

University of Nottingham
UK | CHINA | MALAYSIA

Integrative Molecular and Pathologic Profiling of Triple Negative Breast Cancer

Mansour Abdualaziz Alsaleem

B.Sc., M.Sc.

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

January 2021

School of Medicine

ABSTRACT

Introduction

Breast cancer (BC) is a heterogeneous disease presenting with variable morphologies, molecular profiles, clinical behaviours and responses to therapy. Triple negative breast cancer (TNBC), characterised by the absent expression of the Oestrogen Receptor (ER), Progesterone Receptor (PR) and HER2 overexpression, is an aggressive class of BC for which targeted therapies are lacking. Patients with TNBC comprise approximately 17% of BC patients and show variable responses to chemotherapy, which reflects the transcriptomic and genomic heterogeneity of the disease. In this study, we hypothesised that deciphering TNBC molecular heterogeneity can pave the way for the identification of novel therapeutic targets and the refinement of classification. We investigated large, well annotated cohorts of TNBC using a variety of molecular techniques to: 1) identify genetic predictors of clinical outcomes in TNBC, 2) investigate the mechanistic and prognostic role of novel targets in TNBC including periplakin (PPL), 3) further evaluate the mechanisms of cell adhesion in TNBC.

Methods

Next generation sequencing was applied to a group of TNBC cases (n=126) with long term follow-up data. Data mining using the supervised neural artificial network (ANN) was performed to identify genes or pathways associated with distant metastasis-free survival (DMFS) and breast cancer-specific survival (BCSS). The prognostic ability of these genes was validated using multiple public domain datasets such as the Breast Cancer Gene-Expression Miner v4. 0 and Genotype 2 outcome datasets. Multivariate Cox regression analysis identified a prognostic gene signature that was independently associated with poor prognosis. Tissue microarrays (TMAs) were constructed to examine the protein expression using immunohistochemistry (IHC).

PPL was identified by ANN analysis as a strong prognostic marker with respect to DMFS and BCSS. The prognostic ability of PPL was also validated using multiple public domain datasets such as the METABRIC, Breast Cancer Gene-Expression Miner v4.0, KM plotter and Genotype 2 outcome datasets. This was followed by the evaluation of its protein expression using IHC. The transient knockdown of PPL using siRNA was used to assess its role in

proliferation, migration and invasive capabilities of TNBC cell lines (MB-MDA 231 and MB-MDA 436).

E-cadherin protein expression was determined using TMAs (n=813). IHC for E-cadherin staining using full-face sections was carried out to confirm the identity of E-cadherin and its role on the global gene expression profile. RNA-seq analysis was completed on invasive ductal carcinoma of the breast (IDC) tumours with either E-cadherin positive or negative/low E-cadherin expression. Genes that differentially expressed E-cadherin were assessed using the Robina implementation of EdgeR.

Results

ANN identified high expression of a set of genes (*ACSM4, SPDYC, PPL, DCTN1-AS1, RP11-29H23.5, RPS10P18, AC079305.10* and *PAXBP1-AS1)* that can individually be used as indicators of poor prognosis and shorter outcomes in TNBC ($p<0.05$). Adjusting for clinicopathological factors including the patient age, grade, nodal stage, tumour size and lymphovascular invasion using multivariate Cox regression analysis yielded a two-gene prognostic signature (*ACSM4* and *SPDYC)* which was associated with poor prognosis ($p<0.05$), independent of other prognostic variables. To validate these findings, the protein expression of these two genes was significantly associated with patient outcomes in both an independent and a combined manner ($p<0.05$).

High *PPL* mRNA expression was significantly associated with shorter BCSS and DMFS (both $P<0.01$). *PPL* mRNA expression was also an independent, poor prognostic predictor for BCSS and DMFS regardless of clinicopathological factors including patient age, grade, nodal stage, tumour size and lymphovascular invasion. High PPL protein expression was significantly associated with poor BCSS and DMFS, independent of other clinicopathological factors (both $P<0.01$). Modulating expression levels of PPL conferred reduced cell proliferation and disrupted TNBC cell migration and invasion.

Differential gene expression identified a total of 2143 genes differentially expressed between IDCs showing a reduced or loss of expression and normal or above threshold E-cadherin expression. IDCs with reduced or lost E-cadherin expression showed a dysregulation of genes regulating the Wnt signalling pathway (*FZD2, GNG5, HLTF, WNT2* and *CER1*) and the PIK3-AKT signalling pathway (*FGFR2, GNF5, GNGT1, IFNA17* and *IGF1*). Interestingly, key

genes reported to be differentially expressed between lobular and ductal tumours did not show associations with E-cadherin loss in IDC.

Conclusion

The study of independent, large cohorts of clinically well annotated TNBC, using various molecular techniques and bioinformatics approaches, provides evidence of the molecular heterogeneity of TNBC and improves our understanding of the disease. This study has identified novel specific genes and proteins that were strongly predictive of TNBC outcomes including ACSM4, SPDYC and PPL, which are potential therapeutic targets. Also, the methodological approaches applied in this study can be used to identify genetic panels predicting patients' outcomes and to investigate molecular functions in other subtypes of BC as well as different types of tumours. We provide further insights on the mechanism of the altered adhesion process and the impact of reduced or lost E-cadherin expression in TNBC. Further studies on the therapeutic values of the identified novel genes are warranted.

Peer-Reviewed Publications Arising from This Study

1- **Mansour Alsaleem**, Michael S Toss, Chitra Joseph, Mohammed Aleskandarany, Sasagu Kurozumi, Ibrahim Alshankyty, Angela Ogden, Padmashree CG Rida, Ian O Ellis, Ritu Aneja, Andrew R Green, Nigel P Mongan, Emad A Rakha. "The molecular mechanisms underlying reduced E-cadherin expression in invasive ductal carcinoma of the breast: high throughput analysis of large cohorts". Modern Pathology (2019).

2- **Mansour Alsaleem**, Graham Ball, Michael S Toss, Sara Raafat, Mohammed Aleskandarany, Chitra Joseph, Angela Ogden, Shristi Bhattarai, Padmashree CG Rida, Francesca Khani, Melissa Davis, Olivier Elemento, Ritu Aneja, Ian O Ellis, Andrew Green, Nigel P Mongan, Emad Rakha. "A novel prognostic two-gene signature for triple negative breast cancer". Modern Pathology (2020).

3- Teresa Gagliano, Kalpit Shah, Sofia Gargani, Liyan Lao, **Mansour Alsaleem**, Jianing Chen, Vasileios Ntafis, Penghan Huang, Angeliki Ditsiou, Viviana Vella, Kritika Yadav, Kamila Bienkowska, Giulia Bresciani, Kai Kang, Leping Li, Philip Carter, Graeme Benstead-Hume, Timothy O'Hanlon, Michael Dean, Frances MG Pearl, Soo-Chin Lee, Emad A Rakha, Andrew R Green, Dimitris L Kontoyiannis, Erwei Song, Justin Stebbing, Georgios Giamas. "PIK3Cδ expression by fibroblasts promotes triple-negative breast cancer progression". The Journal of Clinical Investigation (2020).

4- Angela Ogden, Shristi Bhattarai, Bikram Sahoo, Nigel P Mongan, **Mansour Alsaleem**, Andrew R Green, Mohammed Aleskandarany, Ian O Ellis, Sonal Pattni, Xiaoxian Bill Li, Carlos S Moreno, Uma Krishnamurti, Emiel A Janssen, Kristin Jonsdottir, Emad Rakha, Padmashree Rida, Ritu Aneja. "Combined HER3-EGFR score in triple-negative breast cancer provides prognostic and predictive significance superior to individual biomarkers". Scientific reports (2020).

5- **Mansour Alsaleem**, Chitra Joseph, Yousif A Kariri ,Pavan L Narasimha, Michael S Toss, Jennie Jeyapalan, Daisy B Haigh, Corinne L Woodcock, Anna E Harris, Jennifer Lothion-Roy, Mohammed Aleskandarany, Padmashree CG Rida, Ritu Aneja, Graham Ball, Ian O Ellis, Andrew Green, Nigel P Mongan, Emad Rakha. "Mechanistic and clinical evidence supports a key role for Periplakin (PPL) as a novel independent predictor of outcome in Triple Negative Breast Cancer". post-viva publication.

6- **Mansour Alsaleem**, Chitra Joseph, Michael S Toss, Helen E Bryant, Angela Cox, Ian O Ellis, Andrew Green, Nigel P Mongan, Emad Rakha. "Immunohistochemical assessment of DNA repair Mechanism in TNBC". post-viva publication.

Scientific Contributions in Peer-Reviewed Publications

1- Chitra Joseph, Athanasia Papadaki, Maryam Althobiti, **Mansour Alsaleem**, Mohammed A Aleskandarany, Emad A Rakha. "Breast cancer intratumour heterogeneity: current status and clinical implications". (Histopathology 2018).

2- Michael S Toss, Islam Miligy, Abdubaqi Al-Kawaz, **Mansour Alsaleem**, Hazem Khout, Padmashree C Rida, Ritu Aneja, Andrew R Green, Ian O Ellis, Emad A Rakha. "Prognostic significance of tumour-infiltrating lymphocytes in ductal carcinoma in situ of the breast". Modern Pathology (2018).

3- Sasagu Kurozumi, Chitra Joseph, Sultan Sonbul, Mohammed A Aleskandarany, Marian Pigera**, Mansour Alsaleem**, Sami Alsaeed, Yousif Kariri, Christopher C Nolan, Maria Diez-Rodriguez, Simon Johnston, Nigel P Mongan, Takaaki Fujii, Ken Shirabe, Stewart G Martin, Ian O Ellis, Andrew R Green, Emad A Rakha. "Clinicopathological and prognostic significance of Ras association and pleckstrin homology domains 1 (RAPH1) in breast cancer". Breast Cancer Research and Treatment (2018).

4- Sasagu Kurozumi, Chitra Joseph, Sultan Sonbul, Sami Alsaeed, Yousif Kariri, Abrar Aljohani, Sara Raafat, **Mansour Alsaleem**, Angela Ogden, Simon J Johnston, Mohammed A Aleskandarany, Takaaki Fujii, Ken Shirabe, Carlos Caldas, Ibraheem Ashankyty, Leslie Dalton, Ian O Ellis, Christine Desmedt, Andrew R Green, Nigel P Mongan, Emad A Rakha. "A key genomic subtype associated with lymphovascular invasion in invasive breast cancer". Breast Cancer Research and Treatment (2019).

5- Sasagu Kurozumi, Chitra Joseph, Sara Raafat, Sultan Sonbul, Yousif Kariri, Sami Alsaeed, Marian Pigera, **Mansour Alsaleem**, Christopher C Nolan, Simon J Johnston, Mohammed A Aleskandarany, Angela Ogden, Takaaki Fujii, Ken Shirabe, Stewart G Martin, Ibraheem Alshankyty, Nigel P Mongan, Ian O Ellis, Andrew R Green, Emad A Rakha. "Utility of ankyrin 3 as a prognostic marker in androgen-receptor-positive breast cancer". Breast Cancer Research and Treatment (2019).

6- Chitra Joseph, Sara Al-Izzi, **Mansour Alsaleem**, Sasagu Kurozumi, Michael S Toss, Maariya Arshad, Fang Qin Goh, Ibraheem M Alshankyty, Mohammed A Aleskandarany, Simak Ali, Ian O Ellis, Nigel P Mongan, Andrew R Green, Emad A Rakha. "Retinoid X receptor gamma (RXRG) is an independent prognostic biomarker in ER-positive invasive breast cancer". Breast Cancer Research and Treatment (2019).

7- Yousif A Kariri, Chitra Joseph, Sasagu Kurozumi, Michael S Toss, **Mansour Alsaleem**, Sara Raafat, Nigel P Mongan, Mohammed A Aleskandarany, Andrew R Green, Emad A Rakha. "Prognostic significance of KN motif and ankyrin repeat domains 1 (KANK1) in invasive breast cancer". Breast Cancer Research and Treatment (2020).

8- Chitra Joseph, **Mansour Alsaleem**, Nnamdi Orah, Pavan L Narasimha, Islam M Miligy, Sasagu Kurozumi, Ian O Ellis, Nigel P Mongan, Andrew R Green, Emad A Rakha. “Elevated MMP9 expression in breast cancer is a predictor of shorter patient survival”. Breast Cancer Research and Treatment (2020).

