered biomarkers in their approach. VDAC2 showed the best prognostic value with a hazard ratio of 4.17, and it was an excellent predictor of BC recurrence, distant metastasis, and treatment failure [394]. In a different study, VDAC2 was found to bind to and inhibit the pro-apoptotic B-cell lymphoma 2 (Bcl-2) protein, Bak, during mitochondrial apoptosis pathways [395], and it reduced the permeabilisation of the mitochondrial outer membrane, keeping apoptosis signalling molecules in the mitochondrial matrix, attenuating cellular apoptosis, and prolonging BC survival [396]. In an earlier study, tumour lymphovascular emboli were claimed to be the source of embryonal stem cells (ESC) that typify recurrent, metastatic, drug-resistant, and anti-apoptotic cancers [87].

From the indications of our results, and according to the aforementioned studies regarding the altered expression of *VDAC2*, further systematic investigations at the level of glycolytic enzymes and their associations with the expression of VDAC2 in LVI^+ patients are proposed. The regulatory roles of VDAC2 in anti-apoptotic, stem cell-related, and treatment-resistance pathways and the interactions with the proto-oncogene c-Myc in LVI^+ patients, especially those with LN-, is warranted to develop a meaningful prognostic model for *VDAC2* and to tackle LVI mechanisms in invasive BC.

5.4.2.3. *Pivotal gene expression associations between VDAC2 expression and matrix metalloproteinases*

The gene expression of *VDAC2* showed mutual and positive gene expression correlations with *MMP1*, *MMP12*, and *MMP24* among the patients of the $LVI^{+/-}$ and METABRIC cohorts. The expression of *MMP1* was reported to be upregulated in the higher grades and LVI^+ BC [257, 352], as our results also suggested for *VDAC2*. The expression of *VDAC2* was also found to be highly correlated with the expression of *HSP90AA1*. Adding the positive co-expression of the pro-inflammatory protein macrophage metalloelastase (*MMP12*) to this equation, the qualification of the altered expression of *VDAC2* as a potential LVI driver/associate in invasive BC could be enhanced. MMP12 is an efficient ECM remodelling agent that is activated by TNF-α [397], and it is able to recruit neutrophils [398]. Activated MMP12 by TNF-α could be a sign of the presence of metastasis-promoting mesenchymal stromal cells in tumours [399]. MMP12 was identified as one of the MMPs that was downregulated upon the ablation of Histone-lysine N-methyltransferase (hSETD1A), which is known to be upregulated in metastatic BC [400]. The positive gene expression association between *VDAC2* and *MMP24* may imply a synergistic function for *VDAC2* and *MMP24* in the tissue remodelling process. As *MMP24* is expressed extracellularly by the macrophages and involved in the hydrolysis reactions of the elastin protein, *VDAC2* could be utilised by tumour cells in the recruitment of different TME components and aid in successful tumour migration [401, 402]. Further validation studies need to be performed to assert these findings functionally in the future.

5.4.2.4. *Key gene expression associations between VDAC2 and immune cells*

In the $LVI^{+/-}$ cohort of patients, all of the statistically significant gene expression correlations of *VDAC2* against the assessed CD encoding genes were inverted. The highest opposite gene expression correlations were with *CD3E*, *CD20*, and *CD267*. The same trend of gene expression was observed for *VDAC2* in the METABRIC cohort, except for a unique positive correlation with the expression of *CD80*.

Altered expression of the glycoprotein chains of the T-cell co-receptor CD3 was reported previously as exhibiting independent predictive value in the HNSCC [403]. The CD3 epsilon chain (*CD3E*) is a protein motif that is presented on the T-lymphocyte cellular membrane. The signal transduction of the T-cell activation pathway is activated upon phosphorylation by specific Src family protein tyrosine kinases. Downregulation of CD3E may indicate a depletion in T-cell activation and t-cell dysfunction, and it may indicate lower numbers of infiltrating t-cells. Therefore, there is deterioration in adaptive immune response activity in BC as our results suggested, and as confirmed previously in lung carcinoma by another group of researchers [404]. The low levels of CD20 may also point to a weaker anti-tumour environment as the presence of the $CD20^{+}$ B-cells may enhance the BCSS of invasive BC patients [405]. Finally, the downregulation of the transmembrane-activator-1 and calcium-modulator- and cyclophilin ligand-interactor (*CD267*) could also be a sign of regression in the development of B-lymphocytes [406].

5.4.2.5. *VDAC2 gene expression and tumour microenvironment and progression biomarker*

Interestingly, the *VDAC2* and *HSP90AA1* gene expressions shared almost identical gene expression correlation results with *MKI67*, *CDH2*, *CXCL12*, *THY1*, and *CAV1*. Moreover, the gene expression correlation between *VDAC2* and *HSP90AA1* was positive with a high statistical significance. Plus, *VDAC2* and *HSP90AA1* were the only potential LVI-drivers from the list of T8 molecules the statistical significance of which as continuous variables was increased when dichotomised. These findings suggest a promising integrative biochemical relationship between the molecular products of *VDAC2*, *HSP90AA1*, *MKI67*, and *CDH2*. The possible common pathways of the aforementioned genes/proteins have never been considered in one inclusive study interrogating LVI mechanisms in invasive BC. In order to expand our exploratory study of *VDAC2* in promoting LVI mechanisms, we considered multiple associations of *VDAC2* gene expression against the complete list of the identified LVI-DEM.

5.4.2.6. *Co-expression contexts of phosphoglycerate mutase-1 (PGAM1) and triosephosphate isomerase-1 (TPI1) with VDAC2*

Little is known regarding the involvement of *VDAC2* in the progression of invasive BC. We could not pinpoint any specific in vitro or in vivo studies that inspected the effects of VDAC2 mRNA/protein on LVI in invasive BC. We applied correlation analysis that integrated all of the 56 identified and validated LVI-DEM to assess the gene expression network between them and *VDAC2*. We found 33

mRNA/proteins with the positive direction of co-expression with *VDAC2*. Among these LVI candidates, two molecules suggested biochemical interlinks with *VDAC2* and the altered carbohydrate metabolism pathways, *PGAM1* and *TPI1*. Briefly, we will discuss a few recent studies on various malignancies that tried to construct a study-based link between their molecular functions and biological processes with LVI development.

In order to produce ATP in a glycolysis pathway, glucose is biochemically converted into smaller metabolites. Phosphoglycerate mutase-1 (*PGAM1*) is a key ion-independent enzyme (EC5.4.2.11) that catalyses the transferring of an inorganic phosphate (Pi) from the intracellular metabolic intermediate 3-phosphoglycerate to the free hydroxyl group (OH) of 2-phosphoglycerate at the eighth step of glycolysis. PGAM1 is upregulated and augments tumour proliferation through the modulation of metabolites upon the loss of TP53 [407]. Cancer cells augment aerobic glycolysis machinery to protect themselves from the harsh effects of oxidative stress [408]. In a previous study, the inhibition of the enzymatic activity of PGAM1 resulted in halting the proliferation of the MDA-MB-231 BC cell line [409]. Proteomics research indicated key mediators of ROS-induced migration and invasion, including PGAM1, STMN1, and ITGB3 that are differentially expressed in cancer cells and drive CRC metastasis [410]. Another research group suggested that the upregulation of PGAM1 activity attributed to the loss of *TP53* function, the cellular tumour antigen p53, in cancer cells [411]. Immunohistochemical analysis on 130 glioma specimens revealed novel prognostic relevance for the overexpression of ANXA2 and PGAM1 in comparison to normal brain tissues [412]. A subsequent study revealed VDAC2, PGAM1, TPI1, and STMN1 proteins as members of the altered protein expression signature in young patients, age <45 years, of glioblastoma [413]. A further studied revealed that siRNA knockdown of *PGAM1* restrained glioma cell migration and invasion in vitro [414]. Silencing of TPI1 siRNA reduced the cellular motility and invasion capabilities of primary uveal melanoma cells [415]. Additionally, TPI1 was identified in the matching axillary LN metastases of primary BC in two-dimensional gel electrophoresis (2D gels) as upregulated at 1.3 folds [416].

However, the non-metabolic function of PGAM1 was discovered. The researchers indicated a synergistic interaction between PGAM1 and the smooth muscle actin (ACTA2) in increasing the migrating and cell motility of MDA-MB-231 in vitro [417]. Gene expression data was utilised in functional pathway analysis and the results suggested that the overexpression of *VDAC2* and *STMN1* exhibits involvements with the Wnt signalling pathway [390] that is usually activated by

cancer stem cells (CSCs) in invasive BC [418]. Stathmin protein (STMN1) is essential for invasion of the MDA-MB-231 BC cell line as its signalling complements Rac1, PI3K, and microtubule-end-binding protein (EB1) molecular regulation during lamellipodia formation under the influence of insulin-like growth factor-1 (IGF1) [419].

5.4.2.7. *Inferences of VDAC2*

In this study, we identified that the overexpression of *VDAC2* is strongly associated with LVI definite occurrences and the poor prognosis parameters of tumour grade, LNM, NPI, and tumour size in invasive BC patients. In addition, high *VDAC2* expression was positively correlated with shorter survival of all invasive BC patients in the $LVI^{+/-}$, METABRIC, and external validation cohorts. The strong association between LVI^{+} BC patients and the $VDAC2^{+}$ and $HSP90AA1^{+}$ gene expressions was consistent, and the statistical relevance for both genes increased in the sampling distribution statistics. At the molecular level, our results showed that the high gene expression of *VDAC2* was strongly correlated with the MMP1, MMP12, and MMP24 encoding genes. The latter axes of $VDAC2^{+}$/$proteases^{+}$ could point to a vital involvement for the upregulated *VDAC2* in tissue remodelling and migration/invasion cascade in invasive BC. In addition, the positive expression of *VDAC2* showed no positive correlation with any of the adaptive immunity CD markers, and this may infer that the immune-suppression biological processes are activated in invasive BC patients with $VDAC2^{+}$ profiles. Our study provided strong statistical evidence of the overexpression of *VDAC2* as being associated with the highly proliferative and discohesive/migrating tumour cells as it was reciprocally expressed with *MKI67* and *CDH2*. Invasive BC patients with $VDAC2^{+}$/$MKI67^{+}$/$CDH2^{+}$ may represent a subclass of high-risk patients with LVI^{+}, proliferative, active EMT, migrating, and chemotherapy-resistant characteristics.

Finally, the concordant gene expression associations between *VDAC2*, *HSP90AA1*, *MKI67*, *CDH2*, *PGAM1*, and *TPI1* in the BC patients in this study, and according to the conclusions of several published studies, we propose a vital synchronised regulation of energy production and cellular motility machinery that, ultimately, may endorse the initiation and progression of the LVI mechanism in a subpopulation of invasive BC patients – distinctly, in patients who exhibit intrinsic triple-negative or HER2-enriched subtypes, as our results showed. According to the results in this study, the LVI-regulatory mechanisms by *VDAC2* warrant better systematic investigation. We highly recommend further functional studies on the altered expression of *VDAC2* as it could be a promising prognostic and therapeutic target in the future.

5.4.3. ***ARHGDIA*: A signal transducer that regulates the Rho-GTPase cycle and ER-alpha transactivation, and stimulates axonal growth**

5.4.3.1. *Overview*

The protein-encoding gene of *ARHGDIA* is located at the 17q25.3 chromosomal location, HGNC ID: 678; *ARHGDIA* is also known as *RHOGDI-1* [420]. The translation of the processed *ARHGDIA* mRNA produces the low molecular weight protein of Ras homologous (Rho) guanine diphosphate (GDP) dissociation inhibitor-1 (DI-α): RhoGDI-α. RhoGDI-α is a 204 aa protein (23.2 Da) that belongs to the RhoGDI family of proteins (UniProtKB reference ID: P52565) [186]. The protein name for *ARHGDIA*, RhoGDI-α, will be used in this section from this point forward. RhoGDI-α was reported to be ubiquitously expressed in mammals and, according to the COMPARTMENTS database for proteins' subcellular localisation, RhoGDI-α localises primarily in the cytoplasm [421]. However, some studies stated that the nuclear expression of the RhoGDI-α protein was also detected in a few cell lines. Alongside RhoGDI-α, RhoGDI-β (*ARHGDIB*) and RhoGDI-γ (*ARHGDIG*) comprise the three key regulatory RhoGDI proteins for the Rho family of GTPases, such as Ras-related C3 botulinum toxin substrate 1 (*RAC1*), cell-division control protein 42 (*CDC42*), and transforming protein RhoA (*RHOA*) [422, 423].

The fundamental regulatory role of RhoGDI-α is to switch Rho-GTPase proteins on and off during the Rho-GTPases cycle in the cytoplasm [424]. Activated Rho-GTPases govern myriad downstream signalling pathways in BC [425]. Altered expression of RhoGDI-α may disturb the active and inactive settings of Rho-GTPases and may regulate the development of aggressive and metastatic tumour phenotypes [423]. Rho-GTPases are activated, GTP-bound, when they are phosphorylated and translocated into the cell membrane. A Rho-GTPase is deactivated by protein kinase A (PRKA) that phosphorylates a membrane-integrated Rho-GTPases. The latter reaction extracts Rho GTPases from the cellular membrane and complexes them with RhoGDI-α in the cytosol. RhoGDP-RhoGDI complex will prohibit Rho-GDPase from entering the GTPase cycle; hence, it will be activated until it is released from RhoGDI-α [426]. Upon the specific disassembly of the RhoGDP-RhoGDI complex by specific proteins, such as ezrin, radixin, and moesin, the released Rho-GDPase is now enabled to react with two key regulatory proteins of the GTPase cycle [427]. Proteins of Rho-GDP exchange factors (GEFs), or (ARHGEFs), donate an inorganic phosphate molecule (Pi) to catalyse the activation of a Rho-GDPase to Rho-GTPase. The activation of ARHGEFs is made by specific kinases known as focal adhesion kinase 1 and kalirin, which phosphorylate ARHGEFs [428]. Then, the activated Rho-GTPase will be translocated to the plasma

membrane to mediate downstream biochemical signalling [424]. Alternatively, the activation of Rho-GTPase may be halted by the GTPase-activating proteins or ARHGAPs. The ARHGAPs are phosphorylated by mitogen-activated protein kinase-1 (MAPK), allowing them to enter the GTPase cycle in an active form. Activated ARHGAPs enhance GTP hydrolysis to produce GDP-bound Rho proteins [429]. Therefore, ARHGAPs attenuate the activity of several downstream pathways by abating Rho-GTPases from their effectors [193, 430].

An interesting study discussed the detection of endogenously expressed RhoGDI-α in the nuclear extract of MCF7 cell line applying WB and confirmed it with immunocytochemistry (ICC) [431]. The authors of this study barely detected RhoGDI-α in the nuclei of MDA-MB-231. They stated that nuclear RhoGDI-α is associated with DNA-bound ER-α and may enhance the ER-α mediated signalling. In a different study, RhoGDI-α expression was detected with two-dimensional gel electrophoresis and MS technologies, specifically in MCF7 and in specimens of ductal BC, as one of the majorly expressed proteins in both sample types [432]. The authors also stated that HSP60 and HSP90 were also detected among this cluster, beside RhoGDI-α, and they were more highly expressed in carcinomas than in fibroadenomas.

5.4.3.2. *Clinicopathological relevance of the ARHGDIA altered gene expression*

Recently, RhoGDI-α was found to exhibit oncogenic properties in CRC [433]. The proliferative, migrating, and invasive abilities of CRC cells were depleted upon the silencing of RhoGDI-α expression. Although RhoGDI protein expression was comprehensively discussed in many published studies and in various cancer contexts, we found no focused study regarding the effects of the altered protein expression of RhoGDI-α or its mRNA on LVI initiation or progression in invasive BC patients.

We recognised the positive statistical relevance of the upregulated *ARHGDIA* mRNA, as continuous variables, against LVI definite patients (p-value $\leq$0.004). When the patients were stratified into two clusters around their expression median of *ARHGDIA*, the association with LVI was reduced to a trend of significance in the $LVI^{+/-}$ cohort (p-value $\leq$0.065) and lost in the METABRIC large cohort (p-value $\leq$0.105). Our results of nodal metastasis with the altered expression of *ARHGDIA* revealed no evidence of positive LN invasion in BC patients. However, the results of the BCSS, proportional hazard model, and the preferential expression of *ARHGDIA* in the luminal-B subtype may provide independent indications of poor prognosis for the upregulated *ARHGDIA* in invasive BC. Albeit that the aforementioned findings may increase the scepticism on the prognostic impact of

ARHGDIA mRNA as a crucial molecule for LVI, we could designate exceptional perceptions for *ARHGDIA* and LVI.

Contradictive implications for *ARHGDIA*'s protein were reported previously in a few cancer studies. In one study, both Rho proteins and RhoGDIs were reported to be preferentially overexpressed in breast tumours in comparison to their corresponding normal tissues [425]. In contrast, other researchers investigated both RhoGDI-α and RhoGDI-β proteins and their transcripts in a BC cohort of patients who underwent radical mastectomy, applying IHC and RT-PCR [434]. The conclusion of this study asserted that the low expressions of RhoGDI-α and RhoGDI-β were associated with LN invasion and distant metastasis. In this, the study stated that the overall balance between Rho proteins and RhoGDIs levels of expression may contribute to the clinical prognosis of invasive BC. In parallel to BC, two independent research groups investigated the role of ARHGDIA expression on invasive properties of HCC in the same year. One study declared that the overexpression of *ARHGDIA* was associated with the developed migration and distant metastasis of HCC [435], whereas the other concluded that the loss of *ARHGDIA* induced invasion and metastasis of HCC [436].

These discrepancies in the clinical indications for *ARHGDIA* may arise from imbalanced cohort construction, inappropriate size of the studied cohort, and the selection of statistical methods. In addition, some tumours' cells may express splice variants of *RhoGDI-1* mRNA that, consequently, tend to be involved in the specific pathway(s). We suggest that the biochemical activity of RhoGDI-α appears to be extremely contextualised by tumour molecular types and the unique interactions with the tumour microenvironment of invasion. In addition, our results of LN involvement in the $LVI^{+/-}$ cohort showed a favoured positivity in patients with upregulated *ARHGDIA* mRNA at an early stage of LN positive invasion (61.2%). It was noted that the distribution of patients, according to their expression of *ARHGDIA* mRNA, was equal (50.0%) in the advanced stages of LN positivity. This may propose a progressive role of the upregulated ARHGDIA on BC invasive cells/lesions at early stages.

5.4.3.3. *Pivotal gene expression associations between ARHGDIA expression and matrix metalloproteinases*

Diverse relationships of gene expression were found for *ARHGDIA* against different MMP encoding genes in the METABRIC cohort of invasive BC patients. These associations between *ARHGDIA* and *MMP*s were limited to four *MMP*s in the

$LVI^{+/-}$ cohort of patients. Positive gene expression of *ARHGDIA* was associated with *MMP9* and *MMP12*, and negatively with *MMP3* and *MMP27* expressions.

MMP9 is one of the required proteolytic enzymes that disintegrate basement membranes around tumour cells, allowing them to invade adjacent tissues. A recent study stated that the stimulated secretion of MMP9 in several cancers, including BC, may be involved in the formation of invadopodia [437]. In another study, the silencing of Rho-related GTP-binding protein (RhoC) expression via small interfering RNA (siRNA) led to a significant decrease in the transcription and translation of *MMP9* and *CXCR4* in the iBC cells, and this strategy was proposed as a therapeutic option for the metastatic aggressive phenotypes of invasive BC [438]. The indirect regulation of MMP9 by the induction of the tumour suppressor repressor element-1 (RE-1) silencing transcription (REST) showed an attenuation in the invasion ability of invasive BC [439]. Similar to *MMP9*, *MMP12* showed a positive gene expression associated with *ARHGDIA*. *MMP12* was found as one metastasis-promoting gene for BC that is stimulated through the recruitment of $CXCR2^{+}$ neutrophils by the tumour necrosis factor alpha (TNFα)-activated mesenchymal stromal cells (MSCs) [399]. Another study interrogated the MMPs that are responsible for promoting the invasion attributes of the MDA-MB-231 invasive BC cell line. It was identified that MMP12 displayed individual ability to enhance the invasiveness, but not migration, of these BC cells specifically [440]. A possible mechanism that involves RhoGDI, MMP9, and MMP12 in promoting LVI through the recruitment of several stromal components in the TME should be investigated at the protein level in vitro.

Regarding the negative expression associations between *ARHGDIA* and *MMP27* and *MMP3*, no published studies have explained the biological relations between *MMP27* expression and invasive BC. However, the expression of MMP27 at the protein level was identified in a subset of endometrial macrophages and in B-cells [441], suggesting an immune-modulating role for RhoGDI in invasive BC. An opportunity to discover novel findings for the expression of MMP27 in invasive BC could be obtainable. The knockdown of mouse double minute 2 (*MDM2*) and prostate-specific membrane antigen (*PSMA*) genes increased the *MMP3* levels of expression in the MDA-MB-231 BC cell line. The upregulation of *MMP3* was accompanied by a significant decrease in the adhesion, migration, and invasion abilities of the studied BC cell lines [442]. All of these studies support the malignant effects of *ARHGDIA* positive expression, which may make BC patients susceptible to LVI and other progressive BC biological behaviours by controlling Rho proteins' homeostasis, interacting with TME, and modulating the immune response.

5.4.3.4. *Key gene expression associations between ARHGDIA and immune cells*

The gene expression of *ARHGDIA* showed a single and positive gene expression association with *CD4* in the $LVI^{+/-}$ cohort. *CD4* encodes an integral membrane glycoprotein that serves different molecular immune responses. The function of the CD4 protein may vary according to the protein-expressing cell. The expression of CD4 in the helper T-cell type 2 was found to cause BC to metastasise to the lungs [443]. In the earlier study, it was stated that the pro-inflammatory conditions in the TME may establish the high production of IL-4 and IL-13 by helper T-cell type 2, and the activation of a tumour-associated macrophage (TAM), subsequently. The activated TAM may produce high levels of epidermal growth factor (EGF) that, in turn, facilitates the metastatic spread of BC cells to the lungs. A recent study illustrated that the neutralisation of IL-25, another pro-inflammatory cytokine-favouring helper T-cell type 2 immune response, decreased BC metastasis to the lungs, but not the tumour growth. The inhibition of IL-25 also led to a significant reduction in IL-25-expressing helper T-cell type 2 and macrophages [444]. The elevated ratio of CD4/CD8 was found to have a positive correlation with the unfavourable outcome of BC patients. Our results indicate that the expression of *ARHGDIA* was correlated with the expression of *CD8* negatively. Possible biochemical interrelations between the expression of *ARHGDIA*/RhoGDI and the pro-invasion components of the TME are to be studied.

5.4.3.5. *ARHGDIA gene expression and tumour microenvironment and progression biomarker*

The high expression of *ARHGIDA* may indicate the proliferative BC cells as its expression was associated positively with the Ki67 protein-encoding gene (*MKI67*). Available studies for the relationship between RhoGDI and Ki67 in invasive BC are relatively scarce, especially for the luminal-A subtype. Additionally, the involvement of RhoGDI in the development of BC proliferation and invasion seems controversial. One study stated that the knockdown of RhoGDI increased the growth rate of the MDA-MB-231 BC cell line [445], while another study indicated that both miR-200b-3p and miR-200b-5p miRNAs targeted the RhoGDI signalling pathway and inhibited triple-negative BC metastasis [446]. Further comprehensive studies on the molecular regulation between ARHGDIA/RhoGDI and *MKI67*/ki67 are recommended.

The other two positive gene expression associations for the gene expression of *ARHGDIA* in $LVI^{+/-}$ patients are with *FN1* (fibronectin) and *CDH1*. Overexpression of fibronectin was suggested to increase the proliferation of breast epithelial cells and acinar size as two signs of oncogenic changes in the mammary gland [446]. Fibronectin was also considered as one of the mesenchymal markers beside

vimentin (*VIM*) and β-catenin (*CTNNB1*) in a recent study [447]. The main biological function of the CDH1 protein is to regulate cell-cell adhesion. However, this function may be altered due to pathological stress [447]. In a pooled cohort consisting of 4,124 ER^+ invasive BC patients, the high gene expression of *CDH1* was statistically associated with worse BCSS [188, 189]. These findings may support our claim for *ARHGDIA*'s altered expression in directing LVI mechanisms in invasive BC patients. Functional studies on the varied invasive BC phenotypes are warranted to unravel the biological dilemmas of *ARHGDIA*/RhoGDI.

5.4.3.6. *ARHGDIA: Is it a dependent signalling molecule for invasion?*

Understanding the complexity of *ARHGDIA* mRNA altered expression and its involvement with LVI requires careful consideration. Therefore, we examined the co-expression status between *ARHGDIA* mRNA and all of the identified LVI-DEM in order to find a common link that binds *ARHGDIA* transcript, or protein, with a certain or invasion-related pathway. We found ten transcripts, of 56, that were positively co-expressed with *ARHGDIA*. One is *DYNLRB1*, which belongs to the list of T8 and is ranked second among the ten *ARHGDIA*-associated transcripts statistically (r=0.332, p-value ≤9.5E-6). The remaining nine transcripts are: *ACTG1* (r=-0.360, p-value ≤1.42E-6); *ACTB* (r=0.315, p-value ≤2.9E-5); *UPF1* (r=0.300, p-value ≤7.2E-5); TPI1 (r=0.245, p-value ≤0.001); *RAC1* (r=0.235, p-value ≤0.002); *ILF3* (r=0.206, p-value ≤0.007); *COL6A* (r=0.201, p-value ≤0.008); *PGAM1* (r=0.194, p-value ≤0.011); and *H2AFY* (r=0.169, p-value ≤0.028).

Then, we explored whether there was a congregative pathway(s) for these ten targets, utilising the Reactome pathway knowledge base (ReactomeKB) [183, 448]. Version 63 of this database was released on 18 December 2017. Although ReactomeKB covers 2,179 human pathways, 11,426 reactions, 10,996 proteins, 1,764 small molecules, and 27,694 literature references, no data was found for *COL6A*, *ILF3*, *H2AFY* and *ARHGDIA* on their protein-protein interactions. On the other hand, we identified common pathways for RhoGDI-α with RAC1, ACTB, and ACTG1. Interestingly, there are identifiers for DYNLRB1, PGAM1, RPI1, and UPF1 in ReactomeKB; however, we could not locate any published pathways for them with RhoGDI-α. The interactions of the latter proteins with RhoGDI-α remains questionable. All of the retrieved pathway/reactions from the ReactomeKB for RhoGDI-α are among the signal transduction hub of pathways – namely, in the NGFR-mediated axonal growth stimulation pathway. Identifiers of RhoGDI-α pathways and biochemical reactions in the ReactomeKB are as follows: R-HSA-193704, R-HSA-194840, R-HSA-166520, R-HSA-194315, R-HSA-9006934, R-HSA-162582, R-HSA-209563, R-HSA-193634, and R-HSA-193697.

5.4.3.7. *ARHGDIA: A cross-linker in cancer-nerve invasion?*

Our bioinformatics search on ReactomeKB revealed a unique complex between nerve growth factor (NGF) and its receptor (NGFR) with RhoGDI-α and GDP-bound RhoA (RhoA-GDP:RhoGDI-α:NGFR:NGF), in the deactivation biochemical reaction of the RhoA protein. In the latter protein complex, NGF is located at the extracellular region, while the remaining molecules are integrated through the cellular membrane; a complex has never been found in any other pathway. The bond between NGF and NGFR in this complex fosters the RhoGDI activity; therefore, it inhibits the dissociation of GDP from RhoA and prevents its replacement with a GTP molecule via phosphorylation. Keeping RhoA in a deactivated state, GDP-bound, may affect the rigidity of the actin cytoskeleton, allowing the elongation of filopodia and axonal growth to occur [449]. Co-expression analysis of our transcriptomic data revealed a strong association between *ARHGDIA and RHOA* (r=0.1, p-value $\leq$0.00001) and *NGFR* (r=0.118, p-value $\leq$1.3E-7). Both *RHOA* and *NGFR* were positively associated as well (r=0.050, p-value $\leq$0.025).

There is accumulating evidence that supports the associations between cancer progression and the presence of nerve fibres in the tumour microenvironment of invasion, besides other stromal components. A recent study stated that neural invasion is common in grade-III BC, poor prognosis, cancer proliferation, and metastasis. The expression of the neuropeptide NGF by the propagated BC cells around adjacent nerve fibres not only influences the feeling of pain but could also lead eventually to perineural invasion (PNI), the same authors elaborated [450]. Invaded nerve fibres may help in the dissemination of cancer cells to distant regions submissively [451]. Other researchers applied IHC methods to 352 BC specimens and reported that 68.5% of nerve fibres were found at the invasive front of the examined BC sections. Furthermore, the close-up contact between nerve fibres and the invasive front of BC has been identified as being correlated with positive LN metastasis and poor prognosis [452]. A different study also suggested that the presence of nerve fibres is associated with LN invasion in BC, and the sprouting of these nerve fibres around a tumour is mainly triggered by MCF7 BC cells that secrete high amounts of NGF [453]. The axes between RHOA, RhoGDI, NGF, and NGFR in the context of PNI seem interesting and warrant more in-depth investigation.

5.4.3.8. *Inferences of ARHGDIA*

In this study, we identified that the positive expression of *ARHGDIA* could be associated with poor prognosis in invasive BC patients. The overexpression of

ARHGDIA could show a trend of a poorer survival in $LVI^{+/-}$ patients after five years from initial diagnosis; regardless, *ARHGDIA*'s high expression was associated with poor BCSS in the METABRIC and the external pooled cohorts assuredly. Clinically, we showed that the invasive BC patients who expressed less *ARHGDIA* showed fewer LVI^{+} events in the $LVI^{+/-}$ cohort, as well as low tumour grade, small tumour size, and good NPI index. The altered gene expression of *ARHGDIA* exhibited an independent prognostic hallmark in this study.

At the molecular level, our gene correlation analyses illustrated that the high gene expression of *ARHGDIA* was exceptionally co-upregulated with the MMP9 and MMP12 encoding genes that were previously confirmed to promote invasion mechanisms in the $ER^{-}/HER2^{mixed}$ and the $ER^{-}/HER2^{+}$ invasive BC, respectively. Our results showed a possible immune-suppression function for *ARHGDIA* when overly expressed. The silencing of the immune response could be one of the immune-evasion mechanisms utilised by BC cells. The positive expression of *ARHGDIA* with *MKI67*, *CDH1*, and *FN1* may interrelate with proliferative, adhesive, and anchorage-dependent BC cells that underwent EMT successfully.

Based upon our aforementioned findings, we suggest that the inhibition of cytoskeleton rigidity by sequestering the function of RhoA and by upregulated and activated RhoGDI-α signalling is an essential mechanism that may allow solid tumours such as BC to acquire higher motility prior to invading their surroundings by approaching an adjacent lymphatic vessel or nerve fibre. Interrogating the roles of *ARHGDIA*/RhoGDI in the EMT, modulating immune responses, and permitting BC cells to invade their adjacent lymphatic vasculature or bordering perineural niche through LVI and PNI mechanisms is necessary.

5.4.4. DYNLRB1: A member of the dynein superfamily of proteins that is crucial for epithelial cells polarisation

5.4.4.1. *Overview*

Dynein light chain roadblock-type 1 (DYNLRB1) is a (10,922 Da) protein with a lighter isoform of (5,060 Da) [326]. The DYNLRB1 protein was referred to as Km23-1, mLC7-1, robl-1, or Dnlc2a, DYRB1 in several previous publications [454]. Km23-1 will denote the protein name of *DYNLRB1* mRNA from this point forward to avoid any misunderstanding of mRNA/protein nomenclatures. The protein-encoding gene of Km23-1 is located at the cytogenetic band of 20q11.22, and it was found to be ubiquitously expressed in human normal tissue – specifically, upregulated in brain and kidney normal tissues [216].

Km23-1 is a cytoskeletal motor protein that belongs to the dynein superfamily of proteins [455]. Km23-1 is one of the cytoplasmic dynein 1 complex constituents that is responsible for the intracellular retrograde motility of transporting vesicles and organelles along microtubules of eukaryotic cells. In contrast to the reverse transportation of protein vesicles, from the membrane toward the nuclear direction or the minus end, proteins of the kinesins superfamily represent the positive-directed vesicular transporters of cargos. Dyneins and kinesins are microtubule-dependent motor proteins; in contrast, myosins represent the members of the third superfamily of the cytoskeletal motors for vesicular transportation. Myosins are considered the prime transporters of cargos along actin filaments [456].

5.4.4.2. *Clinicopathological relevance of the DYNLRB1 altered gene expression*

Six informative BC studies concluded associations between the amplification of *DYNLRB1* and cancer progression according to the cBioPortal platform for cancer genomics [322, 323]. We identified a statistical difference in the ranked distributions of the overly expressed *DYNLRB1* probes among the LVI^{+} group of BC patients by applying non-parametric statistics. Nonetheless, we could not define any clinicopathological significance, including LVI, when the expression scores of *DYNLRB*1 mRNA were stratified as positive/negative groups around the median cut-off in the $LVI^{+/-}$ cohort. The analyses of BCSS assigned no statistical relevance for the altered expression of *DYNLRB1* in both cohorts of this study, $LVI^{+/-}$ and METABRIC. However, the strong association between the upregulated *DYNLRB1* and poor BCSS was validated in an external pooled cohort of 2,711 invasive BC patients applying the exhaustive prognostic analytical tool [188, 189]. Those patients exhibited the distinctive molecular and clinical features of ER^{+}/metastatic relapse/$LNM^{+/-}$. Such findings may indicate that the prognostic implications of *DYNLRB1* and its association with LVI mechanisms are dependent on the clinicopathological features and the composition of the molecular subtypes for the studied samples. We assume that the preferential upregulation of *DYNLRB1* and the interactions with LVI pathways are restricted to specific molecular classes of invasive BC patients. Diminishing of statistical significance of *DYNLRB1* expression against LVI in the $LVI^{+/-}$, upon sample grouping, may imply a heterogeneous expression for *DYNLRB1*. We suggest that there are different factors that may affect the expression of *DYNLRB1*, therefore endorsing LVI.

Hence, we assessed the contribution of the different tumour grades on the expression of *DYNLRB1* regardless of LVI status. We found out that the mean expression of *DYNLRB1* reached its peak in patients with tumour grade-II (μ=9.4, n=770). The expression means of *DYNLRB1* for patients of tumour grade-I and

grad-III were (μ=9.3, n=171) and (μ=9.3, n=952), consequently. We identified a logical increasing of *DYNLRB1* expression means in tumour grade-II and grade-III (μ=9.6) in comparison to tumour grade-I (μ=9.3) when we exclusively scrutinised the differential expression of *DYNLRB1* for LVI^{+} patients among the three tumour grades. The latter mean of expression was equal to its counterpart of LVI^{-}/LN^{-} and tumour grade-III patients. The distribution of *DYNLRB1* expression between the LVI categorised groups could be affected by the different tumour grades. It appears that the expression of *DYNLRB1* is optimised by cancer cells according to specific molecular and histological contexts. Understanding of the conditional expression of *DYNLRB1*/Km23-1 and how they could be recruited to promote LVI mechanisms in invasive BC patients appears challenging.

As a different standpoint, the upregulation of *DYNLRB*1 mRNA was strongly associated with the luminal-B subtype according to PAM50 classification in the $LVI^{+/-}$ cohort (p-value $\leq$2.0E-5) and the METABRIC cohort (p-value $\leq$0.009). We tried to evaluate this finding in an external cohort of 4,770 invasive BC patients and identified that the high expression of *DYNLRB1* was strongly associated with the PAM50 subtypes of luminal-B, but not basal-like/triple-negative BC. Luminal-B patients were reported with LVI occurrences more often [226]. Patients in the basal molecular class, on the other hand, scored the highest percentage of the downregulated *DYNLRB1* profiles among the two investigated cohorts. Patients of the basal subtype accounted for the least number of BC patients with LVI^{+} diagnosis [104].

5.4.4.3. *Pivotal gene expression associations between DYNLRB1 expression and matrix metalloproteinases*

Of all investigated MMP encoding genes, an inversed gene expression association was identified between *DYNLRB1* and the MMP2 encoding gene in the $LVI^{+/-}$ patients. On the other hand, various gene expression associations were found for the *DYNLRB1* altered expression against/with several MMP encoding genes in the METABRIC patients. This vast difference in the *DYNLRB1* differential expression between these two cohorts may suggest a secondary biological interrelation between the *DYNLRB1* altered expression and some proteolytic enzymes in the development of LVI in invasive BC. As we discussed previously in section 12.4.1.3, the upregulation of *MMP2* mRNA could be frequently detected in the luminal-A phenotypes of invasive BC [355]. However, the conclusion from that study may have been influenced by the molecular subtype composition of the studied cohort – 76% luminal-A, as the authors stated. In our study, luminal-A was also the dominant molecular subtype; however, we showed that the expression of *MMP2* is

downregulated when *DYNLRB1* is upregulated. We propose that the upregulation of *DYNLRB1* could contribute to enhancing the tumour cells motility through the induction of specific molecular arrangements into their actin cytoskeletons, thus improving their migration abilities [457]. The altered expression of *DYNLRB1* could be an interesting subject to study the metastatic potentials of the invasive malignancies, but not their mechanisms of ECM destruction and LVI in particular.

5.4.4.4. *Key gene expression associations between DYNLRB1 and immune cells*

Remarkably, almost all gene expressions between *DYNLRB1* and the CD encoding genes were inversely correlated in the $LVI^{+/-}$ and METABRIC cohorts. The remaining studied CD encoding genes showed no statistical associations with the gene expression of *DYNLRB1* in the same cohorts. In a previous study, it was stated that the Km23-1 protein modulates Ras/MAPK activities in the TGFβ-sensitive epithelial cells of human CRC cells [454]. A subsequent study indicated that the genetic and transcriptome alterations in the activity of the Ras/MAPK pathway is correlated with the lower presence of tumour-infiltrating lymphocytes (TILs) in triple-negative BC after neoadjuvant chemotherapy [458]. Our results showed that the upregulation of the *DYNLRB1* gene could be a secondary event in LVI^{+} patients, but no direct links with invasion mechanisms were found. We suggest that invasive BC patients with upregulated profiles of DYNLRB1/Km23-1 may endure highly migrating and metastatic tumours via endorsing immune suppression and EMT types of machinery through the activation of TGFβ expression that modulates the Ras/MAPK signalling and enriches TGFβ/JNK pathway [458-460]. *DYNLRB1*/Km23-1 may be a novel biomarker that inhibits TILs and augments EMT in the residual disease of invasive BC. Further functional studies are required in the future.

5.4.4.5. *DYNLRB1 gene expression and tumour microenvironment and progression biomarker*

Different gene expression associations were identified for *DYNLRB1* among the BC patients of the $LVI^{+/-}$ and METABRIC cohorts. A unique positive gene expression association was found between *DYNLRB1* and *CTNNB1* in the $LVI^{+/-}$ patients but not in the METABRIC patients. The positive association with *CDH1* was present in both cohorts. The overexpression of β-catenin could initiate carcinogenesis [461] and enhance the EMT in the iBC [462]. Moreover, it was stated that the upregulation of *CDH1* is involved in the E-cadherin/β-catenin axis that could be associated with the modulation of the actin cytoskeleton and the stimulation of the lymphovascular emboli in the iBC phenotype [373]. Our results may provide logical and accumulating evidence for *DYNLRB1* as one of the key players in the progression of the EMT-invasion partnership, influenced by the MAPK downstream

signalling pathway, as it is also expressed positively with *CTNNB1*/β-catenin, *CDH1*/E-cadherin, ezrin, and *HSP90AA1*/HSP90-α [369-371, 454]. We are fully aware that the statistical relevance does not necessarily reflect clinical relevance. Therefore, the actual conclusions for the altered *DYNLRB1* expression and its molecular interactions must be interrogated at the protein level in focused and well-designed functional studies in the future.

5.4.4.6. *Km23-1: A mediator of cellular motility, migration, and invasion*

In this study, we identified that the altered expression of *DYNLRB1* could be one of the promising migration/invasion molecular targets. Nevertheless, we could not find a direct explanation for its contribution to LVI specifically. In order to find a comprehensive relationship between Km23-1 and LVI mechanisms, we sought expanded interactions of Km23-1 with other invasion-related proteins in BC and in any other malignancies.

Proteins of the dynein superfamily were reported to exhibit crucial regulatory roles on cells with polarised structures, such as epithelial cells and neurones [463]. The dynein light chain (DLC) was recounted as the subunit that is responsible for instructing the proper cargo selection, then binding to other transporting and receptor systems. Previously, Km23-1 has been identified as one of the transforming growth factor-β (TGFβ) receptor interacting proteins [464]. A research group of biochemists and molecular biologists identified Km23-1 as a molecular adapter that activates Ras/JNK/ERK pathways by linking TGFβ receptors in a TGFβ-dependent manner in vitro [460]. The authors also confirmed the necessity of Km23-1 for TGFβ auto-stimulation of the TGFβ1 mRNA in TGFβ-sensitive epithelial cells. The activation of the TGFβ receptor, in epithelial cells, regulates several downstream pathways that are associated with proliferation, cell differentiation, apoptosis, interactions with extracellular components, and the regulation of cellular motility through the cytoskeleton [465]. However, TGFβ was found to exhibit tumour-suppressing characteristics in cancers at early stages. The oncogenic features of TGFβ appear as cancer progresses, and it may stimulate the induction of epithelial-to-mesenchymal transition (EMT), migration, invasion, and metastasis. This molecular phenomenon of TGFβ dual functionality is known as the TGFβ paradox [466]. From our findings and the conclusions of the last study, we propose mutual biochemical axes of interaction between Km23-1 and the effective pleiotropic TGFβ. The continued influence of Km23-1 on the auto-stimulation of TGFβ may lead to metastatic epithelial-mesenchymal transitions [466]. Finally, the reduction of km23-1 signalling by siRNA repressed ERK activation, TGF promoter

transactivation, TGF secretion, cell migration, cell invasion, and the expression of metastasis-associated scaffold protein, ezrin, in CRC patients [454]. Our initial results, besides the limited investigations on the altered expression of *DYNRLB1*/Km23-1 in invasive BC patients, encourage focused efforts to unveil the molecular and clinical implications for *DYNLRB1*/Km23-1.

5.4.4.7. *Inferences of DYNLRB1*

In this study, we identified that the mean of sum rank for *DYNLRB1* mRNA expression was statistically higher in the LVI^+ invasive BC patients in comparison to the LVI^- equivalents. Our results indicated that the upregulated expression of *DYNLRB1* was not useful in determining whether it is associated with good or poor BCSS in our studied cohorts. We could not identify a specific clinicopathological feature that could label the invasive BC patients with $DYNLRB1^{+/-}$ expressions with good or poor prognosis. However, the high expression of *DYNLRB1* could be useful in the prognosis of a distinctive subclass of invasive BC patients with the luminal-B subtype and metastatic relapse. At the molecular level, *DYNLRB1* was inversely expressed against the MMP2 and MMP27 encoding genes as two proper candidates for additional functional studies on their role in promoting invasion mechanisms in invasive BC patients. Our results also indicated that patients with $DYNLRB1^+$ expression profiles could suffer from weak immune responses against the disease. Finally, the positive overexpression of *DYNLRB1* with *CTNNB1* and *CDH1* may boost the oncogenic properties, EMT processes, and migration, but not the invasion, capacities of the malignant BC cells through the activation of the down-streaming pathway MAPK/ERK and its major signalling cassettes such as the JNK pathway.

In conclusion, we spotlighted initial evidence of a possible biological dependence between the upregulated expression of either *DYNLRB1*/Km23-1 and the high potential of metastasis. We propose hierarchical regulatory roles of Km23-1 in promoting a migration-invasion cascade, especially in the luminal-B subtype of invasive BC patients. Further thorough in vitro investigations are required in the future.

5.4.5. **FBLN1: A family member of the basement membrane-associated glycoproteins that suppress BC cellular motility**

5.4.5.1. *Overview*

Fibulin-1 is an ECM secreted glycoprotein that is localised at the basement membranes of the epithelial cells, the elastic fibres, and is also a component of the fibrinogen complex in human blood plasma [467, 468]. The fibulin-1 protein has a

molecular weight of 77,214 Da and is identified as isoform-D. Isoform-A at 61,552 Da, isoform-B at 65,443 Da, and isoform-C at 74,434 Da are the other three isoforms of fibulin-1; their aa sequences differ from that of isoform-D at positions 567–703. Isoform-A and isoform-B are expressed in the placenta, while isoform-C and isoform-D are expressed in various tissues and cell lines [469]. For this study, we identified *FBLN1* as one of the overexpressed genes of good prognosis that may inhibit LVI progression. Fibulin-1 is a key player in diverse of biological processes [470].

5.4.5.2. *Clinicopathological relevance of the FBLN1 altered gene expression*

In this study, we identified that the overexpression of *FBLN1* could imply a good prognosis for the $LVI^{+/-}$ and METABRIC cohorts of invasive BC patients. Fibulin-1 protein was reported previously to be highly expressed in the cells undergoing EMT [471]. It was also stated that fibulin-1 was found at high levels in the ECM of Hodgkin lymphoma (HL), which is characterised by an abundant fibrous stroma [472]. In BC, the protein localisation of fibulin-1 in the cytoplasm and at the cell surface of the MCF7, SKBR3, and MDA-MB-157 BC cell lines was previously reported to enhance the anti-tumour immune environment and improve survival. The provoking of an immune response by fibulin-1^{+} BC could be attributed specifically to the isoform-D of the fibulin-1 protein that could be accumulated in the stroma due to lack of interaction with other targeted components in the ECM [473]. The elevated level of fibulin-1 was detected in human BC and ovarian cancer, though varieties of immunoreactive fibulin-1 polypeptides ranging from 100 kDa to 25 kDa were identified, and their implications in BC were varied as well [474]. In this study, we provided statistically consistent clinicopathological results for the altered expression of *FBLN1* in invasive BC. However, the different impacts of the isoform-D of the fibulin-1 protein may illuminate the contradictory positive associations for *FBLN1* with some of the invasion-driver genes in invasive BC patients. Further details on the results of the gene correlation analyses will be discussed in the following sections.

The diminishing of the statistical significance of *FBLN1* expression against LVI status in the METABRIC cohort was a remarkable difference for the altered gene expression of *FBLN1* between the two investigated cohorts in this study. In the METABRIC cohort, patients with LVI^{-} status were LN^{mixed}. The downregulation of *FBLN1* was statistically relevant with the positive nodal involvement in the $LVI^{+/-}$ patients, but not in the METRABRIC patients. The identification of fibulin-1 as an LNM-associated protein and a prognostic factor was confirmed in a previous study on lung adenocarcinoma [470]. In another study, the low expression of fibulin-1

protein was associated with nodal involvement and distant metastasis in CRC [475]. Earlier, the different intensities of fibulin-1 expression were detected immunohistochemically in human ovarian metastatic cancer. The authors of the latter study stated that fibulin-1 is a secreted stromal protein that is found to be accumulated in the stroma. The authors concluded that the gene expression of *FBLN1* was very low by stromal cells in comparison to a tumour's epithelial cells in OC [476]. This could prove that the expression of fibulin-1 is epithelial in nature. As a secreted protein, the nature of the biochemical interactions of fibulin-1 with the various cellular components of the ECM, and not only its level of expression, is exciting and could dictate diverse biological contexts of tumour progression or suppression. The good prognostic implications for the overexpression of fibulin-1 in invasive BC was suggested by this study, and confirmed by other research in different malignancies, including BC [477-479]. However, we expect that our good-prognosis claims for *FBLN1* in invasive BC could raise several questions and criticism. We have tried to answer these questions in light of the current findings of this study.

The defining of a logical and scientifically proven mechanism for *FBLN1*/fibulin-1 as a potential LVI-driver in invasive BC patients could lead to a BC prognostic biomarker and possibly a therapeutic target, with novel properties. The accurate explanation of the effects of *FBLN1* expression in invasive BC patients requires deeper gene-gene interaction and protein-protein interaction analyses. In the next few sections, we will discuss the results of the gene expression correlations for *FBLN1* against the different protein-encoding genes related to invasion mechanisms in order to produce further insights into the involvements of *FBLN1*/fibulin-1 in the regulation of LVI in invasive BC.

5.4.5.3. *Pivotal gene expression associations between FBLN1 expression and matrix metalloproteinases*

Although we have labelled *FBLN1* with the characteristics of a good prognostic biomarker based on the clinicopathological data analysis, the gene expression of *FBLN1* shared positive and negative gene expression correlations with several MMP encoding genes in the $LVI^{+/-}$ and in the METABRIC cohorts. The positive gene expressions between *FBLN1* and the *MMP2*, *MMP3*, *MMP10*, *MMP11*, *MMP13*, *MMP14*, *MMP27*, and *MMP28* in the $LVI^{+/-}$ and the METABRIC cohorts were mutual. Our results showed no specific MMP encoding gene that was upregulated in the $LVI^{+/-}$ patients exclusively.

The shared localisation site of the extracellular region makes the fibulin-1 protein vulnerable to enzymatic degradation reactions by specific MMPs. In a previous study, it was found that the expression of the fibulin-1 protein is essential in the suppression of cell migration signalling through the interaction with fibronectin (*FN1*) and syndecan-4 (*SDC4*). It was also stated that the fibulin-1 protein was downregulated in invasive HCC and PC [480], as our results showed in the studied invasive BC cohorts. The successful cellular migration by the tumour could be dependent on the cleavage of fibulin-1 by MMP13, and this would optimise the conditions for a tumour to metastasise later. The authors of the last reference discovered that calumenin (*CALU*), an ECM protein with 15 different isoforms, bound to fibulin-1 and shielded it from the proteolytic activity of MMP13. In addition, it was found that fibulin-1, calumenin, and fibronectin modulated the ERK1/2 signalling pathway and inhibited tumour migration [480]. We propose that the upregulation of *FBLN1* in invasive BC does not necessarily downregulate the expression of a specific MMP encoding gene. In fact, it is the protein-protein interaction that should be considered for fibulin-1 in the study of invasive BC and LVI mechanisms. Based on the available mRNA data, we can nominate several novel fibulin-1/MMP(X) interactions for further investigation in the future.

The noteworthy finding of the altered gene expression for *FBLN1* against MMP encoding genes in the $LVI^{+/-}$ patients is the negative gene expression correlation between *FBLN1* and *MMP1*. This correlation is the only unique gene expression correlation that differentiates the $LVI^{+/-}$ cohort from the METABRIC regarding *FBLN1* gene expression. In the METABRIC cohort, the *MMP1* gene was expressed positively with *HSP90AA1*, *VDAC2, DYNLRB1, ARHGDIA,* and *PPIA.* The first two genes from the previous succession are our candidates for LVI initiation/progression. In the same cohort, the gene expression of *MMP1* showed no significant gene expression associations with *FBLN1*, *RPS18*, and *RPL10A* – the genes claimed by this study to have good prognostic features. In the $LVI^{+/-}$ cohort of patients, *MMP1* was expressed positively with the gene expression of *VDAC2* and *PPIA*, but negatively with *FBLN1* specifically. As discussed earlier, the upregulated expression of *MMP1* denotes higher grades/LVI^{+} tumours [257]. To date, there are no specific studies that interrogate the co-expression and biological interactions between the fibulin-1 and MMP1 proteins in promoting LVI mechanisms in invasive BC patients. Alternatively, the proteolytic degradation of the fibulin-5 protein in ECM occurred due to the cancer-recruited macrophages that upregulate the expression of MMP7. The cleavage of fibulin-5 to a smaller protein product was associated with an enhanced local metastasis and cell adhesion in OC, whereas the intact fibulin-5 in the ECM hindered the dissemination of the migrating and invasive cancer cells [481, 482].

Our findings suggest a similar LVI-regulatory mechanism that may involve fibulin-1 and MMP1 proteins in invasive BC.

5.4.5.4. *Key gene expression associations between FBLN1 and immune cells*

Several variations in the gene expression correlations between *FBLN1* and the CD encoding genes of the immune cells were identified between the $LVI^{+/-}$ and the METABRIC cohorts. The altered gene expression of *FBLN1* showed several positive correlations with the different protein-encoding genes of the T-cell receptor (TCR)-CD3 complex, *CD4*, and *CD8* in the METABRIC cohort. However, no positive associations of gene expression were found for *FBLN1* in all of the investigated CDs in the $LVI^{+/-}$. In the $LVI^{+/-}$ cohort, the positive expression of *FBLN1* was inversely correlated with the gene expression of *CD80* and *CD20*, and this relationship may indicate a lack in the activation of B-cells, dendritic cells, and macrophages in the TME.

In addition to fibulin-1's role as an ECM stabiliser, it may be involved in the regulation of tumour-related immune responses. A previous study stated that the monocyte-derived dendritic cells, presenting fibulin-1 as an antigen, induced the activation of distinct subsets of fibulin-1-specific $CD4^+$ T cells under inflammatory conditions [473]. As our results did not verify any *FBLN1*$^+$-induced $CD4^+$ T cells in the $LVI^{+/-}$ cohort of patients, this immune response could be exclusively valid for invasive BC patients in the METABRIC cohort. The secreted fibulin-1 in ECM may attenuate the inflammation environment that lines tumour progression. Our results also suggest that the positive expression of fibulin-1 in the ECM could provide an anti-inflammatory environment that reduces invasive BC migration and invasion abilities by blocking the recruitment of immune cells in ECM [483]. A recent study indicated that the recently discovered member of the fibulin family of proteins, fibulin-7, inhibited the differentiation of the macrophage colony-stimulating factor (MCSF)-induced monocyte, and reduced the infiltration of immune cells to the inflamed tissues [484]. The limited presence of the intratumoural immune cells may cut the recruitment line of tumour-associated myeloid cells by cancer cells; however, this effect could also bring unfavourable and weak anti-tumour immunity by acting as a double-edged sword mechanism in alternative biological and molecular circumstances. Well-designed functional and in vivo studies should be considered to understand the influential roles of *FBLN1* on the immune cells in the TME of invasive BC patients.

5.4.5.5. *FBLN1 gene expression and tumour microenvironment and progression biomarker*

The gene expression of *FBLN1* in the $LVI^{+/-}$ patients was correlated positively with the expression of nine protein-encoding genes co-enriched in gene ontology (GO) terms of cellular adhesion/motility/polarity/chemotaxis. These identified genes are *CXCL12, COL3A1, CAV1, THY1, SNAI2, VIM, SDC1*, *ETS1*, and *FN1*. These genes were ordered according to their statistical significance with the gene expression of *FBLN1*. On the other hand, two negative gene expression associations for *FBLN1* against *MKI67* and *CDH1* were identified. The CDH1 encoding gene is another adhesion/motility/polarity/chemotaxis-related protein-encoding gene. Our results had also revealed three other positive gene expression correlations for *FBLN1* in the $LVI^{+/-}$ patients: *FAP*, a protease-encoding gene; *PDGFC*, a growth factor-encoding gene; and *S100A4*, a low molecular weight calcium-binding protein. In the METABRIC cohort, *FBLN1* was associated with the same list of genes as in the $LVI^{+/-}$ cohort, and in the same direction of gene expression. For the gene expression correlation analysis of the T8 molecules in the METABRIC cohort, extra genes were positively correlated with *FBLN1* in comparison to those in the $LVI^{+/-}$ cohort. *FBLN1* was co-expressed with *CTNNB1*, *CDH2*, *CD24*, and *CD44*.

As previously mentioned, the stromal expressions of *CXCL12*, *CXCR4*, and *VIM* were found to be upregulated in the aggressive BC subtypes such as luminal-B and associated with a high population of TAM and LNM [375]. The activation of TAM may reflect a pro-inflammatory conditioned TME that is enriched with inflammatory interleukins and developed BC dissemination signalling [443]. Nevertheless, the upregulation of *CXCL12* may indicate good prognostic features for $ER^{+}/HER2^{-}$ BC patients, as also discussed earlier in section 12.4.1.5. [374]. A recent study identified that the CXCL12 protein was differentially expressed among ER^{+} and ER^{-} BC specimens, and the upregulation of the CXCL12 protein in ER^{-} BC specifically was associated with a worse prognosis [485]. Our results of the intrinsic subtypes and the integrated classification clusters, PAM50 and IntClusts, for the $LVI^{+/-}$ invasive BC patients showed that the positive expression of *FBLN1* was statistically associated with a higher percentage of patients in normal-like, luminal-A, IntClust-3, IntClust-4, and IntClust-8. All of the previous BC intrinsic subtypes were characterised with good prognosis, and all of the three IntClusts were primarily composed of luminal-A invasive BC patients. For this study, our results could point toward an interactive biological role for *FBLN1* in modulating the TME and immune response in invasive BC. This type of modulation may be subtype governed. We suggest that the mutual upregulation of *FBLN1* and *CXCL12* may improve the immune response in TME in a favourable fashion and help in postponing LVI progression. Studying the *FBLN1*/*CXCL12* axis in invasive BC patients as a

promising prognostic and therapeutic option for invasive BC patients in the future is highly encouraged.

The next gene expression correlation for *FBLN1* was with the collagen-α-1(III) chain-encoding gene (*COL3A1*). The collagen-α-1(III) chain is a precursor to the collagen-III protein that is abundant in connective tissues. In a recent study, *COL3A1* and the cartilage oligomeric matrix protein-encoding gene (*COMP*) were identified as two novel prognostic biomarkers that were associated with early-stage papillary thyroid carcinoma (PCL). The altered expression of *COL3A1* in the last reference was stated as having a strong association with vasculature development and was related to the development of LNM in PCL when co-expressed with *COMP* [486]. The prognostic value of *COL3A1* was also claimed to be in partnership with the *FN1* gene in a recent DGE/IHC advanced rectal cancer study. The collagen-α-1(III) chain and fibronectin proteins were the best independent predictors of the pathological response to neoadjuvant chemotherapy in rectal cancer. These two stromal proteins were suggested as a developed immunohistochemical classifiers that label a subclass of rectal cancer patients with an absent response to neoadjuvant treatment [487].

Another recent study linked the intracellular expression of fibronectin by the MCF-10A BC cell line with no metastatic properties. The autocrine production of fibronectin by mesenchymal BC cells hindered their metastatic capacity and led to the retrieval of the epithelial features upon the reduction of fibronectin [488]. However, the stimulated expression of both collagen-III and fibronectin by the transforming growth factor-beta 1 (TGF-β1) was correlated with metastatic GC [489]. The high expression of TGF-β1 in invasive BC was reported to reduce BCSS and promote invasion and migration in triple-negative BC by modulating EMT rates [490]. The TGF-β1 isoform was not the highest expressed TGF-β isoform in the luminal-A subtype; in fact, the high expression of TGF-β1 was strongly associated with the aggressive subtype of basal-like BC [491]. Our results indicated that basal-like and IntClust-10 were enriched in patients with *FBLN1*$^{-}$ profiles. Investigating the expression levels of *FBLN1* in the ER^{-}/LVI^{+} subclass of patients may add further clinicopathological insights to *FBLN1* in high-risk BC patients and may help them to receive enhanced disease management.

There are several co-factors, such as TGF-β isoforms, that may alter the expression of specific genes/proteins based on different biological contexts. The family of miR-29, small RNA molecules, were suggested to elicit a tumour-supressing mechanism, and their downregulation was associated with aggressive cancer phenotypes and poor prognosis [492]. In contrast, a recent study concluded

that the expression of miR-29b may act as an enhancer for PC invasion. The inhibition of miR-29b led to a significant increase of the *COL3A1* mRNA levels in the DU145 cell lines, and the production of collagens may control PC metastasis [493]. In earlier cancer research studies, miR-29b was found to be downregulated in the patients of chronic lymphocytic leukaemia (CLL) and acute myeloid leukaemia (AML) [494, 495]. In invasive BC, miR-29b was also downregulated in the ER^-/PR^- patients but was upregulated in the ER^+/PR^+ patients. Moreover, the median expression of the miR-29a, another member of the miR-29 family, was upregulated in LVI^- invasive BC patients [144]. Another study stated that miR-29b may regulate the expression of podoplanin (*PDPN*), leading to the inhibition of invasion in glioblastoma patients [496]. These accumulating findings spotlight the complex involvement and the multi-layer biochemical networks of the proteins in TME that dictate the invasion/metastasis cascade of events. The sharing of the extracellular region by the fibulin-1, fibronectin, and collagen-α-1(III) chain as secreted proteins requires comprehensive functional studies. The co-expression of these three proteins may suggest different settings for invasive BC patients according to their distinct associations with BC molecular subtypes and other gene/protein networks.

In the previous section, we discussed how fibulin-1 protein collaboratively inhibits the migration of tumour cells with calumenin and fibronectin by halting the ERK1/2 pathway and evading proteolytic degradation by MMP13 [480]. By recalling our comments on the fibulin-1/MMPs axes from the previous section and the points discussed in the current section, we can propose that the prognostic implications of the positive expression of *FBLN1*/fibulin-1 could be influenced by the presence of several interacting proteins in ECM. The gene expression regulators for *FBLN1* and the localisation of the fibulin-1 protein emphasise the necessity of studying the molecular interactions of *FBLN1*/fibulin-1 rather than their levels of expression in invasive BC patients.

As discussed above in section 12.4.1.5., the stromal expression of the caveolin-1 protein exhibited tumour-suppressing features and improved the outcomes of invasive BC patients [376]. On the one hand, the stromal expression of caveolin-1 protein in ER^+ invasive BC showed good prognostic implications. On the other hand, the accumulation of caveolin-1 protein in the cytoplasm was reported to be associated with the positive expression of EGFR, and this aberrant localisation may activate EGFR-mediated mitosis [497]. Another study stated that the upregulation of caveolin-1 was found to reduce the cell surface expression of the calcium-activated potassium channel subunit alpha-1 (*KCNMA1*) and inhibited the proliferation and invasion of the MCF7 BC cell line [498]. The amplification of

KCNMA1 was associated with high-proliferative tumour cells based on the ki67 index [499].

However, the positive expression of caveolin-1 may not always be a good prognostic sign for invasive BC patients; in fact, caveolin-1 could mediate invasion in specific molecular subtypes of invasive BC through alternative pathways. The downregulation of caveolin-1 by siRNA led to less phosphorylation of the RhoC protein and decreased activation of the RAC-alpha serine/threonine-protein kinase (AKT1) in iBC patients [500]. Previous investigations revealed that caveolin-1 is highly expressed in the iBC phenotype, a high-grade and aggressive aneuploid BC that invades the dermal lymphatic vessels and is characterised by frequent p53 mutations and high expression of E-cadherin [501, 502]. We stated earlier that the tumour cells of the iBC phenotype tended to form florid and cohesive tumour emboli in the cavities of the lymphatic vessels [373]. The aberrant expression of *CAV1*/caveolin-1 may indicate good prognostic and fewer LVI occurrences in patients of the ER^+ phenotypes, and this suggestion may support our hypothesis of the good prognostic properties for *FBLN1* in invasive BC patients. Studying the combined localisation and interactions between fibulin-1 and caveolin-1 proteins for the different molecular subtypes of invasive BC patients in the future is suggested.

The CD90 or Thy-1 membrane glycoprotein (*THY1*) is a glycophosphatidylinositol-anchored cell surface protein that plays key roles in cancer cell adhesion, cytoskeletal organisation, angiogenesis, metastasis, and the activation of GTPases [503]. The overexpression of Thy-1 protein was described recently as a poor indicator of the survival of BC patients and was also associated with metastasis in the basal-like cell lines [504]. The positive expression of the Thy-1 protein was confirmed in a subpopulation of $CD44^+$ BC cells localised on the invasive periphery of basal-like invasive BC. With the apocrine signalling of the $CD90^+$ stromal cancer-associated fibroblast (CAF) cells, the $CD90^+/CD44^+$ BC cells at the invasive front demonstrated high invasion properties [505]. The involvement of $CAF^{Thy\text{-}1+}$ was studied recently in human NSCLC tissues. The confocal microscopic images of the examined tumours showed that the $NSCLC^{Thy\text{-}1+}$ was stained and localised inside the tumour interstitium, on the tumour-invasive margins, and at the perivascular stromal region [506]. In a different study, the overexpression of Thy-1 was identified as decreasing the activity of ALDH, sphere formation, and stemness ability of the adenocarcinoma in the OC cell line, SKOV3. The latter study concluded that the overexpression of Thy-1 predicts good prognosis for OC patients [507]. It was stated that the gene expression of *THY1* is downregulated in two-thirds of nasopharyngeal carcinoma (NPC) patients [508]. The upregulation of *THY1*

inhibited colony formation and hindered the invasion of NPC in vitro [509]. In order to clarify the inequality of the differential gene expression of *THY1* in the different molecular phenotypes of invasive BC, we assessed the prognostic values of the altered gene expression of *THY1* in a large external cohort of invasive BC patients. The overexpression of the *THY1* gene was specifically linked to poor survival in the ER^- subtypes of invasive BC patients, [188, 189], but not in the ER^+ equivalents.

The $LVI^{+/-}$ cohort of this study was composed exclusively of invasive BC patients with the luminal-A subtype. It appeared that the malicious effects of upregulated *THY1*/Thy-1 in promoting invasion mechanisms in invasive BC were associated with a specific biological context that could be absent in the patients of the luminal-A subtype. The malignancy of Thy-1 may be governed by a cluster of proteins in a phenotype-conditioned ECM. As Thy-1 is involved in cell matrix regulatory mechanisms, and as fibulin-1 acts as an ECM organiser/stabiliser, the cross-talk between fibulin-1 and Thy-1 in promoting/inhibiting LVI and recruiting the stromal components in the different molecular subtypes of invasive BC is interesting area for further research. Our current results for the *FBLN1*/*THY1* axis could unveil novel prognostic features for invasive BC patients.

The next gene expression association of *FBLN1* was with the plasma membrane/secreted protein-encoding gene of *FAP*. The expression of prolyl endopeptidase, or the fibroblast activation protein (FAP), promotes ECM degradation, tissue remodelling, fibrosis, inflammation, cell adhesion, wound healing, migration, and invasion of cancer cells [510]. The associations between the overexpression of FAP and human cancers have been reported previously. The upregulation of FAP was correlated with poor prognosis in oesophageal adenocarcinoma patients and was one of five genes of a stromal gene signature that distinguished invasive tumours from pre-invasive tumours [511]. The flow cytometry analyses of specimens of freshly disassociated single-cell suspension of human invasive BC revealed that the expression of FAP was not restricted to CAF in the tumour stroma. In fact, a subpopulation of $CD45^+$ cells was also detected and indicated that FAP could also be expressed by fibrocyte or leukocyte cells. The authors also found that the $FAP^+/CD45^+$ subtype of cells showed a positive expression of CD14, which may suggest a novel involvement of TAM in the expression of FAP, besides CAF, in invasive BC [512].

From a different perspective, the expression of FAP was visualised in the MCF7 and MDA-MB-231 BC cell lines by applying immunofluorescence staining in another study [513]. The upregulation of FAP enhanced growth in the MCF7 cell line, but not MDA-MB-231. The growth of MDA-MB231 could be not modulated by FAP. It

was suggested that the increased expression of the prolyl endopeptidase FAP protein could decrease BC cells' adhesion to ECM and migration abilities through the deactivation of the focal adhesion kinase (FAK) protein; however, the exact effects of the FAP protein in invasive BC remain vague [513]. The FAK pathway was reported previously as promoting invasion and migration in several cancer studies, and targeting the FAK-mediated pathways by several agents inhibited tumour cell growth, migration, and invasion in invasive BC [514], glioma [515], and HNSCC [516]. In an earlier study, the high expression of FAP protein in the MDA-MB-231 cells was reported to be highly associated with rapid growth and a significant increase in tumour microvessel density [517]. A later study identified that the activation of FAK is required to induce EMT, migration, and invasion in triple-negative BC [518]. Another invasive BC study applied IHC methods to investigate the protein expression of FAP in 642 BC patients. The assessment of FAP expression was performed under the consideration of different molecular subtypes and stromal histological patterns. The expression of CAF-related proteins, including FAP, were at their highest levels in the HER2 subtype, whereas the luminal-A subtype was associated with the lowest expression levels. Regarding the stromal classes, FAP was differentially expressed according to stromata type. FAP was overexpressed in the inflammatory stroma, while the reduced expression of FAP was seen in the sclerotic type [519]. The overexpression of FAP may be associated with inflammatory cells and cytokines-enriched stroma in the ER^- tumours.

In this study, approximately 70.6% of the $LVI^{+/-}$ invasive BC cohort, which was utilised for the RNA microarray study, comprised mainly ER^+ patients defined by IHC (LVI^+, 63.3%; LVI^-, 78.8%). The $LVI^{+/-}$ cohort in the MS experiment consisted entirely of the luminal-A subtype. Therefore, the identified LVI-DEM in this study are most likely to be related to the ER^+ biochemical networks. Nonetheless, the expressed genetic profiles of the luminal-B, HER2-enriched, and basal-like subtypes were distributed randomly among those patients, as the results of the PAM50 classifier showed. Thus, the prognostic implementations of the T8 LVI-candidate biomarkers may vary dramatically. The positive expression association between *FBLN1* and *FAP* could propose good prognostic features for the positive expression of *FAP* in the luminal-A subtype, as the results of the PAM50 and IntClusts classifications for *FBLN1* showed. In the $LVI^{+/-}$ cohort, the gene expression relationship between *FBLN1* and *FAP* was unique, and no other molecules from the list of T8 showed gene expression correlation with *FAP*, except for *FBLN1*. Moreover, the expression relationship between *FBLN1* and *FAP* was inversed in the METABRIC cohort. Focused studies on the pivotal interactions of the

fibulin-1/prolyl endopeptidase axis may reveal unique subclasses of invasive BC patients with better therapeutic options.

Another cell surface and ECM-secreted protein that was statistically correlated with the gene expression of *FBLN1* was *SDC1*. Syndecan-1 is a proteoglycan protein that is attached to exchangeable heparan sulphate chains that support its functions and structure [520]. The knockdown of *SDC1* expression reduced the migration and invasion in the A172 and U87 human glioblastoma cell lines, and downregulated the expressions of *MMP9* and the proliferating cell nuclear antigen-encoding gene *PCNA* [521]. The upregulation of *SDC1* was reported to be associated with BC and poor prognosis but not with normal tissues. In addition, *SDC1* was co-expressed with the urokinase-type plasminogen activator-encoding gene *PLAU* [522]. In a recent study, the syndecan-1 protein band was visualised at 67 KDa after the elimination of the glycosaminoglycan chains by adding heparinase to the cell lysates of the MCF7 and MDA-MB-231 BC cell lines. The MCF7 cell lines expressed an abundant amount of the syndecan-1 protein in comparison to the MDA-MB-231 cell line, and this expression was associated with lower cellular mobility and invasiveness of the MCF7 cells. It was stated that the membrane-bound syndecan-1 acted as an inhibitor for BC invasion; conversely, the secreted syndecan-1 played an invasion-promoting role [523].

The protein expression of syndecan-1 was scrutinised in the different molecular subtypes of invasive BC by applying IHC methods to invasive BC TMA sections. It was determined that the high expression of syndecan-1 was not associated with the luminal subtypes. In fact, high-grade tumours and the subtypes of HER2 and triple-negative BC were strongly related to the elevated levels of syndecan-1 protein in the examined specimens. The localisation of syncedan-1 was both cytoplasmic and membranous, and stromal. The stromal expression of syndecan-1 was observed in the cytoplasm of the stromal fibroblasts and was associated with the patients of unfavourable phenotypes of invasive BC [524]. In a subsequent study, the siRNA-silencing of the *SDC1* expression reduced the expression of ALDH1 in the MDA-MB-231 and MCF7 invasive BC cell lines. The knockdown of *SDC1* altered the expressions of E-cadherin, vimentin, interleukin-6, interleukin-6 receptor subunit α, and C-C motif chemokine 20 in the MDA-MB-231 cells. It was proven that syndecan-1 is pivotal for the development of EMT and CSC properties in iBC patients [525].

Interestingly, the stromal expression and the accumulation of fibronectin was suggested to be regulated by syndecan-1 [526]. We stated earlier that another syndecan protein, syndecan-4, may collaboratively inhibit tumour cell migration

signalling with fibulin-1 and fibronectin in PC and HCC [480]. The altered stromal expression of syndecan-1 may promote directional migration of invasive BC cells via an extensive modulation of ECM; however, fibronectin was proposed as supporting the attachment of invasive BC cells in ECM but not their migration or invasion. We validated the proposed prognostic features of the co-expressed *FBLN1*/*SDC1* in a larger cohort of invasive BC patients with different molecular subtypes and LNM status. The exhaustive validation results showed that *SDC1*$^{+}$ patients with poor survival were ER^{-}/LN$^{+/-}$, but not ER^{+}/LN$^{+/-}$, invasive BC patients [188, 189].

5.4.5.6. *Inferences of FBLN1*

In this study, we identified that the positive expression of *FBLN1* is strongly correlated with good prognosis and lower LVI incidences in LVI$^{+/-}$ invasive BC patients. The positive gene expression of *FBLN1* also showed a trend toward significance in the LVI^{-} patients in the METABRIC cohort, although this subset included LNM$^{+/-}$ patients. In both cohorts, *FBLN1*$^{+}$ patients were associated with low tumour grade and good NPI. In the LVI$^{+/-}$ cohort, *FBLN1*$^{-}$ patients displayed a significant number of LNM^{+} events, whereas no associations were found for the METABRIC patients. Our results of the survival analysis in this study showed that the patients of the *FBLN1*$^{+}$ expression profile survived for much longer than their *FBLN1*$^{-}$ counterparts in the LVI$^{+/-}$, METABRIC, and external validation cohorts. However, we could not identify independent prognostic features for *FBLN1*, and this may underscore the several key players that could dictate the behaviour of *FBLN1*$^{+/-}$ invasive BC.

At the molecular level, the differential expression of *FBLN1* in LVI$^{+/-}$ patients was exceptionally controversial. The good prognostic features of *FBLN1* could be rejected due to the several positive correlations with a list of MMP encoding genes. It appears that the protein localisation and PPIs for the fibulin-1 protein could be modulated by several other factors, and this may negatively impact the good-prognosis feature of fibulin-1. The immune-interrelations for *FBLN1* among the patients of the studied cohorts were extremely different. The positive expression of *FBLN1* was associated inversely with *CD20* and *CD80* in the LVI$^{+/-}$ patients, and all of the positive associations that were presented in the METABRIC cohort were totally absent in the LVI$^{+/-}$ patients. Our results suggest that the positive expression of *FBLN1* may block the BC recruitment of TAM and other myeloid cells in TME. The *FBLN1*-modulated immune responses require further investigation and better understanding. Regarding the associations between *FBLN1* and EMT and invasion-related genes, our results indicated that the altered gene expression of *FBLN1* may

plot several cross-talk scenarios with various ECM-secreted proteins and cancer-associated stromal cells. We suggest a preferential good-prognostic characteristic for the positive gene expression of *FBLN1* in $ER^{+}/LNM^{+/-}$ patients. The overexpression of *FBLN1* may attenuate the proliferation, EMT, migration, and invasion capacities of malignant BC cells in $ER^{+}/LNM^{+/-}$ patients. However, these gene expression correlations of *FBLN1* could be contextualised into different prognostic suggestions in patients of the ER^{-} and $ER^{+}/HER2^{+}$ phenotypes, especially *CDH1* $^{+}$, *SNAI2*$^{+}$, *VIM*$^{+}$, and *MKI67*$^{+}$ iBC/triple-negative BC patients.