9- Emad A Rakha, **Mansour Alsaleem**, Khloud A ElSharawy, Michael S Toss, Sara Raafat, Raluca Mihai, Fayyaz A Minhas, Andrew R Green, Nasir M Rajpoot, Leslie W Dalton, Nigel P Mongan. “Visual histological assessment of morphological features reflects the underlying molecular profile in invasive breast cancer: a morphomolecular study”. Histopathology (2020).

10- Yousif A Kariri, **Mansour Alsaleem**, Chitra Joseph, Sami Alsaeed, Abrar Aljohani, Sho Shiino, Omar J. Mohammed, Michael S Toss, Andrew R Green, Emad A Rakha. “The prognostic significance of Interferon Stimulated Gene 15 (ISG15) in invasive breast cancer”. Breast Cancer Research and Treatment (2020).

11- Sasagu Kurozumi, **Mansour Alsaleem**, Cíntia J. Monteiro, Kartikeya Bhardwaj, Stacey E. P. Joosten, Takaaki Fujii, Ken Shirabe, Andrew R. Green, Ian O. Ellis, Emad A. Rakha, Nigel P. Mongan, David M. Heery, Wilbert Zwart, Steffi Oesterreich & Simon J. Johnston. “Targetable ERBB2 mutation status is an independent marker of adverse prognosis in estrogen receptor positive, ERBB2 non-amplified primary lobular breast carcinoma: a retrospective in silico analysis of public datasets”. Breast Cancer Research and Treatment (2020).

12- Chitra Joseph, **Mansour Alsaleem**, Michael S. Toss, Yousif A. Kariri, Maryam Althobiti, Sami Alsaeed, Abrar I. Aljohani, Pavan L. Narasimha, Nigel P. Mongan, Andrew R. Green, Emad A. Rakha. “The ITIM-containing Receptor: Leukocyte-associated immunoglobulin-like receptor-1 (LAIR-1) modulates immune response and confers poor prognosis in invasive breast carcinoma”. Cancers (2020).

13- Bhopal C. Mohapatra, Sameer Mirza, Aditya Bele1, C. B. Gurumurthy, Irfana Saleem, Matthew D Storck, Surendra Shukla, Siddesh Southekal, Nitish K Mishra, Kay U. Wagner, Fang Qiu, Subodh Lele, Sara Abuelmaaty, **Mansour Alsaleem**, Emad A Rakha, Chittibabu Guda, Pankaj K. Singh, Robert D. Cardiff, Hamid Band and Vimla Band. “Ecdysoneless overexpression in breast cancer is an independent prognostic marker, and its mammary specific overexpression promotes oncogenesis by regulating myc and wnt signaling pathways”. (Molecular Cancer, under review).

14- Reem Ali, Adel Alblihy, Islam Miligy, Muslim Alabdullah, **Mansour Alsaleem**, Michael S Toss, Meshael Algethami, Tarek Abdel-Fattah, Paul Moseley, Stephen Chan, Nigel P Mongan, Satya Narayan, Emad A Rakha and Srinivasan Madhusudan. "Molecular disruption of DNA polymerase β for platinum sensitization and synthetic lethality in epithelial ovarian cancers"(Oncogene, under review).

15- Adrian Sanz-Moreno; Sonia Palomeras, Kim Pedersen; Beatriz Morancho, Tomas Pascual, Patricia Galván, Sandra Benítez, Jorge Gomez-Miragaya, Marina Ciscar; Sonia Pernas, Anna Petit, Maria Teresa Soler-Monsó, Gemma Viñas, **Mansour Alsaleem**, Emad A Rakha, Andrew R Green, Patricia G Santamaria, Celine Mulder, Simone Lemeer, Joaquin Arribas, Aleix Prat, Teresa Puig and Eva Gonzalez Suarez. . "RANK signaling increases after anti-HER2 therapy contributing to the emergence of resistance in HER2-positive breast cancer". (Breast Cancer Research and Treatment, under review).

16- Sami Alsaeed, Michael S Toss, **Mansour Alsaleem**, Mohammed A Aleskandarany, Chitra Joseph, Sasagu Kurozumi, Graham Ball, Nigel P Mongan, Andrew R Green and Emad A Rakha. "The prognostic significance of Heat Shock Protein 90AA1 (HSP90α) in invasive breast cancer". (Journal of Clinical Pathology, under review).

17- Pavan Narasimha, **Mansour Alsaleem**, Maryam Althobiti, Omar Mohammed, Andrew Green, Chitra Joseph, and Emad Rakha. "CD276 Overexpression is Prognostic of Poor Survival in Invasive Breast Cancer and linked with Immune Biomarkers". post-viva publication.

18- Chitra Joseph, **Mansour Alsaleem**, Michael Toss, Sasagu Kurozumi, Sameer Mirza , Nigel Mongan , Anna Grabowska , Andrew Green , Emad Rakha. "The Role Of Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Zeta (YWHAZ) In Breast Cancer: A Key Driver Of Lymphovascular Invasion". post-viva publication).

19- Sameer Mirza, Irfana Saleem, Achyuth Kalluchi, Mohsin Raza, Dhananjaya Pal, Bhopal Mohapatra, Subodh Lele3, Fang Qiu, Lulu Yu, Zhi-Ming Zheng, Ying Zhang, **Mansour A Alsaleem**, Emad A Rakha, , Hamid Band , M Jordan Rowley , and Vimla Band "Ecdysoneless protein regulates viral and cellular mRNA splicing to promote cervical oncogenesis" post-viva publication.

20- Yousif A Kariri, **Mansour Alsaleem**, Chitra Joseph, Sami Alsaeed, Abrar Aljohani, Sho Shiino, Michael S Toss, Andrew R Green, Emad A Rakha "The prognostic significance of Cell Division Cycle Associated 5 (CDCA5) in Breast Cancer" post-viva publication.

Conference Presentations

As the main presenter:

1. Poster presentation "Deciphering the molecular landscape of E-cadherin in ductal carcinoma of the breast by next-generation sequencing." The 1st UK Interdisciplinary Breast Cancer Symposium—15th–16th January 2018, Manchester Central Convention Centre, Manchester, United Kingdom.

2. Poster presentation "Potential predictor genes for clinical outcome in triple negative breast cancer by next generation sequencing." The 1st UK Interdisciplinary Breast Cancer Symposium,15th–16th January 2018, Manchester Central Convention Centre, Manchester, United Kingdom.

3. Poster presentation " Transcriptomic Profiling of Triple Negative Breast Cancer Identifies SPDYC as a Novel Independent Predictor of Outcome ." 12th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 2 – 4 July 2019, Harrogate Convention Centre, Harrogate, North Yorkshire, United Kingdom.

4. Poster presentation "Comprehensive transcriptomic profiling reveals Perplakin as a novel independent predictor of outcome in triple-negative breast cancer." The 2nd UK Interdisciplinary Breast Cancer Symposium 2020, 27th-28th January 2020, International convention centre, Birmingham, United Kingdom.

5. Poster presentation "Targetable *ERBB2* mutation status is an independent marker of adverse prognosis in estrogen receptor positive, *ERBB2* non-amplified primary lobular breast carcinoma: validation using a novel gene signature of HER2 activation**". The 2020 San Antonio Breast Cancer Symposium®, December 8-12, 2020, The Henry B. Gonzalez Convention Center, 900 E. Market Street, San Antonio, TX 78205 USA.**

Acknowledgment

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

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

I would also like to convey my appreciation to my supervisor, Professor Nigel Mongan, for his unparalleled supervision, inspiration and huge passion for science that has guided me throughout this journey. My heartfelt appreciation goes to Professor Mongan for his extreme support, enthusiasm and for always being available to offer his advice. His rapid responses and support during the writing process are similarly appreciated.

Further thanks must go to my supervisors, Professor Ian Ellis and Dr Andrew Green, for the instrumental advice, guidance and scientific contributions they provided throughout this intense but thoroughly rewarding journey. There could not have been a superior combination of mentors and supervisors for me during my PhD studies. My gratitude to them is immense, and I hope to continue to collaborate with these outstanding scholars.

I am deeply indebted and grateful to Dr Chitra Joseph, for her invaluable friendship and mentorship during this journey. Her graceful patience, guidance and instruction in the laboratory, especially with *in vitro* work, have been instrumental to this achievement and my progress.

I would also like express my gratitude to Dr Michael Toss, whose immeasurable knowledge of pathology and friendship over these four years have encouraged me to think outside the box and accomplish comprehensive investigations.

I must also convey my thanks to the members of the Cancer and Stem Cells Division and Histopathology Department, for their technical, academic and deeply appreciated moral support.

I would also like to thank Dr Abduallah Almayouf, Meshari and Mohammad Alsaleem and Rami Alharthi. Their steadfast friendship and support have proven invaluable during my studies abroad.

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

Dedication

Words cannot express my esteemed admiration and respect for the heroes in my life, my mother, Loloah, and my father, Abdulaziz, who themselves have faced their own struggles against breast cancer and chronic myelogenous leukaemia and have shown unending strength in their battles against cancer. I know the both of you have remained strong, not only for yourselves, but for all of us. Mum and Dad, I keep pushing because of you, and you are the reason I can be brave through my struggles, hardships and pain. I love the both of you so much!

Special appreciation and thanks must be given to Hanaa, my wife. Her love, dedication and constant confidence in me have alleviated all my burdens. Your presence and consistent encouragement have been a source of motivation during our life abroad and the demanding process of completing my research. While you have struggled too, it has been a wonderful comfort and relief knowing that you were willing to support me in any way you could.

This is also for my beloved brothers, Ryan and Azam, my sister, Nouf and her children, Abdulaziz, Meshari and the latest addition to our family, Sara. Your love and continued support have always propelled me.

I would like also to dedicate this to my inspiration in academia and research, my dear uncle, Prof Yousif Alsaleem. Prof Yousif, I always remembered your words of encouragement and advocacy for tenacity throughout my adventure abroad. Your strength and perseverance ignited my career in pursuing my true passion in cancer research.

Ethical Approval

This work obtained ethics approval by the This work obtained ethics approval by the North West – Greater Manchester Central Research Ethics Committee under the title; Nottingham Health Science Biobank (NHSB), reference number 15/NW/0685. Informed consent was obtained from all individuals prior to surgery to use their tissue materials in research. All samples used in this study were pseudo-anonymized and collected prior to 2006 and stored in compliance with the UK Human Tissue Act.

List of Abbreviations

µg	Microgram
µg/ml	Microgram per millilitre
µl	Microlitre
µm	Micrometre
mm	Millimetre
l	Litre
AB	Antibody
AC	Anthracycline
ACC	Acinic Cell Carcinoma
ADCC	Adenoid Cystic Carcinoma
ADH	Atypical Ductal Hyperplasia
ALH	Atypical Lobular Hyperplasia
AMGA	Atypical Microgranular Adenosis
ANN	Artificial Neural Network
AR	Androgen Receptor
ASCO	The American Society of Clinical Oncology
ATM	Ataxia Telangiectasia Mutated
AUC	Area Under the Curve
BC	Breast Cancer
bcGenExMiner	Breast Cancer Gene Expression Miner
BCSS	Breast Cancer Specific Survival
BL1	Triple Negative Basal Like 1 Breast Cancer Tumours

BL2	Triple Negative Basal Like 2 Breast Cancer Tumours
bp	Base Pair
BRCA1	Breast Cancer Type-1 Susceptibility Protein
BRCA2	Breast Cancer Type-2 Susceptibility Protein
BSA	Bovine Serum Albumin
CAP	The College Of American Pathologists
CDH1	E Cadherin
CI	Confidence Interval
CK5/6	Cytoskeletal 5 And 6
CMF	Cyclophosphamide, Methotrexate, Fluorouracil
CNA	Copy Number Alteration
DAB	3,3’-Diaminobenzidine
DCIS	Ductal Carcinoma In Situ
DEG	Differntially Expressed Genes
DFS	Disease Free Survival
DGE	Differntial Gene Expression
DMFS	Distant Mestastasis Free Survival
DMSO	Dimethyl Sulfoxide
DPX	Dibutylphthalate Polystyrene Xylene
EGFR	Epidermal Growth Factor Receptor
EMT	Epethilal To Mesenchymal Transition
ER	Oestrogen
ERK	Kinase Signalling Pathway