According to the results in this study, we encourage further examination of the interactive associations of *FBLN1*/fibulin-1 in invasive BC specimens with robustly defined LVI status. The different molecular subtypes of these invasive BC samples and their histological types of stromata must also be considered. The inclusion of the CSC and EMT biomarkers (e.g. *CD44*, *CD24*, *CDH1*, *CDH2*, and *VIM*) in future scrutinisation of the gene expression hub and the protein-protein interactions of *FBLN1*/fibulin-1 in $LVI^{+/-}$ invasive BC patients is also recommended.

5.5. Molecular architecture of LVI in invasive BC patients: A chapter conclusion

Invasive BC is one of the lethal diseases that lead to death in women worldwide if not detected early and treated. The identification of several molecular signatures that may enhance the clinical interventions has been achieved by many research groups and for different malignancies. However, the idealised molecular signature for LVI has remained undiscovered. In this study, we utilised the high-throughput molecular biology techniques of RNA microarray and MS to identify a suggested molecular signature consisting of differentially expressed genes and proteins for invasive BC patients with robustly defined LVI status.

The proposed gene signature for LVI in this study included an overlying list of 56 mRNA/protein identifiers, the LVI-DEM. We assessed the gene expression network for these LVI-DEM on several online gene networking platforms. We also conducted a comprehensive gene network interrogation of all of the 56 identified LVI-DEM to create a novel molecular network of interactor nodes and interaction edges. We nominated a subset of eight mRNAs/proteins, the T8 list, that were statistically and positively correlated with LVI occurrence and LVI absence. The T8 list was divided into five highly expressed genes/proteins with LVI^{+} (*HSP90AA1*, *ARHGDIA*, *DYNLRB1*, *VDAC2*, and *PPIA*) and three highly expressed genes/proteins with LVI^{-} (*FBLN1*, *RPS18*, and *RPL10A*). The distribution statistical analyses shed light on the aberrant expressions of *HSP90AA1* and *FBLN1* as potential prognostic

biomarkers for which the high expressions were associated respectively with LVI^{+} and poor survival and LVI^{-} and good survival in invasive BC patients of the $LVI^{+/-}$ cohort. The molecular interactions for all of the T8 molecules, principally *HSP90AA1*, *VDAC2*, and *FBLN1*, differed vastly between the $LVI^{+/-}$ and METABRIC studied cohorts. We can suggest that the initiation and progression of LVI in invasive BC are dictated by unique phenotype-driven and TME-stimulated pathways.

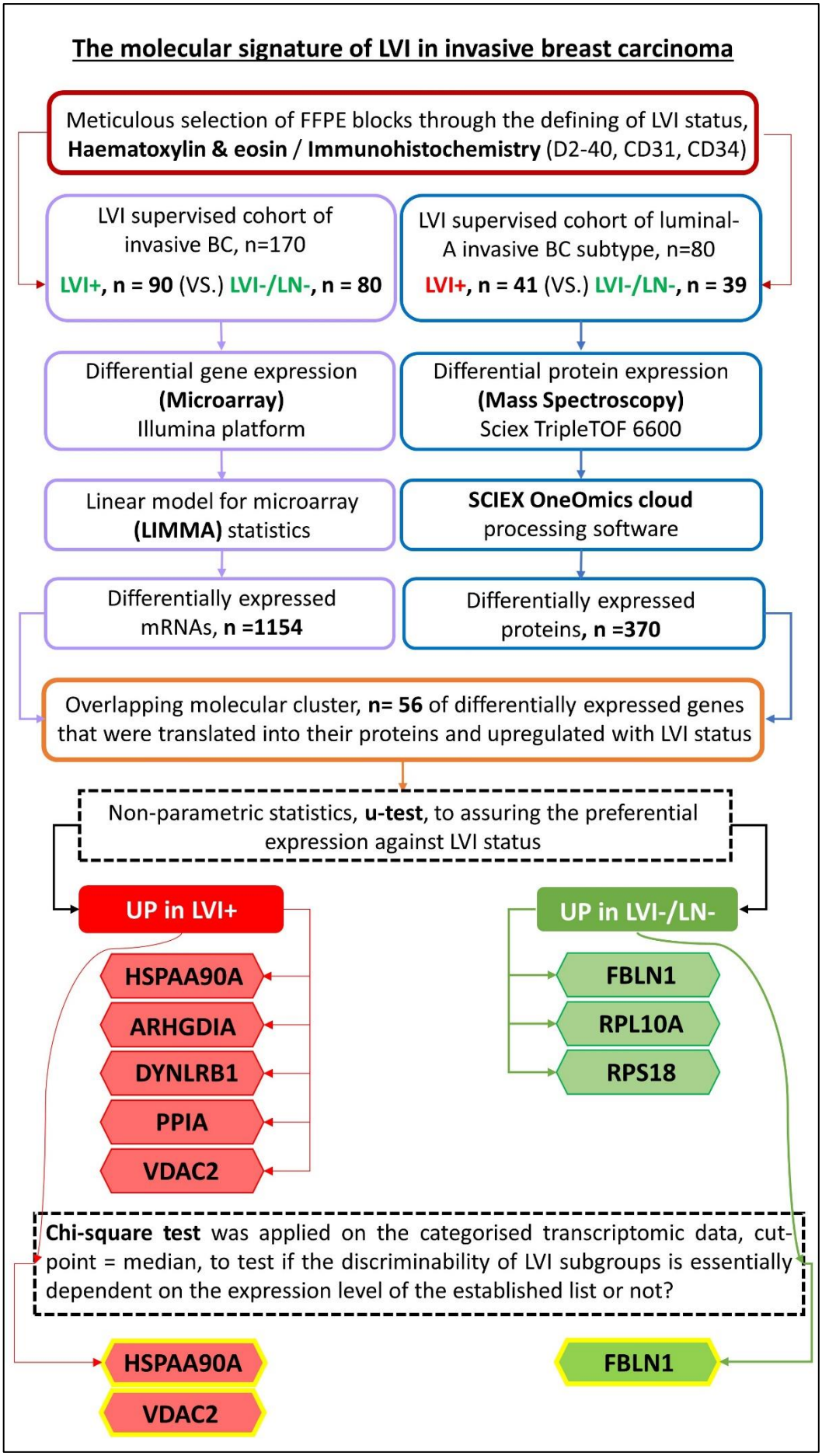
Specific molecular phenotypes or intrinsic subclasses of invasive BC are enriched for stem cell, high energy demand, and fast proliferation biological processes, whereas others are enriched for less metabolic, migration, invasion, and high drug-sensitivity qualities. As we stated in the introduction to this study, the unveiling of a conclusive molecular signature for LVI in invasive BC patients could be an interesting, yet difficult task. The positive gene expression correlation between *HSP90AA1*/heat shock protein HSP 90-α and *VDAC2*/voltage-dependent anion-selective channel protein 2, besides their strong association with LVI definite events in all statistical approaches in this study, may suggest an LVI^{+} molecular hub that places them at the centre. We identified that the subclasses of invasive BC patients most likely to upregulate *HSP90AA1* and *VDAC2* and possibly be at higher susceptibility to LVI initiation and progression are those that belong to the luminal-B and HER2-enriched intrinsic subtypes. Moreover, the vast majority of *HSP90AA1*+ patients were aggregated in IntClust-1 and IntClust-9, which mainly comprised $ER^{+/-}$/$HER2^{+}$ patients. On the other hand, an LVI^{-} molecular hub that could be functionally effective in lowering LVI incidences comprised *FBLN1*/fibulin-1 at its core. The *FBLN1*$^{+}$ patients were the least invasive BC patients with LVI^{+} events in this study. A relevant number of *FBLN1*$^{+}$ patients were clustered in the normal-like and luminal-A intrinsic subtypes and in IntClust-3 and IntClust-4 in the $LVI^{+/-}$ and METABRIC cohorts. The comprehensive and systematic scrutinisation of the interrelating genes and interacting proteins surrounding the molecular hubs of HSP90AA1/heat shock protein HSP 90-α, VDAC2/voltage-dependent anion-selective channel protein 2, and *FBLN1*/fibulin-1 was the essence of the aim of this study.

In this study, our results could facilitate a better understanding of the cellular and stromal factors that could regulate LVI mechanisms in invasive BC. In order to decipher the exact molecular regulation for LVI in invasive BC, the altered gene expression level, aberrant protein localisation, and protein-protein interactions for each expressed protein from our prospective LVI drivers/inhibitors are critical subjects to be included in future LVI studies.

A summarised list of LVI-differentially expressed and overlapped genes/proteins produced by RNA microarray and HPLC-MS protocols

Red label = upregulated in LVI^{+}; **Green label** = upregulated in LVI^{-}/LN^{-}

Differentiall expressed and overlapped proteins (DEP) in gene ID format	Full name	Adjusted p-value (LIMMA)	p-value (Mann-Whitney U test)	p-value (Chi square test)
ACTB	Actin, cytoplasmic 1	0.0064	0.5509	NA
ACTG1	Actin, cytoplasmic 2	0.0017	0.4351	NA
ANP32A	Acidic leucine-rich nuclear phosphoprotein 32 family member A	0.0016	0.9726	NA
ARHGDIA	**Rho GDP-dissociation inhibitor 1**	**0.0332**	**0.0043**	0.065
ARPC5	Actin-related protein 2/3 complex subunit 5	0.0203	0.2038	NA
CCT8	T-complex protein 1 subunit theta	0.0013	0.1389	NA
COL6A2	Collagen alpha-2(VI) chain;COL6A2	0.0007	0.7644	NA
COMT	Catechol O-methyltransferase	0.0041	0.9875	NA
DBI	Acyl-CoA-binding protein	0.0114	0.7811	NA
DDX17	Probable ATP-dependent RNA helicase DDX17	0.0079	0.9179	NA
DYNLRB1	**Dynein light chain roadblock-type 1**	**0.0014**	**0.018**	0.357
EIF4E	Eukaryotic translation initiation factor 4E	0.0020	0.2222	NA
ERH	Enhancer of rudimentary homolog	0.0270	0.7359	NA
FBLN1	**Fibulin-1**	**0.0015**	**0.004**	**0.006**
H2AFY	Core histone macro-H2A.1	0.0036	0.9738	NA
HMGB2	High mobility group protein B2	0.0079	0.221	NA
HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1	0.0264	0.9502	NA
HNRNPH3	Heterogeneous nuclear ribonucleoprotein H3	0.0032	0.908	NA
HNRNPK	Heterogeneous nuclear ribonucleoprotein K	0.0280	0.504	NA
HSP90AA1	**Heat shock protein HSP 90-alpha**	**0.0153**	**0.0001**	**0.00000017**
HSPA4	Heat shock 70 kDa protein 4	0.0008	0.2111	NA
ILF3	Interleukin enhancer-binding factor 3	0.0063	0.9055	NA
KCTD12	BTB/POZ domain-containing protein KCTD12	0.0415	0.0677	NA
LGALS3	Galectin-3	0.0231	0.2506	NA
LMNA	Prelamin-A/C	0.0047	0.9402	NA
OLA1	Obg-like ATPase 1	0.0082	0.7859	NA
PGAM1	Phosphoglycerate mutase 1	0.0141	0.9402	NA
PPIA	**Peptidyl-prolyl cis-trans isomerase A**	**0.0079**	**0.047**	0.124
PRRC1	Protein PRRC1;PRRC1	0.0073	0.3315	NA
PSMA6	Proteasome subunit alpha type-6	0.0408	0.2584	NA
PTGES3	Prostaglandin E synthase 3	0.0122	0.16	NA
PTMA	Prothymosin alpha	0.0022	0.2257	NA
RAC1	Ras-related C3 botulinum toxin substrate 1	0.0043	0.4098	NA
RBBP4	Histone-binding protein RBBP4	0.0015	0.5509	NA
RPL10A	**60S ribosomal protein L10a**	**0.0232**	**0.0457**	0.124
RPL7	60S ribosomal protein L7	0.0035	0.8221	NA
RPL9	60S ribosomal protein L9	0.0093	0.4133	NA
RPLP1	60S acidic ribosomal protein P1	0.0093	0.9278	NA
RPS11	40S ribosomal protein S11	0.0079	0.28	NA
RPS18	**40S ribosomal protein S18**	**0.0473**	**0.037**	0.219
RPS2	40S ribosomal protein S2	0.0101	0.3521	NA
RPS28	40S ribosomal protein S28	0.0228	0.2293	NA
RPS4X	40S ribosomal protein S4, X isoform	0.0095	0.259	NA
RPS7	40S ribosomal protein S7	0.0195	0.995	NA
RPSA	40S ribosomal protein SA	0.0032	0.098	NA
SAR1A	GTP-binding protein SAR1a	0.0072	0.743	NA
SEC61B	Protein transport protein Sec61 subunit beta	0.0362	0.3073	NA
SNRPE	Small nuclear ribonucleoprotein E	0.0085	0.9701	NA
SRSF3	Serine/arginine-rich splicing factor 3	0.0024	0.0641	NA
SUB1	Activated RNA polymerase II transcriptional coactivator p15	0.0176	0.1012	NA
TPI1	Triosephosphate isomerase	0.0072	0.1666	NA
TUBB6	Tubulin beta-6 chain	0.0157	0.3752	NA
UFM1	Ubiquitin-fold modifier 1	0.0010	0.8318	NA
UPF1	Regulator of nonsense transcripts 1	0.0420	0.3735	NA
VDAC2	**Voltage-dependent anion-selective channel protein 2**	**0.0426**	**0.0302**	**0.006**
YWHAE	14-3-3 protein epsilon	0.0019	0.6174	NA


The molecular signature of LVI in invasive breast carcinoma
Meticulous selection of FFPE blocks through the defining of LVI status, Haematoxylin & eosin / Immunohistochemistry (D2-40, CD31, CD34)
LVI supervised cohort of invasive BC, n=170
LVI+, n = 90 (VS.) LVI-/LN-, n = 80
LVI supervised cohort of luminal-A invasive BC subtype, n=80
LVI+, n = 41 (VS.) LVI-/LN-, n = 39
Differential gene expression (Microarray) Illumina platform
Differential protein expression (Mass Spectroscopy) Sciex TripleTOF 6600
Linear model for microarray (LIMMA) statistics
SCIEX OneOmics cloud processing software
Differentially expressed mRNAs, n =1154
Differentially expressed proteins, n =370
Overlapping molecular cluster, n= 56 of differentially expressed genes that were translated into their proteins and upregulated with LVI status
Non-parametric statistics, u-test, to assuring the preferential expression against LVI status
UP in LVI+
HSPAA90A
ARHGDIA
DYNLRB1
PPIA
VDAC2
UP in LVI-/LN-
FBLN1
RPL10A
RPS18
Chi-square test was applied on the categorised transcriptomic data, cut-point = median, to test if the discriminability of LVI subgroups is essentially dependent on the expression level of the established list or not?
HSPAA90A
VDAC2
FBLN1

Chapter 6: General Discussion & Conclusion

6.1. General discussion

As we described and discussed in the previous chapters, invasive BC is one of the most frequently diagnosed malignancies among women. Invasive BC is also a leading cause of female mortality worldwide. Hence, enhancement of the current prevention, diagnosis, prognosis, and therapeutic management of invasive BC is of crucial significance for female health and well-being. This task is challenging because BC is a heterogeneous disease that manifests myriad morphological and molecular entities, and it has been demonstrated that each BC patient can exhibit a unique molecular architecture, inter-tumour heterogeneity, or, even within the same tumour, intra-tumour heterogeneity [527, 528].

The progression of BC to a metastatic disease is highly associated with worse survival in invasive BC patients, and it requires certain molecular settings that include different proteins and converging pathways [70-72]. The key molecular players of BC progression can modulate BC growth, metabolism, cellular adhesion, migration, and invasion. In clinical studies, LVI is stated as one of the strongly correlated prognostic factors with BCSS [529]. At the molecular level, BC cells can employ specific biological reactions, facilitating their invasion-metastatic cascade of events through LVI [53, 73-75]. The exploration of DEGs and DEPs related to LVI signalling pathways is crucial to finding LVI-guiding molecules, and it can produce novel prognostic factors and therapeutic candidates for invasive BC patients. The interest in studying cancer-related pathways and identifying clusters of DEGs and DEPs through the employment of genomic, transcriptomic, and proteomic-based methodologies has increased recently. Regardless, the governing molecular mechanisms and LVI-driver genes/proteins in invasive BC remain ambiguous.

We used an exploratory design for this research, pursuing reliable LVI prognostic candidates for invasive BC patients. The LVI signature in invasive BC was poorly studied. Therefore, we focused on gaining a large number of possible LVI-related molecules utilising transcriptomic and proteomic high-throughput techniques. The identification of LVI candidates by transcriptomic techniques may provide a large number of identifiers from which to select; however, functional proteins are the key to deciphering LVI pathways. Biologically, a successful LVI event by cancer cells demands the recruitment of various cellular, biophysical, and biochemical machineries.

The regulation of LVI mechanisms in invasive BC patients can require specific genes and protein expression profiles, as we discussed thoroughly in section 1.3 in the first chapter. Few LVI gene-signature studies were conducted previously on BC

[142], HCC [143], and endometrial cancer [141], and their results suggested specific clusters of aberrantly expressed genes that are associated with LVI mechanisms. Therefore, our hypothesis stated that the manifestation of LVI in invasive BC patients is influenced by specific molecular expression profiles that can be associated with distinctive classes of invasive BC patients. To verify this hypothesis, we designated a list of experimental objectives in the current study. To explore LVI mechanisms in invasive BC patients, we tested two meticulously described LVI-supervised cohorts for their differential transcriptomic profiles. In one experimental design, the selection of candidate DEGs was based on their statistical association with $LVI^{+/-}$ (p. value <0.05), CNA statistical analysis, (p. value <0.05), and calculated CNA gain (ratio > 2.5). Several transcripts were tested at the protein level by applying IHC methods [43, 190, 192, 193]. In this study, and according to the gene selection criteria mentioned above, the associations between LVI and the aberrant expression of MCFD2 and CCT8 were interrogated in two large cohorts of invasive BC patients at the mRNA and protein levels. Prognostic implications and patients' outcomes were also considered for MCFD2 and CCT8. A separate approach was also used to identify associated genes with MCFD2 and CCT8. The investigation of GSEA studies and the MSigDB unveiled SNAP23 as a co-expressed gene with MCFD2 and CCT8. The clinicopathological implications, including LVI, for the SNAP23 protein were also investigated with IHC and presented in this study.

A second exhaustive experimental design was based on the complete inclusion of the RNA microarray transcriptome produced by the DGE. The selection of accurately LVI-associated transcripts was based on their statistical association with LVI status, log2FC >1.0, positive AvE values, and perfect quality of probe matching. This experimental design produced 283 LVI-associated transcripts. The assessment of individual transcripts against the histological LVI status of a subset of invasive BC patients separated the identified transcripts into three unique groups: 1) a total of 13 upregulated transcripts with LVI^{+}; 2) a total of 22 upregulated transcripts with LVI^{-}; and 3) a cluster with no LVI associations. Further analysis showed SEC14L1 to be an excellent LVI^{+} candidate. The aberrant protein expression of SEC14L1 was also tested by applying IHC methods [191]. For MCFD2 and CCT8, clinical impact and BC outcome were also measured for SEC14L1. The first and second experimental designs yielded different results. Therefore, we aimed to utilise HPLC/MS as another high-throughput technique to identify DEPs associated with LVI in a third experimental design.

For this purpose, we constructed a different cohort of invasive BC patients with accurately recorded LVI/LN data. This cohort consisted of 80 $ER^{+}/HER2^{-}$

tumours because a relevant number of samples was required for the HPLC/MS experiment. The identified DEPs in the $LVI^{+/-}$ patients were matched against the previously identified DEGs by RNA microarray to discover the overlapping identifiers. The 56 matched identifiers of genes/proteins for the $LVI^{+/-}$ patients were assessed by applying u-test statistics against the histological LVI status. This assessment produced a total of eight genes/proteins associated with the LVI status of the studied invasive BC patients. The protein-encoding genes of *HSP90AA1*, *ARHGDAI*, *DYNLRB1*, *VDAC2*, and *PPIA* were upregulated in the LVI^{+} invasive BC patients, but *FBLN1*, *RPS18*, and *RPL10* were upregulated in the LVI^{-} invasive BC patients. However, the application of the goodness-of-fit statistical model showed that *HSP90AA1*, *VDAC2*, and *FBLN1* were of greater consistence with LVI incidence. Currently, our group is investigating the LVI driver molecules generated by this study. HSP90-α protein has been IHC stained by a first-year PhD student in our group, and the clinicopathological analysis revealed that HSP90-α protein has strong potentialities in driving LVI mechanisms at the protein level in invasive BC patients.

6.2. Preeminent findings of the study

To recap, the first method we followed to identify LVI-related genes was based on the results of the RNA microarray DGE experiment and the analysis of CNA. Among the top 100 DEGs, MCFD2 and CCT8 were the only transcripts that showed significant association with LVI, significant CNA difference, and a CNA gain ratio >2.5. The altered mRNA expressions of MCFD2 and CCT8 did not show clinicopathological relevance, although they were precisely selected based on their statistical results against LVI and CNA. In contrast, the altered protein expressions for MCFD2 and CCT8 demonstrated a strong association with LVI. However, the relationships between LVI and the altered protein expression of MCFD2 and CCT8 were inversed and conditioned by the protein sub-cellular localisations. The opposite direction of expression and the clinicopathological implications for MCFD2 were mentioned previously in a testicular seminoma study [202]. The MCFD2 gene expression was downregulated in testicular seminoma according to the microarray results, but the IHC results showed that the MCFD2 protein was overexpressed in the germ cells of the examined tissues. The justification for this difference was attributed to post-transcriptional changes and mRNA splicing. Our results showed that MCFD2 is preferentially upregulated in the basal intrinsic subtype and in IntClust-10 at the mRNA level. However, MCFD2 protein expression was generally associated with ER^{+} tumours. The MCFD2 protein also showed independent prognostic features of disease recurrence when upregulated. These findings indicate

a possible association between MCFD2 positive expression and the presence of CSC and recurrence in aggressive BC phenotypes. A recent study indicated that MCFD2 is a key protein that enhances metastasis in OSCC [200]. Based on the limited number of studies available, MCFD2 appears to be an interesting prognostic target in invasive BC. Although the positive expression of the MCFD2 protein exhibited an overall good prognosis in invasive BC and positive association with the LVI^-/LN^- patients, the interrogation of MCFD2 positive expression in the CSC-enriched and intra-tumoural heterogeneous BC phenotypes, such as triple-negative [530, 531], is warranted.

To some extent, the results of CCT8 were similar to those of MCFD2. The overexpression of CCT8 in the nuclear sub-cellular compartment exhibited tumour-supressing qualities, and the comparison between nuclear and cytoplasmic CCT8 expressions regarding their prognostic impacts in invasive BC patients was extremely different. All the tested clinicopathological parameters were inversely and significantly associated with the nuclear positive expression of CCT8, but the cytoplasmic expression was not. The nuclear positive expression of CCT8 was majorly associated with ER^+ tumours, and the largest proportion of tumours that exhibited a significant loss of the nuclear CCT8 expression were the triple-negative. Our results showed that a sub-class of invasive BC patients with downregulated levels of nuclear CCT8 was susceptible to developing metastatic disease, BC recurrence, and poor survival.

Existing studies that interrogated the altered expression of CCT8 in different human cancers could not identify a nuclear expression for CCT8 in the IHC-examined tissues; instead, the cytosolic expression of CCT8 was stated to be correlated with poor outcomes and promotion of tumour proliferation, migration, invasion, and metastasis in ESCC [205], HCC [203, 239], and glioma [204]. According to our results, CCT8 was one of the LVI-DEGs in the microarray experiment, a DEP by IHC staining, and one of the 56 overlapping gene/protein identifiers by HPLC/MS procedure. Currently, there are no studies investigating the associations between CCT8 and LVI in invasive BC. The novel association between CCT8 and LVI backed up with presented clinicopathological results encourages scrutiny of the CCT8 protein in well-designed functional studies in the future.

The expression of SNAP23 also showed two distinctive sub-cellular localisations, cytosolic and membranous. Membranous positive expression of SNAP23 showed significant associations with LVI definite occurrences and tumour grade-I and grade-II. The SNAP23 protein functions as a vital constituent of the membrane fusion machinery and a crucial modulator of vesicular trafficking and

vesicle docking and fusion. This may suggest that SNAP23 is an important sensor protein that mediates the conveying of other proteins related to the invasion mechanism, especially on the peripheral invasive front [295]. LVI is not strongly associated with low histological grades [532]. Being relatively associated with lower tumour grades, the positive membranous expression of SNAP23 could be a beneficial predictive tool for the detection of LVI development in the early stages.

The last IHC-examined protein was SEC14L1. In this study, the selection of SEC14L1 was sophisticated compared to the selection of MCFD2, CCT8, and SNAP23. We managed to identify a 35 LVI-gene signature by utilising the complete METABRIC transcriptomic data. SEC14L1 mRNA was one of the top positively expressed molecules in LVI^{+} patients and was nominated for IHC staining. The overexpression of SEC14L1 was strongly associated with LVI definite incidences in invasive BC patients. The statistical figures presented in this study differ from the published versions because of altered statistical conditions. Nevertheless, the association between LVI^{+} tumours and SEC14L1 upregulation was conserved. The SEC14L1 was IHC stained in two cohorts during this study: the $LVI^{+/-}$ cohort, to which the presented statistical figures refer, and the recently constructed TMA set derived from the new cohort of Nottingham-invasive BC patients, diagnosed between 1998 and 2006. By the time of this study, the clinicopathological data of the latter cohort were not completed; the HER2 status was not validated by in situ hybridisation methods; and the LVI status was not differentiated from BVI by D2-40, CD31, and CD34 IHC staining. Better insights may be obtained by functional investigations of the roles of SEC14L1 in promoting LVI mechanisms in invasive BC patients.

From the results of the IHC experiments, the present study identified four promising proteins of clinicopathological importance and a strong association with LVI. The cytosolic MCFD2 and nuclear CCT8 represent two novel good prognosis candidates for LVI absence. The membranous SNAP23 and cytosolic SEC14L1 were associated with poor outcomes and LVI definite occurrence in invasive BC patients. As we discussed in detail in the previous discussion sections, MCFD2, CCT8, SNAP23, and SEC14L1 can be correlated with the Golgi-mediated secretory pathway, microtubule-directional vesicular transport system, and membrane trafficking pathway [205, 231, 271, 273, 294, 295, 301, 533].

The third experiment involved a comparison between two LVI signatures obtained by RNA microarray and HPLC/MS techniques. HSP90-α, VDAC2, and fibulin-1 were three top LVI-DEPs that were also differentially expressed at the mRNA level against LVI.

$HER2^+$ BC tumours, like CCT8 [331]. Our transcriptomic analysis of HSP90AA1 and VDAC2 showed strong positive associations between the upregulated levels of HSP90AA1 and VDAC2, with LVI positivity, large tumour size, high tumour grade, poor NPI, and a preferential upregulation in HER2-amplified tumours. The upregulated expressions of HSP90AA1 and VDAC2 were also associated with unfavourable outcomes in the $LVI^{+/-}$ and METABRIC cohorts of invasive BC patients. The expression of HSP90AA1 was recently suggested as a part of the PathCR for BC patients [338], and VDAC2 was reported to be an excellent predictor of BC recurrence, distant metastasis, and treatment failure [394]. The strong association between LVI^+ BC patients and the expression profiles of $VDAC2^+$ and $HSP90AA1^+$ represents the cardinal finding for RNA microarray/HPLC-MS integrative studies. In contrast, the upregulation in FBLN1 was correlated with better outcomes in the $LVI^{+/-}$ and METABRIC cohorts and LVI negativity in the $LVI^{+/-}$ cohort. FBLN1 was also one of the identified DEGs in the UP-22/LVI^- LVI signature.

From the findings of the HPLC/MS experiment, the present study suggests three potential LVI drivers that display statistically reliable correlations with LVI and prognostic relevance. The progressive mechanisms of LVI in invasive BC patients may recruit proteins that modulate the HER2 signalling pathway; mitochondrial transport; positive regulation of nitric oxide biosynthetic process; interactions with VEGF receptor, such as HSP90-α; and trafficking of hydrophilic metabolites across the outer membrane of mitochondria, such as VDAC2. Additionally, the disruption of basement membrane proteins, such as fibulin-1, appears necessary for progressive LVI mechanisms and in the mediation of particular ECM stromal components.

6.3. Strengths and limitations of the study

6.3.1. **Potential strengths**

1. The interrogation of LVI DEGs and DEPs on large cohorts of primary invasive BC patients and the proposal of a novel LVI-gene signature.
2. The identification of novel prognostic proteins associated with LVI occurrence in invasive BC patients, such as MCFD2, CCT8, and SEC14L1.
3. The consistent identification of FBLN1 and CCT8 as two LVI-differentially expressed molecules on the mRNA and protein levels in all of the studied cohorts and applied methods.
4. The utilisation of large and well-characterised primary invasive BC cohorts, such as NPS, with available IHC and clinicopathological data relevant to BC.
5. The inclusion of controls with each IHC run to ensure genuine staining.
6. The determination of antibody-antigen binding specificity with WB and IPB.

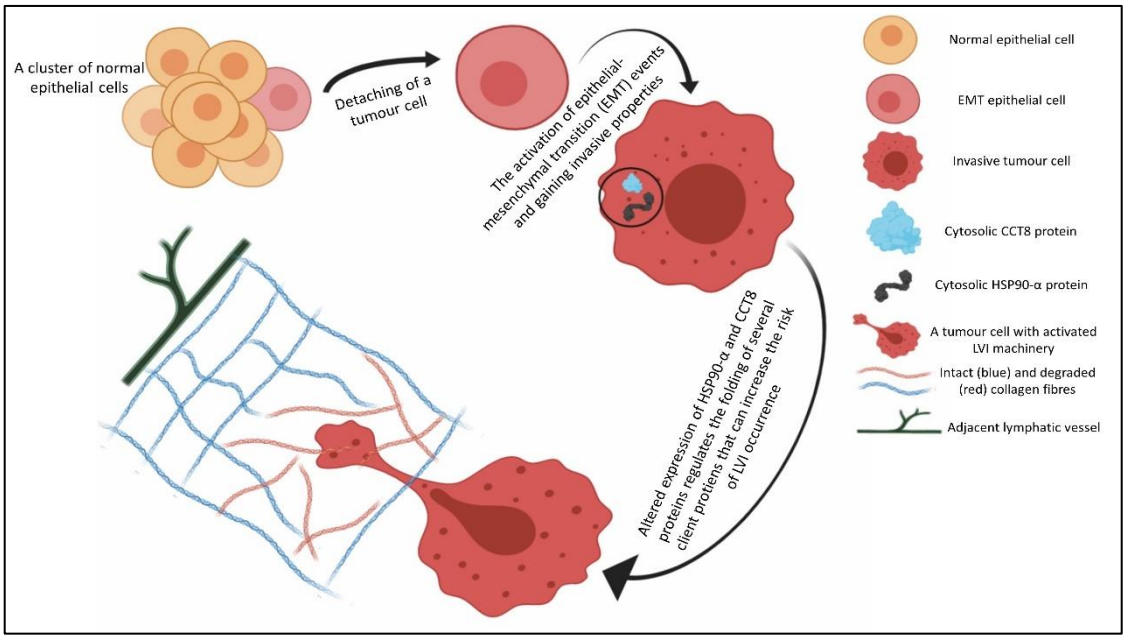
7. The utilisation of the transcriptomic data of the METABRIC cohort.
8. The inclusion of the ten genome-driven integrative clusters of the METABRIC.
9. The provision of research targets for other undergraduate and postgraduate students in our group.
10. Several published studies were derived from this study.
11. The consideration of promising LVI candidates from this study for functional studies in the future.

6.3.2. **Potential limitations**

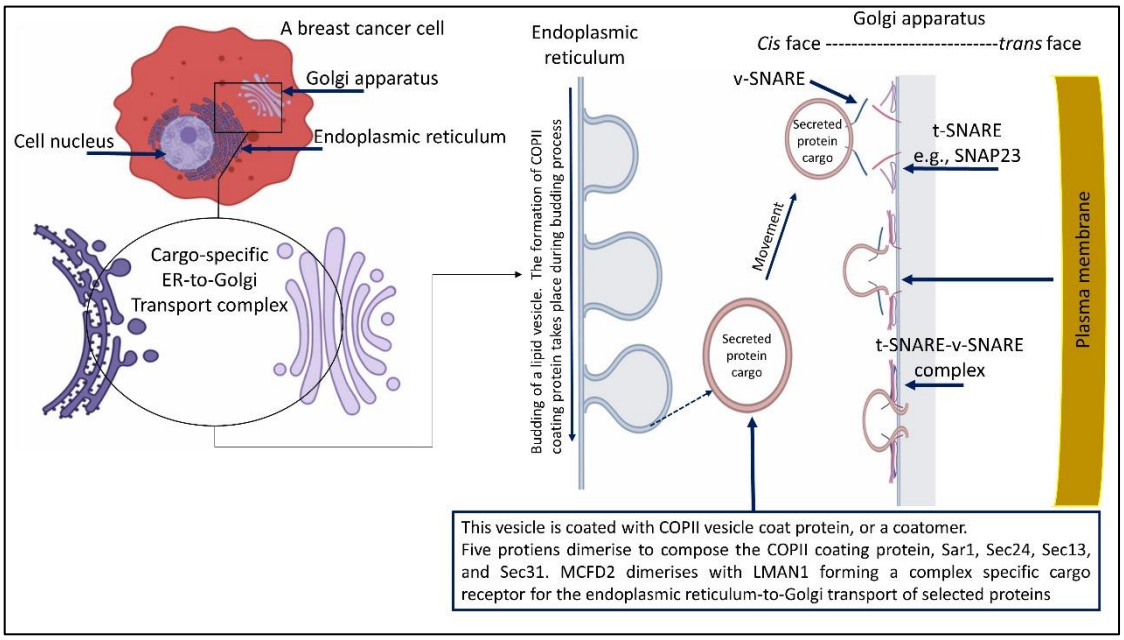
1. For the exploratory research approach, the limited number of published studies interrogating LVI in invasive BC and other malignancies required us to apply several experimental methods rather than scrutinising a specific pathway or a group of LVI-related molecules.
2. The RT-qPCR method was not practicable because we were investigating a large number of LVI candidates.
3. The lack of detailed transcriptomic data at the beginning of this study delayed our progress to some extent.
4. The BC TMA blocks used in this study were historic, and some TMA cores were missing or the tumour burden was marginal.
5. IHC H-scoring is a semi-quantitative method that can be influenced by the scorer's subjective judgment, in spite of which blind double scoring was considered.
6. Although the protein expression patterns for the IHC-stained proteins were homogenous, and the use of TMA on a large scale was acknowledged by different studies, it is worth using at least 2–3 cores per case for an enhanced assessment of these proteins.
7. Cell lysates were not always available in our lab because of non-technical issues.
8. Laser-capture microdissection was the intended procedure to harvest tumour tissues in the HPLC/MS; however, after starting this step, the machine malfunctioned several times, and we decided to apply scalpel macrodissection instead. This modification required full-face staining, assessment of the tumour burden, and marking of the tumour lesions for all of the 80 FFPE blocks used in the HPLC/MS experiment.
9. The incompleteness of the clinicopathological data of the new Nottingham primary invasive BC series prevented the utilisation of such a large cohort.
10. The retrieval and validation of the METABRIC transcriptomic and clinical data took a considerable amount of time.

6.3.3. **Potential hypotheses for future testing**

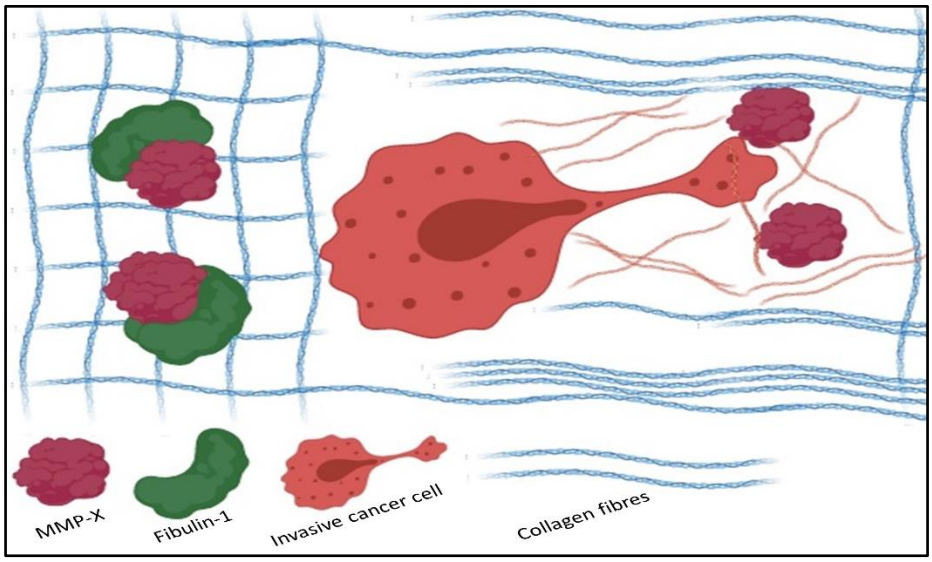
1- The development of LVI in invasive BC require altered expression and functionality of HSP90-α and CCT8 proteins in the chaperonin-mediated protein folding pathways in the cytoplasm; especially for the HER2^{+} tumours.



2- The upregulated protein expression of vesicular trafficking, vesicle docking and fusion, and Golgi-mediated secretory pathways by tumour cells at the invasive front, such as SNAP23, MCFD2, and SEC14L1, is an essential step to promote LVI activity.



3- The fibulins protein family member fibulin-1 can be a crucial inhibitor for the proteolytic degradation of the extracellular matrix by MMPs in invasive BC.



6.4. Future work

1. Investigating the protein expression of SEC14L1 in the new large cohort of primary invasive BC patients.
2. Requesting the original non-normalised transcriptomic data of the METABRIC to design a modified DGE analysis related to LVI.
3. Requesting the LVI status data for the locally assessed Addenbrookes cases.
4. Assessing the molecular regulations of LVI in a large cohort of triple-negative BC patients.
5. Designing functional studies for a selected cluster of LVI associated proteins, including optimum selection of BC cell lines based on the current molecular intrinsic subtyping results.
6. Assessing sub-cellular localisation and co-expression for current LVI candidates in vitro.
7. Getting involved in and producing several publications about LVI molecular regulations.

6.5. Final conclusions

In conclusion, the present study suggests that LVI molecular mechanisms are significantly governed by a specific cluster of molecules at the transcriptomic and proteomic levels. Various genes/proteins were identified as being associated with LVI in invasive BC patients. Of these, MCFD2, CCT8, SNAP23, SEC14L1, HSP90AA1, VDAC2, and FBLN1 were highly ranked LVI candidates in this study. The comprehensive statistical analysis of the identified LVI candidates suggested that the LVI mechanisms in invasive BC patients are differentially regulated according

to the primary BC phenotype. By applying IHC, the LVI implications of particular proteins, such as CCT8 and SNAP23, were found to be dependent on their aberrant sub-cellular localisation. The initiation and progression of LVI mechanisms in invasive BC can also be dictated by intersecting pathways based on the TME conditions and interactions with several stromal components in the ECM, including the cells of the immune system. The use of functional studies is highly recommended for further validation.

REFERENCES

1. Hassiotou, F. and D. Geddes, *Anatomy of the human mammary gland: Current status of knowledge.* Clin Anat, 2013. **26**(1): p. 29-48.
2. Hartmann, P.E., *The lactating breast: an overview from down under.* Breastfeed Med, 2007. **2**(1): p. 3-9.
3. Medina, D., *The mammary gland: a unique organ for the study of development and tumorigenesis.* J Mammary Gland Biol Neoplasia, 1996. **1**(1): p. 5-19.
4. Menes, T.S., et al., *Subsequent Breast Cancer Risk Following Diagnosis of Atypical Ductal Hyperplasia on Needle Biopsy.* JAMA Oncol, 2017. **3**(1): p. 36-41.
5. Maxwell, A.J., et al., *Risk factors for the development of invasive cancer in unresected ductal carcinoma in situ.* European Journal of Surgical Oncology, 2018. **44**(4): p. 429-435.
6. Stingl, J. and C. Caldas, *Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis.* Nat Rev Cancer, 2007. **7**(10): p. 791-9.
7. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
8. Sjostrom, M., et al., *Stem cell biomarker ALDH1A1 in breast cancer shows an association with prognosis and clinicopathological variables that is highly cut-off dependent.* J Clin Pathol, 2015. **68**(12): p. 1012-9.
9. Wu, H., et al., *Can CD44+/CD24- Tumor Cells Be Used to Determine the Extent of Breast Cancer Invasion Following Neoadjuvant Chemotherapy?* J Breast Cancer, 2011. **14**(3): p. 175-80.
10. Braunstein, L.Z. and A.G. Taghian, *Molecular Phenotype, Multigene Assays, and the Locoregional Management of Breast Cancer.* Semin Radiat Oncol, 2016. **26**(1): p. 9-16.
11. Sabiani, L., et al., *Breast cancer in young women: Pathologic features and molecular phenotype.* Breast, 2016. **29**: p. 109-16.
12. Goksu, S.S., et al., *Clinicopathologic features and molecular subtypes of breast cancer in young women (age </=35).* Asian Pac J Cancer Prev, 2014. **15**(16): p. 6665-8.
13. Youlden, D.R., et al., *The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality.* Cancer Epidemiol, 2012. **36**(3): p. 237-48.
14. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.
15. Bray, F. and I. Soerjomataram, *The Changing Global Burden of Cancer: Transitions in Human Development and Implications for Cancer Prevention and Control*, in *Cancer: Disease Control Priorities, Third Edition (Volume 3)*, H. Gelband, et al., Editors. 2015, The International Bank for Reconstruction and Development / The World Bank

(c) 2015 International Bank for Reconstruction and Development / The World Bank.: Washington (DC).

16. Saggu, S., et al., *Recent incidence and descriptive epidemiological survey of breast cancer in Saudi Arabia.* Saudi Med J, 2015. **36**(10): p. 1176-80.
17. El Bcheraoui, C., et al., *Breast cancer screening in Saudi Arabia: free but almost no takers.* PLoS One, 2015. **10**(3): p. e0119051.
18. Gierisch, J.M., et al., *Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: a systematic review.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(11): p. 1931-43.
19. Crandall, C.J., 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): p. 275-85.

20. Russo, J., et al., *The protective role of pregnancy in breast cancer.* Breast cancer research : BCR, 2005. **7**(3): p. 131-142.
21. Zhou, Y., et al., *Association between breastfeeding and breast cancer risk: evidence from a meta-analysis.* Breastfeed Med, 2015. **10**(3): p. 175-82.
22. Assi, H.A., et al., *Epidemiology and prognosis of breast cancer in young women.* J Thorac Dis, 2013. **5 Suppl 1**: p. S2-8.
23. Slaoui, M., et al., *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): p. e0164841.
24. Honrado, E., J. Benitez, and J. Palacios, *The molecular pathology of hereditary breast cancer: genetic testing and therapeutic implications.* Mod Pathol, 2005. **18**(10): p. 1305-20.
25. Larsen, M.J., et al., *Hereditary breast cancer: clinical, pathological and molecular characteristics.* Breast Cancer (Auckl), 2014. **8**: p. 145-55.
26. Marchina, E., et al., *BRCA1 and BRCA2 genetic test in high risk patients and families: counselling and management.* Oncol Rep, 2010. **24**(6): p. 1661-7.
27. Shroff, R.T., et al., *Rucaparib Monotherapy in Patients With Pancreatic Cancer and a Known Deleterious BRCA Mutation.* JCO Precis Oncol, 2018. **2018**.
28. Wang, G.H., et al., *BRCA1 and BRCA2 expression patterns and prognostic significance in digestive system cancers.* Hum Pathol, 2018. **71**: p. 135-144.
29. Liu, Y., N. Nguyen, and G.A. Colditz, *Links between Alcohol Consumption and Breast Cancer: A Look at the Evidence.* Women's Health, 2015. **11**(1): p. 65-77.
30. Flom, J.D., et al., *Alcohol intake over the life course and mammographic density.* Breast Cancer Res Treat, 2009. **117**(3): p. 643-51.
31. Dossus, L., et al., *Active and passive cigarette smoking and breast cancer risk: results from the EPIC cohort.* Int J Cancer, 2014. **134**(8): p. 1871-88.
32. Neuhouser, M.L., et al., *Overweight, Obesity, and Postmenopausal Invasive Breast Cancer Risk: A Secondary Analysis of the Women's Health Initiative Randomized Clinical Trials.* JAMA Oncol, 2015. **1**(5): p. 611-21.
33. McDonnell, D.P., et al., *Obesity, cholesterol metabolism, and breast cancer pathogenesis.* Cancer Res, 2014. **74**(18): p. 4976-82.
34. Baglietto, L., et al., *Associations of mammographic dense and nondense areas and body mass index with risk of breast cancer.* Am J Epidemiol, 2014. **179**(4): p. 475-83.
35. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
36. Desmedt, C., C. Sotiriou, and M.J. Piccart-Gebhart, *Development and validation of gene expression profile signatures in early-stage breast cancer.* Cancer Invest, 2009. **27**(1): p. 1-10.
37. Rakha, E.A., et al., *Triple-Negative Breast Cancer: Distinguishing between Basal and Nonbasal Subtypes.* Clinical Cancer Research, 2009. **15**(7): p. 2302-2310.
38. Curtis, C., et al., *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.* Nature, 2012. **486**(7403): p. 346-52.
39. Dawson, S.J., et al., *A new genome-driven integrated classification of breast cancer and its implications.* Embo j, 2013. **32**(5): p. 617-28.
40. Rakha, E.A., et al., *Prognostic significance of Nottingham histologic grade in invasive breast carcinoma.* J Clin Oncol, 2008. **26**(19): p. 3153-8.
41. Elston, C.W. and I.O. Ellis, *Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up.* Histopathology, 1991. **19**(5): p. 403-10.

42. Lee, A.H. and I.O. Ellis, *The Nottingham prognostic index for invasive carcinoma of the breast.* Pathol Oncol Res, 2008. **14**(2): p. 113-5.
43. Joseph, C., et al., *Mediator complex (MED) 7: a biomarker associated with good prognosis in invasive breast cancer, especially ER+ luminal subtypes.* Br J Cancer, 2018. **118**(8): p. 1142-1151.
44. El Ansari, R., et al., *The amino acid transporter SLC7A5 confers a poor prognosis in the highly proliferative breast cancer subtypes and is a key therapeutic target in luminal B tumours.* Breast Cancer Res, 2018. **20**(1): p. 21.
45. Rakha, E.A., et al., *Prognostic factors in metaplastic carcinoma of the breast: a multi-institutional study.* Br J Cancer, 2015. **112**(2): p. 283-9.
46. Mojarad, S., 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 Res, 2013. **33**(9): p. 3925-33.
47. Saadatmand, S., et al., *Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients.* Bmj, 2015. **351**: p. h4901.
48. Cianfrocca, M. and L.J. Goldstein, *Prognostic and predictive factors in early-stage breast cancer.* Oncologist, 2004. **9**(6): p. 606-16.
49. Galea, M.H., et al., *The Nottingham Prognostic Index in primary breast cancer.* Breast Cancer Res Treat, 1992. **22**(3): p. 207-19.
50. Kurshumliu, F., et al., *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): p. 243.
51. Blamey, R.W., et al., *Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999.* Eur J Cancer, 2007. **43**(10): p. 1548-55.
52. Rakha, E.A., et al., *The prognostic significance of lymphovascular invasion in invasive breast carcinoma.* Cancer, 2012. **118**(15): p. 3670-80.
53. Stacker, S.A., et al., *Lymphangiogenesis and lymphatic vessel remodelling in cancer.* Nat Rev Cancer, 2014. **14**(3): p. 159-72.
54. Al-Rohil, R.N., et al., *Intratumoral and peritumoral lymphovascular invasion detected by D2-40 immunohistochemistry correlates with metastasis in primary cutaneous Merkel cell carcinoma.* Hum Pathol, 2018. **77**: p. 98-107.
55. Payne, S.J., et al., *Predictive markers in breast cancer--the present.* Histopathology, 2008. **52**(1): p. 82-90.
56. Hammond, M.E., et al., *American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer.* J Oncol Pract, 2010. **6**(4): p. 195-7.
57. Lee, E.S., et al., *Factors associated with late recurrence after completion of 5-year adjuvant tamoxifen in estrogen receptor positive breast cancer.* BMC Cancer, 2016. **16**: p. 430.
58. Rakha, E.A., J.S. Reis-Filho, and I.O. Ellis, *Combinatorial biomarker expression in breast cancer.* Breast Cancer Res Treat, 2010. **120**(2): p. 293-308.
59. Cancello, G. and E. Montagna, *Treatment of breast cancer in young women: do we need more aggressive therapies?* J Thorac Dis, 2013. **5 Suppl 1**: p. S47-54.
60. Zhou, X.-L., et al., *The Clinical Significance of PR, ER, NF-κB, and TNF-α in Breast Cancer.* Disease Markers, 2014. **2014**: p. 7.
61. Figueroa-Magalhaes, M.C., et al., *Treatment of HER2-positive breast cancer.* Breast, 2014. **23**(2): p. 128-136.
62. Ingthorsson, S., et al., *HER2 induced EMT and tumorigenicity in breast epithelial progenitor cells is inhibited by coexpression of EGFR.* Oncogene, 2016. **35**(32): p. 4244-55.

63. Mohammed, Z.M.A., 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): p. 31.
64. Feigin, M.E. and S.K. Muthuswamy, *Polarity proteins regulate mammalian cell-cell junctions and cancer pathogenesis.* Curr Opin Cell Biol, 2009. **21**(5): p. 694-700.
65. Ahmed, S., A. Sami, and J. Xiang, *HER2-directed therapy: current treatment options for HER2-positive breast cancer.* Breast Cancer, 2015. **22**(2): p. 101-16.
66. Hamy, A.S., et al., *Lymphovascular invasion after neoadjuvant chemotherapy is strongly associated with poor prognosis in breast carcinoma.* Breast Cancer Res Treat, 2018. **169**(2): p. 295-304.
67. *Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials.* Lancet, 2005. **365**(9472): p. 1687-717.
68. Kennecke, H., et al., *Metastatic behavior of breast cancer subtypes.* J Clin Oncol, 2010. **28**(20): p. 3271-7.
69. Savci-Heijink, C.D., et al., *Retrospective analysis of metastatic behaviour of breast cancer subtypes.* Breast Cancer Res Treat, 2015. **150**(3): p. 547-57.
70. Sun, Y., et al., *Inhibitory effect of emodin on migration, invasion and metastasis of human breast cancer MDA-MB-231 cells in vitro and in vivo.* Oncol Rep, 2015. **33**(1): p. 338-46.
71. Fazilaty, H., et al., *Crosstalk between breast cancer stem cells and metastatic niche: emerging molecular metastasis pathway?* Tumour Biol, 2013. **34**(4): p. 2019-30.
72. Gonzalez de Vega, R., et al., *MMP-11 as a biomarker for metastatic breast cancer by immunohistochemical-assisted imaging mass spectrometry.* Anal Bioanal Chem, 2018.
73. Lee, C.C., et al., *Macrophage-secreted interleukin-35 regulates cancer cell plasticity to facilitate metastatic colonization.* Nat Commun, 2018. **9**(1): p. 3763.
74. Insua-Rodriguez, J., et al., *Stress signaling in breast cancer cells induces matrix components that promote chemoresistant metastasis.* EMBO Mol Med, 2018.
75. Shinde, A., et al., *Pyruvate carboxylase supports the pulmonary tropism of metastatic breast cancer.* Breast Cancer Res, 2018. **20**(1): p. 76.
76. Bray, K., et al., *Cdc42 overexpression induces hyperbranching in the developing mammary gland by enhancing cell migration.* Breast Cancer Res, 2013. **15**(5): p. R91.
77. Kandemir, N.O., et al., *Can lymphatic vascular density be used in determining metastatic spreading potential of tumor in invasive ductal carcinomas?* Pathol Oncol Res, 2012. **18**(2): p. 253-62.
78. Mohammed, R.A., 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.* J Pathol, 2011. **223**(3): p. 358-65.
79. Hishida, T., 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?* Eur J Cardiothorac Surg, 2013. **43**(4): p. 772-7.
80. Agarwal, D., 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): p. Zc86-9.
81. Rack, B., et al., *Circulating tumor cells predict survival in early average-to-high risk breast cancer patients.* J Natl Cancer Inst, 2014. **106**(5).

82. Kuo, M.C., et al., *Cancer stemness in bone marrow micrometastases of human breast cancer.* Surgery, 2018. **163**(2): p. 330-335.
83. Hartkopf, A.D., et al., *Prognostic relevance of disseminated tumour cells from the bone marrow of early stage breast cancer patients - results from a large single-centre analysis.* Eur J Cancer, 2014. **50**(15): p. 2550-9.
84. Robertson, F.M., et al., *Inflammatory breast cancer: the disease, the biology, the treatment.* CA Cancer J Clin, 2010. **60**(6): p. 351-75.
85. Manfrin, E., et al., *Comparison Between Invasive Breast Cancer With Extensive Peritumoral Vascular Invasion and Inflammatory Breast CarcinomaA Clinicopathologic Study of 161 Cases.* American Journal of Clinical Pathology, 2014. **142**(3): p. 299-306.
86. Xiao, Y., et al., *The lymphovascular embolus of inflammatory breast cancer exhibits a Notch 3 addiction.* Oncogene, 2011. **30**(3): p. 287-300.
87. Xiao, Y., et al., *The lymphovascular embolus of inflammatory breast cancer expresses a stem cell-like phenotype.* Am J Pathol, 2008. **173**(2): p. 561-74.
88. Alpaugh, M.L., et al., *Cooperative role of E-cadherin and sialyl-Lewis X/A-deficient MUC1 in the passive dissemination of tumor emboli in inflammatory breast carcinoma.* Oncogene, 2002. **21**(22): p. 3631-43.
89. Pinho, S.S., et al., *Molecular plasticity of E-cadherin and sialyl lewis x expression, in two comparative models of mammary tumorigenesis.* PLoS One, 2009. **4**(8): p. e6636.
90. Tang, S.-L., et al., *Clinicopathologic study of invasive micropapillary carcinoma of the breast.* Oncotarget, 2017. **8**(26): p. 42455-42465.
91. Gruel, N., et al., *Polarity gene alterations in pure invasive micropapillary carcinomas of the breast.* Breast Cancer Res, 2014. **16**(3): p. R46.
92. Ni, Y.B. and G.M. Tse, *Pathological criteria and practical issues in papillary lesions of the breast - a review.* Histopathology, 2016. **68**(1): p. 22-32.
93. Laser, J., et al., *Invasive lobular carcinoma of the breast: role of endothelial lymphatic marker D2-40.* Ann Clin Lab Sci, 2008. **38**(2): p. 99-104.
94. Fernandez, B., et al., *Lymph-node metastases in invasive lobular carcinoma are different from those in ductal carcinoma of the breast.* J Clin Pathol, 2011. **64**(11): p. 995-1000.
95. Bentz, J.S., N. Yassa, and F. Clayton, *Pleomorphic lobular carcinoma of the breast: clinicopathologic features of 12 cases.* Mod Pathol, 1998. **11**(9): p. 814-22.
96. Buchanan, C.L., et al., *Is pleomorphic lobular carcinoma really a distinct clinical entity?* J Surg Oncol, 2008. **98**(5): p. 314-7.
97. Sahin, S., et al., *Invasive Pleomorphic Lobular Histology Is an Adverse Prognostic Factor on Survival in Patients with Breast Cancer.* Am Surg, 2017. **83**(4): p. 359-364.
98. Ueng, S.H., T. Mezzetti, and F.A. Tavassoli, *Papillary neoplasms of the breast: a review.* Arch Pathol Lab Med, 2009. **133**(6): p. 893-907.
99. Song, Y.J., et al., *The role of lymphovascular invasion as a prognostic factor in patients with lymph node-positive operable invasive breast cancer.* J Breast Cancer, 2011. **14**(3): p. 198-203.
100. Truong, P.T., et al., *Lymphovascular invasion is associated with reduced locoregional control and survival in women with node-negative breast cancer treated with mastectomy and systemic therapy.* J Am Coll Surg, 2005. **200**(6): p. 912-21.
101. Guarnieri, A., et al., *Prediction of lymph node status by analysis of prognostic factors and possible indications for elective axillary dissection in T1 breast cancers.* Eur J Surg, 2001. **167**(4): p. 255-9.
102. Choi, S.H., S.H. Barsky, and H.R. Chang, *Clinicopathologic analysis of sentinel lymph node mapping in early breast cancer.* Breast J, 2003. **9**(3): p. 153-62.

103. Chen, H.L. and A. Ding, *Comparison of invasive micropapillary and triple negative invasive ductal carcinoma of the breast.* Breast, 2015. **24**(6): p. 723-31.
104. Ugras, S., et al., *Estrogen receptor, progesterone receptor, and HER2 status predict lymphovascular invasion and lymph node involvement.* Ann Surg Oncol, 2014. **21**(12): p. 3780-6.
105. Sabatier, R., et al., *Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization.* Mol Cancer, 2014. **13**: p. 228.
106. Holliday, D.L. and V. Speirs, *Choosing the right cell line for breast cancer research.* Breast Cancer Res, 2011. **13**(4): p. 215.
107. Sato, M., et al., *Blood Vessel Invasion Is a Strong Predictor of Postoperative Recurrence in Endometrial Cancer.* Int J Gynecol Cancer, 2018. **28**(5): p. 875-881.
108. Wu, J.K., et al., *Aberrant lymphatic endothelial progenitors in lymphatic malformation development.* PLoS One, 2015. **10**(2): p. e0117352.
109. Dadiani, M., et al., *Real-time imaging of lymphogenic metastasis in orthotopic human breast cancer.* Cancer Res, 2006. **66**(16): p. 8037-41.
110. Shields, J.D., et al., *Autologous chemotaxis as a mechanism of tumor cell homing to lymphatics via interstitial flow and autocrine CCR7 signaling.* Cancer Cell, 2007. **11**(6): p. 526-38.
111. Wong, S.Y. and R.O. Hynes, *Lymphatic or hematogenous dissemination: how does a metastatic tumor cell decide?* Cell Cycle, 2006. **5**(8): p. 812-7.
112. Steinskog, E.S., et al., *Impaired lymphatic function accelerates cancer growth.* Oncotarget, 2016. **7**(29): p. 45789-45802.
113. Ran, S., et al., *Lymphangiogenesis and lymphatic metastasis in breast cancer.* Pathophysiology, 2010. **17**(4): p. 229-51.
114. Shi, S.J., et al., *LncRNA-ATB promotes trastuzumab resistance and invasion-metastasis cascade in breast cancer.* Oncotarget, 2015. **6**(13): p. 11652-63.
115. Bhat-Nakshatri, P., et al., *SLUG/SNAI2 and tumor necrosis factor generate breast cells with CD44+/CD24- phenotype.* BMC Cancer, 2010. **10**: p. 411.
116. Balkwill, F.R., M. Capasso, and T. Hagemann, *The tumor microenvironment at a glance.* J Cell Sci, 2012. **125**(Pt 23): p. 5591-6.
117. Yousef, E.M., et al., *MMP-9 expression varies according to molecular subtypes of breast cancer.* BMC Cancer, 2014. **14**: p. 609.
118. Wu, Q.W., et al., *Expression and clinical significance of matrix metalloproteinase-9 in lymphatic invasiveness and metastasis of breast cancer.* PLoS One, 2014. **9**(5): p. e97804.
119. Fidalgo, F., et al., *Lymphovascular invasion and histologic grade are associated with specific genomic profiles in invasive carcinomas of the breast.* Tumour Biol, 2015. **36**(3): p. 1835-48.
120. Hotary, K., et al., *A cancer cell metalloprotease triad regulates the basement membrane transmigration program.* Genes Dev, 2006. **20**(19): p. 2673-86.
121. Perentes, J.Y., et al., *Cancer cell-associated MT1-MMP promotes blood vessel invasion and distant metastasis in triple-negative mammary tumors.* Cancer Res, 2011. **71**(13): p. 4527-38.
122. Pula, B., et al., *Impact of different tumour stroma assessment methods regarding podoplanin expression on clinical outcome in patients with invasive ductal breast carcinoma.* Anticancer Res, 2013. **33**(4): p. 1447-55.
123. Pula, B., et al., *Podoplanin expression in cancer-associated fibroblasts correlates with VEGF-C expression in cancer cells of invasive ductal breast carcinoma.* Neoplasma, 2013. **60**(5): p. 516-24.
124. Issa, A., et al., *Vascular endothelial growth factor-C and C-C chemokine receptor 7 in tumor cell-lymphatic cross-talk promote invasive phenotype.* Cancer Res, 2009. **69**(1): p. 349-57.

125. Wyckoff, J., et al., *A paracrine loop between tumor cells and macrophages is required for tumor cell migration in mammary tumors.* Cancer Res, 2004. **64**(19): p. 7022-9.
126. Cheang, M.C., et al., *Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype.* Clin Cancer Res, 2008. **14**(5): p. 1368-76.
127. Huo, Z., et al., *Metastases of basal-like breast invasive ductal carcinoma to the endometrium: A case report and review of the literature.* Thorac Cancer, 2015. **6**(4): p. 548-52.
128. Joimel, U., et al., *Stimulation of angiogenesis resulting from cooperation between macrophages and MDA-MB-231 breast cancer cells: proposed molecular mechanism and effect of tetrathiomolybdate.* BMC Cancer, 2010. **10**: p. 375.
129. Cursiefen, C., et al., *VEGF-A stimulates lymphangiogenesis and hemangiogenesis in inflammatory neovascularization via macrophage recruitment.* J Clin Invest, 2004. **113**(7): p. 1040-50.
130. Robinson, B.D., et al., *Tumor microenvironment of metastasis in human breast carcinoma: a potential prognostic marker linked to hematogenous dissemination.* Clin Cancer Res, 2009. **15**(7): p. 2433-41.
131. Cheung, K.J. and A.J. Ewald, *Illuminating breast cancer invasion: diverse roles for cell-cell interactions.* Curr Opin Cell Biol, 2014. **30**: p. 99-111.
132. Alizadeh, D., et al., *Doxorubicin eliminates myeloid-derived suppressor cells and enhances the efficacy of adoptive T-cell transfer in breast cancer.* Cancer Res, 2014. **74**(1): p. 104-18.
133. Maniotis, A.J., et al., *Vascular channel formation by human melanoma cells in vivo and in vitro: vasculogenic mimicry.* Am J Pathol, 1999. **155**(3): p. 739-52.
134. Kirschmann, D.A., et al., *Molecular pathways: vasculogenic mimicry in tumor cells: diagnostic and therapeutic implications.* Clin Cancer Res, 2012. **18**(10): p. 2726-32.
135. Shirakawa, K., et al., *Absence of endothelial cells, central necrosis, and fibrosis are associated with aggressive inflammatory breast cancer.* Cancer Res, 2001. **61**(2): p. 445-51.
136. Shirakawa, K., et al., *Hemodynamics in vasculogenic mimicry and angiogenesis of inflammatory breast cancer xenograft.* Cancer Res, 2002. **62**(2): p. 560-6.
137. Liu, T., et al., *OCT4 expression and vasculogenic mimicry formation positively correlate with poor prognosis in human breast cancer.* Int J Mol Sci, 2014. **15**(11): p. 19634-49.
138. Paulis, Y.W., et al., *Signalling pathways in vasculogenic mimicry.* Biochim Biophys Acta, 2010. **1806**(1): p. 18-28.
139. Qiao, L., et al., *Advanced research on vasculogenic mimicry in cancer.* J Cell Mol Med, 2015. **19**(2): p. 315-26.
140. Rakha, E.A., et al., *Expression profiling technology: its contribution to our understanding of breast cancer.* Histopathology, 2008. **52**(1): p. 67-81.
141. Mannelqvist, M., et al., *Gene expression patterns related to vascular invasion and aggressive features in endometrial cancer.* Am J Pathol, 2011. **178**(2): p. 861-71.
142. Mannelqvist, M., et al., *An 18-gene signature for vascular invasion is associated with aggressive features and reduced survival in breast cancer.* PLoS One, 2014. **9**(6): p. e98787.
143. Minguez, B., et al., *Gene-expression signature of vascular invasion in hepatocellular carcinoma.* J Hepatol, 2011. **55**(6): p. 1325-31.
144. Iorio, M.V., et al., *MicroRNA gene expression deregulation in human breast cancer.* Cancer Res, 2005. **65**(16): p. 7065-70.
145. Fiannaca, A., et al., *Analysis of miRNA expression profiles in breast cancer using biclustering.* BMC Bioinformatics, 2015. **16 Suppl 4**: p. S7.

146. Ma, L., J. Teruya-Feldstein, and R.A. Weinberg, *Tumour invasion and metastasis initiated by microRNA-10b in breast cancer.* Nature, 2007. **449**(7163): p. 682-8.
147. Mohammed, R.A., et al., *Improved methods of detection of lymphovascular invasion demonstrate that it is the predominant method of vascular invasion in breast cancer and has important clinical consequences.* Am J Surg Pathol, 2007. **31**(12): p. 1825-33.
148. Zhao, Y.C., et al., *Tumor-derived VEGF-C, but not VEGF-D, promotes sentinel lymph node lymphangiogenesis prior to metastasis in breast cancer patients.* Med Oncol, 2012. **29**(4): p. 2594-600.
149. Zhang, X.H., et al., *Coexpression of VEGF-C and COX-2 and its association with lymphangiogenesis in human breast cancer.* BMC Cancer, 2008. **8**: p. 4.
150. Chiavenna, S.M., J.P. Jaworski, and A. Vendrell *State of the art in anti-cancer mAbs*. Journal of biomedical science, 2017. **24**, 15 DOI: 10.1186/s12929-016-0311-y.
151. Storr, S.J., et al., *Calpastatin is associated with lymphovascular invasion in breast cancer.* Breast, 2011. **20**(5): p. 413-8.
152. Storr, S.J., et al., *The calpain system and cancer.* Nat Rev Cancer, 2011. **11**(5): p. 364-74.
153. Jerjees, D.A., et al., *ERK1/2 is related to oestrogen receptor and predicts outcome in hormone-treated breast cancer.* Breast Cancer Res Treat, 2014. **147**(1): p. 25-37.
154. Marinho, V.F., et al., *Lymph vascular invasion in invasive mammary carcinomas identified by the endothelial lymphatic marker D2-40 is associated with other indicators of poor prognosis.* BMC Cancer, 2008. **8**: p. 64.
155. Alshareeda, A.T., et al., *Clinicopathological significance of KU70/KU80, a key DNA damage repair protein in breast cancer.* Breast Cancer Res Treat, 2013. **139**(2): p. 301-10.
156. Alshareeda, A.T., et al., *SUMOylation proteins in breast cancer.* Breast Cancer Res Treat, 2014. **144**(3): p. 519-30.
157. Green, A.R., et al., *Biological and clinical significance of PARP1 protein expression in breast cancer.* Breast Cancer Res Treat, 2015. **149**(2): p. 353-62.
158. Al-Kaabi, M.M., et al., *Checkpoint kinase1 (CHK1) is an important biomarker in breast cancer having a role in chemotherapy response.* Br J Cancer, 2015. **112**(5): p. 901-11.
159. Nguyen-Ngoc, K.V., et al., *ECM microenvironment regulates collective migration and local dissemination in normal and malignant mammary epithelium.* Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2595-604.
160. Abd El-Rehim, D.M., 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): p. 1532-42.
161. Mahmoud, S.M., et al., *An evaluation of the clinical significance of FOXP3+ infiltrating cells in human breast cancer.* Breast Cancer Res Treat, 2011. **127**(1): p. 99-108.
162. Mahmoud, S.M., et al., *Tumour-infiltrating macrophages and clinical outcome in breast cancer.* J Clin Pathol, 2012. **65**(2): p. 159-63.
163. Rakha, E.A., et al., *Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes.* Clin Cancer Res, 2009. **15**(7): p. 2302-10.
164. Pereira, B., et al., *Erratum: The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes.* Nat Commun, 2016. **7**: p. 11908.
165. Towbin, H., T. Staehelin, and J. Gordon, *Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications.* Proc Natl Acad Sci U S A, 1979. **76**(9): p. 4350-4.

166. Coons, A.H., H.J. Creech, and R.N. Jones, *Immunological Properties of an Antibody Containing a Fluorescent Group.* Proceedings of the Society for Experimental Biology and Medicine, 1941. **47**(2): p. 200-202.
167. Rabban, J.T., et al., *Immunophenotypic overlap between adenoid cystic carcinoma and collagenous spherulosis of the breast: potential diagnostic pitfalls using myoepithelial markers.* Mod Pathol, 2006. **19**(10): p. 1351-7.
168. Nakai, T., et al., *The unique luminal staining pattern of cytokeratin 5/6 in adenoid cystic carcinoma of the breast may aid in differentiating it from its mimickers.* Virchows Arch, 2016. **469**(2): p. 213-22.
169. Reed, A.E.M., et al., *Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics.* Breast Cancer Research, 2015. **17**(1): p. 12.
170. Lee, H.C., et al., *Expression of Immunohistochemical Markers before and after Neoadjuvant Chemotherapy in Breast Carcinoma, and Their Use as Predictors of Response.* J Breast Cancer, 2013. **16**(4): p. 395-403.
171. Pereira, B., et al., *The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes.* Nat Commun, 2016. **7**: p. 11479.
172. Compton, S.J. and C.G. Jones, *Mechanism of dye response and interference in the Bradford protein assay.* Anal Biochem, 1985. **151**(2): p. 369-74.
173. Faktor, J., et al., *Comparison of targeted proteomics approaches for detecting and quantifying proteins derived from human cancer tissues.* Proteomics, 2017. **17**(5).
174. Xie, W., et al., *A practical strategy for the characterization of ponicidin metabolites in vivo and in vitro by UHPLC-Q-TOF-MS based on nontargeted SWATH data acquisition.* J Pharm Biomed Anal, 2017. **145**: p. 865-878.
175. Turashvili, G. and E. Brogi, *Tumor Heterogeneity in Breast Cancer.* Front Med (Lausanne), 2017. **4**: p. 227.
176. Wang, Y., et al., *Integrated bioinformatics analysis reveals key candidate genes and pathways in breast cancer.* Mol Med Rep, 2018. **17**(6): p. 8091-8100.
177. Gao, H., M. Yang, and X. Zhang, *Investigating a multigene prognostic assay based on significant pathways for Luminal A breast cancer through gene expression profile analysis.* Oncol Lett, 2018. **15**(4): p. 5027-5033.
178. Klahan, S., et al., *Identification of genes and pathways related to lymphovascular invasion in breast cancer patients: A bioinformatics analysis of gene expression profiles.* Tumour Biol, 2017. **39**(6): p. 1010428317705573.
179. Wu, H.-T., I. Hajirasouliha, and B.J. Raphael, *Detecting independent and recurrent copy number aberrations using interval graphs.* Bioinformatics, 2014. **30**(12): p. i195-i203.
180. Li, W., A. Lee, and P.K. Gregersen, *Copy-number-variation and copy-number-alteration region detection by cumulative plots.* BMC Bioinformatics, 2009. **10 Suppl 1**: p. S67.
181. Mi, H., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.* Nucleic Acids Research, 2017. **45**(D1): p. D183-D189.
182. Gaudet, P., et al., *Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium.* Brief Bioinform, 2011. **12**(5): p. 449-62.
183. Fabregat, A., et al., *The Reactome Pathway Knowledgebase.* Nucleic Acids Res, 2018. **46**(D1): p. D649-d655.
184. Chatr-Aryamontri, A., et al., *The BioGRID interaction database: 2017 update.* Nucleic Acids Res, 2017. **45**(D1): p. D369-d379.
185. The UniProt Consortium, *UniProt: the universal protein knowledgebase.* Nucleic Acids Research, 2017. **45**(D1): p. D158-D169.
186. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.* Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20.