ESE	Epithelial Specific Enhancer
FAC	Fluorouracil-Adriamycin -Cyclophosphamide
FEC	Fluorouracil, Epirubicin, Cyclophosphamide
FFPE	Formalin Fixed Paraffin Embedded
GATA3	GATA Binding Protein 3
GPG	Good Prognosis Group
GSEA	Gene Set Enrichment Analysis
H&E	Haematoxylin And Eosin
H2O2	Hydrogen Peroxide
HCl	Hydrochloric Acid
HER2	Human Epidermal Growth Factor 2
HIER	Heat-Induced Epitope Retrieval
HR	Hazard Ratio
HRD	Homologous Recombination Deficiency
HRE	Hormonal Receptor Expression
HRP	Horse-Radish Peroxide
H-score	Histological Score
HTA	Health Technology Assessment
HUT	Hyperplasia of Usual Type
IARC	International Agency for Research On Cancer
IBC	Invasive Breast Cancer
ICC	Intra-Class Correlation Coefficient
IDC	Ductal Carcinoma

IDC-NST	Invasive Ductal Carcinoma Not Otherwise Specified
IGFR	Insulin Like Growth Factor 1
IHC	Immunohistochemistry
ILC	Invasive Lobular Carcinoma
IM	Triple Negative Immunomodulatory Breast Cancer Tumours
IntClust	Inter Cluster
IPB	Immunising Peptide Blocking
KD	Knock Down
kDa	Kilodalton
LAR-TNBC	Triple Negative Luminal Androgen Receptor Breast Cancer Tumours
LCIS	Lobular Carcinoma In Situ
LN	Lymph Node
log2FC	The Logarithm To Base 2 Fold Change
LVI	Lymphovascular Invasion
MBC	Metaplastic Breast Carcinoma
METABRIC	Molecular Taxonomy Of Breast Cancer International Consortium
MGA	Micro Glandular Adenosis
MMP2	Metalloproteinase 2
MPG	Moderate Prognosis Group
MSL	Triple Negative Mesenchymal Like Breast Cancer Tumours
M-TNBC	Triple Negative Mesenchymal Breast Cancer Tumours
NGS	Next Generation Sequencing
NHSB	Nottingham Health Science Biobank

NPI	Nottingham Prognostic Index
NPV	Negative Predictive Value
PAM50	Prediction Analysis Of Microarrays (Pam50) Gene Panel
PARP	Poly-Adp Ribose Polymerase
PBS	Phosphate-Buffered Saline
pcr	Complete Pathological Response
PDGF	Platelet-Derived Growth Factor
PD-L1	Programmed Cell Death Ligand 1
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PLC	Pregnancy-Lactation Cycle
PPG	Poor Prognosis Group
PPV	Positive Predictive Value
PR	Progesterone
REMARK	Tumour Marker Prognostic Studies Criteria
ROC	Receiver Operative Characteristic
rpm	Rotation Per Minute
R/T	Room Temperature
RT	Radiotherapy
RT-PCR	Quantitative Reverse Transcriptase Polymerase Chain Reaction
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SES	Socioeconomic Status
SNP/SNVs	Single Nucleotide Polymorphism/Variance

SPCRP	Solid Papillary Carcinomas with Reverse Polarity
TCGA	The Cancer Genome Atlas
TDLU	Terminal Duct Lobular Unit
TF	Transcription Factors
TMA	Tissue Microarray
TN	Triple Negative Breast Cancer
TNBC	Triple Negative Breast Cancer
TWIST	Twist-Related Protein 2
WB	Western Blotting
WHO	World Health Organisation
ZEB1	Zinc Finger E-Box-Binding Homeobox 1

Table of Contents

ABSTRACT ..... **ii**

Peer-Reviewed Publications Arising from This Study ..... **v**

Scientific Contributions in Peer-Reviewed Publications ..... **vi**

Conference Presentations ..... **ix**

Acknowledgment ..... **x**

Dedication ..... **xii**

Ethical Approval ..... **xiii**

List of Abbreviations ..... **xiv**

Table of Contents ..... **xx**

List of Tables ..... **xxx**

List of Figures ..... **xxxiv**

Chapter 1 : Introduction **38**

1.1 Cancer **39**

1.2 Hallmarks of Cancer **40**

1.3 E-cadherin **42**

1.3.1 Mechanisms of E-cadherin Loss **43**

1.4 Tumour Metastasis **45**

1.5 Transcriptome Technologies **46**

1.6 Anatomy and Histology of Normal Breast Tissue 48

1.7 Development of Invasive Breast Cancer 50

1.7.1 Breast Cancer: Prevalence and Mortality 51

1.7.2 Risk Factors for Breast Cancer 52

1.7.2.1 Exposure to Oestrogen 52

1.7.2.2 Age 53

1.7.2.3 Family History and Genetic Predisposition 53

1.7.2.4 Alcohol Consumption, Smoking, and Obesity 54

1.7.2.5 Breast Density 54

1.7.3 Prognostic Factors 55

1.7.3.1 Tumour Grade 55

1.7.3.2 Tumour Size 56

1.7.3.3 Lymph Node Stage 56

1.7.3.4 Nottingham Prognostic Index 57

1.7.3.5 Lymphovascular Invasion 58

1.7.4 Breast Cancer Predictive Factors 58

1.7.4.1 Hormone Receptors ER and PR 58

1.7.4.2 HER2 59

1.7.4.3 Breast Cancer Molecular Classification by Microarray 59

1.8 Triple-Negative Breast Cancer (TNBC) 62

1.8.1 Pathogenesis, Incidence, and Risk Factors: 62

1.8.2 Prognostic and Predictive Factors in TNBC 62

1.8.3 TNBC Molecular Classification by Microarray 63

1.8.3.1 Basal-Like 1 (BL1) and Basal-Like 2 (BL2) TNBCs.....63

1.8.3.2 Immunomodulatory TNBC.....63

1.8.3.3 Mesenchymal and Mesenchymal Stem-Like TNBC.....64

1.8.3.4 Luminal Androgen Receptor (LAR) TNBC.....64

1.8.4 Gene Expression Analysis of TNBCs by Next-Generation Sequencing.....68

1.8.4.1 Copy Number Aberrations in TNBC.....70

1.8.4.2 Somatically Acquired Mutations in TNBC.....70

1.8.5 Morphological Heterogeneity of TNBC.....71

1.8.5.1 Low-Grade TNBC.....72

1.8.5.2 Salivary Gland-Like Tumours of the Breast.....74

1.8.5.3 High-Grade Histological Types of TNBC.....74

1.8.5.3.1 Invasive Ductal Carcinoma Not Otherwise Specified.....74

1.8.5.3.2 Infiltrating Lobular Carcinoma.....75

1.8.5.3.3 Metaplastic Breast Carcinoma.....75

1.8.5.3.4 Carcinoma of the Breast with Medullary Features.....76

1.8.5.3.5 Carcinoma of the Breast with Apocrine Features.....77

1.8.6 TNBC Heterogeneity: Consequences for Treatment Decisions and Future Targeted Therapeutic Approaches.....77

1.9 Hypothesis and Aims of the Work.....81

Chapter 2 : Materials and Methods.....82

2.1 Cohort Identification and Data Collection.....83

2.2 Next-Generation Sequencing (NGS).....87

2.2.1 RNA Extraction, Library Construction, and NGS Analysis.....87

2.2.2 Differential gene expression (DGE) analysis. 90

2.2.3 Pathway Analysis 90

2.3 Tissue Microarray (TMA) Construction 90

2.3.1 Tissue Handling, Processing, and FFPE Block Retrieval from Nottingham Health Science Biobank (NHSB) 92

2.3.2 TMA Construction Protocol for TNBC Cases 93

2.3.3 Digital Image Marking 93

2.3.4 Recipient Block Construction 93

2.3.5 TMA Layout Design 94

2.3.6 Coring Process 94

2.3.7 TMA Block Annealing and Sectioning 95

2.4 Optimisation of Antibodies Used in the Study 96

2.4.1 Cell Lines and Culture Media 96

2.4.2 Cell Passage Procedure 97

2.4.2.1 Preparation of Cell Lysates 98

2.4.2.2 Cryopreservation of Cell Lines 99

2.5 Specificity of Antibodies by Western Blot 100

2.5.1 Background Principles 100

2.5.2 Polyacrylamide Gel Electrophoresis and Protein Transfer 102

2.5.3 Sample Preparation 103

2.5.4 Loading Samples and Separation of Proteins Mixture 103

2.5.5 Transferring Protein Bands onto Nitrocellulose Membrane/PVDF 104

2.5.6 Visualisation of Protein Bands 105

2.5.7 Immunising Peptide Blocking (IPB).....105

2.6 Immunohistochemistry (IHC).....106

2.6.1 Background Principles 106

2.6.2 Deparaffinisation and Rehydration of Tissue Sections 108

2.6.3 Heat-Induced Epitope Retrieval: Citrate Buffer, pH 6.....108

2.6.4 Preparation of Sodium Citrate Buffer, pH 6.....108

2.6.5 Slide Assembly onto Shandon™ Sequenza Coverplates™.....109

2.6.6 Blocking of Endogenous Peroxidases.....109

2.6.7 Blocking Non-Specific Reacting Proteins 109

2.6.8 Optimisation of Antibodies Used for IHC 110

2.6.9 Applying Post-Primary Antibody.....110

2.6.10 Applying Novolink™ Polymer 111

2.6.11 Blotting with 3,3'-diaminobenzidine (DAB).....111

2.6.12 Counterstaining with Haematoxylin and Tissue Dehydration 111

2.6.13 Assessment of Immunoreactivity Prior to Immunohistochemical Scoring .112

2.6.14 Whole Slide Imaging.....112

2.7 H-Score.....112

2.8 Determination of Cut-Off Points.....113

2.9 Statistical Analysis 115

2.9.1 Introduction.....115

2.9.2 Intra-Class Correlation Coefficient (ICC).....115

2.9.3 Chi-Squared Test (Goodness of Fit).....115

2.9.4 Mann-Whitney U test (Median Analysis) 116

2.9.5 Student's t-Test and One-Way Analysis of Variance (ANOVA; Mean Analysis) **116**

2.9.6 Spearman Rank Test (Model/Line of Best Fit) **116**

2.9.7 Kaplan-Meier Survival Analysis and Log-Rank test **117**

2.9.8 Proportional Hazards Model (Cox Regression Model) **117**

2.9.9 Calculation of Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value **117**

2.9.10 Validation of the Findings **119**

2.9.11 The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) **120**

2.9.12 In-House Biomarker Consortium Data **122**

Chapter 3 : Identification of potential prognostic genes in TNBC **123**

3.1 Background **124**

3.2 Available Prognostic gene signatures **125**

3.3 Advancement on transcriptomic technologies **125**

3.4 Hypothesis and Aim **126**

3.5 Materials and Methods **126**

3.5.1 Study Cohort **126**

3.5.2 Transcriptomic and Pathway Analysis **129**

3.5.3 Prognostic Gene Signature Score **132**

3.5.4 External Validation of Transcriptomic Data **132**

3.5.5 Immunohistochemistry and western blotting analyses **133**

3.5.6 Statistical Analysis **135**

3.6 RESULTS 135

3.6.1 Gene Selection 135

3.6.2 Prognostic Two-Gene Signature 148

3.6.3 External Validation of Genomic Findings 153

3.6.4 Immunohistochemistry of the Prognostic Two-Gene Signature 155

3.6.5 Testing the performance of the prognostic two-gene at the transcriptomic and protein Levels 165

3.7 Discussion 167

3.8 Limitation of the study 172

3.9 Conclusion 173

Chapter 4 : Comprehensive Transcriptomic Profiling Reveals PPL as a Novel, Independent Predictor of Outcomes in TNBC 174

4.1 Background 175

4.2 Role of Periplakin in Cancer 175

4.3 Hypothesis and Aims of the study 176

4.4 Material and Methods 177

4.4.1 Study cohort 177

4.4.2 External Validation of Transcriptomic Data 178

4.4.3 Immunohistochemistry and western blotting analyses 179

4.4.4 Evaluation of the mechanistic and phenotypic functions of PPL in TNBC cell lines 181

4.4.5 siRNA transfection 181

4.4.6 Proliferation Assay.....184

4.4.7 Migration and Invasion assays 184

4.4.8 BC cell line RNA-seq and bioinformatic analyses 185

4.4.9 Differential Gene Expression Analysis.....185

4.4.10 Pathway Analysis 186

4.4.11 RNA isolation and quantitative real-time PCR (qRT-PCR) analysis.....186

4.4.12 Statistical Analysis 187

4.5 RESULTS 188

4.5.1 Clinical significance of PPL expression in TNBC at both transcriptomic and protein level 188

4.5.2 PPL expression prognostic value in TNBC chemotherapy treated patients .194

4.5.3 External Validation of PPL expression in TNBC cohorts.....197

4.5.4 Testing the performance of PPL expression to predict distant metastasis in TNBC cohorts at the transcriptomic and protein level 200