187. Belinky, F., et al., *Non-redundant compendium of human ncRNA genes in GeneCards.* Bioinformatics, 2013. **29**(2): p. 255-61.
188. Jezequel, P., et al., *bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer.* Breast Cancer Res Treat, 2012. **131**(3): p. 765-75.
189. Jezequel, P., et al., *bc-GenExMiner 3.0: new mining module computes breast cancer gene expression correlation analyses.* Database (Oxford), 2013. **2013**: p. bas060.
190. Kurozumi, S., et al., *Clinicopathological and prognostic significance of Ras association and pleckstrin homology domains 1 (RAPH1) in breast cancer.* Breast Cancer Res Treat, 2018.
191. Sonbul, S.N., et al., *Saccharomyces cerevisiae-like 1 (SEC14L1) is a prognostic factor in breast cancer associated with lymphovascular invasion.* Mod Pathol, 2018.
192. Kurozumi, S., et al., *Clinical and biological roles of Kelch-like family member 7 in breast cancer: a marker of poor prognosis.* Breast Cancer Res Treat, 2018. **170**(3): p. 525-533.
193. Aleskandarany, M.A., et al., *Rho-GTPase activating-protein 18: a biomarker associated with good prognosis in invasive breast cancer.* Br J Cancer, 2017. **117**(8): p. 1176-1184.
194. Zhou, C., et al., *Integrated Analysis of Copy Number Variations and Gene Expression Profiling in Hepatocellular carcinoma.* Scientific Reports, 2017. **7**(1): p. 10570.
195. Gatza, M.L., et al., *An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer.* Nature Genetics, 2014. **46**: p. 1051.
196. Ohshima, K., et al., *Integrated analysis of gene expression and copy number identified potential cancer driver genes with amplification-dependent overexpression in 1,454 solid tumors.* Scientific Reports, 2017. **7**(1): p. 641.
197. Takekoshi, T., et al., *CCR7-expressing B16 melanoma cells downregulate interferon-gamma-mediated inflammation and increase lymphangiogenesis in the tumor microenvironment.* Oncogenesis, 2012. **1**: p. e9.
198. Emmett, M.S., et al., *CCR7 mediates directed growth of melanomas towards lymphatics.* Microcirculation, 2011. **18**(3): p. 172-82.
199. Sonbul, S.N., et al., *Chemokine (C-C motif) receptor 7 (CCR7) associates with the tumour immune microenvironment but not progression in invasive breast carcinoma.* J Pathol Clin Res, 2017. **3**(2): p. 105-114.
200. Fukamachi, M., et al., *Multiple coagulation factor deficiency protein 2 as a crucial component in metastasis of human oral cancer.* Exp Cell Res, 2018. **368**(1): p. 119-125.
201. Salo-Mullen, E.E., et al., *Contiguous gene deletion of chromosome 2p16.3-p21 as a cause of Lynch syndrome.* Fam Cancer, 2018. **17**(1): p. 71-77.
202. Gashaw, I., et al., *Novel germ cell markers characterize testicular seminoma and fetal testis.* Mol Hum Reprod, 2007. **13**(10): p. 721-7.
203. Huang, X., et al., *Chaperonin containing TCP1, subunit 8 (CCT8) is upregulated in hepatocellular carcinoma and promotes HCC proliferation.* Apmis, 2014. **122**(11): p. 1070-9.
204. Qiu, X., et al., *Overexpression of CCT8 and its significance for tumor cell proliferation, migration and invasion in glioma.* Pathol Res Pract, 2015. **211**(10): p. 717-25.
205. Yang, X., et al., *Chaperonin-containing Tcomplex protein 1 subunit 8 promotes cell migration and invasion in human esophageal squamous cell carcinoma by regulating alpha-actin and beta-tubulin expression.* Int J Oncol, 2018. **52**(6): p. 2021-2030.
206. Kalikin, L.M., et al., *Genomic characterization of human SEC14L1 splice variants within a 17q25 candidate tumor suppressor gene region and*

identification of an unrelated embedded expressed sequence tag. Mamm Genome, 2001. **12**(12): p. 925-9.
207. Sinclair, C.S., et al., *The 17q23 amplicon and breast cancer.* Breast Cancer Res Treat, 2003. **78**(3): p. 313-22.
208. Agell, L., et al., *A 12-gene expression signature is associated with aggressive histological in prostate cancer: SEC14L1 and TCEB1 genes are potential markers of progression.* Am J Pathol, 2012. **181**(5): p. 1585-94.
209. Zhang, S., et al., *Integrative transcriptome and proteome study to identify the signaling network regulated by POPX2 phosphatase.* J Proteome Res, 2013. **12**(6): p. 2525-36.
210. Obayashi, S., et al., *Stathmin1 expression is associated with aggressive phenotypes and cancer stem cell marker expression in breast cancer patients.* Int J Oncol, 2017. **51**(3): p. 781-790.
211. Szabo, L., et al., *Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development.* Genome Biol, 2015. **16**: p. 126.
212. Mackay, A., et al., *cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells.* Oncogene, 2003. **22**(17): p. 2680-8.
213. Xiong, G., et al., *Prolyl-4-hydroxylase alpha subunit 2 promotes breast cancer progression and metastasis by regulating collagen deposition.* BMC Cancer, 2014. **14**: p. 1.
214. Liu, M., et al., *Transcriptional profiles of different states of cancer stem cells in triple-negative breast cancer.* Mol Cancer, 2018. **17**(1): p. 65.
215. Moon, H.G., et al., *Prognostic and functional importance of the engraftment-associated genes in the patient-derived xenograft models of triple-negative breast cancers.* Breast Cancer Res Treat, 2015. **154**(1): p. 13-22.
216. Fagerberg, L., et al., *Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics.* Mol Cell Proteomics, 2014. **13**(2): p. 397-406.
217. Wu, W., et al., *Prognostic significance of CXCL12, CXCR4, and CXCR7 in patients with breast cancer.* Int J Clin Exp Pathol, 2015. **8**(10): p. 13217-24.
218. Sun, Y., et al., *CXCL12-CXCR4 axis promotes the natural selection of breast cancer cell metastasis.* Tumour Biol, 2014. **35**(8): p. 7765-73.
219. Lv, Z.D., et al., *CXCL12 chemokine expression suppresses human breast cancer growth and metastasis in vitro and in vivo.* Int J Clin Exp Pathol, 2014. **7**(10): p. 6671-8.
220. Fang, X., et al., *Lysophospholipid growth factors in the initiation, progression, metastases, and management of ovarian cancer.* Ann N Y Acad Sci, 2000. **905**: p. 188-208.
221. Horak, C.E., et al., *Nm23-H1 Suppresses Tumor Cell Motility by Down-regulating the Lysophosphatidic Acid Receptor EDG2.* Cancer Research, 2007. **67**(15): p. 7238-7246.
222. Hirane, M., et al., *Inhibitory effects of LPA1 on cell motile activities stimulated by hydrogen peroxide and 2,3-dimethoxy-1,4-naphthoquinone in fibroblast 3T3 cells.* Biochem Biophys Res Commun, 2013. **441**(1): p. 47-52.
223. Takahashi, K., et al., *Involvement of LPA signaling via LPA receptor-2 in the promotion of malignant properties in osteosarcoma cells.* Exp Cell Res, 2018. **369**(2): p. 316-324.
224. Murph, M.M., et al., *Sharpening the edges of understanding the structure/function of the LPA1 receptor: expression in cancer and mechanisms of regulation.* Biochim Biophys Acta, 2008. **1781**(9): p. 547-57.
225. Talmadge, J.E. and I.J. Fidler, *AACR centennial series: the biology of cancer metastasis: historical perspective.* Cancer Res, 2010. **70**(14): p. 5649-69.

226. Aleskandarany, M.A., et al., *Molecular Mechanisms Underlying Lymphovascular Invasion in Invasive Breast Cancer.* Pathobiology, 2015. **82**(3-4): p. 113-23.
227. Rippe, V., et al., *Identification of a gene rearranged by 2p21 aberrations in thyroid adenomas.* Oncogene, 0000. **22**(38): p. 6111-6114.
228. Han, S.S., et al., *The chromosome 2p21 region harbors a complex genetic architecture for association with risk for renal cell carcinoma.* Hum Mol Genet, 2012. **21**(5): p. 1190-200.
229. Liu, H., et al., *Multiple coagulation factor deficiency protein 2 contains the ability to support stem cell self-renewal.* Faseb j, 2013. **27**(8): p. 3298-305.
230. Nishio, M., et al., *Structural basis for the cooperative interplay between the two causative gene products of combined factor V and factor VIII deficiency.* Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4034-9.
231. Zhang, B., et al., *Bleeding due to disruption of a cargo-specific ER-to-Golgi transport complex.* Nat Genet, 2003. **34**(2): p. 220-5.
232. Joachimiak, L.A., et al., *The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT.* Cell, 2014. **159**(5): p. 1042-1055.
233. Bouhouche, A., et al., *Mutation in the epsilon subunit of the cytosolic chaperonin-containing t-complex peptide-1 (Cct5) gene causes autosomal recessive mutilating sensory neuropathy with spastic paraplegia.* J Med Genet, 2006. **43**(5): p. 441-3.
234. Kasembeli, M., et al., *Modulation of STAT3 folding and function by TRiC/CCT chaperonin.* PLoS Biol, 2014. **12**(4): p. e1001844.
235. Trinidad, A.G., et al., *Interaction of p53 with the CCT complex promotes protein folding and wild-type p53 activity.* Mol Cell, 2013. **50**(6): p. 805-17.
236. Walerych, D., et al., *Hsp70 molecular chaperones are required to support p53 tumor suppressor activity under stress conditions.* Oncogene, 2009. **28**(48): p. 4284-94.
237. Feldman, D.E., et al., *Tumorigenic mutations in VHL disrupt folding in vivo by interfering with chaperonin binding.* Mol Cell, 2003. **12**(5): p. 1213-24.
238. Yin, H., et al., *The role of the Chaperonin containing t-complex polypeptide 1, subunit 8 (CCT8) in B-cell non-Hodgkin's lymphoma.* Leuk Res, 2016. **45**: p. 59-67.
239. Wei, P.L., et al., *Glucose-regulated protein 94 mediates metastasis by CCT8 and the JNK pathway in hepatocellular carcinoma.* Tumour Biol, 2016. **37**(6): p. 8219-27.
240. Özdemir, A., et al., *Identification of the T-complex protein as a binding partner for newly synthesized cytoplasmic dynein intermediate chain 2.* Biochemical and Biophysical Research Communications, 2016. **469**(1): p. 126-131.
241. Subramanian, A., et al., *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences, 2005. **102**(43): p. 15545-15550.
242. Mootha, V.K., et al., *PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nature Genetics, 2003. **34**: p. 267.
243. Ren, Y.X., et al., *Mouse mesenchymal stem cells expressing PAX-FKHR form alveolar rhabdomyosarcomas by cooperating with secondary mutations.* Cancer Res, 2008. **68**(16): p. 6587-97.
244. Graessmann, M., et al., *Chemotherapy resistance of mouse WAP-SVT/t breast cancer cells is mediated by osteopontin, inhibiting apoptosis downstream of caspase-3.* Oncogene, 2007. **26**(20): p. 2840-50.
245. Zhang, Y., et al., *Long intergenic non-coding RNA expression signature in human breast cancer.* Sci Rep, 2016. **6**: p. 37821.
246. Gao, R., et al., *Punctuated copy number evolution and clonal stasis in triple-negative breast cancer.* Nat Genet, 2016. **48**(10): p. 1119-30.

247. Nyfeler, B., et al., *Cargo selectivity of the ERGIC-53/MCFD2 transport receptor complex.* Traffic, 2006. **7**(11): p. 1473-81.
248. Torun, D., et al., *Two new mutations at ERGIC-53 gene in a Turkish family.* Clin Appl Thromb Hemost, 2011. **17**(3): p. 248-50.
249. Hamza, A., et al., *Unveiling the unfolding pathway of F5F8D disorder-associated D81H/V100D mutant of MCFD2 via multiple molecular dynamics simulations.* J Biomol Struct Dyn, 2012. **29**(4): p. 699-714.
250. Zhang, B., et al., *Genotype-phenotype correlation in combined deficiency of factor V and factor VIII.* Blood, 2008. **111**(12): p. 5592-600.
251. Ivaskevicius, V., et al., *The first case of combined coagulation factor V and coagulation factor VIII deficiency in Poland due to a novel p.Tyr135Asn missense mutation in the MCFD2 gene.* Blood Coagul Fibrinolysis, 2008. **19**(6): p. 531-534.
252. Zhang, B., et al., *Combined deficiency of factor V and factor VIII is due to mutations in either LMAN1 or MCFD2.* Blood, 2006. **107**(5): p. 1903-7.
253. Kajiho, H., Y. Kajiho, and G. Scita, *Harnessing membrane trafficking to promote cancer spreading and invasion: The case of RAB2A.* Small GTPases, 2017: p. 1-6.
254. Roeckel, N., et al., *High frequency of LMAN1 abnormalities in colorectal tumors with microsatellite instability.* Cancer Res, 2009. **69**(1): p. 292-9.
255. Agarwal, S., A. Singh, and P.K. Bagga, *Immunohistochemical evaluation of lymphovascular invasion in carcinoma breast with CD34 and D2-40 and its correlation with other prognostic markers.* Indian J Pathol Microbiol, 2018. **61**(1): p. 39-44.
256. Zhang, S., et al., *The relationship of lymphatic vessel density, lymphovascular invasion, and lymph node metastasis in breast cancer: a systematic review and meta-analysis.* Oncotarget, 2016.
257. Fidalgo, F., et al., *Lymphovascular invasion and histologic grade are associated with specific genomic profiles in invasive carcinomas of the breast.* Tumour Biology, 2015. **36**(3): p. 1835-1848.
258. Toda, H., et al., *Stem cell-derived neural stem/progenitor cell supporting factor is an autocrine/paracrine survival factor for adult neural stem/progenitor cells.* J Biol Chem, 2003. **278**(37): p. 35491-500.
259. Kraby, M.R., et al., *The prognostic value of androgen receptors in breast cancer subtypes.* Breast Cancer Res Treat, 2018.
260. Tanaka, K., et al., *The relationship between the expression of FOXA1 and GATA3 and the efficacy of neoadjuvant endocrine therapy.* Breast Cancer, 2016.
261. Kong, S.Y., et al., *The ELK3-GATA3 axis orchestrates invasion and metastasis of breast cancer cells in vitro and in vivo.* Oncotarget, 2016.
262. Wiesner, F.G., et al., *Ki-67 as a prognostic molecular marker in routine clinical use in breast cancer patients.* Breast, 2009. **18**(2): p. 135-41.
263. Li, L.T., et al., *Ki67 is a promising molecular target in the diagnosis of cancer (review).* Mol Med Rep, 2015. **11**(3): p. 1566-72.
264. Inwald, E.C., et al., *Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry.* Breast Cancer Res Treat, 2013. **139**(2): p. 539-52.
265. Chao, Y.L., C.R. Shepard, and A. Wells, *Breast carcinoma cells re-express E-cadherin during mesenchymal to epithelial reverting transition.* Mol Cancer, 2010. **9**: p. 179.
266. Hesse, K., et al., *Characterisation of Prognosis and Invasion of Cutaneous Squamous Cell Carcinoma by Podoplanin and E-Cadherin Expression.* Dermatology, 2016.
267. Tang, X., et al., *Cystine addiction of triple-negative breast cancer associated with EMT augmented death signaling.* Oncogene, 2016.
268. Oh, E., et al., *Overexpression of angiotensin II type 1 receptor in breast cancer cells induces epithelial-mesenchymal transition and promotes tumor*

growth and angiogenesis. Biochim Biophys Acta, 2016. **1863**(6 Pt A): p. 1071-81.
269. Gruber, R., M. Levitt, and A. Horovitz, *Sequential allosteric mechanism of ATP hydrolysis by the CCT/TRiC chaperone is revealed through Arrhenius analysis.* Proc Natl Acad Sci U S A, 2017. **114**(20): p. 5189-5194.
270. Freund, A., et al., *Proteostatic control of telomerase function through TRiC-mediated folding of TCAB1.* Cell, 2014. **159**(6): p. 1389-403.
271. Seo, S., et al., *BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly.* Proc Natl Acad Sci U S A, 2010. **107**(4): p. 1488-93.
272. Echbarthi, M., J. Vallin, and J. Grantham, *Interactions between monomeric CCTdelta and p150(Glued): A novel function for CCTdelta at the cell periphery distinct from the protein folding activity of the molecular chaperone CCT.* Exp Cell Res, 2018. **370**(1): p. 137-149.
273. Ozdemir, A., et al., *Identification of the T-complex protein as a binding partner for newly synthesized cytoplasmic dynein intermediate chain 2.* Biochem Biophys Res Commun, 2016. **469**(1): p. 126-131.
274. Ahmed, M.A., et al., *A CD44(-)/CD24(+) phenotype is a poor prognostic marker in early invasive breast cancer.* Breast Cancer Res Treat, 2012. **133**(3): p. 979-95.
275. Weichert, W., et al., *Cytoplasmic CD24 expression in colorectal cancer independently correlates with shortened patient survival.* Clin Cancer Res, 2005. **11**(18): p. 6574-81.
276. Song, S.P., et al., *NDRG2 down-regulation and CD24 up-regulation promote tumor aggravation and poor survival in patients with gallbladder carcinoma.* Med Oncol, 2012. **29**(3): p. 1879-85.
277. Yang, X.R., et al., *CD24 is a novel predictor for poor prognosis of hepatocellular carcinoma after surgery.* Clin Cancer Res, 2009. **15**(17): p. 5518-27.
278. Lee, S.H., et al., *CD24 and S100A4 expression in resectable pancreatic cancers with earlier disease recurrence and poor survival.* Pancreas, 2014. **43**(3): p. 380-8.
279. Deng, J., et al., *CD24 expression as a marker for predicting clinical outcome in human gliomas.* J Biomed Biotechnol, 2012. **2012**: p. 517172.
280. Ma, Z.L., et al., *Knockdown of CD24 inhibits proliferation, invasion and sensitizes breast cancer MCF-7 cells to tamoxifen in vitro.* Eur Rev Med Pharmacol Sci, 2015. **19**(13): p. 2394-9.
281. Ma, F., et al., *Enriched CD44(+)/CD24(-) population drives the aggressive phenotypes presented in triple-negative breast cancer (TNBC).* Cancer Lett, 2014. **353**(2): p. 153-9.
282. Kean, M.J., et al., *VAMP3, syntaxin-13 and SNAP23 are involved in secretion of matrix metalloproteinases, degradation of the extracellular matrix and cell invasion.* J Cell Sci, 2009. **122**(Pt 22): p. 4089-98.
283. Williams, K.C. and M.G. Coppolino, *SNARE-dependent interaction of Src, EGFR and beta1 integrin regulates invadopodia formation and tumor cell invasion.* J Cell Sci, 2014. **127**(Pt 8): p. 1712-25.
284. Hepp, R., et al., *Phosphorylation of SNAP-23 regulates exocytosis from mast cells.* J Biol Chem, 2005. **280**(8): p. 6610-20.
285. Reis, C.R., et al., *Crosstalk between Akt/GSK3beta signaling and dynamin-1 regulates clathrin-mediated endocytosis.* Embo j, 2015. **34**(16): p. 2132-46.
286. Ganguly, A., et al., *A dynamic formin-dependent deep F-actin network in axons.* J Cell Biol, 2015. **210**(3): p. 401-17.
287. Petersen, J.D., S. Kaech, and G. Banker, *Selective microtubule-based transport of dendritic membrane proteins arises in concert with axon specification.* J Neurosci, 2014. **34**(12): p. 4135-47.

288. Johnson, J.L., et al., *Munc13-4 Is a Rab11-binding Protein That Regulates Rab11-positive Vesicle Trafficking and Docking at the Plasma Membrane.* J Biol Chem, 2016. **291**(7): p. 3423-38.
289. Fath, S., et al., *Structure and organization of coat proteins in the COPII cage.* Cell, 2007. **129**(7): p. 1325-36.
290. Stornaiuolo, M., et al., *KDEL and KKXX retrieval signals appended to the same reporter protein determine different trafficking between endoplasmic reticulum, intermediate compartment, and Golgi complex.* Mol Biol Cell, 2003. **14**(3): p. 889-902.
291. Strauss, J.A., et al., *Immunofluorescence microscopy of SNAP23 in human skeletal muscle reveals colocalization with plasma membrane, lipid droplets, and mitochondria.* Physiol Rep, 2016. **4**(1).
292. Sun, Q., et al., *SNAP23 promotes the malignant process of ovarian cancer.* Journal of Ovarian Research, 2016. **9**(1): p. 80.
293. Zhu, B., et al., *SNAP23 suppresses cervical cancer progression via modulating the cell cycle.* Gene, 2018. **673**: p. 217-224.
294. Williams, K.C. and M.G. Coppolino, *SNARE-dependent interaction of Src, EGFR and β₁ integrin regulates invadopodia formation and tumor cell invasion.* Journal of Cell Science, 2014. **127**(8): p. 1712-1725.
295. Williams, K.C., R.E. McNeilly, and M.G. Coppolino, *SNAP23, Syntaxin4, and vesicle-associated membrane protein 7 (VAMP7) mediate trafficking of membrane type 1-matrix metalloproteinase (MT1-MMP) during invadopodium formation and tumor cell invasion.* Mol Biol Cell, 2014. **25**(13): p. 2061-70.
296. Brasher, M.I., et al., *Interaction of Munc18c and syntaxin4 facilitates invadopodium formation and extracellular matrix invasion of tumor cells.* Journal of Biological Chemistry, 2017. **292**(39): p. 16199-16210.
297. Yu, H., et al., *Comparative studies of Munc18c and Munc18-1 reveal conserved and divergent mechanisms of Sec1/Munc18 proteins.* Proceedings of the National Academy of Sciences, 2013. **110**(35): p. E3271-E3280.
298. Zhu, D., et al., *Munc18c mediates exocytosis of pre-docked and newcomer insulin granules underlying biphasic glucose stimulated insulin secretion in human pancreatic beta-cells.* Molecular Metabolism, 2015. **4**(5): p. 418-426.
299. Khazen, R., et al., *Melanoma cell lysosome secretory burst neutralizes the CTL-mediated cytotoxicity at the lytic synapse.* Nature Communications, 2016. **7**: p. 10823.
300. Camp, R.L., M. Dolled-Filhart, and D.L. Rimm, *X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization.* Clin Cancer Res, 2004. **10**(21): p. 7252-9.
301. Mousley, C.J., et al., *The Sec14-superfamily and the regulatory interface between phospholipid metabolism and membrane trafficking.* Biochimica et biophysica acta, 2007. **1771**(6): p. 727-736.
302. Mousley, C.J., J.M. Davison, and V.A. Bankaitis, *Sec14 like PITPs couple lipid metabolism with phosphoinositide synthesis to regulate Golgi functionality.* Subcell Biochem, 2012. **59**: p. 271-87.
303. Ribeiro, F.M., et al., *SEC14-like protein 1 interacts with cholinergic transporters.* Neurochem Int, 2007. **50**(2): p. 356-64.
304. Li, M.-T., et al., *Negative Regulation of RIG-I-Mediated Innate Antiviral Signaling by SEC14L1.* Journal of Virology, 2013. **87**(18): p. 10037-10046.
305. Burdelski, C., et al., *Saccharomyces cerevisiae-like 1 overexpression is frequent in prostate cancer and has markedly different effects in Ets-related gene fusion-positive and fusion-negative cancers.* Hum Pathol, 2015. **46**(4): p. 514-23.
306. Carver, B.S., et al., *Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate.* Nature Genetics, 2009. **41**: p. 619.

307. Puc, J., et al., *Lack of PTEN sequesters CHK1 and initiates genetic instability.* Cancer Cell, 2005. **7**(2): p. 193-204.
308. Stern, H.M., et al., *PTEN Loss Is Associated with Worse Outcome in HER2-Amplified Breast Cancer Patients but Is Not Associated with Trastuzumab Resistance.* Clin Cancer Res, 2015. **21**(9): p. 2065-74.
309. Yu, J., et al., *An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression.* Cancer Cell, 2010. **17**(5): p. 443-54.
310. Shah, A.V., G.M. Birdsey, and A.M. Randi, *Regulation of endothelial homeostasis, vascular development and angiogenesis by the transcription factor ERG.* Vascul Pharmacol, 2016. **86**: p. 3-13.
311. Polson, E.S., et al., *Monoallelic expression of TMPRSS2/ERG in prostate cancer stem cells.* Nat Commun, 2013. **4**: p. 1623.
312. Zhao, S., et al., *Cellular Retinaldehyde-Binding Protein-Like (CRALBPL), a Novel Human Sec14p-Like Gene That Is Upregulated in Human Hepatocellular Carcinomas, May Be Used as a Marker for Human Hepatocellular Carcinomas.* DNA and Cell Biology, 2008. **27**(3): p. 159-163.
313. Chen, J., et al., *Gene-expression signature predicts survival benefit from postoperative chemoradiotherapy in head and neck squamous cell carcinoma.* Oncol Lett, 2018. **16**(2): p. 2565-2578.
314. Ji, Y., et al., *Developing a radiosensitivity gene signature for Caucasian patients with breast cancer.* Oncol Rep, 2018. **40**(3): p. 1695-1705.
315. Singh, R., et al., *Significant alterations of the novel 15 gene signature identified from macrophage-tumor interactions in breast cancer.* Biochim Biophys Acta, 2018. **1862**(3): p. 669-683.
316. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
317. Wang, D.Y., et al., *A new gene expression signature, the ClinicoMolecular Triad Classification, may improve prediction and prognostication of breast cancer at the time of diagnosis.* Breast Cancer Res, 2011. **13**(5): p. R92.
318. Kuo, W.H., et al., *Molecular characteristics and metastasis predictor genes of triple-negative breast cancer: a clinical study of triple-negative breast carcinomas.* PLoS One, 2012. **7**(9): p. e45831.
319. Nagalla, S., et al., *Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis.* Genome Biol, 2013. **14**(4): p. R34.
320. Sabatier, R., et al., *A gene expression signature identifies two prognostic subgroups of basal breast cancer.* Breast Cancer Res Treat, 2011. **126**(2): p. 407-20.
321. Kao, K.J., et al., *Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization.* BMC Cancer, 2011. **11**: p. 143.
322. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
323. Cerami, E., et al., *The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data.* Cancer Discovery, 2012. **2**(5): p. 401-404.
324. Chen, C., H. Huang, and C.H. Wu, *Protein Bioinformatics Databases and Resources.* Methods Mol Biol, 2017. **1558**: p. 3-39.
325. Nightingale, A., et al., *The Proteins API: accessing key integrated protein and genome information.* Nucleic Acids Research, 2017. **45**(W1): p. W539-W544.
326. Poux, S., et al., *On expert curation and sustainability: UniProtKB/Swiss-Prot as a case study.* bioRxiv, 2016.
327. Higgins, R., et al., *The absence of specific yeast heat-shock proteins leads to abnormal aggregation and compromised autophagic clearance of mutant Huntingtin proteins.* PLoS One, 2018. **13**(1): p. e0191490.

328. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
329. *Expansion of the Gene Ontology knowledgebase and resources.* Nucleic Acids Res, 2017. **45**(D1): p. D331-d338.
330. *Database resources of the National Center for Biotechnology Information.* Nucleic Acids Res, 2018. **46**(D1): p. D8-d13.
331. Xu, W., et al., *Surface charge and hydrophobicity determine ErbB2 binding to the Hsp90 chaperone complex.* Nat Struct Mol Biol, 2005. **12**(2): p. 120-6.
332. Xu, W., et al., *Loss of Hsp90 association up-regulates Src-dependent ErbB2 activity.* Mol Cell Biol, 2007. **27**(1): p. 220-8.
333. Citri, A., et al., *Hsp90 restrains ErbB-2/HER2 signalling by limiting heterodimer formation.* EMBO Rep, 2004. **5**(12): p. 1165-70.
334. Bria, E., et al., *Human epidermal growth factor receptor 2-positive breast cancer: heat shock protein 90 overexpression, Ki67 proliferative index, and topoisomerase II-alpha co-amplification as predictors of pathologic complete response to neoadjuvant chemotherapy with trastuzumab and docetaxel.* Clin Breast Cancer, 2015. **15**(1): p. 16-23.
335. Wuerstlein, R. and N. Harbeck, *Neoadjuvant Therapy for HER2-positive Breast Cancer.* Rev Recent Clin Trials, 2017. **12**(2): p. 81-92.
336. Muftah, A.A., et al., *Ki67 expression in invasive breast cancer: the use of tissue microarrays compared with whole tissue sections.* Breast Cancer Res Treat, 2017. **164**(2): p. 341-348.
337. Chen, V.E., et al., *Pathologic response after neoadjuvant chemotherapy predicts locoregional control in patients with triple negative breast cancer.* Adv Radiat Oncol, 2017. **2**(2): p. 105-109.
338. Jarzab, M., et al., *Ratio of proliferation markers and HSP90 gene expression as a predictor of pathological complete response in breast cancer neoadjuvant chemotherapy.* Folia Histochem Cytobiol, 2016. **54**(4): p. 202-209.
339. Cheng, Q., et al., *Amplification and high-level expression of heat shock protein 90 marks aggressive phenotypes of human epidermal growth factor receptor 2 negative breast cancer.* Breast Cancer Res, 2012. **14**(2): p. R62.
340. Daoud, A., et al., *Molecular and functional crosstalk between extracellular Hsp90 and ephrin A1 signaling.* Oncotarget, 2017. **8**(63): p. 106807-106819.
341. McCready, J., et al., *Secretion of extracellular hsp90alpha via exosomes increases cancer cell motility: a role for plasminogen activation.* BMC Cancer, 2010. **10**: p. 294.
342. Singh, R., et al., *Significant alterations of the novel 15 gene signature identified from macrophage-tumor interactions in breast cancer.* Biochim Biophys Acta, 2017. **1862**(3): p. 669-683.
343. Bode, W., et al., *The metzincin-superfamily of zinc-peptidases.* Adv Exp Med Biol, 1996. **389**: p. 1-11.
344. Xu, Y., et al., *Genetic variants in the metzincin metallopeptidase family genes predict melanoma survival.* Mol Carcinog, 2018. **57**(1): p. 22-31.
345. Shi, Y., et al., *Propoxur enhances MMP-2 expression and the corresponding invasion of human breast cancer cells via the ERK/Nrf2 signaling pathway.* Oncotarget, 2017. **8**(50): p. 87107-87123.
346. Singh, R., et al., *Positive Correlation between Matrix Metalloproteinases and Epithelial-to-Mesenchymal Transition and its Association with Clinical Outcome in Bladder Cancer Patients.* Cancer Microenviron, 2018.
347. Huang, M., et al., *Fatty acid binding protein 4 enhances prostate cancer progression by upregulating matrix metalloproteinases and stromal cell cytokine production.* Oncotarget, 2017. **8**(67): p. 111780-111794.

348. Wang, J., et al., *Matrix metalloproteinase-1 expression in breast carcinoma: a marker for unfavorable prognosis.* Oncotarget, 2017. **8**(53): p. 91379-91390.
349. Klassen, L.M.B., et al., *MMP9 gene expression regulation by intragenic epigenetic modifications in breast cancer.* Gene, 2018. **642**: p. 461-466.
350. Bai, X.Y., et al., *Kruppel-like factor 9 down-regulates matrix metalloproteinase 9 transcription and suppresses human breast cancer invasion.* Cancer Lett, 2018. **412**: p. 224-235.
351. Orive-Ramos, A., et al., *Regulation of the prometastatic neuregulin-MMP13 axis by SRC family kinases: therapeutic implications.* Mol Oncol, 2017. **11**(12): p. 1788-1805.
352. Merdad, A., et al., *Expression of matrix metalloproteinases (MMPs) in primary human breast cancer: MMP-9 as a potential biomarker for cancer invasion and metastasis.* Anticancer Res, 2014. **34**(3): p. 1355-66.
353. Gari, H.H., et al., *PRL-3 engages the focal adhesion pathway in triple-negative breast cancer cells to alter actin structure and substrate adhesion properties critical for cell migration and invasion.* Cancer Lett, 2016. **380**(2): p. 505-12.
354. Delassus, G.S., H. Cho, and G.L. Eliceiri, *New signaling pathways from cancer progression modulators to mRNA expression of matrix metalloproteinases in breast cancer cells.* J Cell Physiol, 2011. **226**(12): p. 3378-84.
355. Decock, J., et al., *Matrix metalloproteinase expression patterns in luminal A type breast carcinomas.* Dis Markers, 2007. **23**(3): p. 189-96.
356. Sims, J.D., J. McCready, and D.G. Jay, *Extracellular heat shock protein (Hsp)70 and Hsp90alpha assist in matrix metalloproteinase-2 activation and breast cancer cell migration and invasion.* PLoS One, 2011. **6**(4): p. e18848.
357. Lorenc, Z., et al., *Profile of Expression of Genes Encoding Matrix Metallopeptidase 9 (MMP9), Matrix Metallopeptidase 28 (MMP28) and TIMP Metallopeptidase Inhibitor 1 (TIMP1) in Colorectal Cancer: Assessment of the Role in Diagnosis and Prognostication.* Med Sci Monit, 2017. **23**: p. 1305-1311.
358. Bister, V.O., et al., *Differential expression of three matrix metalloproteinases, MMP-19, MMP-26, and MMP-28, in normal and inflamed intestine and colon cancer.* Dig Dis Sci, 2004. **49**(4): p. 653-61.
359. Zhang, J., et al., *Overexpression of MMP21 and MMP28 is associated with gastric cancer progression and poor prognosis.* Oncol Lett, 2018. **15**(5): p. 7776-7782.
360. Yu, K.-D., et al., *Identification of Prognosis-Relevant Subgroups in Patients with Chemoresistant Triple-Negative Breast Cancer.* Clinical Cancer Research, 2013. **19**(10): p. 2723-2733.
361. Leifeld, L., et al., *Enhanced expression of CD80 (B7-1), CD86 (B7-2), and CD40 and their ligands CD28 and CD154 in fulminant hepatic failure.* Am J Pathol, 1999. **154**(6): p. 1711-20.
362. Kontani, K., et al., *Involvement of granzyme B and perforin in suppressing nodal metastasis of cancer cells in breast and lung cancers.* Eur J Surg Oncol, 2001. **27**(2): p. 180-6.
363. Liu, Q., et al., *Targeting interlukin-6 to relieve immunosuppression in tumor microenvironment.* Tumour Biol, 2017. **39**(6): p. 1010428317712445.
364. Stuart-Harris, R., et al., *Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients.* Breast, 2008. **17**(4): p. 323-34.
365. Moazed, V., et al., *Prognostic Significance of Reduction in Ki67 Index After Neoadjuvant Chemotherapy in Patients With Breast Cancer in Kerman Between 2009 And 2014.* Iranian Journal of Pathology, 2018. **13**(1): p. 71-77.

366. Chen, L., et al., *Co-targeting HSP90 and its client proteins for treatment of prostate cancer.* Molecular cancer therapeutics, 2016. **15**(9): p. 2107-2118.
367. Sebastiá, R., et al., *Immunohistochemical detection of Hsp90 and Ki-67 in pterygium.* Diagnostic Pathology, 2013. **8**: p. 32-32.
368. Ramasamy, B., et al., *Temporal pterygium: benign or not?* The British Journal of Ophthalmology, 2005. **89**(11): p. 1533-1534.
369. Whitesell, L. and S.L. Lindquist, *HSP90 and the chaperoning of cancer.* Nature Reviews Cancer, 2005. **5**: p. 761.
370. Bjorge, J.D., et al., *Simultaneous siRNA targeting of Src and downstream signaling molecules inhibit tumor formation and metastasis of a human model breast cancer cell line.* PLoS One, 2011. **6**(4): p. e19309.
371. Ghaffari, A., et al., *A novel role for ezrin in breast cancer angio/lymphangiogenesis.* Breast Cancer Research : BCR, 2014. **16**: p. 438.
372. Ignacio, R.-C., et al., *Multi-scale modelling of cancer cell intravasation: the role of cadherins in metastasis.* Physical Biology, 2009. **6**(1): p. 016008.
373. Tomlinson, J.S., M.L. Alpaugh, and S.H. Barsky, *An intact overexpressed E-cadherin/alpha,beta-catenin axis characterizes the lymphovascular emboli of inflammatory breast carcinoma.* Cancer Res, 2001. **61**(13): p. 5231-41.
374. Liu, H., et al., *Prognostic and clinicopathological value of CXCL12/SDF1 expression in breast cancer: A meta-analysis.* Clin Chim Acta, 2018. **484**: p. 72-80.
375. Raschioni, C., et al., *CXCR4/CXCL12 Signaling and Protumor Macrophages in Primary Tumors and Sentinel Lymph Nodes Are Involved in Luminal B Breast Cancer Progression.* Dis Markers, 2018. **2018**: p. 5018671.
376. Diaz-Valdivia, N.I., et al., *Anti-neoplastic drugs increase caveolin-1-dependent migration, invasion and metastasis of cancer cells.* Oncotarget, 2017. **8**(67): p. 111943-111965.
377. Saier, M.H., *A functional-phylogenetic classification system for transmembrane solute transporters.* Microbiology and Molecular Biology Reviews : MMBR, 2000. **64**(2): p. 354-411.
378. Bay, D.C., et al., *Phylogenetic and coevolutionary analysis of the beta-barrel protein family comprised of mitochondrial porin (VDAC) and Tom40.* Biochim Biophys Acta, 2012. **1818**(6): p. 1502-19.
379. Huret, J.L., et al., *An Atlas on genes and chromosomes in oncology and haematology.* Cell Mol Biol (Noisy-le-grand), 2004. **50**(7): p. 805-7.
380. Blachly-Dyson, E. and M. Forte, *VDAC channels.* IUBMB Life, 2001. **52**(3-5): p. 113-8.
381. Cotterill SJ. *Chromosome 10*. 2018 [cited 2018 Sun Apr 22]; Available from: http://www.cancerindex.org/geneweb/clinkc10.htm.
382. Wang, T., et al., *Ankyrin G expression is associated with androgen receptor stability, invasiveness, and lethal outcome in prostate cancer patients.* J Mol Med (Berl), 2016. **94**(12): p. 1411-1422.
383. Miyake, M., et al., *Collagen type IV alpha 1 (COL4A1) and collagen type XIII alpha 1 (COL13A1) produced in cancer cells promote tumor budding at the invasion front in human urothelial carcinoma of the bladder.* Oncotarget, 2017. **8**(22): p. 36099-36114.
384. Ren, T., et al., *MCU-dependent mitochondrial Ca(2+) inhibits NAD(+)/SIRT3/SOD2 pathway to promote ROS production and metastasis of HCC cells.* Oncogene, 2017. **36**(42): p. 5897-5909.
385. Yu, C., et al., *Mitochondrial calcium uniporter as a target of microRNA-340 and promoter of metastasis via enhancing the Warburg effect.* Oncotarget, 2017. **8**(48): p. 83831-83844.
386. Wang, Y.P. and Q.Y. Lei, *Perspectives of Reprogramming Breast Cancer Metabolism.* Adv Exp Med Biol, 2017. **1026**: p. 217-232.
387. Cheng, Y.A.O., et al., *High glucose-induced resistance to 5-fluorouracil in pancreatic cancer cells alleviated by 2-deoxy-D-glucose.* Biomedical Reports, 2014. **2**(2): p. 188-192.

388. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**(5930): p. 1029-33.
389. Chang, C.W., et al., *ROS-independent ER stress-mediated NRF2 activation promotes warburg effect to maintain stemness-associated properties of cancer-initiating cells.* Cell Death Dis, 2018. **9**(2): p. 194.
390. Oishi, Y., et al., *Functional pathway characterized by gene expression analysis of supraclavicular lymph node metastasis-positive breast cancer.* J Hum Genet, 2007. **52**(3): p. 271-9.
391. Kakumu, T., et al., *Identification of proteasomal catalytic subunit PSMA6 as a therapeutic target for lung cancer.* Cancer Sci, 2017. **108**(4): p. 732-743.
392. Robey, I.F., et al., *Regulation of the Warburg effect in early-passage breast cancer cells.* Neoplasia, 2008. **10**(8): p. 745-56.
393. Lee, K.-m., et al., *MYC and MCL1 Cooperatively Promote Chemotherapy-Resistant Breast Cancer Stem Cells via Regulation of Mitochondrial Oxidative Phosphorylation.* Cell Metabolism, 2017. **26**(4): p. 633-647.e7.
394. Sotgia, F., M. Fiorillo, and M.P. Lisanti, *Mitochondrial markers predict recurrence, metastasis and tamoxifen-resistance in breast cancer patients: Early detection of treatment failure with companion diagnostics.* Oncotarget, 2017. **8**(40): p. 68730-68745.
395. Ren, D., et al., *The VDAC2-BAK rheostat controls thymocyte survival.* Sci Signal, 2009. **2**(85): p. ra48.
396. Lazarou, M., et al., *Inhibition of Bak activation by VDAC2 is dependent on the Bak transmembrane anchor.* J Biol Chem, 2010. **285**(47): p. 36876-83.
397. Churg, A., et al., *Macrophage metalloelastase mediates acute cigarette smoke-induced inflammation via tumor necrosis factor-alpha release.* Am J Respir Crit Care Med, 2003. **167**(8): p. 1083-9.
398. Dean, R.A., et al., *Macrophage-specific metalloelastase (MMP-12) truncates and inactivates ELR+ CXC chemokines and generates CCL2, -7, -8, and -13 antagonists: potential role of the macrophage in terminating polymorphonuclear leukocyte influx.* Blood, 2008. **112**(8): p. 3455-64.
399. Yu, P.F., et al., *TNFalpha-activated mesenchymal stromal cells promote breast cancer metastasis by recruiting CXCR2(+) neutrophils.* Oncogene, 2017. **36**(4): p. 482-490.
400. Salz, T., et al., *Histone Methyltransferase hSETD1A Is a Novel Regulator of Metastasis in Breast Cancer.* Mol Cancer Res, 2015. **13**(3): p. 461-9.
401. de la Pena, S., et al., *Expression of the matrix metalloproteases 2, 14, 24, and 25 and tissue inhibitor 3 as potential molecular markers in advanced human gastric cancer.* Dis Markers, 2014. **2014**: p. 285906.
402. Benson, C.S., et al., *Expression of matrix metalloproteinases in human breast cancer tissues.* Dis Markers, 2013. **34**(6): p. 395-405.
403. Upreti, D., et al., *Change in CD3zeta-chain expression is an independent predictor of disease status in head and neck cancer patients.* Int J Cancer, 2016. **139**(1): p. 122-9.
404. Prado-Garcia, H., et al., *Lung carcinomas do not induce T-cell apoptosis via the Fas/Fas ligand pathway but down-regulate CD3 epsilon expression.* Cancer Immunol Immunother, 2008. **57**(3): p. 325-36.
405. Arias-Pulido, H., et al., *The combined presence of CD20 + B cells and PD-L1 + tumor-infiltrating lymphocytes in inflammatory breast cancer is prognostic of improved patient outcome.* Breast Cancer Res Treat, 2018.
406. Langat, D.L., et al., *Signaling pathways for B cell-activating factor (BAFF) and a proliferation-inducing ligand (APRIL) in human placenta.* Am J Pathol, 2008. **172**(5): p. 1303-11.
407. Hitosugi, T., et al., *Phosphoglycerate mutase 1 coordinates glycolysis and biosynthesis to promote tumor growth.* Cancer Cell, 2012. **22**(5): p. 585-600.

408. Kondoh, H., et al., *Protection from oxidative stress by enhanced glycolysis; a possible mechanism of cellular immortalization.* Histol Histopathol, 2007. **22**(1): p. 85-90.
409. Evans, M.J., et al., *Target discovery in small-molecule cell-based screens by in situ proteome reactivity profiling.* Nat Biotechnol, 2005. **23**(10): p. 1303-7.
410. Lei, Y., et al., *Proteomics identification of ITGB3 as a key regulator in reactive oxygen species-induced migration and invasion of colorectal cancer cells.* Mol Cell Proteomics, 2011. **10**(10): p. M110.005397.
411. Hitosugi, T., et al., *Phosphoglycerate Mutase 1 Coordinates Glycolysis and Biosynthesis to Promote Tumor Growth.* Cancer Cell, 2012. **22**(5): p. 585-600.
412. Gao, H., et al., *Correlation of expression levels of ANXA2, PGAM1, and CALR with glioma grade and prognosis.* J Neurosurg, 2013. **118**(4): p. 846-53.
413. Deighton, R.F., et al., *The proteomic response in glioblastoma in young patients.* Journal of Neuro-Oncology, 2014. **119**(1): p. 79-89.
414. Xu, Z., et al., *The diagnostic value and functional roles of phosphoglycerate mutase 1 in glioma.* Oncol Rep, 2016. **36**(4): p. 2236-44.
415. Linge, A., et al., *Differential expression of fourteen proteins between uveal melanoma from patients who subsequently developed distant metastases versus those who did Not.* Invest Ophthalmol Vis Sci, 2012. **53**(8): p. 4634-43.
416. Thongwatchara, P., et al., *Differential protein expression in primary breast cancer and matched axillary node metastasis.* Oncol Rep, 2011. **26**(1): p. 185-91.
417. Zhang, D., et al., *Phosphoglycerate mutase 1 promotes cancer cell migration independent of its metabolic activity.* Oncogene, 2016. **36**: p. 2900.
418. Hou, M.F., P.M. Chen, and P.Y. Chu, *LGR5 overexpression confers poor relapse-free survival in breast cancer patients.* BMC Cancer, 2018. **18**(1): p. 219.
419. *Rac1 and Stathmin but Not EB1 Are Required for Invasion of Breast Cancer Cells in Response to IGF-I.* International Journal of Cell Biology, 2011. **2011**.
420. Gray, K.A., et al., *Genenames.org: the HGNC resources in 2015.* Nucleic Acids Res, 2015. **43**(Database issue): p. D1079-85.
421. Binder, J.X., et al., *COMPARTMENTS: unification and visualization of protein subcellular localization evidence.* Database, 2014. **2014**: p. bau012-bau012.
422. Hall, A., *Rho family GTPases.* Biochem Soc Trans, 2012. **40**(6): p. 1378-82.
423. Cho, H.J., et al., *Protein phosphatase 1B dephosphorylates Rho guanine nucleotide dissociation inhibitor 1 and suppresses cancer cell migration and invasion.* Cancer Lett, 2018. **417**: p. 141-151.
424. Lartey, J. and A. Lopez Bernal, *RHO protein regulation of contraction in the human uterus.* Reproduction, 2009. **138**(3): p. 407-24.
425. Fritz, G., et al., *Rho GTPases in human breast tumours: expression and mutation analyses and correlation with clinical parameters.* British Journal of Cancer, 2002. **87**(6): p. 635-644.
426. DerMardirossian, C. and G.M. Bokoch, *GDIs: central regulatory molecules in Rho GTPase activation.* Trends Cell Biol, 2005. **15**(7): p. 356-63.
427. Takahashi, K., et al., *Direct Interaction of the Rho GDP Dissociation Inhibitor with Ezrin/Radixin/Moesin Initiates the Activation of the Rho Small G Protein.* Journal of Biological Chemistry, 1997. **272**(37): p. 23371-23375.
428. Penzes, P., et al., *Rapid induction of dendritic spine morphogenesis by trans-synaptic ephrinB-EphB receptor activation of the Rho-GEF kalirin.* Neuron, 2003. **37**(2): p. 263-74.
429. Lamarche, N. and A. Hall, *GAPs for rho-related GTPases.* Trends Genet, 1994. **10**(12): p. 436-40.

430. Thompson, W.R., et al., *LARG GEF and ARHGAP18 orchestrate RhoA activity to control mesenchymal stem cell lineage.* Bone, 2018. **107**: p. 172-180.
431. El Marzouk, S., et al., *Rho GDP dissociation inhibitor alpha interacts with estrogen receptor alpha and influences estrogen responsiveness.* J Mol Endocrinol, 2007. **39**(4): p. 249-59.
432. Hondermarck, H., et al., *Proteomics of breast cancer for marker discovery and signal pathway profiling.* Proteomics, 2001. **1**(10): p. 1216-32.
433. Huang, D., et al., *Rho GDP-dissociation inhibitor alpha is a potential prognostic biomarker and controls telomere regulation in colorectal cancer.* Cancer Sci, 2017. **108**(7): p. 1293-1302.
434. Jiang, W.G., et al., *Prognostic value of rho GTPases and rho guanine nucleotide dissociation inhibitors in human breast cancers.* Clin Cancer Res, 2003. **9**(17): p. 6432-40.
435. Wang, H., et al., *Overexpression of RhoGDI, a novel predictor of distant metastasis, promotes cell proliferation and migration in hepatocellular carcinoma.* FEBS Lett, 2014. **588**(3): p. 503-8.
436. Liang, L., et al., *Loss of ARHGDIA expression is associated with poor prognosis in HCC and promotes invasion and metastasis of HCC cells.* Int J Oncol, 2014. **45**(2): p. 659-66.
437. Grauzam, S., et al., *NEDD9 stimulated MMP9 secretion is required for invadopodia formation in oral squamous cell carcinoma.* Oncotarget, 2018. **9**(39): p. 25503-25516.
438. Xu, X.D., et al., *Anti-RhoC siRNAs inhibit the proliferation and invasiveness of breast cancer cells via modulating the KAI1, MMP9, and CXCR4 expression.* Onco Targets Ther, 2017. **10**: p. 1827-1834.
439. Liu, Y., et al., *Demethylation of Repressor Element-1 Silencing Transcription (REST) Suppresses the Malignant Phenotype of Breast Cancer via MMP9.* Oncol Res, 2017. **25**(3): p. 445-454.
440. Hegedus, L., et al., *Additional MDA-MB-231 breast cancer cell matrix metalloproteinases promote invasiveness.* J Cell Physiol, 2008. **216**(2): p. 480-5.
441. Cominelli, A., et al., *Matrix metalloproteinase-27 is expressed in CD163+/CD206+ M2 macrophages in the cycling human endometrium and in superficial endometriotic lesions.* Mol Hum Reprod, 2014. **20**(8): p. 767-75.
442. Bradbury, R., W.G. Jiang, and Y.X. Cui, *MDM2 and PSMA Play Inhibitory Roles in Metastatic Breast Cancer Cells Through Regulation of Matrix Metalloproteinases.* Anticancer Res, 2016. **36**(3): p. 1143-51.
443. DeNardo, D.G., et al., *CD4⁺ T Cells Regulate Pulmonary Metastasis of Mammary Carcinomas by Enhancing Protumor Properties of Macrophages.* Cancer Cell, 2009. **16**(2): p. 91-102.
444. Jiang, Z., et al., *IL-25 blockade inhibits metastasis in breast cancer.* Protein Cell, 2017. **8**(3): p. 191-201.
445. Bozza, W.P., et al., *RhoGDI deficiency induces constitutive activation of Rho GTPases and COX-2 pathways in association with breast cancer progression.* Oncotarget, 2015. **6**(32): p. 32723-36.
446. Rhodes, L.V., et al., *Dual regulation by microRNA-200b-3p and microRNA-200b-5p in the inhibition of epithelial-to-mesenchymal transition in triple-negative breast cancer.* Oncotarget, 2015. **6**(18): p. 16638-52.
447. Na, Y., et al., *Stress chaperone mortalin contributes to epithelial-mesenchymal transition and cancer metastasis.* Cancer Res, 2016.
448. Croft, D., et al., *The Reactome pathway knowledgebase.* Nucleic Acids Res, 2014. **42**(Database issue): p. D472-7.
449. Gehler, S., et al., *p75 neurotrophin receptor signaling regulates growth cone filopodial dynamics through modulating RhoA activity.* J Neurosci, 2004. **24**(18): p. 4363-72.

450. Austin, M., et al., *Breast cancer induced nociceptor aberrant growth and collateral sensory axonal branching.* Oncotarget, 2017. **8**(44): p. 76606-76621.
451. Jobling, P., et al., *Nerve-Cancer Cell Cross-talk: A Novel Promoter of Tumor Progression.* Cancer Res, 2015. **75**(9): p. 1777-81.
452. Huang, D., et al., *Nerve fibers in breast cancer tissues indicate aggressive tumor progression.* Medicine (Baltimore), 2014. **93**(27): p. e172.
453. Pundavela, J., et al., *Nerve fibers infiltrate the tumor microenvironment and are associated with nerve growth factor production and lymph node invasion in breast cancer.* Mol Oncol, 2015. **9**(8): p. 1626-35.
454. Jin, Q., et al., *Decreased tumor progression and invasion by a novel anti-cell motility target for human colorectal carcinoma cells.* PLoS One, 2013. **8**(6): p. e66439.
455. King, S.M., *The dynein microtubule motor.* Biochim Biophys Acta, 2000. **1496**(1): p. 60-75.
456. Jiang, J., et al., *Identification of two novel human dynein light chain genes, DNLC2A and DNLC2B, and their expression changes in hepatocellular carcinoma tissues from 68 Chinese patients.* Gene, 2001. **281**(1-2): p. 103-13.
457. Jin, Q., et al., *Role of km23-1 in RhoA/actin-based cell migration.* Biochem Biophys Res Commun, 2012. **428**(3): p. 333-8.
458. Loi, S., et al., *RAS/MAPK Activation Is Associated with Reduced Tumor-Infiltrating Lymphocytes in Triple-Negative Breast Cancer: Therapeutic Cooperation Between MEK and PD-1/PD-L1 Immune Checkpoint Inhibitors.* Clin Cancer Res, 2016. **22**(6): p. 1499-509.
459. Epstein Shochet, G., et al., *Placenta-breast cancer cell interactions promote cancer cell epithelial mesenchymal transition via TGFbeta/JNK pathway.* Clin Exp Metastasis, 2014. **31**(8): p. 961-75.
460. Jin, Q., W. Ding, and K.M. Mulder, *The TGFbeta receptor-interacting protein km23-1/DYNLRB1 plays an adaptor role in TGFbeta1 autoinduction via its association with Ras.* J Biol Chem, 2012. **287**(31): p. 26453-63.
461. Liu, J.J., et al., *Shp2 deletion in hepatocytes suppresses hepatocarcinogenesis driven by oncogenic beta-Catenin, PIK3CA and MET.* J Hepatol, 2018. **69**(1): p. 79-88.
462. Fatima, I., et al., *The natural compound Jatrophone interferes with Wnt/beta-catenin signaling and inhibits proliferation and EMT in human triple-negative breast cancer.* PLoS One, 2017. **12**(12): p. e0189864.
463. Hirokawa, N., *Kinesin and dynein superfamily proteins and the mechanism of organelle transport.* Science, 1998. **279**(5350): p. 519-26.
464. Tang, Q., et al., *A novel transforming growth factor-beta receptor-interacting protein that is also a light chain of the motor protein dynein.* Mol Biol Cell, 2002. **13**(12): p. 4484-96.
465. Vivekanandhan, S. and D. Mukhopadhyay, *Genetic status of KRAS influences Transforming Growth Factor-beta (TGF-beta) signaling: An insight into Neuropilin-1 (NRP1) mediated tumorigenesis.* Semin Cancer Biol, 2018.
466. Wendt, M.K., M. Tian, and W.P. Schiemann, *Deconstructing the mechanisms and consequences of TGF-β-induced EMT during cancer progression.* Cell and Tissue Research, 2012. **347**(1): p. 85-101.
467. DiScipio, R.G., R.C. Liddington, and I.U. Schraufstatter, *Fibulin-1 purification from human plasma using affinity chromatography on Factor H-Sepharose.* Protein Expr Purif, 2016. **121**: p. 118-24.
468. Halper, J. and M. Kjaer, *Basic components of connective tissues and extracellular matrix: elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins.* Adv Exp Med Biol, 2014. **802**: p. 31-47.

469. Tran, H., et al., *Human fibulin-1D: molecular cloning, expression and similarity with S1-5 protein, a new member of the fibulin gene family.* Matrix Biol, 1997. **15**(7): p. 479-93.
470. Cui, Y., et al., *Fibulin-1 functions as a prognostic factor in lung adenocarcinoma.* Jpn J Clin Oncol, 2015. **45**(9): p. 854-9.
471. de Vega, S., T. Iwamoto, and Y. Yamada, *Fibulins: Multiple roles in matrix structures and tissue functions.* Cellular and Molecular Life Sciences, 2009. **66**(11): p. 1890-1902.
472. Kischel, P., et al., *Identification of stromal proteins overexpressed in nodular sclerosis Hodgkin lymphoma.* Proteome Sci, 2011. **9**(1): p. 63.
473. Pupa, S.M., et al., *Immunological and pathobiological roles of fibulin-1 in breast cancer.* Oncogene, 2004. **23**(12): p. 2153-60.
474. Greene, L.M., et al., *Elevated expression and altered processing of fibulin-1 protein in human breast cancer.* British Journal Of Cancer, 2003. **88**: p. 871.
475. Zhu, J., et al., *Expression of fibulin-1 predicted good prognosis in patients with colorectal cancer.* Am J Transl Res, 2015. **7**(2): p. 339-47.
476. Roger, P., et al., *Increased immunostaining of fibulin-1, an estrogen-regulated protein in the stroma of human ovarian epithelial tumors.* Am J Pathol, 1998. **153**(5): p. 1579-88.
477. Lertsuwan, K., et al., *Identification of Fibulin-1 as a Human Bone Marrow Stromal (HS-5) Cell-Derived Factor That Induces Human Prostate Cancer Cell Death.* Prostate, 2017. **77**(7): p. 729-742.
478. Feng, L., et al., *Low expression of fibulin-1 correlates with unfavorable prognosis in gastric cancer.* Tumour Biol, 2016. **37**(7): p. 9399-410.
479. Xiao, W., et al., *Fibulin-1 is epigenetically down-regulated and related with bladder cancer recurrence.* BMC Cancer, 2014. **14**: p. 677.
480. Wang, Q., et al., *Extracellular calumenin suppresses ERK1/2 signaling and cell migration by protecting fibulin-1 from MMP-13-mediated proteolysis.* Oncogene, 2015. **34**(8): p. 1006-18.
481. Manders, D.B., et al., *Dysregulation of fibulin-5 and matrix metalloproteases in epithelial ovarian cancer.* Oncotarget, 2018. **9**(18): p. 14251-14267.
482. Heo, J.H., et al., *Fibulin-5 is a tumour suppressor inhibiting cell migration and invasion in ovarian cancer.* J Clin Pathol, 2016. **69**(2): p. 109-16.
483. Sica, A., et al., *Origin and Functions of Tumor-Associated Myeloid Cells (TAMCs).* Cancer Microenviron, 2012. **5**(2): p. 133-49.
484. Sarangi, P.P., et al., *Cell adhesion protein fibulin-7 and its C-terminal fragment negatively regulate monocyte and macrophage migration and functions in vitro and in vivo.* Faseb j, 2018: p. fj201700686RRR.
485. Wu, W., et al., *[Expression of chemokine CXCL12 and its receptor (CXCR4 and CXCR7) in different molecular subtypes of human breast carcinoma and the clinical significance].* Zhong Nan Da Xue Xue Bao Yi Xue Ban, 2017. **42**(2): p. 147-153.
486. Han, J., et al., *Identification of Biomarkers Based on Differentially Expressed Genes in Papillary Thyroid Carcinoma.* Sci Rep, 2018. **8**(1): p. 9912.
487. Goncalves-Ribeiro, S., et al., *Prediction of pathological response to neoadjuvant treatment in rectal cancer with a two-protein immunohistochemical score derived from stromal gene-profiling.* Ann Oncol, 2017. **28**(9): p. 2160-2168.
488. Shinde, A., et al., *Autocrine Fibronectin Inhibits Breast Cancer Metastasis.* Mol Cancer Res, 2018.
489. Lu, Z.D., et al., *[Effect of peritoneal fibrosis induced by transforming growth factor-beta 1 on the adhesion of gastric cancer cell].* Zhonghua Yi Xue Za Zhi, 2012. **92**(24): p. 1698-701.
490. Ding, M.J., et al., *Association between transforming growth factor-beta1 expression and the clinical features of triple negative breast cancer.* Oncol Lett, 2016. **11**(6): p. 4040-4044.

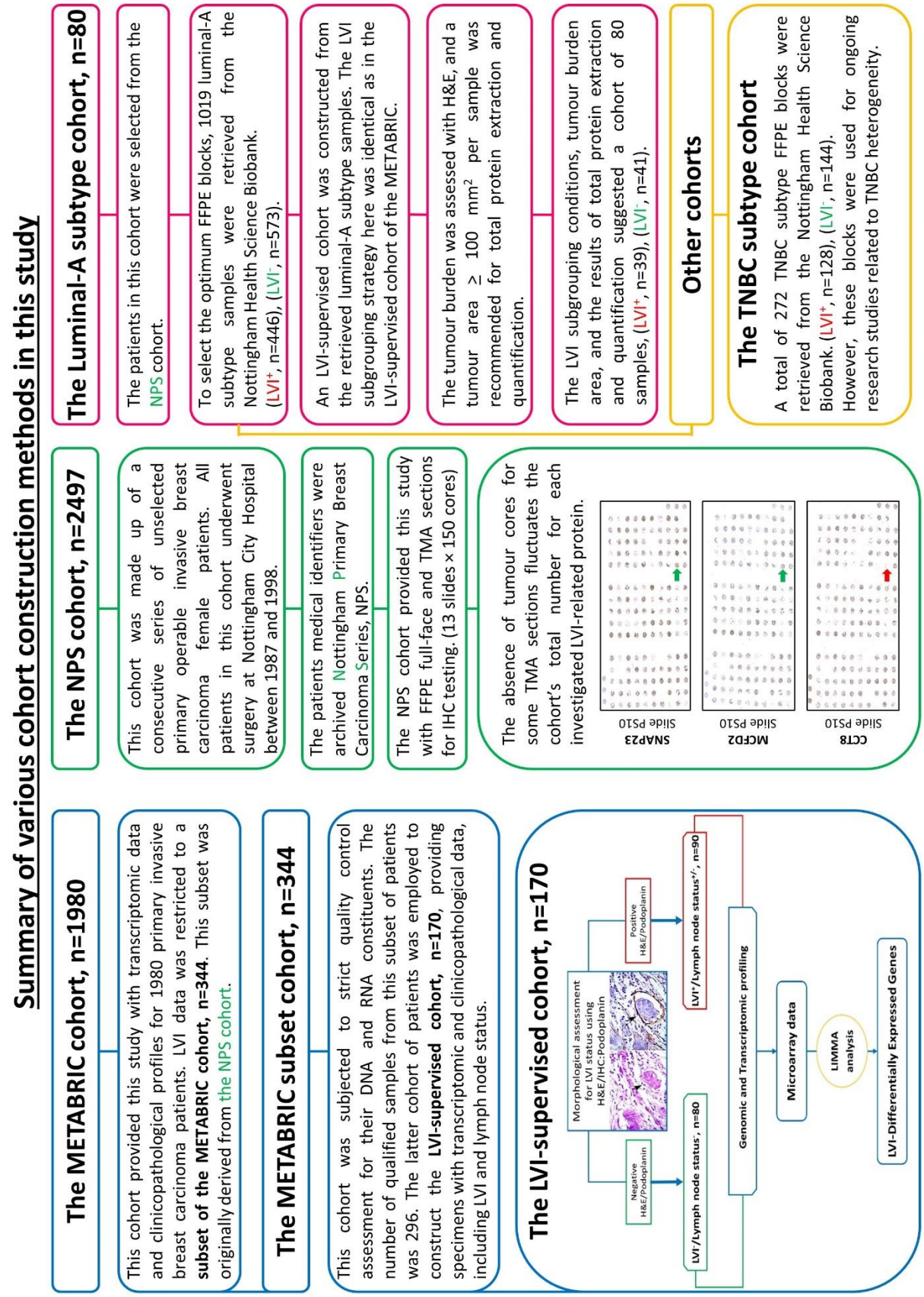
491. Hachim, M.Y., et al., *Differential expression of TGFbeta isoforms in breast cancer highlights different roles during breast cancer progression.* Tumour Biol, 2018. **40**(1): p. 1010428317748254.
492. Qi, Y., et al., *Prognostic value of the MicroRNA-29 family in multiple human cancers: A meta-analysis and systematic review.* Clin Exp Pharmacol Physiol, 2017. **44**(4): p. 441-454.
493. Ivanovic, R.F., et al., *miR-29b enhances prostate cancer cell invasion independently of MMP-2 expression.* Cancer Cell Int, 2018. **18**: p. 18.
494. Pekarsky, Y., et al., *Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181.* Cancer Res, 2006. **66**(24): p. 11590-3.
495. Garzon, R., et al., *MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia.* Blood, 2008. **111**(6): p. 3183-9.
496. Cortez, M.A., et al., *miR-29b and miR-125a regulate podoplanin and suppress invasion in glioblastoma.* Genes Chromosomes Cancer, 2010. **49**(11): p. 981-90.
497. Liang, Y.N., et al., *Combined caveolin-1 and epidermal growth factor receptor expression as a prognostic marker for breast cancer.* Oncol Lett, 2018. **15**(6): p. 9271-9282.
498. Du, C., et al., *Caveolin-1 limits the contribution of BKCa channel to MCF-7 breast cancer cell proliferation and invasion.* Int J Mol Sci, 2014. **15**(11): p. 20706-22.
499. Oeggerli, M., et al., *Role of KCNMA1 in breast cancer.* PLoS One, 2012. **7**(8): p. e41664.
500. Joglekar, M., et al., *Caveolin-1 mediates inflammatory breast cancer cell invasion via the Akt1 pathway and RhoC GTPase.* J Cell Biochem, 2015. **116**(6): p. 923-33.
501. Van den Eynden, G.G., et al., *Overexpression of caveolin-1 and -2 in cell lines and in human samples of inflammatory breast cancer.* Breast Cancer Res Treat, 2006. **95**(3): p. 219-28.
502. Colpaert, C.G., et al., *Inflammatory breast cancer shows angiogenesis with high endothelial proliferation rate and strong E-cadherin expression.* Br J Cancer, 2003. **88**(5): p. 718-25.
503. Kumar, A., et al., *Multiple roles of CD90 in cancer.* Tumour Biol, 2016. **37**(9): p. 11611-11622.
504. Lobba, A.R.M., et al., *High CD90 (THY-1) expression positively correlates with cell transformation and worse prognosis in basal-like breast cancer tumors.* PLoS One, 2018. **13**(6): p. e0199254.
505. Donnenberg, V.S., et al., *Localization of CD44 and CD90 positive cells to the invasive front of breast tumors.* Cytometry B Clin Cytom, 2010. **78**(5): p. 287-301.
506. Schliekelman, M.J., et al., *Thy-1(+) Cancer-associated Fibroblasts Adversely Impact Lung Cancer Prognosis.* Sci Rep, 2017. **7**(1): p. 6478.
507. Chen, W.C., et al., *Cancer stem cell marker CD90 inhibits ovarian cancer formation via beta3 integrin.* Int J Oncol, 2016. **49**(5): p. 1881-1889.
508. Lung, H.L., et al., *THY1 is a candidate tumour suppressor gene with decreased expression in metastatic nasopharyngeal carcinoma.* Oncogene, 2005. **24**(43): p. 6525-32.
509. Lung, H.L., et al., *Functional characterization of THY1 as a tumor suppressor gene with antiinvasive activity in nasopharyngeal carcinoma.* Int J Cancer, 2010. **127**(2): p. 304-12.
510. Osborne, B., et al., *A rare variant in human fibroblast activation protein associated with ER stress, loss of enzymatic function and loss of cell surface localisation.* Biochim Biophys Acta, 2014. **1844**(7): p. 1248-59.
511. Saadi, A., et al., *Stromal genes discriminate preinvasive from invasive disease, predict outcome, and highlight inflammatory pathways in digestive cancers.* Proc Natl Acad Sci U S A, 2010. **107**(5): p. 2177-82.

512. Tchou, J., et al., *Fibroblast activation protein expression by stromal cells and tumor-associated macrophages in human breast cancer.* Hum Pathol, 2013. **44**(11): p. 2549-57.
513. Jia, J., et al., *FAP-alpha (Fibroblast activation protein-alpha) is involved in the control of human breast cancer cell line growth and motility via the FAK pathway.* BMC Cell Biol, 2014. **15**: p. 16.
514. Ying, X., et al., *Lycorine inhibits breast cancer growth and metastasis via inducing apoptosis and blocking Src/FAK-involved pathway.* Sci China Life Sci, 2017. **60**(4): p. 417-428.
515. Lee, C.S., et al., *Isothiocyanates inhibit the invasion and migration of C6 glioma cells by blocking FAK/JNK-mediated MMP-9 expression.* Oncol Rep, 2015. **34**(6): p. 2901-8.
516. Kim, S.A., et al., *5'-Nitro-indirubinoxime, an indirubin derivative, suppresses metastatic ability of human head and neck cancer cells through the inhibition of Integrin beta1/FAK/Akt signaling.* Cancer Lett, 2011. **306**(2): p. 197-204.
517. Huang, Y., S. Wang, and T. Kelly, *Seprase promotes rapid tumor growth and increased microvessel density in a mouse model of human breast cancer.* Cancer Res, 2004. **64**(8): p. 2712-6.
518. Taliaferro-Smith, L., et al., *FAK activation is required for IGF1R-mediated regulation of EMT, migration, and invasion in mesenchymal triple negative breast cancer cells.* Oncotarget, 2015. **6**(7): p. 4757-72.
519. Park, S.Y., H.M. Kim, and J.S. Koo, *Differential expression of cancer-associated fibroblast-related proteins according to molecular subtype and stromal histology in breast cancer.* Breast Cancer Res Treat, 2015. **149**(3): p. 727-41.
520. Lambaerts, K., S.A. Wilcox-Adelman, and P. Zimmermann, *The signalling mechanisms of syndecan heparan sulphate proteoglycans.* Current opinion in cell biology, 2009. **21**(5): p. 662-669.
521. Shi, S., et al., *Syndecan-1 knockdown inhibits glioma cell proliferation and invasion by deregulating a c-src/FAK-associated signaling pathway.* Oncotarget, 2017. **8**(25): p. 40922-40934.
522. Cui, X., et al., *Clinicopathological and prognostic significance of SDC1 overexpression in breast cancer.* Oncotarget, 2017. **8**(67): p. 111444-111455.
523. Uen, Y., et al., *Mining of potential microRNAs with clinical correlation - regulation of syndecan-1 expression by miR-122-5p altered mobility of breast cancer cells and possible correlation with liver injury.* Oncotarget, 2018. **9**(46): p. 28165-28175.
524. Nguyen, T.L., et al., *Syndecan-1 overexpression is associated with nonluminal subtypes and poor prognosis in advanced breast cancer.* Am J Clin Pathol, 2013. **140**(4): p. 468-74.
525. Ibrahim, S.A., et al., *Syndecan-1 (CD138) modulates triple-negative breast cancer stem cell properties via regulation of LRP-6 and IL-6-mediated STAT3 signaling.* PLoS One, 2013. **8**(12): p. e85737.
526. Yang, N., et al., *Syndecan-1 in breast cancer stroma fibroblasts regulates extracellular matrix fiber organization and carcinoma cell motility.* Am J Pathol, 2011. **178**(1): p. 325-35.
527. Lindstrom, L.S., et al., *Intratumor Heterogeneity of the Estrogen Receptor and the Long-term Risk of Fatal Breast Cancer.* J Natl Cancer Inst, 2018. **110**(7): p. 726-733.
528. Hong, D., et al., *Epithelial-to-mesenchymal transition and cancer stem cells contribute to breast cancer heterogeneity.* Journal of Cellular Physiology. **0**(0).
529. Garcia-Fernandez, A., et al., *Factors predictive of mortality in a cohort of women surgically treated for breast cancer from 1997 to 2014.* Int J Gynaecol Obstet, 2016. **134**(2): p. 212-6.

530. Thiagarajan, P.S., et al., *Cx26 drives self-renewal in triple-negative breast cancer via interaction with NANOG and focal adhesion kinase.* Nat Commun, 2018. **9**(1): p. 578.
531. O'Conor, C.J., et al., *Cancer stem cells in triple-negative breast cancer: a potential target and prognostic marker.* Biomark Med, 2018. **12**(7): p. 813-820.
532. Shen, S.D., et al., *Correlation of lymphovascular invasion with clinicopathological factors in invasive breast cancer: a meta-analysis.* Int J Clin Exp Med, 2015. **8**(10): p. 17789-95.
533. Mousley, C.J., J.M. Davison, and V.A. Bankaitis, *Sec14 Like PITPs Couple Lipid Metabolism with Phosphoinositide Synthesis to Regulate Golgi Functionality*, in *Phosphoinositides II: The Diverse Biological Functions*, T. Balla, M. Wymann, and J.D. York, Editors. 2012, Springer Netherlands: Dordrecht. p. 271-287.

Appendices

APPENDIX A: SUPPLEMENTARY INFORMATION FOR CHAPTER 2



APPENDIX B: SUPPLEMENTARY INFORMATION FOR CHAPTER 3

WOLTERS KLUWER HEALTH, INC. LICENSE TERMS AND CONDITIONS

Dec 30, 2018

This Agreement between Mr. Sultan Sonbul ("You") and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc.") consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.

License Number	4444781031820
License date	Oct 09, 2018
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	American Journal of Surgical Pathology, The
Licensed Content Title	Improved Methods of Detection of Lymphovascular Invasion Demonstrate That It is the Predominant Method of Vascular Invasion in Breast Cancer and has Important Clinical Consequences
Licensed Content Author	Rabab Mohammed, Stewart Martin, Muhammad Gill, et al
Licensed Content Date	Dec 1, 2007
Licensed Content Volume	31
Licensed Content Issue	12
Type of Use	Dissertation/Thesis
Requestor type	Individual
STM publisher name	
Portion	Figures/table/illustration
Number of figures/tables/illustrations	1
Figures/tables/illustrations used	figure 1
Author of this Wolters Kluwer article	No
Title of your thesis / dissertation	Deciphering molecular portrait of lymphovascular invasion in invasive breast cancer
Expected completion date	Dec 2018
Estimated size(pages)	400
Requestor Location	Mr. Sultan Sonbul Flat 6, Thomas Forman Court, Stanhope Avenue Nottingham, Nottinghamshire NG51QY United Kingdom Attn: Mr. Sultan Sonbul
Publisher Tax ID	EU826013006
Customer VAT ID	GB690391225
Billing Type	Invoice
Billing Address	Mr. Sultan Sonbul Flat 6, Thomas Forman Court, Stanhope Avenue Nottingham, United Kingdom NG51QY Attn: Mr. Sultan Sonbul
Total	0.00 GBP

Protein information for MCFD2

MCFD2 The MCFD2-LMAN1 complex forms a specific cargo receptor for the ER-to-Golgi transport of selected proteins. Plays a role in the secretion of coagulation factors. Defects in MCFD2 are a cause of factor V and factor VIII combined deficiency type 2 (F5F8D2); also known as multiple coagulation factor deficiency 2 (MCFD2). F5F8D2 is a blood coagulation disorder characterized by bleeding symptoms similar to those in hemophilia or parahemophilia, that are caused by single deficiency of FV or FVIII, respectively. The most common symptoms are epistaxis, menorrhagia, and excessive bleeding during or after trauma. Plasma levels of coagulation factors V and VIII are in the range of 5 to 30% of normal. 3 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Endoplasmic reticulum

Chromosomal Location of Human Ortholog: 2p21

Cellular Component: endoplasmic reticulum membrane; endoplasmic reticulum-Golgi intermediate compartment membrane; Golgi membrane view less <

Molecular Function: calcium ion binding

Biological Process: COPII vesicle coating; ER to Golgi vesicle-mediated transport; protein N-linked glycosylation via asparagine; protein transport view less <

Disease: Factor V And Factor Viii, Combined Deficiency Of, 2

Reference #: Q8NI22 (UniProtKB)

Alt. Names/Synonyms: DKFZp686G21263; F5F8D; LMAN1IP; MCFD2; multiple coagulation factor deficiency 2; Multiple coagulation factor deficiency protein 2; neural stem cell derived neuronal survival protein; Neural stem cell-derived neuronal survival protein; SDNSF

Gene Symbols: MCFD2

Molecular weight: 16,390 Da

Basal Isoelectric point: 4.51 Predict pI for various phosphorylation states

Protein information for CCT8

CCT8 a chaperone of the TCP-1 chaperonin family. A part of a hetero-oligomeric complex of about 850 to 900 kDa that forms two stacked rings, 12 to 16 nm in diameter. Note: This description may include information from UniProtKB.