4.5.5 PPL gene silencing reduces proliferation, cell migration and invasion 201

4.5.6 Clinical significance of PPL expression in unselective BC cohorts at transcriptomic level 204

4.5.7 Changes in TNBC transcriptome upon PPL knockdown 208

4.6 Discussion.....217

4.7 Limitations of the study.....219

4.8 Conclusion 221

Chapter 5 : The Molecular Landscape of E-Cadherin in Ductal Carcinoma of The Breast.....222

5.1 Background 223

5.2 Prognostic Significance of E-cadherin expression in BC 223

5.3 Hypothesis and Aims 224

5.4 Materials and Methods 225

5.4.1 Study cohorts 225

5.4.2 Immunohistochemistry and western blotting analyses 228

5.4.3 Statistical analysis 230

5.4.4 Bioinformatics investigation and Pathway analysis 231

5.5 RESULTS 232

5.5.1 Evaluation of E-cadherin protein expression in the high-grade IDC protein cohort (n=813) 232

5.5.2 E-cadherin copy number alteration in high-grade IDC tumours 234

5.5.3 Expression of E-cadherin suppressor transcription factors 237

5.5.4 Proteins associated with E-cadherin expression in invasive TNBC-IDC 242

5.5.5 E-cadherin loss and expression of genes deferentially expressed between ILC and IDC within the TNBC IDC cohort. 246

5.5.6 Bioinformatics investigation and pathway analysis 249

5.6 Discussion 252

5.7 Limitation of the study 255

5.8 Conclusion 256

Chapter 6 : General Discussion 257

6.1 Background 258

6.2 Hypothesis, Aims and Methods of the Study 259

6.3 Key Findings of the Study 260

6.3.1 Identification of Potential Prognostic Genes in TNBC 260

6.3.2 Comprehensive Transcriptomic Profiling Reveals PPL as a Novel, Independent Predictor of Outcomes in TNBC 261

6.3.3 The Molecular Landscape of E-Cadherin Expression in Ductal Carcinoma of the Breast 263

6.4 Concluding remarks 264

6.5 Strengths of the study 264

6.6 Limitations of the study 265

6.7 Future Plan 266

References 268

Appendices 313

7.1 Chapter 1: Copyright approval 313

7.2 Chapter 3: Identification of potential prognostic genes in TNBC 319

7.3 Chapter 4: Comprehensive Transcriptomic Profiling Reveals PPL as a Novel, Independent Predictor of Outcomes in TNBC 341

7.4 Chapter 5: The Molecular Landscape Of E-Cadherin in Ductal Carcinoma of The Breast 352

List of Tables

Table 1-1: Features of the Integrative Clusters 61

Table 2-1: Patients cohort 86

Table 2-2: Cell Lines Used in this Study 97

Table 2-3: Western Blotting Evaluation of Antibodies Used in the Study 101

Table 2-4: Immunohistochemistry Evaluation of Antibodies Used in the Study 107

Table 2-5: Expression localisation and cut-off details of the stained antibodies. 114

Table 2-6: METABRIC-TNBC patients' cohort 121

Table 3-1: Patients cohorts 128

Table 3-2:Pathway analysis of genes strongly predictive of both Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112) 137

Table 3-3:Univariate analyses for the 21 genes identified by Venny tool that potentially predictive of both Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112) 140

Table 3-4:Multivariate Cox regression analysis for genes potentially associated with both Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112) 145

Table 3-5:Combined multivariate Cox regression analysis of genes potentially associated with both Breast Cancer-Specific Survival and Distant Metastasis-Free Survival (Transcriptomic Cohort, n=112) 149

Table 3-6:Multivariate Cox regression analysis to generate a prognostic score for the two gene signature predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112) 150

Table 3-7:Multivariate Cox regression analysis for the two gene signature predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112) 152

Table 3-8:Multivariate Cox regression analysis for Individual Potential proteins associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333) 158

Table 3-9:Multivariate Cox regression analysis for Individual Potential proteins associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333) 159
Table 3-10:Combined multivariate Cox regression analysis for potential proteins associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333) 161
Table 3-11:Multivariate Cox regression analysis to build prognostic index for the two gene signature predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333).....162
Table 3-12: Multivariate Cox regression analysis for the protein expression prognostic score predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333).....164
Table 3-13:Distribution of cases within transcriptomic and IHC cohorts to test the prognostic index with reference to clinical evidence of distant metastasis.....166
Table 4-1:The siRNA oligonucleotides used in this study.....182
Table 4-2:Multivariate Cox regression analysis for PPL expression associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic cohorts).....189
Table 4-3:PPL copy number alteration (CNA) in TNBC-METABRIC cohort (n=313) .190
Table 4-4:Multivariate Cox regression analysis for PPL expression associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC cohort, n=245).....193
Table 4-5:Distribution of cases within NGS transcriptomic and IHC cohorts to test the prognostic index of PPL expression with reference to clinical evidence of distant metastasis. 200
Table 4-6:Multivariate Cox regression analysis for PPL expression associated with Breast Cancer Specific Survival (BCSS) (METABRIC cohort).....207
Table 4-7:The top regulated pathway as defined by both differential gene expression and pathway analysis on siRNAs depleted endogenous PPL MDA-MB-231 TNBC cell line using the over-representation analysis (ORA) for identification of gene ontologies.....210

Table 4-8:The top regulated pathway as defined by both differential gene expression and pathway analysis on siRNAs depleted endogenous PPL on MDA-MB-436 TNBC cell line using the over-representation analysis (ORA) for identification of gene ontologies.....212
Table 4-9:The top regulated pathway as defined by both differential gene expression and pathway analysis on the 76 overlapping down regulated genes resulted the comparison of scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies.....213
Table 4-10: The Biological process category by the Gene Ontology Consortium shows the 76 core common genes that are significantly downregulated among scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies. 214
Table 4-11: The Molecular function category by the Gene Ontology Consortium shows the 76 core common genes that are significantly downregulated among scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies. 215
Table 5-5-1:Patients cohort 227
Table 5-2:Biomarkers included in this study to examine the potential association with altered E-cadherin expression in high grade IDC protein cohort (n=813).....231
Table 5-3:Correlation between level of proteins associated with altered E-cadherin expression in high-grade IDC cohort (n=813) 233
Table 5-4:CDH1 copy number alteration (CNA) in whole IDC METABRIC cohort (n=883) 235
Table 5-5:Frequency of genes associated with reduced E-cadherin expression in the whole IDC METABRIC cohort (n= 208) 235
Table 5-6:CDH1 copy number alteration (CNA) in the high grade TNBC IDC cohort within the METABRIC cohort (n=235) 236
Table 5-7:Frequency of genes associated with reduced E-cadherin expression in the high grade TNBC IDC cohort within the METABRIC cohort (n=90) 236
Table 5-8:Copy number alteration (CNA) within other chromosomes in the whole IDC METABRIC cohort with reduced/loss CDH1 mRNA expression. (n=208).....237

Table 5-9:known Transcriptional factors associated with reduced E-cadherin expression examined in the reduced/loss of CDH1 mRNA expression IDC METABRIC cases (n=208) 238
Table 5-10:Correlation between mRNA levels of the genes associated with altered E-cadherin expression in BC in the whole IDC METABRIC cohort (n=883) 239
Table 5-11:Genes associated with reduced E-cadherin expression in the TNBC IDC cohort within the METABRIC cohort (n=235) 240
Table 5-12:Correlation between mRNA level of the genes associated with E-cadherin expression in the high grade TNBC IDC cohort within the METABRIC cohort (n=235) 241
Table 5-13:Correlation between level of proteins known to control E-cadherin expression using triple negative -IDC cohort (n=106) 243
Table 5-14:Correlation between level mRNA expression of other genes known to control E-cadherin expression using triple negative invasive breast carcinoma cohort (n=106) .244
Table 5-15:Genes differentially expressed between lobular versus ductal breast carcinomas in triple negative breast cancer cohort (n=106).....247
Table 5-16:Pathway analysis results using Web gestalt to identify differentially regulated canonical pathways in the triple negative breast cancer cohort 250
Table 5-17:Genomic study and pathway analysis based on data generated from next generation RNA sequencing in triple negative breast cancer cohort 251

List of Figures

Figure 1-1: Carcinogenesis in cells 40

Figure 1-2: The Hallmark of Cancer 42

Figure 1-3: Proposed theories for mechanisms of loss of E-cadherin in cancer 44

Figure 1-4: Role of numerous signalling pathways in controlling the expression of E-cadherin 46

Figure 1-5:Summary of current gene expression detection assays for cancer tissue sample 48

Figure 1-6: Anatomy of the female breast 49

Figure 1-7: Historical model of breast cancer development and evolution. 51

Figure 1-8 :TNBC classification in comparison with PAM50 classification..... 67

Figure 1-9: Prevalence of gene mutation affecting TNBC genome profile 69

Figure 1-10: Spectrum of TNBC histological subtypes 72

Figure 1-11: Spectrum of current clinical trials for TNBC treatment..... 80

Figure 2-1: Descriptive explanation of used cohorts in the thesis 85

Figure 2-2 :Next generation sequencing flowchart 89

Figure 2-3:TMA construction Protocol..... 91

Figure 3-1:Flowchart of Artificial Neural Network analysis 131

Figure 3-2:Immunohistochemical expression of SPDYC and ACSM4 in FFPE TMA cores from TNBC samples (IHC cohort, n=333) 134

Figure 3-3:Receiver Operative Characteristic curve depicting the sensitivity and specificity of the predicted 21 gene panel associated with both **Breast cancer-specific survival (BCSS) and Distant metastasis-free survival (DMFS) (Transcriptomic cohort, n=112).** 139

Figure 3-4:The output of Venny diagram tool showing the overlapping genes between the top 200 genes associated with Breast Cancer Specific Survival (BCSS) and the top 200 genes associated with Distant Metastasis Free Survival (DMFS) (Transcriptomic cohort, n=112) 139

Figure 3-5:Univariate Kaplan-Meier survival analyses of the transcripts identified by Artificial Neural Network analysis to be strongly predictive of Breast Cancer-Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) using RNA seq matrices 144

Figure 3-6:Univariate Kaplan-Meier survival analysis of the prognostic two gene signature predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic cohort, n=112) 151

Figure 3-7:Univariate Kaplan Meier survival analysis of our proposed combinatorial two gene signature predicting overall Survival (public domain datasets) 154

Figure 3-8:Univariate Kaplan Meier survival analysis for ACSM4 and SPDYC protein expression for association with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333) 157

Figure 3-9:Violin plots demonstrating a positive correlation between protein expressions of SPDYC and ACSM4 (Correlation Coefficient, r=0.29, P=0.00001) (IHC cohort, n=333) 160

Figure 3-10:Univariate Kaplan Meier survival analysis of the protein expression of the two gene signature score predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333) 163

Figure 4-1:Flowchart of Analytical and Experimental Methodology used to explore the prognostic value of PPL expression in TNBC 178

Figure 4-2:Western blot and immunohistochemical expression of PPL in TNBC 180
Figure 4-3:Differential expression of PPL protein in breast cancer cell lines. PPL protein levels were quantified by densitometry and normalised to b-actin / GAPDH levels.....183
Figure 4-4:Univariate Kaplan-Meier survival analysis of PPL expression predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic cohorts).....191
Figure 4-5:Univariate Kaplan-Meier survival analysis of PPL expression predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC cohort). 192
Figure 4-6:Univariate Kaplan-Meier survival analysis of PPL expression predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) in TNBC patients who received chemotherapy treatment only (Transcriptomic cohorts) 195
Figure 4-7:Univariate Kaplan-Meier survival analysis of PPL expression predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) in TNBC patients who received chemotherapy treatment only (IHC cohort) 196
Figure 4-8:Analysis of PPL expression association with chemotherapy response in TNBC restrictive cohort using public domain data 197
Figure 4-9:Univariate Kaplan-Meier survival analysis of PPL expression predicting patients outcome using public domain data 198
Figure 4-10:Mechanistic evaluation of PPL expression 202
Figure 4-11:Univariate Kaplan-Meier survival analysis of PPL expression association with Breast cancer outcome using public domain data 205
Figure 4-12:Investigation of PPL expression association with endocrine and Chemotherapy response using public domain data 207

Figure 4-13:Differential gene expression analysis on scrambled TNBC cell lines vs siRNA depletion of PPL TNBC cells.....209

Figure 4-14:The Biological process category by the Gene Ontology Consortium shows the 76 core common genes that are significantly downregulated among scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies. 216

Figure 4-15:The Molecular function category by the Gene Ontology Consortium shows the 76 core common genes that are significantly downregulated among scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies. 216

Figure 5-5-1:Flowchart of Analytical and Experimental Methodology of the molecular landscape of E-cadherin in ductal carcinoma of the breast study 226

Figure 5-2:Western blotting & Immunohistochemistry.....229

Figure 5-3:Frequency of Transcription factors suppressing E-cadherin expression in Cases showing reduced/loss CDH1 mRNA expression in the triple negative breast cancer cohort.....245

Figure 5-4:Differentially expressed genes between lobular versus ductal breast carcinomas in the triple negative breast cancer cohort 248

Chapter 1 : Introduction

1.1 **Cancer**

Cancer is a complex disease arising from the deleterious modifications of genes that alter the behaviour, growth, and differentiation of transformed cells [1]. In 2015, 14 million people were diagnosed with cancer, and 8.8 million died because of the disease around the globe. Therefore, cancer accounted for one in every six deaths, making it one of the highest causes of mortality worldwide [2]. While progress in medicine and cancer research has been considerable, over the next fifteen years, the incidence of cancer is projected to increase by approximately 70% with a consequent increase in cancer mortality [3].

During carcinogenesis, mutated genes allow cells to adopt characteristics that enable them to evade the cellular systems that would ordinarily shut down cell proliferation. Instead, these transformed cells can proliferate uncontrollably, spreading to tissues and organs in other parts of the body. For cancer to develop, a number of systems need to fail. These include DNA repair mechanisms and the cellular mechanisms that control cellular proliferation, differentiation, apoptosis, and the cell cycle (**Figure 1-1).** Even though it is accepted that cellular abnormalities accumulate before cancer develops, the exact nature and sequence of events that have to take place before the cell transforms into a cancerous cell remains poorly understood [4–6].

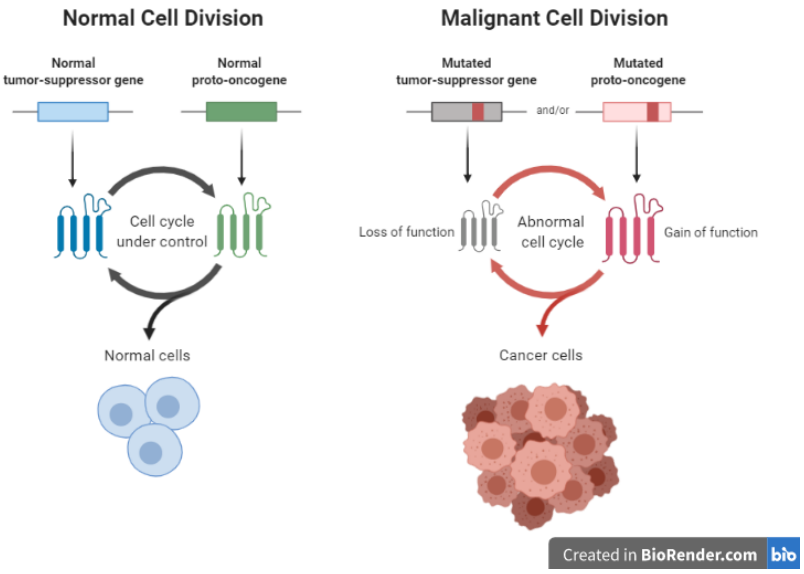


Figure 1-1: Carcinogenesis in cells

- This figure illustrated the differences between “Tumour Suppressor Genes and Proto-oncogenes" in controlling cell division.

1.2 **Hallmarks of Cancer**

Tumourigenesis can simply be defined as the abnormal growth of cells. Characteristic features of malignant tumours include evading apoptosis and growth suppressors, prolonged proliferative signalling, stimulating angiogenesis, triggering invasion and metastasis, and facilitating replicative immortality [5]. These characteristics were defined in Hanahan and Weinberg’s seminal framework, which summarised cellular features that all malignant tumorigenic cells are believed to acquire [5]. Specifically, these are the potential for limitless replication, the capacity to evade apoptosis, the ability for sustained angiogenesis, autonomous growth signals (i.e., not dependent upon external growth signals), insensitivity to signals to inhibit growth, and the ability to invade proximal tissues

and metastasise, spreading to other areas of the body [5]. As discussed later, this is often related to the loss of expression of key genes, such as E-cadherin (*CDH1*). Following advances in cancer research, Hanahan and Weinberg updated their framework in 2011 to include the capacity to evade the immune system, thereby avoiding detection and destruction, and the cells' ability to self-modulate their metabolic processes [1] (**Figure 1-2**).

Hanahan and Weinberg's 2011 review also defined three tumorigenesis-supporting traits that are characteristic of cancer. These are genomic instability and the accumulation of DNA mutation in the tumour microenvironment, inflammation in the tumour, and the abundance of endothelial cells, immune cells, pericytes, and stem cells in the tumour microenvironment. These characteristics enable the tumour to acquire features required for malignancy [1]. Functional, genomic, and genetic approaches have enabled the identification of genes that can protect against cancer (tumour suppressor genes), such as *TP53*, *RB1*, *NF2*, *CDH1*, and *LKB*, or promote susceptibility to cancer (oncogenes), such as *RAS*, *RAF*, and *MYC* [7]. Cell cycle processes are influenced by oncogenes gaining functionality and tumour suppressor genes losing functionality. This in turn influences tumourigenesis [7].

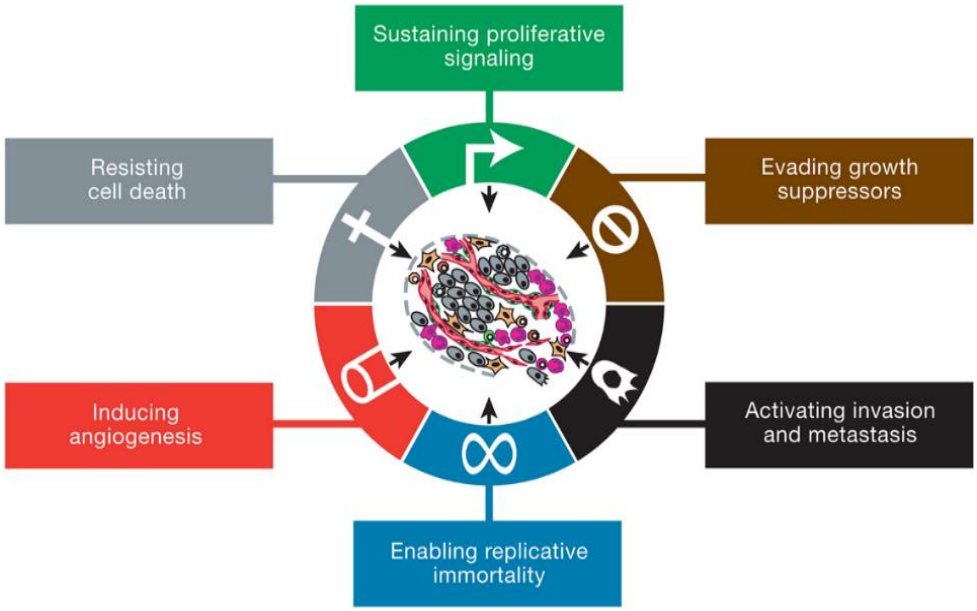


Figure 1-2: The Hallmark of Cancer

- This figure summarises the six hallmarks of cancer. Taken from Hanahan and Weinberg (2011)[1].

1.3 **E-cadherin**

The cadherins are a superfamily of calcium-modulated cell adhesion glycoproteins [8]. There are four groups of cadherins; classical, desmosomal, protocadherin, and unconventional [9] but all are synthesised as pro-proteins that undergo post-translational modifications before they are transported to cell membranes [10]. Cadherins have three domains: intracellular, transmembrane, and extracellular [11]. Adherens junctions, which occur in epithelial and endothelial tissues, describe the connections formed between the intracellular domains of cadherins and actin networks [12]. The proto-typical E-cadherin is a type I transmembrane protein. Its ‘E’ prefix reflects the epithelial expression of this type of cadherin, distinguishing it from those expressed by nerve tissue (N-cadherins) and the placenta (P-cadherins) [13]. The expression of E-cadherin has been studied thoroughly in

both normal and abnormal tissues. This important protein is essential to embryogenesis and the development and maintenance of epithelial tissues [14].

1.3.1 Mechanisms of E-cadherin Loss

As E-cadherin is central to maintaining close contact between adjacent cells in epithelial tissues, it also has a role in preventing the invasion of intrusive and metastatic phenotypes [15]. Cancer can metastasise and progress more readily when cell adhesion falters as a consequence of the partial or complete down-regulation of E-cadherin [16]. According to Van Roy and Berx [14], in malignant states, abnormal endocytosis, mutations, epigenetic silencing, and the up-regulation of non-epithelial cadherins are all able to suppress E-cadherin, thereby promoting malignancy (**Figure 1-3**).

E-cadherin is transcribed actively, the positive regulators within the promoter are GC-rich regions, a CCAAT box, and an epithelial-specific enhancer (ESE) box. In contrast, E-box binding sites are sequences that suppress E-cadherin expression when bound to repressors [17]. Factors from various downstream signalling cascades, such as E47, SIP1/ZEB1, SLUG, SNAIL, and TWIST, can bind to the E-boxes to silence expression [18].

Under normal conditions of histone acetylation of the E-cadherin gene promoter region, transcription factors and co-factors are able to access the promoter, enabling normal levels of transcription [19]. However, in invasion and metastasis, dense methylation of the CpG dinucleotides (CpG Islands) in the CDH1 proximal region severely limits access to the gene. Thus, one mechanism for silencing the expression of E-cadherin is through DNA hypermethylation [20].

Another method of silencing E-cadherin expression is through mutation. Although comparatively infrequent, a number of cancers, including breast, endometrial, gastric, ovarian, and thyroid cancers, are linked to *CDH1* mutations [21]. More commonly, the mutation-associated loss or dysfunction of E-cadherin is associated with *TP53* and *PIK3CA*

mutations [22]. Though the types of mutation vary, loss-of-function mutations for *CDH1* have been observed frequently in diffuse gastric and lobular breast cancers (BCs). While the majority of gastric tumour mutations take the form of in-frame deletions that are caused by skipping exons, the mutations in lobular BCs are predominantly of the out-of-frame variety [23].

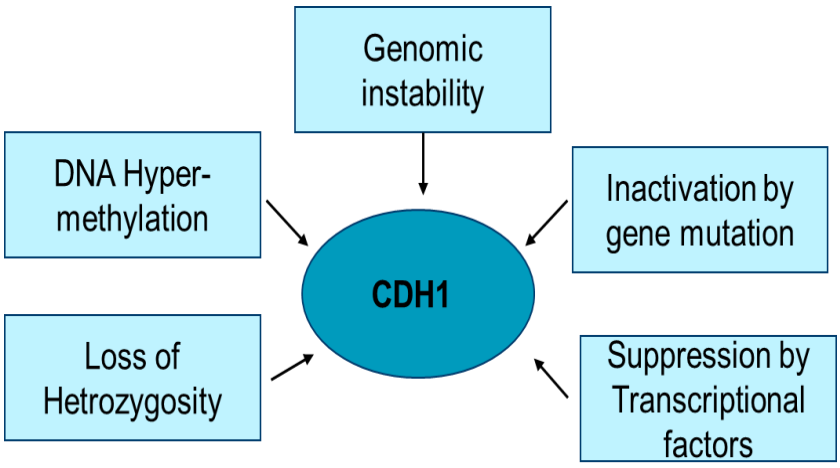


Figure 1-3: Proposed theories for mechanisms of loss of E-cadherin in cancer

- Schematic illustration of various theories causing E cadherin loss of expression

1.4 **Tumour Metastasis**

Metastasis describes the mechanisms whereby cancerous cells from the primary tumour invade, disseminate and spread, and establish new tumours in other parts of the body. For this to occur, tumour cells are believed to degrade the local extracellular matrix to invade surrounding tissue, intravasate into blood or lymph systems, escape detection by immune systems, extravasate, and successfully colonise distant organs. The patient's prognosis deteriorates significantly once the tumour has metastasised, and metastatic cancers are considered incurable [24,25]. Metastatic tumour cells develop the capacity to migrate and invade other tissues. One central feature of metastatic cells is that their epithelial phenotype transitions to being mesenchymal-like (epithelial-to-mesenchymal transition, EMT). Through the EMT mechanism, epithelial cells in one part of the body break away from the epithelium and then travel to different organs [26]. Individual epithelial cells in normal tissue are connected to adjacent epithelial cells by intercellular junctions mediated by E-cadherin, thereby forming sheets of cells. Thus, E-cadherin is central to the epithelial state. The current understanding is that E-cadherin dysfunction is required for cancer cells to invade proximal tissue and metastasise, which commonly manifests as loss of expression [27]. Several transcription factors (TFs) contribute to EMT mechanisms. These include SLUG, SNAIL, TWIST, ZEB1, and ZEB2. They suppress the expression of epithelial markers and promote the expression of mesenchymal-associated markers [28]. Among the actions of a number of these TFs is the direct suppression of E-cadherin expression [28] (**Figure 1-4**).