Protein type: Chaperone; Motility/polarity/chemotaxis

Chromosomal Location of Human Ortholog: 21q21.3

Cellular Component: cell body; centrosome; chaperonin-containing T-complex; cilium; cytoplasm; cytosol; extracellular re... view more >

Molecular Function: ATP binding; ATPase activity, coupled; protein binding

Biological Process: binding of sperm to zona pellucida; neutrophil degranulation; pore complex assembly; positive regula... view more >

Reference #: P50990 (UniProtKB)

Alt. Names/Synonyms: C21orf112; CCT-theta; CCT8; CCTQ; chaperonin containing TCP1, subunit 8 (theta); D21S246; KIAA0002; PRED71; Renal carcinoma antigen NY-REN-15; T-complex protein 1 subunit theta; T-complex protein 1, theta subunit; TCP-1-theta; TCPQ

Gene Symbols: CCT8

Molecular weight: 59,621 Da

Basal Isoelectric point: 5.42 Predict pI for various phosphorylation states

Protein information for PPM1G

PPM1G is a member of the PP2C family of Ser/Thr protein phosphatases. PP2C family members are known to be negative regulators of cell stress response pathways. This phosphatase is found to be responsible for the dephosphorylation of Pre-mRNA splicing factors, which is important for the formation of functional spliceosome. Studies of a similar gene in mice suggested a role of this phosphatase in regulating cell cycle progression. [provided by RefSeq, Apr 2010]

Protein type: EC 3.1.3.16; Motility/polarity/chemotaxis; Protein phosphatase, Ser/Thr (non-receptor); Spliceosome

Chromosomal Location of Human Ortholog: 2p23.3

Cellular Component: cytoplasm; membrane; nucleoplasm; nucleus

Molecular Function: metal ion binding; protein binding; protein serine/threonine phosphatase activity

Biological Process: cell cycle arrest; peptidyl-threonine dephosphorylation; protein amino acid dephosphorylation

Reference #: O15355 (UniProtKB)

Alt. Names/Synonyms: MGC1675; MGC2870; PP2C, gamma; PP2C-gamma; PP2CG; PP2CGAMMA; PPM1C; PPM1G; PPP2CG; Protein phosphatase 1C; Protein phosphatase 1G; protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform; protein phosphatase 2, catalytic subunit, gamma isoform; protein phosphatase 2C gamma isoform; Protein phosphatase 2C isoform gamma; Protein phosphatase magnesium-dependent 1 gamma; protein phosphatase, Mg2+/Mn2+ dependent, 1G

Gene Symbols: PPM1G

Molecular weight: 59,272 Da

Basal Isoelectric point: 4.28 **Predict pI for various phosphorylation states**

Protein information for SEC14L1

SEC14L1 a cytoplasmic protein that may play a role in innate immunity by inhibiting the antiviral RIG-I signaling pathway. Functions as a negative regulator of DDX58, the cytoplasmic sensor of viral nucleic acids. Prevents the interaction of DDX58 with MAVS, an important step in signal propagation. May also regulate the SLC18A3 and SLC5A7 cholinergic transporters. Belongs to the SEC14 cytosolic factor family. Three isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Chromosomal Location of Human Ortholog: 17q25.2-q25.3

Cellular Component: cytoplasm; Golgi apparatus; integral component of membrane

Molecular Function: protein binding

Biological Process: choline transport; innate immune response; regulation of molecular function

Reference #: Q92503 (UniProtKB)

Alt. Names/Synonyms: DKFZp686C06176; PRELID4A; S14L1; SEC14-like 1 (S. cerevisiae); SEC14-like protein 1; SEC14L; SEC14L1

Gene Symbols: SEC14L1

Molecular weight: 81,250 Da

Basal Isoelectric point: 6.01 **Predict pI for various phosphorylation states**

Protein information for P4HA2

P4HA2 Catalyzes the post-translational formation of 4- hydroxyproline in -Xaa-Pro-Gly- sequences in collagens and other proteins. Belongs to the P4HA family. 2 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Amino Acid Metabolism - arginine and proline; EC 1.14.11.2; Oxidoreductase; Secreted; Secreted, signal peptide

Chromosomal Location of Human Ortholog: **5q31.1**

Cellular Component: cytosol; endoplasmic reticulum; endoplasmic reticulum lumen; intracellular membrane-bound organelle; nucleoplasm **view less <**

Molecular Function: electron transfer activity; iron ion binding; L-ascorbic acid binding; oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen; procollagen-proline 4-dioxygenase activity **view less <**

Biological Process: peptidyl-proline hydroxylation

Disease: **Myopia 25, Autosomal Dominant**

Reference #: **O15460** (UniProtKB)

Alt. Names/Synonyms: 4-PH alpha 2; 4-PH alpha-2; C-P4Halpha(II); collagen prolyl 4-hydroxylase alpha(II); P4HA2; procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide II; Procollagen-proline,2-oxoglutarate-4-dioxygenase subunit alpha-2; Prolyl 4-hydroxylase subunit alpha-2; prolyl 4-hydroxylase, alpha polypeptide II

Gene Symbols: P4HA2

Molecular weight: 60,902 Da

Basal Isoelectric point: 5.49 **Predict pI for various phosphorylation states**

Protein information for CTSD

CTSD Acid protease active in intracellular protein breakdown. Involved in the pathogenesis of several diseases such as breast cancer and possibly Alzheimer disease. Consists of a light chain and a heavy chain. Belongs to the peptidase A1 family. Note: This description may include information from UniProtKB.

Protein type: Autophagy; EC 3.4.23.5; Motility/polarity/chemotaxis; Protease

Chromosomal Location of Human Ortholog: **11p15.5**

Cellular Component: extracellular matrix; extracellular region; extracellular space; lysosomal lumen; lysosome; melanoso... **view more >**

Molecular Function: aspartic-type endopeptidase activity; cysteine-type endopeptidase activity; protein binding; serine-... **view more >**

Biological Process: antigen processing and presentation of exogenous peptide antigen via MHC class II; autophagy; collag... **view more >**

Disease: **Ceroid Lipofuscinosis, Neuronal, 10**

Reference #: **P07339** (UniProtKB)

Alt. Names/Synonyms: CATD; Cathepsin D; Cathepsin D heavy chain; Cathepsin D light chain; ceroid-lipofuscinosis, neuronal 10; CLN10; CPSD; CTSD; lysosomal aspartyl peptidase; lysosomal aspartyl protease; MGC2311

Gene Symbols: CTSD

Molecular weight: 44,552 Da

Basal Isoelectric point: 6.1 **Predict pI for various phosphorylation states**

Protein information for ITPRIP

ITPRIP Enhances Ca(2+)-mediated inhibition of inositol 1,4,5- triphosphate receptor (ITPR) Ca(2+) release. Belongs to the ITPRIP family. Note: This description may include information from UniProtKB.

Chromosomal Location of Human Ortholog: 10q25.1

Cellular Component: membrane; plasma membrane

Reference #: Q8IWB1 (UniProtKB)

Alt. Names/Synonyms: bA127L20; bA127L20.2; DANGER; inositol 1,4,5-triphosphate receptor interacting protein; Inositol 1,4,5-triphosphate receptor-interacting protein; Inositol 1,4,5-trisphosphate receptor-interacting protein; IPRI; ITPRIP; KIAA1754; Protein DANGER

Gene Symbols: ITPRIP

Molecular weight: 62,060 Da

Basal Isoelectric point: 5.58 **Predict pI for various phosphorylation states**

Protein information for PILRA

PILRA Paired receptors consist of highly related activating and inhibitory receptors and are widely involved in the regulation of the immune system. PILRA is thought to act as a cellular signaling inhibitory receptor by recruiting cytoplasmic phosphatases like PTPN6/SHP-1 and PTPN11/SHP-2 via their SH2 domains that block signal transduction through dephosphorylation of signaling molecules. Receptor for PIANP. 4 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Membrane protein, integral; Receptor, misc.

Chromosomal Location of Human Ortholog: 7q22.1

Cellular Component: integral component of membrane; plasma membrane

Molecular Function: MHC class I protein binding; protein binding

Biological Process: regulation of immune response; signal transduction; transmembrane receptor protein tyrosine kinase signaling pathway; viral process **view less <**

Reference #: Q9UKJ1 (UniProtKB)

Alt. Names/Synonyms: Cell surface receptor FDF03; FDF03; Inhibitory receptor PILR-alpha; inhibitory receptor PILRalpha; paired immunoglobin-like receptor alpha; paired immunoglobin-like type 2 receptor alpha; paired immunoglobulin-like receptor alpha; Paired immunoglobulin-like type 2 receptor alpha; PILRA

Gene Symbols: PILRA

Molecular weight: 34,005 Da

Basal Isoelectric point: 10.22 **Predict pI for various phosphorylation states**

Protein information for ZDHHC9

ZDHHC9 The ZDHHC9-GOLGA7 complex is a palmitoyltransferase specific for HRAS and NRAS. Defects in ZDHHC9 are the cause of mental retardation syndromic X-linked ZDHHC9-related (MRXSZ). A disorder characterized by significantly sub-average general intellectual functioning associated with impairments in adaptative behavior and manifested during the developmental period. Some patients have marfanoid habitus as an additional feature. Belongs to the DHHC palmitoyltransferase family. ERF2/ZDHHC9 subfamily. Note: This description may include information from UniProtKB.

Protein type: EC 2.3.1.225; Membrane protein, integral; Membrane protein, multi-pass; Transferase

Chromosomal Location of Human Ortholog: **Xq26.1**

Cellular Component: cytosol; endoplasmic reticulum; endoplasmic reticulum membrane; Golgi apparatus; integral component of membrane; intrinsic component of Golgi membrane **view less <**

Molecular Function: palmitoyltransferase activity; protein binding; protein-cysteine S-palmitoyltransferase activity; Ras palmitoyltransferase activity **view less <**

Biological Process: peptidyl-L-cysteine S-palmitoylation; protein palmitoylation

Disease: **Mental Retardation, X-linked, Syndromic, Raymond Type**

Reference #: **Q9Y397** (UniProtKB)

Alt. Names/Synonyms: antigen MMSA-1; Asp-His-His-Cys domain containing protein 9; CGI-89; CXorf11; DHHC-9; DHHC9; MMSA-1; MRXSZ; Palmitoyltransferase ZDHHC9; ZDHC9; ZDHHC10; ZDHHC9; Zinc finger DHHC domain-containing protein 9; Zinc finger protein 379; zinc finger, DHHC domain containing 10; zinc finger, DHHC domain containing 9; zinc finger, DHHC-type containing 10; zinc finger, DHHC-type containing 9; ZNF379; ZNF380

Gene Symbols: ZDHHC9

Molecular weight: 40,916 Da

Basal Isoelectric point: 8.12 **Predict pI for various phosphorylation states**

Protein information for PDLIM7

PDLIM7 May function as a scaffold on which the coordinated assembly of proteins can occur. May play a role as an adapter that, via its PDZ domain, localizes LIM-binding proteins to actin filaments of both skeletal muscle and nonmuscle tissues. Involved in both of the two fundamental mechanisms of bone formation, direct bone formation e.g (embryonic flat bones mandible and cranium), and endochondral bone formation e.g (embryonic long bone development). Plays a role during fracture repair. Involved in BMP6 signaling pathway. 6 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Cytoskeletal

Chromosomal Location of Human Ortholog: **5q35.3**

Cellular Component: actin cytoskeleton; cell-cell adherens junction; cytosol; focal adhesion; nucleus; ruffle; stress fiber **view less <**

Molecular Function: metal ion binding; protein binding

Biological Process: actin cytoskeleton organization; axon guidance; ossification; positive regulation of osteoblast differentiation; receptor-mediated endocytosis **view less <**

Reference #: **Q9NR12** (UniProtKB)

Alt. Names/Synonyms: 1110003B01Rik; ENIGMA; LIM domain protein; LIM mineralization protein; LMP; LMP1; PDLI7; PDLIM7; PDZ and LIM domain 7 (enigma); PDZ and LIM domain protein 7; Protein enigma

Gene Symbols: PDLIM7

Molecular weight: 49,845 Da

Basal Isoelectric point: 8.76 **Predict pI for various phosphorylation states**

Protein information for SLC3A2

SLC3A2 a transmembrane solute carrier protein that transports L-type amino acids and regulates intracellular calcium levels. Involved in sodium-independent, high-affinity transport of large neutral amino acids such as phenylalanine, tyrosine, leucine, arginine and tryptophan. Involved in guiding and targeting of LAT1 and LAT2 to the plasma membrane. When associated with SLC7A6 or SLC7A7 acts as an arginine/glutamine exchanger, following an antiport mechanism for amino acid transport, influencing arginine release in exchange for extracellular amino acids. Plays a role in nitric oxide synthesis in human umbilical vein endothelial cells (HUVECs) via transport of L-arginine. Required for normal and neoplastic cell growth. When associated with SLC7A5, is also involved in the transport of L-DOPA across the blood-brain barrier, and that of thyroid hormones triiodothyronine (T3) and thyroxine (T4) across the cell membrane in tissues such as placenta. Involved in the uptake of methylmercury (MeHg) when administered as the L-cysteine or D,L-homocysteine complexes, and hence plays a role in metal ion homeostasis and toxicity. When associated with SLC7A5 or SLC7A8, involved in the cellular activity of small molecular weight nitrosothiols, via the stereoselective transport of L- nitrosocysteine (L-CNSO) across the transmembrane. Together with ICAM1, regulates the transport activity LAT2 in polarized intestinal cells, by generating and delivering intracellular signals. When associated with SLC7A5, plays an important role in transporting L-leucine from the circulating blood to the retina across the inner blood-retinal barrier. Disulfide-linked heterodimer of a glycosylated heavy chain and a non-glycosylated light chain (SLC7A5, SLC7A6, SLCA7A7, SLC7A8, SLC7A10 or SLCA7A11). Colocalizes with cadherins. Interacts with FAM57A and ICAM1. Constitutively and specifically associates with beta-1 integrins (alpha-2/beta-1, alpha-3/beta-1, alpha-5/beta-1 and alpha-6/beta-1), but minimally with alpha-4/beta-1. Expression is induced in resting peripheral blood T- lymphocytes following PHA stimulation. Expression increases at the time of maximal DNA synthesi, in fibroblasts stimulated to divide. Expression and the uptake of leucine is stimulated in mononuclear, cytotrophoblast-like choriocarcinoma cells by combined treatment with PMA and calcium ionophore. Expressed ubiquitously in all tissues tested with highest levels detected in kidney, placenta and testis and weakest level in thymus. During gestation, expression in the placenta was significantly stronger at full-term than at the mid-trimester stage. Belongs to the SLC3A transporter family. 4 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Membrane protein, integral; Transporter

Chromosomal Location of Human Ortholog: 11q12.3

Cellular Component: apical plasma membrane; cell surface; cytosol; integral component of membrane; melanosome; membrane; nucleus; plasma membrane **view less <**

Molecular Function: cadherin binding; calcium:sodium antiporter activity; catalytic activity; double-stranded RNA binding; neutral amino acid transmembrane transporter activity; protein binding; RNA binding **view less <**

Biological Process: amino acid transport; calcium ion transport; carbohydrate metabolic process; cell growth; leukocyte migration; response to exogenous dsRNA; tryptophan transport **view less <**

Reference #: P08195 (UniProtKB)

Alt. Names/Synonyms: 4F2; 4F2 cell-surface antigen heavy chain; 4F2 heavy chain antigen; 4F2hc; 4T2HC; antigen defined by monoclonal antibody 4F2, heavy chain; antigen identified by monoclonal antibodies 4F2, TRA1.10, TROP4, and T43; CD98; CD98 heavy chain; CD98HC; heavy chain; Lymphocyte activation antigen 4F2 large subunit; MDU1; monoclonal antibody 44D7; NACAE; SLC3A2; solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2; solute carrier family 3, member 2

Gene Symbols: SLC3A2

Molecular weight: 67,994 Da

Basal Isoelectric point: 4.89 **Predict pI for various phosphorylation states**

Protein information for RNASE1

RNASE1 Endonuclease that catalyzes the cleavage of RNA on the 3' side of pyrimidine nucleotides. Acts on single stranded and double stranded RNA. Belongs to the pancreatic ribonuclease family. Note: This description may include information from UniProtKB.

Protein type: EC 3.1.27.5; Ribonuclease; Secreted; Secreted, signal peptide

Chromosomal Location of Human Ortholog: 14q11.2

Cellular Component:

Molecular Function: endonuclease activity; nucleic acid binding; protein binding; ribonuclease A activity; ribonuclease activity **view less <**

Biological Process: defense response to virus

Reference #: P07998 (UniProtKB)

Alt. Names/Synonyms: HP-RNase; MGC12408; RIB-1; RIB1; Ribonuclease 1; Ribonuclease A; Ribonuclease pancreatic; ribonuclease, RNase A family, 1 (pancreatic); RNAS1; RNase 1; RNase A; RNase UpI-1; RNASE1; RNS1

Gene Symbols: RNASE1

Molecular weight: 17,644 Da

Basal Isoelectric point: 9.1 **Predict pI for various phosphorylation states**

Protein information for IKZF3

Aiolos a transcription factor of the ikaros C2H2-type zinc-finger protein family. Plays an important role in the regulation of lymphocyte differentiation. Deletions in Aiolos have been observed in a subset of pre-B-cell acute lymphoblastic leukemia (B-ALL) cases. 6 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: C2H2-type zinc finger protein; DNA-binding; Transcription factor

Chromosomal Location of Human Ortholog: 17q12-q21.1

Cellular Component: cytoplasm; cytosol; nucleoplasm; nucleus; plasma membrane

Molecular Function: DNA binding transcription factor activity; identical protein binding; metal ion binding; protein binding; protein heterodimerization activity; protein homodimerization activity; sequence-specific DNA binding; transcription regulatory region DNA binding **view less <**

Biological Process: B cell activation; mesoderm development; positive regulation of transcription from RNA polymerase II promoter; regulation of apoptosis; regulation of B cell differentiation; regulation of B cell proliferation; regulation of lymphocyte differentiation; regulation of transcription from RNA polymerase II promoter; transcription from RNA polymerase II promoter **view less <**

Reference #: Q9UKT9 (UniProtKB)

Alt. Names/Synonyms: AIO; AIOLOS; IKAROS family zinc finger 3 (Aiolos); Ikaros family zinc finger protein 3; IKZF3; Zinc finger protein Aiolos; zinc finger protein, subfamily 1A, 3 (Aiolos); ZNFN1A3

Gene Symbols: IKZF3

Molecular weight: 58,023 Da

Basal Isoelectric point: 6.11 **Predict pI for various phosphorylation states**

Protein information for SIGLEC1

SIGLEC1 Acts as an endocytic receptor mediating clathrin dependent endocytosis. Macrophage-restricted adhesion molecule that mediates sialic-acid dependent binding to lymphocytes, including granulocytes, monocytes, natural killer cells, B-cells and CD8 T-cells. Preferentially binds to alpha-2,3-linked sialic acid. Binds to SPN/CD43 on T-cells. May play a role in hemopoiesis. Belongs to the immunoglobulin superfamily. SIGLEC (sialic acid binding Ig-like lectin) family. 3 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Membrane protein, integral

Chromosomal Location of Human Ortholog: 20p13

Cellular Component: extracellular region; integral component of membrane; plasma membrane

Molecular Function: carbohydrate binding

Biological Process: cell-matrix adhesion; endocytosis; inflammatory response

Reference #: Q9BZZ2 (UniProtKB)

Alt. Names/Synonyms: CD169; dJ1009E24.1; DKFZp667F058; FLJ00051; FLJ00055; FLJ00073; FLJ00411; FLJ32150; sialic acid binding Ig-like lectin 1, sialoadhesin; Sialic acid-binding Ig-like lectin 1; sialic acid-binding immunoglobulin-like lectin 1; Sialoadhesin; Siglec-1; SIGLEC1; SN

Gene Symbols: SIGLEC1

Molecular weight: 182,624 Da

Basal Isoelectric point: 6.15 **Predict pI for various phosphorylation states**

Protein information for SLC6A9

GlyT1 Terminates the action of glycine by its high affinity sodium-dependent reuptake into presynaptic terminals. May play a role in regulation of glycine levels in NMDA receptor-mediated neurotransmission. Belongs to the sodium:neurotransmitter symporter (SNF) (TC 2.A.22) family. SLC6A9 subfamily. 3 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Membrane protein, integral; Membrane protein, multi-pass; Transporter, SLC family

Chromosomal Location of Human Ortholog: 1p34.1

Cellular Component: integral component of plasma membrane; membrane; plasma membrane

Molecular Function: glycine:sodium symporter activity; neurotransmitter:sodium symporter activity; transporter activity

Biological Process: transport

Disease: Glycine Encephalopathy With Normal Serum Glycine

Reference #: P48067 (UniProtKB)

Alt. Names/Synonyms: DKFZp547A1118; GlyT-1; GlyT1; SC6A9; SLC6A9; Sodium- and chloride-dependent glycine transporter 1; solute carrier family 6 (neurotransmitter transporter, glycine), member 9; Solute carrier family 6 member 9

Gene Symbols: SLC6A9

Molecular weight: 78,260 Da

Basal Isoelectric point: 8.21 Predict pI for various phosphorylation states

Protein information for LPAR1

LPAR1 Receptor for lysophosphatidic acid (LPA), a mediator of diverse cellular activities. Seems to be coupled to the G(i)/G(o), G(12)/G(13), and G(q) families of heteromeric G proteins. Stimulates phospholipase C (PLC) activity in a manner that is dependent on RALA activation. Belongs to the G-protein coupled receptor 1 family. Note: This description may include information from UniProtKB.

Protein type: GPCR, family 1; Membrane protein, integral; Membrane protein, multi-pass; Receptor, GPCR

Chromosomal Location of Human Ortholog: 9q31.3

Cellular Component: cell soma; cell surface; dendritic shaft; dendritic spine; endocytic vesicle; endosome; integral component of plasma membrane; plasma membrane view less <

Molecular Function: G-protein alpha-subunit binding; G-protein coupled receptor activity; PDZ domain binding; protein binding view less <

Biological Process: activation of MAPK activity; activation of phospholipase C activity; bleb formation; cerebellum development; corpus callosum development; elevation of cytosolic calcium ion concentration; elevation of cytosolic calcium ion concentration during G-protein signaling, coupled to IP3 second messenger (phospholipase C activating); G-protein coupled receptor signaling pathway; G-protein signaling, adenylate cyclase inhibiting pathway; myelination; negative regulation of cAMP biosynthetic process; oligodendrocyte development; optic nerve development; positive regulation of apoptosis; positive regulation of I-kappaB kinase/NF-kappaB signaling; positive regulation of MAPK cascade; positive regulation of Rho protein signal transduction; positive regulation of stress fiber formation; regulation of cell shape view less <

Reference #: Q92633 (UniProtKB)

Alt. Names/Synonyms: edg-2; EDG2; endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2; Gpcr26; GPR26; LPA receptor 1; LPA-1; LPA1; LPAR1; Lysophosphatidic acid receptor 1; Lysophosphatidic acid receptor Edg-2; Mrec1.3; rec.1.3; ventricular zone gene 1; vzg-1; VZG1

Gene Symbols: LPAR1

Molecular weight: 41,109 Da

Basal Isoelectric point: 8.87 Predict pI for various phosphorylation states

Protein information for CXCL12

CXCL12 Chemoattractant active on T-lymphocytes, monocytes, but not neutrophils. Activates the C-X-C chemokine receptor CXCR4 to induce a rapid and transient rise in the level of intracellular calcium ions and chemotaxis. Also binds to another C-X-C chemokine receptor CXCR7, which activates the beta-arrestin pathway and acts as a scavenger receptor for SDF-1. SDF-1-beta(3-72) and SDF-1- alpha(3-67) show a reduced chemotactic activity. Binding to cell surface proteoglycans seems to inhibit formation of SDF-1-alpha(3- 67) and thus to preserve activity on local sites. Acts as a positive regulator of monocyte migration and a negative regulator of monocyte adhesion via the LYN kinase. Stimulates migration of monocytes and T-lymphocytes through its receptors, CXCR4 and CXCR7, and decreases monocyte adherence to surfaces coated with ICAM-1, a ligand for beta-2 integrins. SDF1A/CXCR4 signaling axis inhibits beta-2 integrin LFA-1 mediated adhesion of monocytes to ICAM-1 through LYN kinase. Inhibits CXCR4-mediated infection by T- cell line-adapted HIV-1. Plays a protective role after myocardial infarction. Induces down-regulation and internalization of CXCR7 expressed in various cells. Has several critical functions during embryonic development; required for B-cell lymphopoiesis, myelopoiesis in bone marrow and heart ventricular septum formation. Monomer or homodimer; in equilibrium. Dimer formation is induced by non acidic pH and the presence of multivalent anions, and by binding to CXCR4 or heparin. Monomeric form is required for full chemotactic activity and resistance to ischemia/reperfusion injury, whereas the dimeric form acts as a partial agonist of CXCR4, stimulating Ca2+ mobilization but with no chemotactic activity and instead acts as a selective antagonist that blocks chemotaxis induced by the monomeric form. Interacts with the N- terminus of CXCR7. Isoform Alpha and isoform Beta are ubiquitously expressed, with highest levels detected in liver, pancreas and spleen. Isoform Gamma is mainly expressed in heart, with weak expression detected in several other tissues. Isoform Delta, isoform Epsilon and isoform Theta have highest expression levels in pancreas, with lower levels detected in heart, kidney, liver and spleen. Belongs to the intercrine alpha (chemokine CxC) family. 6 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Cell development/differentiation; Chemokine; Motility/polarity/chemotaxis; Secreted; Secreted, signal peptide

Chromosomal Location of Human Ortholog: **10q11.21**

Cellular Component: cytoplasm; external side of plasma membrane; extracellular region

Molecular Function: chemoattractant activity; chemokine activity; chemokine receptor binding; CXCR chemokine receptor binding; growth factor activity; receptor binding **view less <**

Biological Process: adult locomotory behavior; axon guidance; blood circulation; cell adhesion; cellular calcium ion homeostasis; chemotaxis; defense response; detection of mechanical stimulus involved in sensory perception of pain; detection of temperature stimulus involved in sensory perception of pain; G-protein coupled receptor signaling pathway; immune response; induction of positive chemotaxis; neuron migration; organ regeneration; positive chemotaxis; positive regulation of axon extension involved in axon guidance; positive regulation of cell adhesion; positive regulation of dopamine secretion; positive regulation of endothelial cell proliferation; regulation of actin polymerization and/or depolymerization; response to heat; response to hypoxia; response to peptide hormone; response to radiation; response to virus; signal transduction; telencephalon cell migration **view less <**

Disease: **Human Immunodeficiency Virus Type 1, Susceptibility To**

Reference #: **P48061** (UniProtKB)

Alt. Names/Synonyms: C-X-C motif chemokine 12; chemokine (C-X-C motif) ligand 12; chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1); CXCL12; hIRH; hSDF-1; Intercrine reduced in hepatomas; IRH; PBSF; Pre-B cell growth-stimulating factor; SCYB12; SDF-1; SDF-1-alpha(3-67); SDF-1-beta(3-72); SDF-1a; SDF-1b; SDF1; SDF1A; SDF1B; Stromal cell-derived factor 1; stromal cell-derived factor 1 delta; stromal cell-derived factor 1 gamma; stromal cell-derived factor 1a; TLSF-a; TLSF-b; TPAR1

Gene Symbols: CXCL12

Molecular weight: 10,666 Da

Basal Isoelectric point: 9.92 **Predict pI for various phosphorylation states**

Protein information for C17orf81/ELP5

DERP6 Acts as subunit of the RNA polymerase II elongator complex, which is a histone acetyltransferase component of the RNA polymerase II (Pol II) holoenzyme and is involved in transcriptional elongation. Elongator may play a role in chromatin remodeling and is involved in acetylation of histones H3 and probably H4. Involved in cell migration. May be involved in TP53-mediated transcriptional regulation. Belongs to the ELP5 family. 4 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Chromosomal Location of Human Ortholog: 17p13.1

Cellular Component: cytoplasm; cytosol; Elongator holoenzyme complex; nucleus

Molecular Function: protein binding

Biological Process: positive regulation of cell migration; regulation of transcription, DNA-templated; transcription, DN... view more >

Reference #: Q8TE02 (UniProtKB)

Alt. Names/Synonyms: C17orf81; chromosome 17 open reading frame 81; dermal papilla derived protein 6; Dermal papilla-derived protein 6; DERP6; Elongator complex protein 5; HSPC002; MST071; MSTP071; S-phase 2 protein

Gene Symbols: ELP5

Molecular weight: 34,841 Da

Basal Isoelectric point: 4.81 **Predict pI for various phosphorylation states**

Protein information for ECM2

ECM2 Promotes matrix assembly and cell adhesiveness. Belongs to the small leucine-rich proteoglycan (SLRP) family. SLRP class I subfamily. 2 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Cell adhesion; Secreted; Secreted, signal peptide

Chromosomal Location of Human Ortholog: 9q22.31

Cellular Component: extracellular space; proteinaceous extracellular matrix

Molecular Function: heparin binding; integrin binding

Biological Process: axonogenesis; cell-matrix adhesion; extracellular matrix organization

Reference #: O94769 (UniProtKB)

Alt. Names/Synonyms: ECM2; Extracellular matrix protein 2; extracellular matrix protein 2, female organ and adipocyte specific; Matrix glycoprotein SC1/ECM2; MGC126355; MGC126356

Gene Symbols: ECM2

Molecular weight: 79,789 Da

Basal Isoelectric point: 5.3 **Predict pI for various phosphorylation states**

Protein information for FBLN1

FBLN1 Incorporated into fibronectin-containing matrix fibers. May play a role in cell adhesion and migration along protein fibers within the extracellular matrix (ECM). Could be important for certain developmental processes and contribute to the supramolecular organization of ECM architecture, in particular to those of basement membranes. Has been implicated in a role in cellular transformation and tumor invasion, it appears to be a tumor suppressor. May play a role in haemostasis and thrombosis owing to its ability to bind fibrinogen and incorporate into clots. Could play a significant role in modulating the neurotrophic activities of APP, particularly soluble APP. A chromosomal aberration involving FBLN1 is found in a complex type of synpolydactyly referred to as 3/3-prime/4 synpolydactyly associated with metacarpal and metatarsal synostoses. Reciprocal translocation t(12;22)(p11.2;q13.3) with C12orf2. Fibroblasts derived from a patient with synpolydactyly displayed alterations in the level of isoform D splice variant incorporated into the ECM and secreted into the conditioned culture medium. By contrast, the expression of isoform C was not perturbed in the patients fibroblasts. Furthermore, no aberrant polypeptides were detected in extracts of cultured patients fibroblasts. The translocation t(12;22) may result in haploinsufficiency of the isoform D splice variant, which could lead to the observed limb malformation. Elevated expression and altered processing of FBLN1 protein is associated with human breast cancer. Belongs to the fibulin family. 4 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Extracellular matrix; Secreted; Secreted, signal peptide

Chromosomal Location of Human Ortholog: 22q13.31

Cellular Component: basement membrane; elastic fiber; extracellular matrix; extracellular region; extracellular space; fibrinogen complex view less <

Molecular Function: calcium ion binding; extracellular matrix structural constituent; fibronectin binding; identical protein binding; protease activator activity; protein C-terminus binding view less <

Biological Process: blood coagulation, fibrin clot formation; embryo implantation; extracellular matrix organization; integrin-mediated signaling pathway; negative regulation of cell adhesion; negative regulation of cell motility; negative regulation of ERK1 and ERK2 cascade; negative regulation of protein phosphorylation; negative regulation of stem cell proliferation; negative regulation of substrate adhesion-dependent cell spreading; positive regulation of peptidase activity; viral process view less <

Disease: Synpolydactyly 2

Reference #: P23142 (UniProtKB)

Alt. Names/Synonyms: FBLN; FBLN1; FIBL-1; FIBL1; fibulin 1; Fibulin-1

Gene Symbols: FBLN1

Molecular weight: 77,214 Da

Basal Isoelectric point: 5.07 Predict pI for various phosphorylation states

Protein information for APCDD1

APCDD1 a single-pass type I membrane protein. Probably plays a role in colorectal tumorigenesis. Directly activated downstream targets of the WNT/beta-catenin pathway. Transcriptional target of beta-catenin/T-cell (Tcf) factor and may enhance cellular proliferation of tumor cells. Down-regulated by wild-type the adenomatosis polyposis coli (APC) protein. Elevated in many primary colon cancer tissues. Note: This description may include information from UniProtKB.

Protein type: Membrane protein, integral

Chromosomal Location of Human Ortholog: 18p11.22

Cellular Component: integral component of plasma membrane

Molecular Function: Wnt-protein binding

Biological Process: astrocyte cell migration; hair follicle development; negative regulation of Wnt signaling pathway; Wnt signaling pathway view less <

Disease: Hypotrichosis 1

Reference #: Q8J025 (UniProtKB)

Alt. Names/Synonyms: adenomatosis polyposis coli down-regulated 1; Adenomatosis polyposis coli down-regulated 1 protein; APCD1; APCDD1; B7323; DRAPC1; FP7019; Protein APCDD1

Gene Symbols: APCDD1

Molecular weight: 58,797 Da

Basal Isoelectric point: 8.43 Predict pI for various phosphorylation states

Protein information for ANXA5

ANXA5 a calcium/phospholipid-binding protein and an anticoagulant protein that acts as an indirect inhibitor of the thromboplastin-specific complex, which is involved in the blood coagulation cascade. The binding of labeled ANXA5 to phosphatidylserine is used as a marker of apoptosis. Annexins are a family of structurally related proteins whose common property is calcium-dependent binding to phospholipids. There are at least ten different annexins in mammalian species. Annexins do not contain signal peptides, yet some annexins (A1, A2 and A5) appear to be secreted in a physiologically regulated fashion. Note: This description may include information from UniProtKB.

Protein type: Apoptosis; Calcium-binding; Inhibitor; Lipid-binding

Chromosomal Location of Human Ortholog: 4q27

Cellular Component: cytoplasm; cytosol; endothelial microparticle; external side of plasma membrane; extracellular region; intracellular view less <

Molecular Function: calcium ion binding; calcium-dependent phospholipid binding; phospholipase inhibitor activity; phospholipid binding; protein binding view less <

Biological Process: blood coagulation; negative regulation of apoptosis; negative regulation of catalytic activity; negative regulation of coagulation; platelet degranulation; response to organic substance; signal transduction view less <

Disease: Pregnancy Loss, Recurrent, Susceptibility To, 3

Reference #: P08758 (UniProtKB)

Alt. Names/Synonyms: Anchorin CII; Annexin A5; Annexin V; Annexin-5; ANX5; anxa 5; ANXA5; Calphobindin I; CBP-I; Endonexin II; ENX2; Lipocortin V; PAP-I; Placental anticoagulant protein 4; Placental anticoagulant protein I; PP4; Thromboplastin inhibitor; VAC-alpha; Vascular anticoagulant-alpha

Gene Symbols: ANXA5

Molecular weight: 35,937 Da

Basal Isoelectric point: 4.94 Predict pI for various phosphorylation states

Protein information for INKA1

FAM212A Belongs to the FAM212 family. Note: This description may include information from UniProtKB.

Chromosomal Location of Human Ortholog: 3p21.31

Cellular Component: cytoplasm; nucleus

Molecular Function: protein binding; protein kinase binding; protein serine/threonine kinase inhibitor activity

Biological Process: neural tube development

Reference #: **Q96EL1** (UniProtKB)

Alt. Names/Synonyms: C3orf54; FAM212A; family with sequence similarity 212, member A; LOC389119; MGC20416; PAK4-inhibitor INKA1

Gene Symbols: FAM212A

Molecular weight: 31,328 Da

Basal Isoelectric point: 5.28 **Predict pI for various phosphorylation states**

Protein information for TMEM98

TMEM98 Belongs to the TMEM98 family. Note: This description may include information from UniProtKB.

Protein type: Membrane protein, integral

Chromosomal Location of Human Ortholog: 17q11.2

Cellular Component: endoplasmic reticulum; integral component of membrane

Disease: **Nanophthalmos 4**

Reference #: **Q9Y2Y6** (UniProtKB)

Alt. Names/Synonyms: DKFZp564K1964; Protein TADA1; TMEM98; TMM98; Transmembrane protein 98

Gene Symbols: TMEM98

Molecular weight: 24,611 Da

Basal Isoelectric point: 4.81 **Predict pI for various phosphorylation states**

Protein information for FEZ1

FEZ1 May be involved in axonal outgrowth as component of the network of molecules that regulate cellular morphology and axon guidance machinery. Able to restore partial locomotion and axonal fasciculation to C.elegans unc-76 mutants in germline transformation experiments. May participate in the transport of mitochondria and other cargos along microtubules. Belongs to the zygin family. 2 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Chromosomal Location of Human Ortholog: 11q24.2

Cellular Component: axon; cell soma; centrosome; cytoplasm; dendrite; Golgi apparatus; growth cone; microtubule; mitochondrion; plasma membrane **view less <**

Molecular Function: gamma-tubulin binding; protein binding; protein kinase C binding; protein N-terminus binding

Biological Process: axon guidance; cell adhesion; cellular response to growth factor stimulus; establishment of cell polarity; establishment of mitochondrion localization; hippocampus development; nervous system development; positive regulation of anterograde axonal transport of mitochondrion; positive regulation of neuron projection development **view less <**

Reference #: Q99689 (UniProtKB)

Alt. Names/Synonyms: fasciculation and elongation protein zeta 1 (zygin I); Fasciculation and elongation protein zeta-1; FEZ1; Zygin I; Zygin-1

Gene Symbols: FEZ1

Molecular weight: 45,119 Da

Basal Isoelectric point: 4.24 **Predict pI for various phosphorylation states**

Protein information for THAP4

THAP4 2 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Chromosomal Location of Human Ortholog: 2q37.3

Cellular Component:

Molecular Function: DNA binding; metal ion binding

Reference #: Q8WY91 (UniProtKB)

Alt. Names/Synonyms: CGI-36; PP238; THAP domain containing 4; THAP domain-containing protein 4; THAP4

Gene Symbols: THAP4

Molecular weight: 62,890 Da

Basal Isoelectric point: 9.39 **Predict pI for various phosphorylation states**

Protein information for TNFRSF19

TNFRSF19 Can mediate activation of JNK and NF-kappa-B. May promote caspase-independent cell death. 2 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Membrane protein, integral

Chromosomal Location of Human Ortholog: 13q12.12

Cellular Component: integral component of membrane; intracellular; plasma membrane

Molecular Function: receptor activity; tumor necrosis factor-activated receptor activity

Biological Process: apoptosis; hair follicle development; JNK cascade; positive regulation of I-kappaB kinase/NF-kappaB signaling; positive regulation of JNK cascade; tissue development; tumor necrosis factor-mediated signaling pathway view less <

Reference #: Q9NS68 (UniProtKB)

Alt. Names/Synonyms: TAJ; TAJ-alpha; TNFRSF19; TNR19; Toxicity and JNK inducer; TRADE; TROY; Tumor necrosis factor receptor superfamily member 19; tumor necrosis factor receptor superfamily, member 19

Gene Symbols: TNFRSF19

Molecular weight: 46,015 Da

Basal Isoelectric point: 5.28 Predict pI for various phosphorylation states

Protein information for NBR1

NBR1 Acts probably as a receptor for selective autophagosomal degradation of ubiquitinated targets. Homooligomer and heterooligomer. Interacts with TRIM55. Interacts with SQSTM1, titin/TTN, RNF29, USP8, SQSTM1, MAP1LC3A, MAP1LC3B, MAP1LC3C, GABARAP, GABARAPL1 and GABARAPL2. Binds to ubiquitin and ubiquitinated proteins. 2 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Ubiquitin conjugating system

Chromosomal Location of Human Ortholog: 17q21.31

Cellular Component: autophagic vacuole; cytosol; intracellular membrane-bound organelle; late endosome; lysosome; M band; membrane; mitochondrion; nuclear body; nucleoplasm view less <

Molecular Function: mitogen-activated protein kinase binding; protein binding; ubiquitin binding; zinc ion binding

Biological Process: macroautophagy; negative regulation of osteoblast differentiation; protein oligomerization; regulation of bone mineralization; regulation of stress-activated MAPK cascade view less <

Reference #: Q14596 (UniProtKB)

Alt. Names/Synonyms: 1A1-3B; 1A13B; Cell migration-inducing gene 19 protein; FLJ55359; FLJ98272; KIAA0049; M17S2; Membrane component chromosome 17 surface marker 2; MIG19; migration-inducing protein 19; NBR1; neighbor of BRCA1 gene 1; Neighbor of BRCA1 gene 1 protein; Next to BRCA1 gene 1 protein; Protein 1A1-3B

Gene Symbols: NBR1

Molecular weight: 107,413 Da

Basal Isoelectric point: 5.03 Predict pI for various phosphorylation states

Protein information for WISP2

WISP2 May play an important role in modulating bone turnover. Promotes the adhesion of osteoblast cells and inhibits the binding of fibrinogen to integrin receptors. In addition, inhibits osteocalcin production. Belongs to the CCN family. Note: This description may include information from UniProtKB.