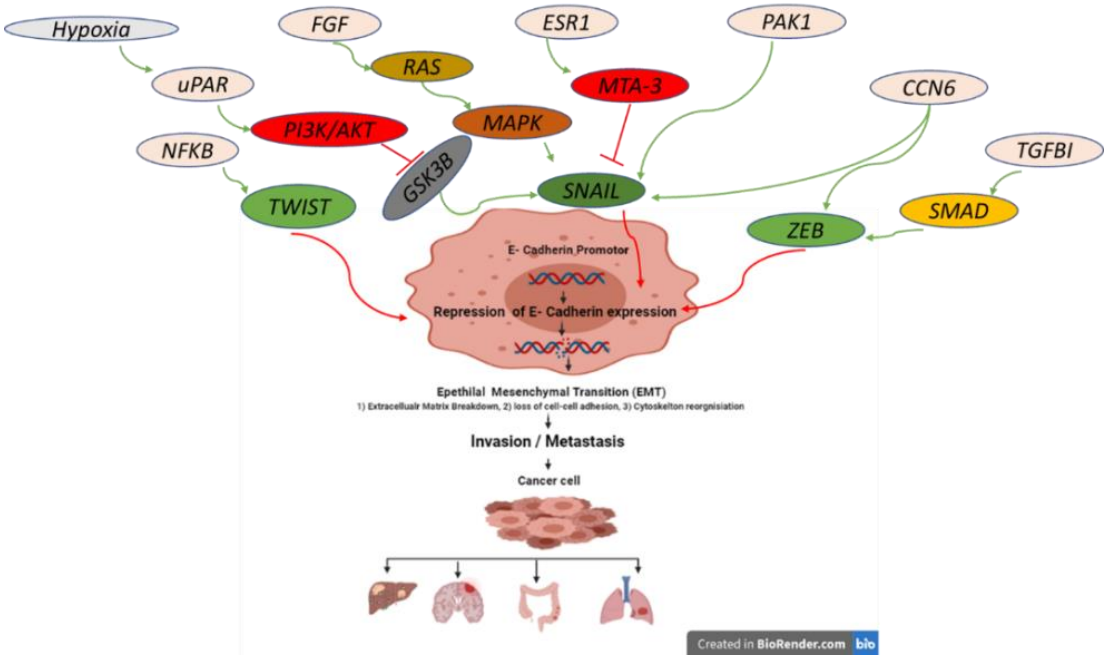


Figure 1-4: Role of numerous signalling pathways in controlling the expression of E-cadherin

- Various E-cadherin signalling pathways regulate the key process of epithelial mesenchymal transition and the progression of breast cancer. Taken from Baranwal and K Alahari [29].

1.5 Transcriptome Technologies

The introduction of genome-wide expression quantification methods (Figure 1-5), such as gene microarrays and massive parallel sequencing/next-generation sequencing (NGS), has paved the way for the molecular classification of many types of cancers [21]. NGS is a high-throughput technique that can be applied to identify complex genomic instabilities, single nucleotide polymorphism/variants (SNP/SNVs), copy number alteration (CNA), short insertions/deletions (indels), and fusions in multiple genes from different patient samples [30,31]. The detection of genomic alterations allows the potential identification of new target therapies, which distinguish tumours that benefit little from traditional cytotoxic chemotherapy response regimens [32]. NGS has shown its high-throughput power in

deciphering the transcriptome of sequenced tumours and identifying potential new diagnostic and therapeutic targets, which are being used to inform and develop new approaches to improve patient outcomes[33–35]. Somatic mutations are believed to usually occur in tumour epithelial cells. It is well established that the tumour microenvironment is critical in tumorigenesis through progression, treatment response, and treatment resistance [36]. Consequently, quantifying the gene expression of tumour and adjacent non-tumour cells may provide important information for cancer progression and treatment responses, as evidenced by multiple studies integrating the transcriptome using gene expression arrays [37,38]. While RNA sequencing through NGS is more complicated in terms of experimental procedures and bioinformatics analysis, it has advantages in supporting a higher dynamic range when compared to gene expression array technologies [39]. NGS technology has proven its ability to accurately quantify gene expression, including those transcripts with either very low or high expression, and thereby overcomes the limitations of other technologies, such as gene expression arrays [40].

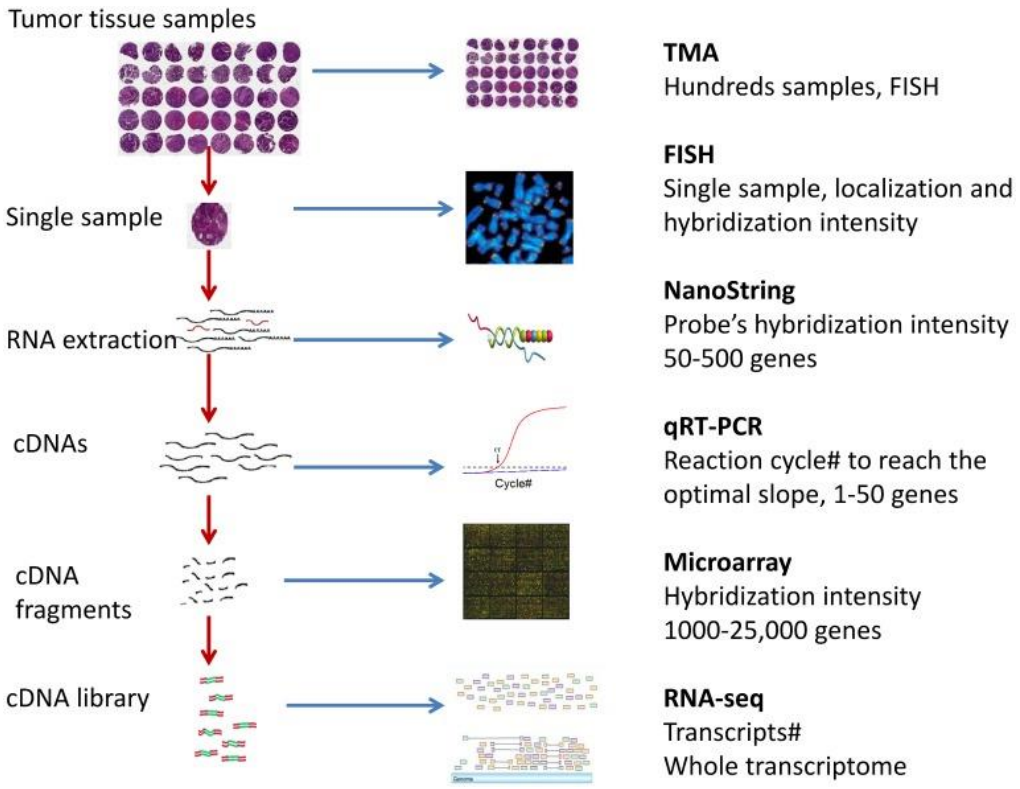


Figure 1-5:Summary of current gene expression detection assays for cancer tissue sample

- A schematic illustration of current detections assays to decipher tumour heterogeneity. Taken from Narrandes and Wayne Xu [41].

1.6 Anatomy and Histology of Normal Breast Tissue

To understand BC, it is necessary to differentiate the histology of non-malignant mammary tissue. Normal female breast development commences early in embryogenesis, when the mammary crest forms, followed by nascent mammary buds. Breasts are rounded, complex, glandular tissues that cover the pectoral muscle at the front of the rib cage. The structural integrity of these milk-producing and secreting glands is maintained by the surrounding fatty and connective tissues. Within each mammary gland, there are between 12 and 20 lobes, which subdivide into several thousands of smaller lobules, and these, in turn, can contain up to 200 acini, which are the milk-secreting pouch-shaped structures. Acini ducts

join together to form a terminal duct; this and its associated lobule form the terminal duct lobular unit (TDLU). 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 [42]. Functional maturity of breast tissue occurs during the pregnancy-lactation cycle (PLC) [43], which is a high-energy state. Puberty, PLC, and post-menopause initiate significant hormonal and structural changes in breast glands [43] (**Figure 1-6**).

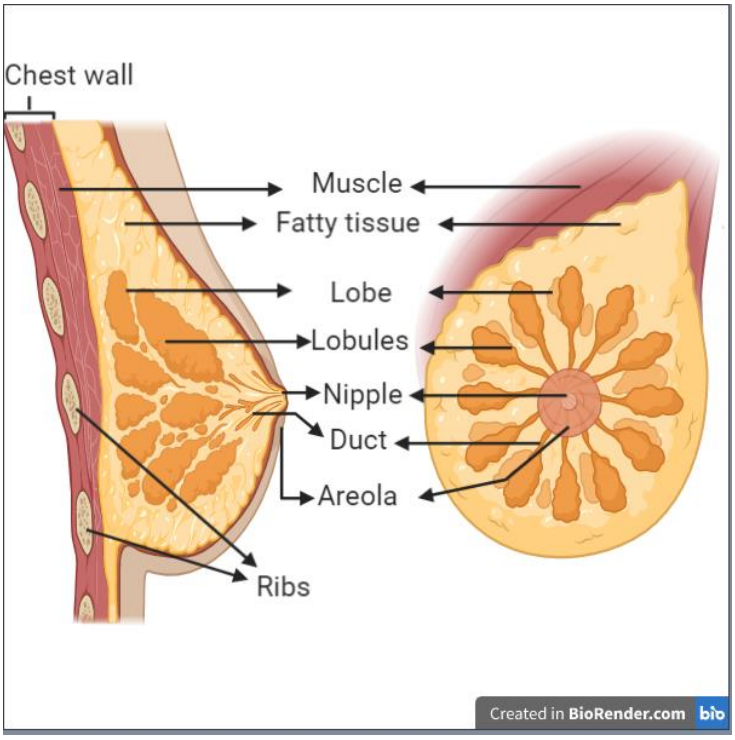


Figure 1-6: Anatomy of the female breast

- The nipple and areola are shown on the outside of the breast. The lobes, lobules, ducts, and other component the breast is also shown.

1.7 Development of Invasive Breast Cancer

One of the first steps in the malignant transformation of breast tissue is the emergence of hyperplasic epithelial cells that exhibit resistance apoptosis [44]. Benign breast lesions may progress to become malignant, with atypical ductal hyperplasia (ADH) and atypical lobular hyperplasia (ALH) being two highly proliferative conditions that in some women are risk factors for cancer [45]. The hyperplasia of the epithelium can progress to cellular atypia, which causes the breast duct to become obstructed and can lead to ductal carcinoma in situ (DCIS). Asymptomatic DCIS is mainly detected by mammographic screening. This is particularly important in young, high-risk DCIS patients, as 'pre-invasive' lesions in normal breast tissue can potentially develop to become locally invasive ductal carcinoma (IDC) [46].

BC is a multi-factorial disease, as there are several highly complex mechanisms that ensure malignant breast cells propagate and survive. Examples of these mechanisms include activating inflammation, avoiding growth suppressors and immune surveillance, genomic instability, invasion, and metastatic pathways [1]. Outcomes of high invasion and poor survival have been linked to BC that has a high number of cancer stem cells [47,48]. However, particular molecular intrinsic subtypes that determine hormonal signalling pathways have been identified in other studies as being responsible for invasive BC (IBC) [49]. Different subtypes of epithelial cells may give rise to different types of BC. Some BCs grow slowly and are sensitive to various treatments, whereas others are aggressive, initiating invasive-metastatic cascades [46]. For non-invasive DCIS to progress to IDC, these tumours must penetrate the myoepithelial lining and basement membrane, which depends upon the presence of genetic, transcriptional, and protein abnormalities [46] (**Figure 1-7**).

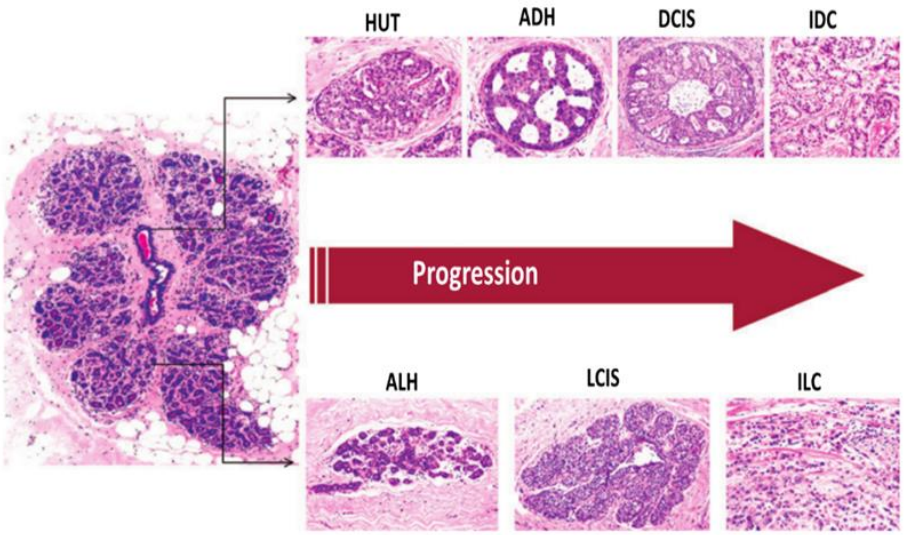


Figure 1-7: Historical model of breast cancer development and evolution.

- ADH:atypical ductal hyperplasia; ALH: atypical lobular hyperplasia; DCIS: ductal carcinoma in situ; HUT: hyperplasia of usual type; IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; LCIS: lobular carcinoma in situ. Taken from Maria A Lopez-Gracia and colleagues [50].

1.7.1 Breast Cancer: Prevalence and Mortality

IBC is a primary cause of female mortality. The global incidence of BC was presented in the World Cancer Report 2014 as 43.3 per 100,000 females. Colorectal carcinoma is the next most common cancer, with an incidence rate of 14.3 per 100,000. The incidence of all cancers, except non-melanoma skin cancer, is highest in high-income countries, including Australia, Belgium, Canada, France, Ireland, Japan, Luxembourg, Monaco, Netherlands, New Zealand, South Korea, the United Kingdom, and the United States. However, the greatest burden of cancer mortality falls on developing regions, with 70% of all cancer-related deaths occurring in Africa, Asia, and Central and South America. It is predicted that by 2025, 20 million new cancer cases could occur, specifically in low-income countries

[51]. In 2008, the number of women with BC was approximately 1.4 million, but within four years it had risen to 1.67 million cases, indicating a growing trend in BC [52]. The World Cancer Report 2014 stated that the respective numbers of cases of BC in women in developed and developing countries were 794,000 and 883,000, respectively. In 2012, the estimated incidence of BC in European females was 464,202 (28.8%), whereas the estimated BC mortality was 131,257 (16.8%). The burden of BC deaths disproportionately affects developing countries, where 324,000 deaths were reported [53]. In Saudi Arabia, the number of female BC patients rose significantly from 1,152 in 2008 to 1,473 in 2010 [54]. The actual incidence of BC in Saudi Arabia could be quite different because the uptake for BC screening is very low [55]. More treatment options may be available if BC is detected early. With effective educational campaigns and lifestyle changes, there could be fewer new cases of BC.

1.7.2 Risk Factors for Breast Cancer

As noted earlier, BC risk is multifactorial, and the precise mechanisms behind BC are not yet completely understood. Environmental, heritable, and lifestyle factors can all predispose an individual to risks of developing BC or its progression. Some of these risk factors are considered in the following sub-sections.

1.7.2.1 Exposure to Oestrogen

The initiation and progression of BC is recognised as being influenced by hormonal factors. The primary sex hormone in females is oestrogen, which is implicated in a multitude of biological functions. Although oestrogen is an endogenous hormone, the risk of BC increases for women who take oral contraceptives or hormone replacement therapy for menopause, as these medications contain oestrogen and can stimulate breast growth [56]. The risk of IBC is reported to be elevated in menopausal women exposed to a combination of oestrogen and medroxyprogesterone, a form of progesterone that regulates ovulation

[57]. As a consequence of pregnancy and longer duration of breastfeeding, lobules differentiate, which confers a protective effect against IBC by reducing exposure to oestrogen, thus reducing the incidence of IBC [58]. The difference in incidence of BC between mothers who have had full-term pregnancies and breast-fed and nulliparous women is significant, with the latter group being much more vulnerable to BC [59].

1.7.2.2 Age

The risk of developing BC increases with age, with women over 50 being more likely to develop the disease compared to younger women. However, when BC occurs in young women, there is a tendency for it to have an aggressive advanced-stage phenotype, such as triple-negative BC (TNBC) or overexpression of human epidermal growth factor receptor-2 (HER2), which are strongly associated with poor survival outcomes [60,61].

1.7.2.3 Family History and Genetic Predisposition

Ten percent of BC cases are attributable to hereditary factors [62]. The risk of developing BC increases for women who have a first- or a second-degree relative (i.e., mother, daughter, or sister) who has suffered from BC. There may also be an increased risk for women who have several relatives who have had ovarian cancer. A relationship has been established between hereditary BC and germline mutations in the genes for BC type-1 susceptibility protein (BRCA1) and BC type-2 susceptibility protein (BRCA2). The results of one study indicated that 25% of familial risk is attributed to *BRCA1/BRCA2* mutations [63]. BCs due to *BRCA1/BRCA2* mutations are generally early-onset and affect both breasts [62]. Women who have *BRCA1/BRCA2* mutations are also more vulnerable to gastric, pancreatic, and colon cancers [64,65].

1.7.2.4 Alcohol Consumption, Smoking, and Obesity

Other factors that increase the risk of developing BC include lifestyle choices. Mammography scans have identified a relationship between the scan density level and regular high-alcohol consumption [66]. The International Agency for Research on Cancer (IARC) stated that the risk of developing BC increased by < 10% for women who drank 10 grams of alcohol per day [67]. Smoking is also a risk factor, with active smokers and those who have smoked for a long time being much more susceptible to BC than passive smokers and non-smokers. Passive smoking is still a risk factor, as non-smoking women exposed to smoke are more likely to develop BC than non-smoking women who are not exposed to smoke [68]. Postmenopausal women who are overweight or obese also have a higher risk of developing BC [69]. There are potentially a few obesity-associated mechanisms that could account for the increased risk of developing BC. For example, adipose tissue expresses aromatase, which converts androgen to oestrogen. Also, there is an association between obesity and elevated levels of circulating insulin, secreted inflammatory cytokines, and insulin-like growth factor I. All of these are coupled with increased risks of developing BC [70].

1.7.2.5 Breast Density

Mammography scans reveal the density of breast tissue. Scans of breast tissue that appear light indicate fatty tissue, whereas dense scans denote glandular and supportive fibrous tissues. There is a robust relationship between the risk of developing BC and the density of the tissue. Women with low-density tissue are less likely to develop BC, and those with dense tissue are much more vulnerable to the disease [71].

1.7.3 Prognostic Factors

BC is histologically and clinically heterogeneous. One major challenge has been the identification of accurate prognostic factors. The prognostic factors used for BC are tumour characteristics, any of which may predict patient outcomes. The latter can be differentiated as BC-specific survival (BCSS), distant metastasis-free survival (DMFS), and disease-free survival (DFS). However, any clinicopathological features of BC may be used as predictive factors to evaluate the probability of a specific treatment eliciting a response. Particular clinicopathological features may demonstrate predictive and prognostic characteristics. Outcomes can potentially be improved by accurately identifying prognostic and predictive features, which can inform the choice of therapeutic intervention to ensure that the treatments that are most likely to be effective are used. The factors most commonly used for BC prognoses and outcome predictions are discussed in the succeeding sub-sections.

1.7.3.1 Tumour Grade

A common prognostic indicator of BC is a three-tiered tumour grading system that compares the appearance of BC cells to normal breast cells and assesses the organisation of BC cells in malignant lesions. The three tumour categories are 1) Grade I: BC cells appear very similar to normal breast cells, are well differentiated, and grow slowly; 2) Grade II: BC cells are differentiated and rapidly growing; therefore, they do not resemble normal breast cells; and 3) Grade III: BC cells are poorly differentiated and rapidly growing, and they bear little resemblance to normal breast cells [72]. When a pathologist is unable to determine the grade, the patient might be given a ‘grade X’ [72]. Some grading systems include Grade IV, which reflects a high grade but undifferentiated cancer. There are three factors used to calculate the grade of a breast tumour. First, the percentage of tubules or acini in the tumour: score = 1 if these make up the majority of the tumour (> 75%); score = 2 if the extent is moderate (10%–75%); and score = 3 if little or none (<

10%). Second is the nuclear pleomorphism: score = 1 if there are small or regular uniform cells; score = 2 if there is a moderate increase in size and variability; and score = 3 if there is pronounced variation. The final aspect is mitotic count: score = 1 if mitotic count is < 9, score = 2 if mitotic count is between 10 and 19, and score = 3 if mitotic count > 20. Mitotic cells are counted using 10 high-power field light microscopy with a field diameter of 0.58 mm. The image area of the digital scans of tumour sections should be 512 × 512 μm2. The tumour grade of a case is determined by summing the scores of the different factors. Therefore, total factor scores of 3–5 are designated tumour grade I; the prognosis for these tumours is good. A total score of 6 or 7 is classed as tumour grade II, and the prognosis is moderate. Total scores of 8 or 9 are grade III tumours, and the prognoses are poor [72,73]. This tumour grading system has been incorporated into the Nottingham Prognostic Index (NPI) [74].

1.7.3.2 Tumour Size

A key factor for predicting BC behaviour and patient outcomes is the size of the tumour. In IBC, the size of the tumour has a significant effect on the prognosis [75]. In particular, it is a robust predictor of nodal metastasis [76]. For this study, tumours were categorised according to their size. Those tumours less than 20 mm were classified as small tumours, and anything exceeding that was deemed to be a large tumour. The size of the tumour is integrated into the NPI [74].

1.7.3.3 Lymph Node Stage

Another prognostic factor for IBC is lymph node (LN) staging. Particularly during the early stages of disease, the correlation between LN stage and DFS and BCSS is robust [77]. For this study, cases were allocated to one of three discrete groups depending on LN staging: 1) LN stage I, no LN was involved; 2) LN stage II, between one and three LNs were involved; and 3) LN stage III, more than three LNs were involved [78,79]. There is a clear

relationship between LN staging and survival outcomes for IBC patients, with lower LN staging being associated with better prospects of survival. Approximately 83% of BC patients with LN stage I survive at least five years [80]. This LN staging system is integrated into the NPI [74].

1.7.3.4 Nottingham Prognostic Index

The NPI is a survival-setting and prognostic stratification tool for the assessment of IBC patients [78]. As previously mentioned, tumour grade and size and LN stage are the prognostic parameters that form the NPI, conferring the tool with robust prognostic power. Each patient's NPI score is based on her clinicopathological tumour and LN scores; the formula used to calculate NPI is as follows:

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

Using the results of the calculation, NPI is categorised as being 1) good, 2) moderate, or 3) poor.

The NPI has been updated. Originally, the good prognosis NPI group was defined as patients with an NPI score of < 3.4, while the moderate prognosis NPI category covered NPI scores of 3.4–5.4, and all scores exceeding 5.4 were the poor prognosis NPI group [74]. However, in the updated version, each of the original three groups is subdivided to give a total of six groups: NPI 2.08–2.4 is the excellent prognostic group (EPG); NPI 2.41–3.40 is the good prognostic group (GPG); NPI 3.41–4.40 is the moderate prognostic group-I (MPG-I); NPI 4.41–5.40 is the moderate prognostic group-II (MPG-II); NPI 5.41–6.40 is the poor prognostic group (PPG); and NPI > 6.40 is the very poor prognostic group (PPG) [81]. This study utilised the original NPI grouping system for all cases.

1.7.3.5 Lymphovascular Invasion

Lymphovascular invasion (LVI) reflects the presence of tumour cells in either or both the blood vessels and lymphatic spaces around the tumour. In IBC patients, LVI correlates strongly with BCSS and DMFS [82]. LVI is identified by microscopic examination of the morphology of the primary tumour that may or may not have endothelial-specific markers. LVI is a marker of metastatic potential if identified in the primary tumour. A number of studies have established the prognostic value of LVI [82–84].

1.7.4 Breast Cancer Predictive Factors

The purpose of evaluating predictive factors or biomarkers is to accurately forecast the prognosis and responsiveness to particular interventions. When used in conjunction with the clinicopathological data obtained from the prognostic parameters described above, predictive biomarkers are key to stratifying BC patients and managing their disease [85].