Protein type: Secreted; Secreted, signal peptide

Chromosomal Location of Human Ortholog: 20q13.12

Cellular Component: extracellular space; proteinaceous extracellular matrix

Molecular Function: heparin binding; insulin-like growth factor binding; integrin binding

Biological Process: cell adhesion; cell-cell signaling; regulation of cell growth; signal transduction

Reference #: O76076 (UniProtKB)

Alt. Names/Synonyms: CCN family member 5; CCN5; Connective tissue growth factor-like protein; Connective tissue growth factor-related protein 58; CT58; CTGF-L; CTGFL; RP11-445H22.2; UNQ228/PRO261; WISP-2; WISP2; WNT1 inducible signaling pathway protein 2; WNT1-inducible-signaling pathway protein 2

Gene Symbols: WISP2

Molecular weight: 26,825 Da

Basal Isoelectric point: 8.32 **Predict pI for various phosphorylation states**

Protein information for CNRIP1

CNRIP1 Isoform 1 suppresses cannabinoid receptor CNR1-mediated tonic inhibition of voltage-gated calcium channels. Isoform 2 does not have this effect. Belongs to the CNRIP family. 2 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Unknown function

Chromosomal Location of Human Ortholog: 2p14

Cellular Component: cytoplasm; myelin sheath; plasma membrane

Molecular Function: protein C-terminus binding; type 1 cannabinoid receptor binding

Reference #: Q96F85 (UniProtKB)

Alt. Names/Synonyms: C2orf32; cannabinoid receptor CB1-interacting protein 1; cannabinoid receptor interacting protein 1; CB1 cannabinoid receptor-interacting protein 1; CNRIP1; CNRP1; CRIP-1; CRIP1; DKFZp566K1924

Gene Symbols: CNRIP1

Molecular weight: 18,648 Da

Basal Isoelectric point: 7.72 **Predict pI for various phosphorylation states**

Protein information for ITSN1

ITSN1 Adapter protein that may provide indirect link between the endocytic membrane traffic and the actin assembly machinery. May regulate the formation of clathrin-coated vesicles. Involved in endocytosis of integrin beta-1 (ITGB1) and transferrin receptor (TFR); internalization of ITGB1 as DAB2-dependent cargo but not TFR may involve association with DAB2. Isoform 1 could be involved in brain-specific synaptic vesicle recycling. Inhibits ARHGAP31 activity toward RAC1. 4 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Adaptor/scaffold; Calcium-binding; GEF; GEF, Rac/Rho

Chromosomal Location of Human Ortholog: 21q22.11

Cellular Component: cell junction; clathrin-coated pit; cytosol; endocytic vesicle; lamellipodium; neuron projection; plasma membrane; synapse view less <

Molecular Function: calcium ion binding; guanyl-nucleotide exchange factor activity; kinase activator activity; proline-rich region binding; protein binding; protein complex scaffold activity; Rho guanyl-nucleotide exchange factor activity view less <

Biological Process: ephrin receptor signaling pathway; G-protein coupled receptor signaling pathway; negative regulation of neuron apoptosis; positive regulation of apoptosis; positive regulation of phosphorylation; positive regulation of protein kinase B signaling; regulation of Rho protein signal transduction; regulation of small GTPase mediated signal transduction; small GTPase mediated signal transduction; synaptic vesicle endocytosis; viral process view less <

Reference #: Q15811 (UniProtKB)

Alt. Names/Synonyms: human intersectin-SH3 domain-containing protein SH3P17; intersectin 1 (SH3 domain protein); intersectin 1 short form variant 11; intersectin 1 short form variant 3; intersectin short variant 12; Intersectin-1; ITSN; ITSN1; MGC134948; MGC134949; SH3 domain protein-1A; SH3 domain-containing protein 1A; SH3D1A; SH3P17; Src homology 3 domain-containing protein

Gene Symbols: ITSN1

Molecular weight: 195,422 Da

Basal Isoelectric point: 7.76 **Predict pI for various phosphorylation states**

Protein information for C1S

C1S C1s B chain is a serine protease that combines with C1q and C1r to form C1, the first component of the classical pathway of the complement system. C1r activates C1s so that it can, in turn, activate C2 and C4. Defects in C1S are the cause of complement component C1s deficiency (C1SD). A rare defect resulting in C1 deficiency and impaired activation of the complement classical pathway. C1 deficiency generally leads to severe immune complex disease with features of systemic lupus erythematosus and glomerulonephritis. Belongs to the peptidase S1 family. Note: This description may include information from UniProtKB.

Protein type: EC 3.4.21.42; Protease

Chromosomal Location of Human Ortholog: 12p13.31

Cellular Component: extracellular region

Molecular Function: calcium ion binding; identical protein binding; protein binding; serine-type endopeptidase activity

Biological Process: complement activation; complement activation, classical pathway; complement activation, lectin pathway; proteolysis; regulation of complement activation view less <

Disease: Complement Component C1s Deficiency; Ehlers-danlos Syndrome, Periodontal Type, 2

Reference #: P09871 (UniProtKB)

Alt. Names/Synonyms: basic proline-rich peptide IB-1; C1 esterase; C1S; Complement C1s subcomponent; Complement C1s subcomponent heavy chain; Complement C1s subcomponent light chain; Complement component 1 subcomponent s; complement component 1, s subcomponent; FLJ44757

Gene Symbols: C1S

Molecular weight: 76,684 Da

Basal Isoelectric point: 4.86 **Predict pI for various phosphorylation states**

Protein information for TNC

TNC Extracellular matrix protein implicated in guidance of migrating neurons as well as axons during development, synaptic plasticity as well as neuronal regeneration. Promotes neurite outgrowth from cortical neurons grown on a monolayer of astrocytes. Ligand for integrins alpha-8/beta-1, alpha-9/beta-1, alpha-V/beta-3 and alpha-V/beta-6. Belongs to the tenascin family. 6 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: Motility/polarity/chemotaxis; Secreted; Secreted, signal peptide

Chromosomal Location of Human Ortholog: 9q33.1

Cellular Component: basement membrane; endoplasmic reticulum lumen; extracellular matrix; extracellular region; extracellular space; focal adhesion; membrane view less <

Molecular Function: syndecan binding

Biological Process: axon regeneration in the peripheral nervous system; cell adhesion; cellular protein metabolic process; extracellular matrix organization; negative regulation of cell adhesion; neuromuscular junction development; odontogenesis of dentine-containing teeth; osteoblast differentiation; positive regulation of cell proliferation; post-translational protein modification; response to ethanol; response to mechanical stimulus; response to wounding; wound healing view less <

Disease: Deafness, Autosomal Dominant 56

Reference #: P24821 (UniProtKB)

Alt. Names/Synonyms: 150-225; Cytotactin; Glioma-associated-extracellular matrix antigen; GMEM; GP; GP 150-225; Hexabrachion; hexabrachion (tenascin); HXB; JI; MGC167029; Myotendinous antigen; Neuronectin; TENA; Tenascin; tenascin C; Tenascin-C; tenascin-C isoform 14/AD1/16; TN; TN-C; TNC

Gene Symbols: TNC

Molecular weight: 240,853 Da

Basal Isoelectric point: 4.79 Predict pI for various phosphorylation states

Protein information for PRRX1

PRRX1 Acts as a transcriptional regulator of muscle creatine kinase (MCK) and so has a role in the establishment of diverse mesodermal muscle types. The protein binds to an A/T-rich element in the muscle creatine enhancer. Defects in PRRX1 are the cause of agnathia-otocephaly complex (AGOTC). AGOTC is a rare condition characterized by mandibular hypoplasia or agnathia, ventromedial auricular malposition (melotia) and/or auricular fusion (synotia), and microstomia with oroglossal hypoplasia or aglossia. Holoprosencephaly is the most commonly identified association, but skeletal, genitourinary, and cardiovascular anomalies, and situs inversus have been reported. The disorder is almost always lethal. Belongs to the paired homeobox family. 2 isoforms of the human protein are produced by alternative splicing. Note: This description may include information from UniProtKB.

Protein type: DNA-binding

Chromosomal Location of Human Ortholog: 1q24.2

Cellular Component: nucleolus; nucleus

Molecular Function: transcription coactivator activity

Biological Process: artery morphogenesis; cartilage development; embryonic cranial skeleton morphogenesis; embryonic limb morphogenesis; inner ear morphogenesis; middle ear morphogenesis; negative regulation of transcription from RNA polymerase II promoter; neuron fate determination; palate development; positive regulation of mesenchymal cell proliferation; positive regulation of smoothened signaling pathway; positive regulation of transcription from RNA polymerase II promoter view less <

Disease: Agnathia-otocephaly Complex

Reference #: P54821 (UniProtKB)

Alt. Names/Synonyms: Homeobox protein PHOX1; paired mesoderm homeo box 1; paired mesoderm homeobox 1 isoform pmx-1b; Paired mesoderm homeobox protein 1; paired related homeobox 1; Paired-related homeobox protein 1; PHOX1; PMX1; PRRX1; PRX-1; PRX1

Gene Symbols: PRRX1

Molecular weight: 27,296 Da

Basal Isoelectric point: 9.48 Predict pI for various phosphorylation states

Protein information for STAT3

STAT3 transcription factor of the STAT family. Phosphorylated and activated by receptor-associated kinases downstream of many cytokines and growth-factor receptors. Constitutively active in a number of human tumors. Forms homo- or heterodimers that translocate into the nucleus where they regulate transcription. Two alternatively spliced isoforms have been described. Note: This description may include information from UniProtKB.

Protein type: DNA-binding; Motility/polarity/chemotaxis; Nuclear receptor co-regulator; Transcription factor

Chromosomal Location of Human Ortholog: 17q21.2

Cellular Component: cytoplasm; cytosol; mitochondrial inner membrane; nuclear chromatin; nucleoplasm; nucleus; plasma membrane; RNA polymerase II transcription factor complex **view less <**

Molecular Function: CCR5 chemokine receptor binding; chromatin DNA binding; DNA binding; DNA binding transcription factor activity; glucocorticoid receptor binding; identical protein binding; ligand-dependent nuclear receptor activity; protein binding; protein dimerization activity; protein homodimerization activity; protein kinase binding; protein phosphatase binding; transcription factor binding; transcription regulatory region DNA binding **view less <**

Biological Process: acute-phase response; aging; astrocyte differentiation; cell proliferation; cellular response to cytokine stimulus; cellular response to hormone stimulus; cellular response to leptin stimulus; cellular response to organic cyclic compound; cytokine-mediated signaling pathway; eating behavior; eye photoreceptor cell differentiation; glucose homeostasis; inflammatory response; interleukin-21-mediated signaling pathway; interleukin-23-mediated signaling pathway; interleukin-27-mediated signaling pathway; interleukin-6-mediated signaling pathway; interleukin-7-mediated signaling pathway; intracellular receptor signaling pathway; JAK-STAT cascade; leptin-mediated signaling pathway; miRNA mediated inhibition of translation; mRNA transcription from RNA polymerase II promoter; negative regulation of apoptosis; negative regulation of cell proliferation; negative regulation of glycolytic process; negative regulation of hydrogen peroxide biosynthetic process; negative regulation of neuron death; negative regulation of neuron migration; negative regulation of stem cell differentiation; negative regulation of transcription from RNA polymerase II promoter; nervous system development; phosphorylation; positive regulation of angiogenesis; positive regulation of ATP biosynthetic process; positive regulation of erythrocyte differentiation; positive regulation of gene expression; positive regulation of gene silencing by miRNA; positive regulation of growth factor dependent skeletal muscle satellite cell proliferation; positive regulation of metalloendopeptidase activity; positive regulation of Notch signaling pathway; positive regulation of transcription from RNA polymerase II promoter; positive regulation of transcription, DNA-templated; positive regulation of tyrosine phosphorylation of STAT protein; positive regulation of vascular endothelial cell proliferation; protein import into nucleus; radial glial cell differentiation; regulation of cell cycle; regulation of mitochondrial membrane permeability; regulation of multicellular organism growth; regulation of transcription from RNA polymerase II promoter; regulation of transcription, DNA-templated; response to estradiol; response to ethanol; response to leptin; sexual reproduction; signal transduction; somatic stem cell maintenance; T-helper 17 cell lineage commitment; thermoregulation; viral process **view less <**

Disease: Autoimmune Disease, Multisystem, Infantile-onset; Hyper-ige Recurrent Infection Syndrome, Autosomal Dominant

Reference #: P40763 (UniProtKB)

Alt. Names/Synonyms: Acute-phase response factor; APRF; DNA-binding protein APRF; FLJ20882; HIES; MGC16063; Signal transducer and activator of transcription 3; signal transducer and activator of transcription 3 (acute-phase response factor); STAT3

Gene Symbols: STAT3

Molecular weight: 88,068 Da

Basal Isoelectric point: 5.94 **Predict pI for various phosphorylation states**

Protein information for TSC1

TSC1 tuberous sclerosis protein 1 (TSC1). May have tumor suppressor activity. Together with tuberin (TSC2), inhibits mammalian target of rapamycin (mTOR)-mediated signaling to eukaryotic initiation factor 4E-binding protein 1 (4E-BP1) and ribosomal protein S6 kinase. Regulated by the phosphatidylinositol 3-kinase/Akt pathway and phosphorylation of tuberin. Phosphorylated by cyclin-dependent kinase 1/cyclin B during the cell cycle. Acts as a co-chaperone for HSP90AA1 facilitating HSP90AA1 chaperoning of protein clients such as kinases, TSC2 and glucocorticoid receptor NR3C1. Increases ATP binding to HSP90AA1 and inhibits HSP90AA1 ATPase activity. Competes with the activating co-chaperone AHSA1 for binding to HSP90AA1, thereby providing a reciprocal regulatory mechanism for chaperoning of client proteins. Recruits TSC2 to HSP90AA1 and stabilizes TSC2. Note: This description may include information from UniProtKB.

Protein type: Motility/polarity/chemotaxis; Tumor suppressor

Chromosomal Location of Human Ortholog: 9q34.13

Cellular Component: actin filament; cell cortex; cytoplasm; cytosol; growth cone; lamellipodium; lipid droplet; membrane; nucleus; perinuclear region of cytoplasm; plasma membrane; TSC1-TSC2 complex view less <

Molecular Function: ATPase inhibitor activity; chaperone binding; GTPase activating protein binding; Hsp70 protein binding; Hsp90 protein binding; protein binding; protein N-terminus binding view less <

Biological Process: adaptive immune response; adult locomotory behavior; cardiac muscle cell differentiation; cell projection organization; cell-matrix adhesion; cerebral cortex development; glucose import; hippocampus development; kidney development; memory T cell differentiation; myelination; negative regulation of ATPase activity; negative regulation of cell proliferation; negative regulation of cell size; negative regulation of insulin receptor signaling pathway; negative regulation of macroautophagy; negative regulation of neuron projection development; negative regulation of TOR signaling; negative regulation of translation; neural tube closure; positive regulation of focal adhesion assembly; positive regulation of macroautophagy; positive regulation of stress fiber formation; potassium ion transport; protein heterooligomerization; protein stabilization; regulation of actin cytoskeleton organization; regulation of cell cycle; regulation of cell-matrix adhesion; regulation of neuron death; regulation of phosphoprotein phosphatase activity; regulation of protein kinase activity; regulation of stress fiber assembly; regulation of translation; response to insulin stimulus; rRNA export from nucleus; synapse organization view less <

Disease: Focal Cortical Dysplasia Of Taylor; Lymphangioleiomyomatosis; Tuberous Sclerosis 1

Reference #: Q92574 (UniProtKB)

Alt. Names/Synonyms: Hamartin; KIAA0243; LAM; MGC86987; TSC; TSC1; tuberous sclerosis 1; Tuberous sclerosis 1 protein; tumor suppressor

Gene Symbols: TSC1

Molecular weight: 129,767 Da

Basal Isoelectric point: 6.02 Predict pI for various phosphorylation states

Protein information for TWIST2

TWIST2 Binds to the E-box consensus sequence 5'-CANNTG-3' as a heterodimer and inhibits transcriptional activation by MYOD1, MYOG, MEF2A and MEF2C. Also represses expression of proinflammatory cytokines such as TNFA and IL1B. Involved in postnatal glycogen storage and energy metabolism. Inhibits the premature or ectopic differentiation of preosteoblast cells during osteogenesis, possibly by changing the internal signal transduction response of osteoblasts to external growth factors. Defects in TWIST2 are the cause of Setleis syndrome (SETLEISS). A focal facial dermal dysplasia characterized by distinctive bitemporal scar-like depressions resembling forceps marks, and additional facial features, including a coarse and leonine appearance, absent eyelashes on both lids or multiple rows on the upper lids, absent Meibomian glands, slanted eyebrows, chin clefting, and hypo- or hyperpigmentation of the skin. Histologically, the bitemporal lesion is an ectodermal dysplasia with near absence of subcutaneous fat, suggesting insufficient migration of neural crest cells into the frontonasal process and the first branchial arch. Note: This description may include information from UniProtKB.

Chromosomal Location of Human Ortholog: 2q37.3

Cellular Component: cytoplasm; nucleolus; nucleus

Molecular Function: DNA binding; protein binding; protein dimerization activity

Biological Process: cell differentiation; multicellular organism development; negative regulation of apoptosis; negative regulation of osteoblast differentiation; negative regulation of transcription, DNA-dependent; transcription, DNA-dependent view less <

Disease: Ablepharon-macrostomia Syndrome; Barber-say Syndrome; Focal Facial Dermal Dysplasia 3, Setleis Type

Reference #: Q8WVJ9 (UniProtKB)

Alt. Names/Synonyms: BHLHA39; Class A basic helix-loop-helix protein 39; Dermis-expressed protein 1; Dermo-1; DERMO1; MGC117334; twist homolog 2 (Drosophila); twist-related bHLH protein Dermo1; Twist-related protein 2; TWIST2; TWST2

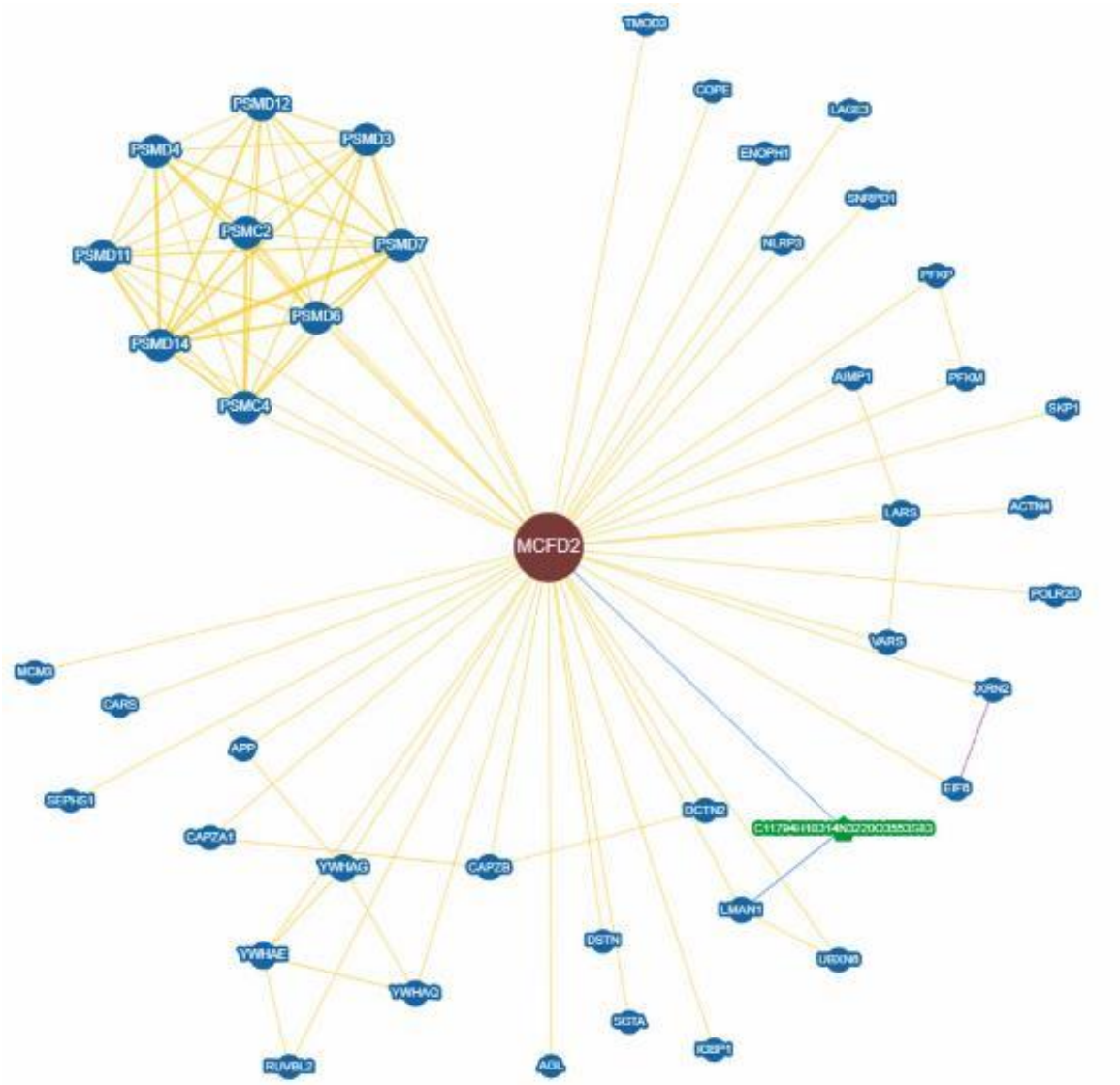
Gene Symbols: TWIST2

Molecular weight: 18,124 Da

Basal Isoelectric point: 9.51 Predict pI for various phosphorylation states

APPENDIX C: SUPPLEMENTARY INFORMATION FOR CHAPTER 4

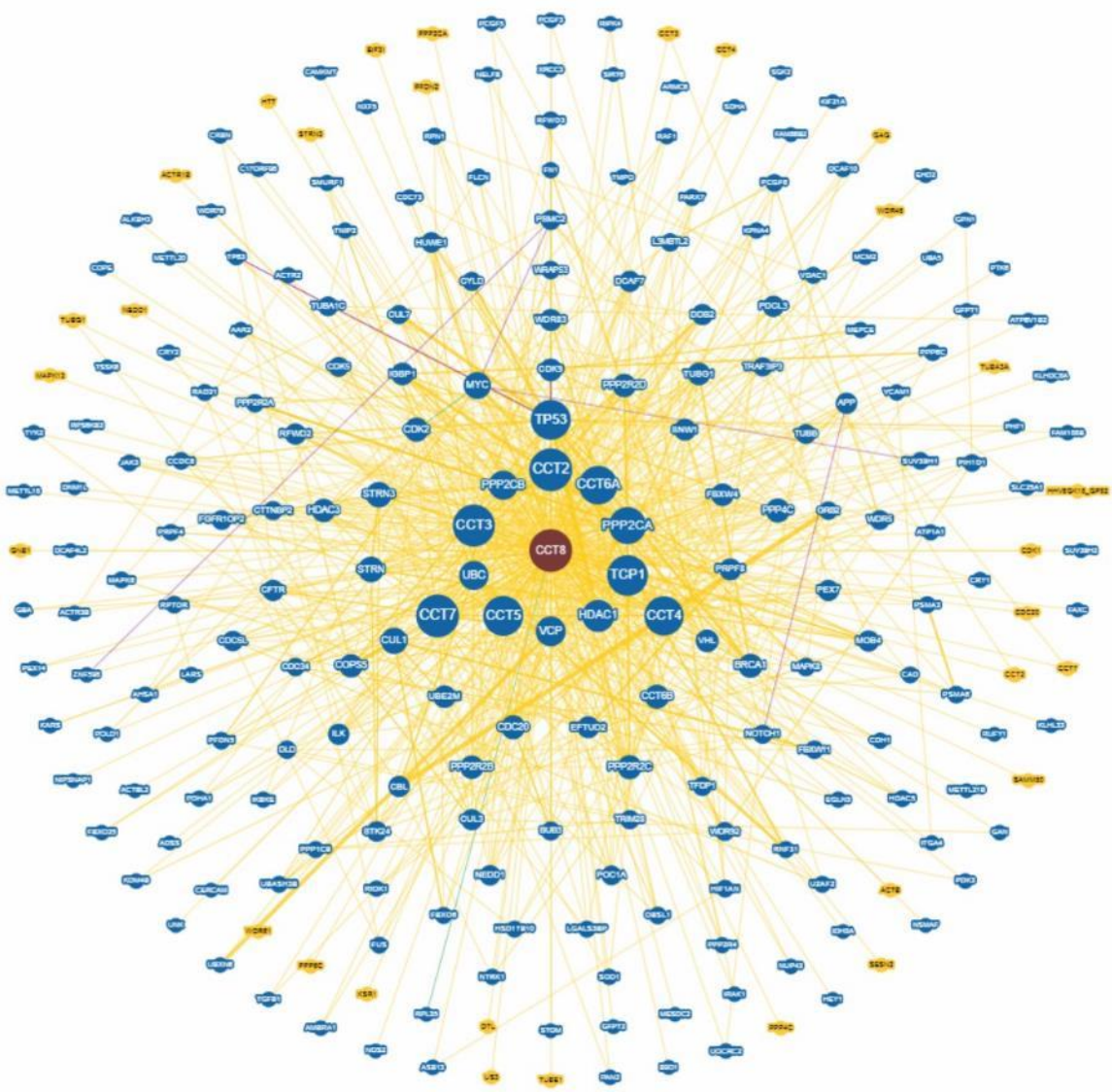
The genetic and protein interaction data for MCFD2 obtained from The Biological General Repository for Interaction Datasets (BioGRID) public database. MCFD2 shows 44 nodes and 92 edges.



https://thebiogrid.org/124712/summary/homo-sapiens/mcfd2.html

Follow the link above then click on the "**Network**" tab in the switch view options for better resolution or more information.

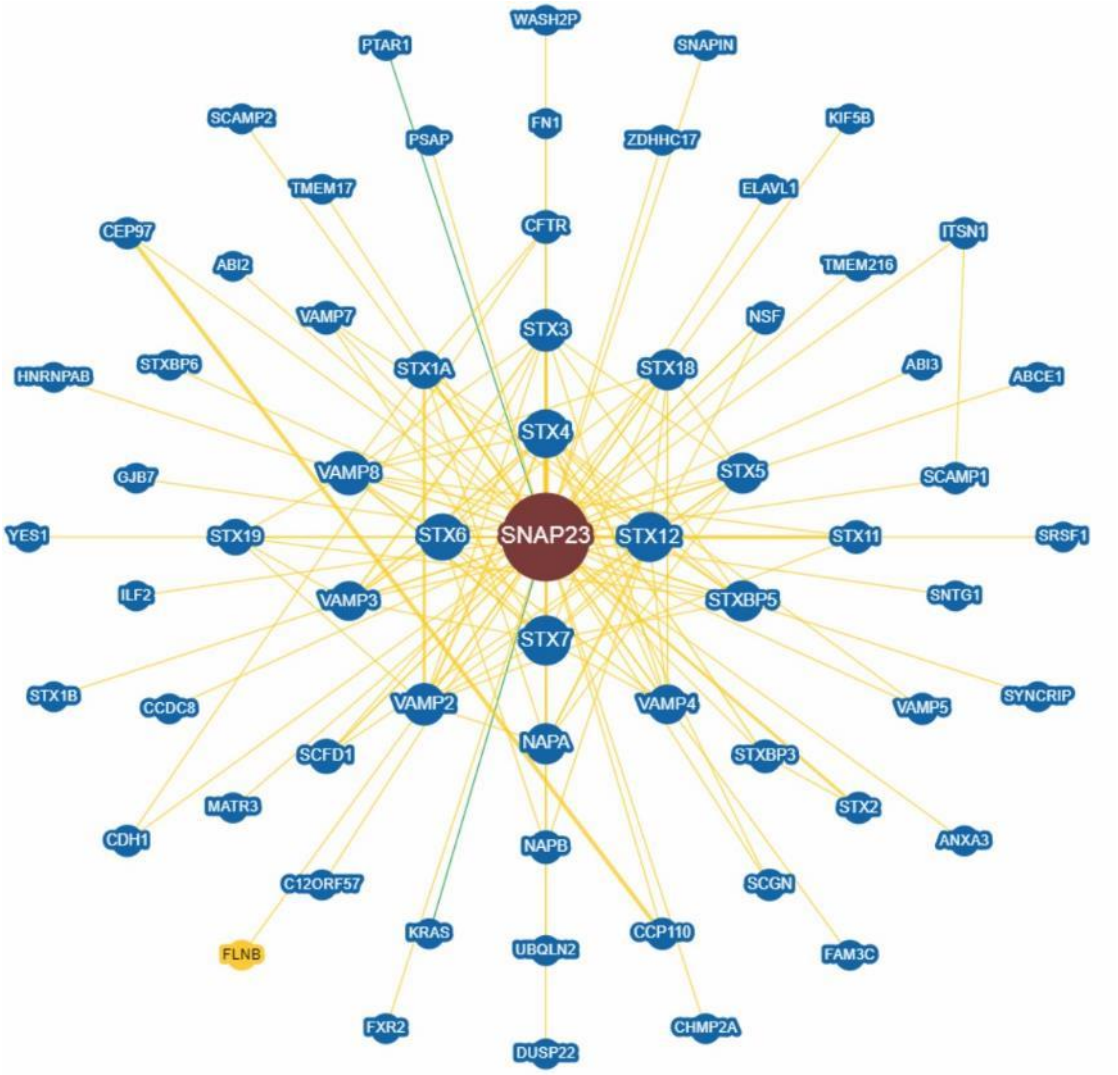
The genetic and protein interaction data for MCFD2 obtained from The Biological General Repository for Interaction Datasets (BioGRID) public database. CCT8 shows 237 nodes and 820 edges.



https://thebiogrid.org/115933/summary/homo-sapiens/cct8.html

Follow the link above then click on the "**Network**" tab in the switch view options for better resolution or more information.

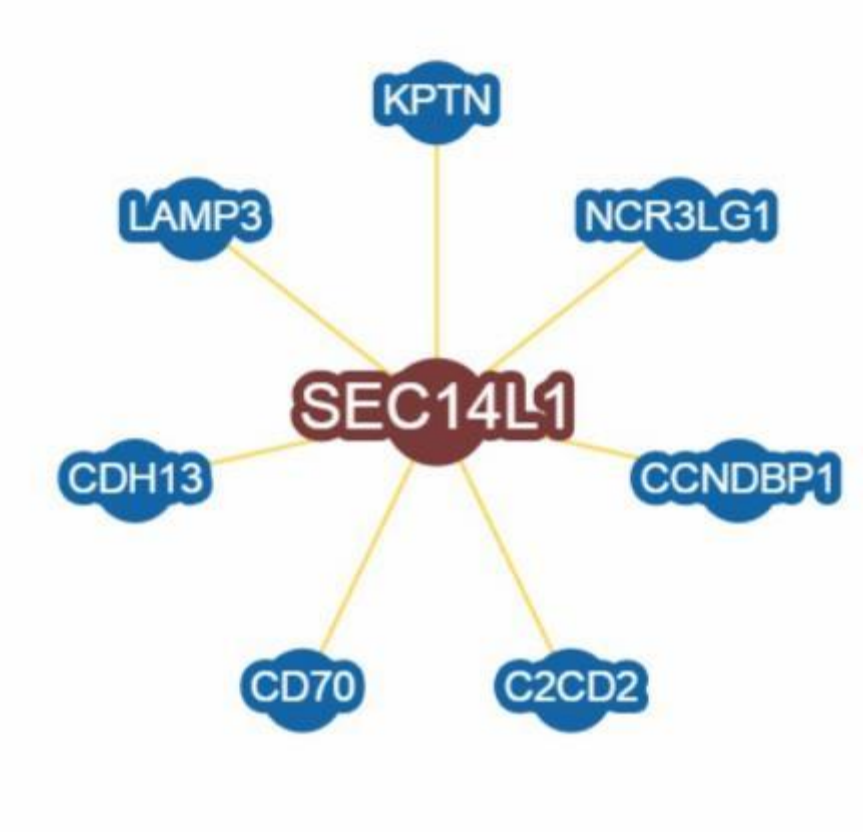
The genetic and protein interaction data for MCFD2 obtained from The Biological General Repository for Interaction Datasets (BioGRID) public database. SNAP23 shows 65 nodes and 137 edges.



https://thebiogrid.org/114303/summary/homo-sapiens/snap23.html

Follow the link above then click on the "**Network**" tab in the switch view options for better resolution or more information.

The genetic and protein interaction data for MCFD2 obtained from The Biological General Repository for Interaction Datasets (BioGRID) public database. SEC14L1 shows 8 nodes and 7 edges.

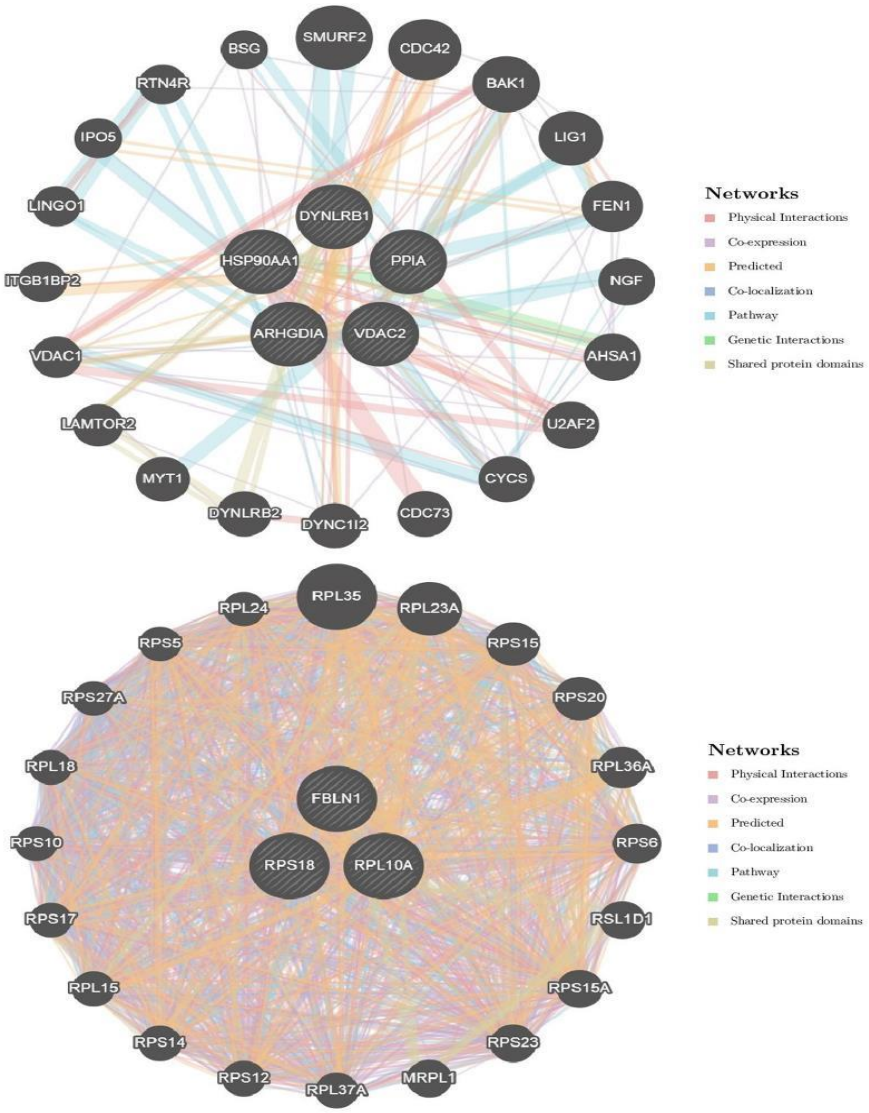


https://thebiogrid.org/112297/summary/homo-sapiens/sec14l1.html

Follow the link above then click on the "**Network**" tab in the switch view options for better resolution or more information.

APPENDIX D: SUPPLEMENTARY INFORMATION FOR CHAPTER 5

The predicted functional associations between the LVI-differentially expressed protiens/genes that are identified with RNA microarray and HPLC/MS experiments.



By utilising the online accessible GeneMania Cytoscape application, protein and genetic interactions, pathways, co-expression, co-localisation and protein domain similarity can be predicted. "GeneMANIA is actively developed at the University of Toronto, in the Donnelly Centre for Cellular and Biomolecular Research, in the labs of Gary Bader and Quaid Morris. GeneMANIA development was originally funded by Genome Canada, through the Ontario Genomics Institute (2007-OGI-TD-05) and is now funded by the Ontario Ministry of Research and Innovation. GeneMANIA is also supported by the National Resource for Network Biology (P41 GM103504)."