1.7.4.1 Hormone Receptors ER and PR

The American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) have issued guidelines that recommend that for IBC patients, the threshold of positive detection for oestrogen receptor (Erα or ER) and progesterone receptor (PR) is 1% of tumour nuclei [86]. ER provides a powerful predictive indication for extended adjuvant endocrine treatment; however, there is a lack of consensus as to the benefit of a lengthy hormone therapy regimen for patients who are ER+ [87]. Because premenopausal women are more susceptible to aggressive BC phenotypes, considering the prognostic relevance of ER positivity is of limited value in these patients. It is likely to be of greater benefit in post-menopausal BC patients. However, by considering ER status, the effect of age as an independent prognostic factor is disregarded, meaning the patient can benefit from intensive chemotherapy [88,89]. Clinical practice routinely tests for the co-expression of PR in ER+ patients. Patients' responses to adjuvant hormonal therapy can be predicted

based upon the co-expression status of ER and PR. While only 50% of ER+/PR- patients respond to hormone therapy, when both receptors are expressed, the response rate rises to 80% 64.

1.7.4.2 HER2

Encoded by the ERBB2 gene, the tyrosine-protein kinase receptor erbB-2 (HER2) is overexpressed in approximately 15% of BC patients. There is a strong correlation between HER2 overexpression and aggressive IBC phenotypes; the survival outcomes are typically poor [90]. HER2 positivity marks discohesive BC cells undergoing EMT and drives oncogenic potential [91]. Manifestations of HER2 positivity in IBC patients include abnormal polarity of cells, invasion, and malignant growth [92]. HER2+ BCs can be treated with trastuzumab (HerceptinTM), which is a humanized monoclonal antibody that targets HER2+ tumours. Thus, trastuzumab is an excellent example of the emerging discipline of precision medicine, where the molecular features of specific tumours can be specifically targeted by therapeutics. However, the effectiveness of trastuzumab declines as the disease progresses and is limited to women with early-stage BC [93]. A further category of BC describes tumours that lack detectable expression of ER and PR hormone receptors and HER2; this is classified as TNBC [94].

1.7.4.3 Breast Cancer Molecular Classification by Microarray

Distinct patterns of gene expression related to the different BC subtypes were first reported in 2001 [95]. Hormonal receptor-positive (HRE+) and hormonal receptor-negative (HRE-) types of BC tumours were identified from DNA microarray experiments using hierarchical clustering analysis using the Prediction Analysis of Microarrays (PAM50) gene panel [95]. Clustering analysis revealed that HRE+ could be sub-categorised according to tumour gene expression profiles that resemble the profiles of luminal epithelial mammary cells. Three classes of tumours were distinguished by clustering analysis of HRE types: 1) luminal 2)

HER2-amplified, 3) basal-like, and 4) normal breast-like. These phenotypes represent the expression profiles of ER, PR, and HER2 as luminal-A (ER+, PR+/-, HER2-), luminal-B (ER+, PR+/-, HER2+), HER2+ (ER-, PR-, HER2+), and basal-like (ER-, PR-, HER2-) [95,96]. Normal-like BC was not shown to be a true subtype. According to Reis-Filho and Tutt [97], ER, PR, or HER2 receptors are expressed in 15%–54% of basal-like tumours, as determined by microarray-based expression profiling studies. This suggests that, even though TNBC and basal-like cancers share a similar histological appearance, they are not necessarily molecularly similar. Therefore, these forms of disease should be distinguished from each other and treated accordingly [97]. Basal-like and non-basal triple-negative basal-like tumours were differentiated by exploring the positive immunohistochemistry (IHC) expression of epidermal growth factor receptor (EGFR+) and/or type II cytoskeletal 5 and 6 (CK5/6+) [98]. The biological features of basal-like BC and non-basal TNBCs differ from each other [98]. In 2007, the classification system was expanded to include the claudin-low-type BC. These tumours are characterised as under-expressing the genes for the proteins involved in cell-cell adhesion and tight junctions in epithelial cells [99]. A study by Dawson and colleagues [100] proposed a new classification of BC that used a different approach to stratify 2,000 samples of BC performed a thorough genome and transcriptome profile, arriving at ten integrative clusters (IntClusts) (**Table 1-1**). The profiles of the clinical and molecular features, genomic patterns, and expression of ER, PR, and HER2 were different for each **IntClust**.

Table 1-1: Features of the Integrative Clusters

IntClust	Defining Molecular Features	Clinical Features	BC Prognosis	Genomic Instability
1	17q23 amplification	High grade tumours	Intermediate	High
2	11q13/14 amplification	No distinct clinical features	Poor	High
3	Lack of CNA changes	Low grade, low LN+	Good	Low
4	CNA devoid	Low grade	Good	Low
5	ERBB2 amplification	Younger age at diagnosis, high grade, high LN+	Poor	Intermediate
6	8p12 amplification	No distinct clinical features	Intermediate	High
7	16p gain, 16q loss, 8q amplification	Older age at diagnosis, low grade	Good	Intermediate
8	1q gain, 16q loss	Older age at diagnosis, low grade	Good	Intermediate
9	8q gain, 20q amplification	High grade tumours	Intermediate	High
10	5q loss, 8q gain, 10p gain, 12p gain	Younger age at diagnosis, high grade, large tumours	Poor	Intermediate

- IntClust, integrative cluster; LN+, lymph node involvement.

1.8 Triple-Negative Breast Cancer (TNBC)

1.8.1 Pathogenesis, Incidence, and Risk Factors:

TNBC is the category of BC where expression of ER, PR, and HER2 are not detected by IHC evaluation. Between 15% and 20% of BC cases are categorised as TNBC [101]. These cancers are typically diagnosed in younger women. Because TNBC is an aggressive form of cancer [101] and is associated with a high frequency of molecular aberrations [102], these tumours are difficult to treat. The five-year survival of patients with TNBC is the worst across all races/ethnicities, whereas ER+/PR+/HER2- BC subtypes have the best prognoses [103]. The demographic factors that affect the incidence of cancer are reported to be highly variable depending upon age, race/ethnicity, and socioeconomic status (SES) [103,104]. Studies revealed that the factors most commonly associated with TNBC include lower SES and age < 40 years [104,105]. According to the results of one study of women under the age of 50 years with BC, almost twice as many had TNBC as the ER+/PR+/HER2- subtype. Furthermore, more stage 3 and 4 tumours and T2 and T3 tumours (tumours more than 5 cm in diameter) were detected in the women with TNBC. This cohort had the lowest incidence of grade 1 tumours [106].

1.8.2 Prognostic and Predictive Factors in TNBC

Because of the typically high-grade nature of TNBC tumours, factors that are usually used to make prognoses, including age, tumour size, grade, and lymph node stage, are of limited predictive value. Furthermore, TNBC tumours are often associated with metastases and recurrence [107]. To improve these outcomes, reliable and precise molecular characterisation of the tumour is essential, as that information can inform the selection of suitable therapeutic interventions.

1.8.3 TNBC Molecular Classification by Microarray

Lehman and colleagues used transcriptomic approaches to identify different subsets of TNBC [108]. The gene expression profiles of TNBC cases (N = 587) from 21 BC data sets were studied. Six discrete TNBC subtypes were detected using cluster analysis: 1) basal-like 1; 2) basal-like 2; 3) immunomodulatory; 4) mesenchymal; 5) mesenchymal stem-like; and 6) luminal androgen receptor. The transcriptomic expression for each of these categories differs from the others. The findings are consistent with TNBC cell line studies [108].

1.8.3.1 Basal-Like 1 (BL1) and Basal-Like 2 (BL2) TNBCs

The basal-like (BL) BC molecular intrinsic subtype constitutes up to 80% of TNBC cases. Using IHC, BL-TNBC demonstrated positive expression for cytokeratin 5 or 14 and EGFR [109,110]. Lehman and colleagues [110] stratified BL-TNBC into two subgroups (BL1 and BL2), both of which include enrichment of the cell cycle component, pathways of DNA replication, DNA damage response, and high expression of proliferation genes, including Ki67. Remarkably, EGFR, insulin-like growth factor receptor (IGFR), and EMT signalling pathways are particularly enriched in the BL2 subtype [110].

1.8.3.2 Immunomodulatory TNBC

The TNBC immunomodulatory (IM) subtype is classified by the enrichment of gene ontologies associated with important immune cell signalling pathways, such as T-cell, B-cell, and natural killer cell signalling, cytokeratin signalling, antigen processing and presentation, and the immune signal transduction pathway [110]. The IM TNBC molecular subtypes predominantly overlap with TNBC tumours with medullary histological differentiation, which is a unique and rare histological subtype of TNBC that, despite depicting a high-grade tumour, presents a favourable prognosis [111].

1.8.3.3 Mesenchymal and Mesenchymal Stem-Like TNBC

Genes involved in cell motility regulation, extracellular matrix receptor communication, and cell differentiation indicating enhanced Wnt/β-catenin and TGF-β/catenin signalling pathways are all enriched in both mesenchymal (M) and mesenchymal-like (MSL) TNBCs. However, the MSL-TNBC subtype differs from the M-TNBC subtype by high expression of genes related to the growth factor pathway (*EGFR* and platelet-derived growth factor (*PDGF*), calcium, and extracellular signal-regulated kinase signalling (ERK). The MSL-TNBC subtype shows an increased expression of genes related to the EMT cascade. The most notable difference between MSL-TNBC and M-TNBC is the deregulation of genes regulating cell proliferation. Studies have shown that EMT induces chemotherapy resistance in cancer cells [112]. Interestingly, TNBCs showing metaplastic histological differentiation are recognised as being highly chemo-resistant. These cancers mostly express M-TNBC and MSL-TNBC genomic profiles [110]. The claudin low subtype transcriptome profile overlaps with the MSL subtype, and both show deregulation of expression for claudin 3, 4, and 7 genes, along with upregulated expression of genes regulating EMT, such as metalloproteinase 2 (*MMP2*) [110].

1.8.3.4 Luminal Androgen Receptor (LAR) TNBC

Although TNBC by definition lacks expression of ER, gene expression profiling of TNBC revealed upregulation of androgen and oestrogen metabolic pathways. Upregulation of the androgen receptor (*AR*) is a major component of steroid-related signalling, which replaces ER expression in TNBC subtype. The downstream LAR-TNBC subtype is supported by positive nuclear expression of the AR using immunohistochemistry. Hierarchical clustering demonstrated that the LAR-TNBC subtype has a luminal-like gene signature resembling luminal ER+ BC. Overlap exists between the LAR molecular subtype and TNBC with

apocrine histological differentiation. Upregulation of phosphoinositide 3-kinase (*PI3KCA*) is also characteristic of the LAR-TNBC molecular subtype [110].

Upon further dissection of the TNBC molecular subtype and comparing the TNBC genome profile with the gene set symbolising the five intrinsic molecular subtypes of BC using the PAM50 gene panel (luminal A, Luminal B, HER2-enriched, basal-Like, and normal-like), Lehman and colleagues,[110] revealed that the TNBC genome profile greatly overlaps with the basal-like profile (81%). Furthermore, 10% correlate with the HER2-enriched subtype, 3% correlate with luminal B, and 1% correlate with luminal A. Further assessment for TNBC molecular subtypes and the PAM50 intrinsic classification showed that the basal-like phenotype mostly comprised each TNBC molecular subtype (BL1 99%, BL2 95%, IM 84%, and M 97%), excluding the LAR and MSL subtypes. The MSL-TNBC subtype has been shown to overlap with the basal-like phenotype by 50%, the normal-like phenotype by 28%, and 14% in luminal B [110]. LAR-TNBC demonstrated a 74% overlap with HER2-enriched and 14% with luminal B [110].

Refinement of the molecular classification of TNBC was achieved following reassessment of stromal and immune cell contributions to the tumour by combining histologic presentation, laser capture micro-dissection, RNA isolation, and gene expression profiling. Lehman and colleagues modified their original TNBC molecular classifications [113]. The modified subtyping provided evidence that tumour-infiltrating lymphocytes and stromal tumour cells predominantly compose the M-TNBC and MSL-TNBC molecular subtypes, respectively. Henceforth, a revised four molecular TNBC subtype classification (BL1, BL2, LAR, and mesenchymal) was presented [113]. The refined molecular classification revealed differences in tumour progression in terms of grade, stage, histopathology diagnosis, and age. The four subtypes exhibited significantly different responses to neo-adjuvant chemotherapy [113]. The study used 300 TNBC patient samples following neo-

adjuvant treatment. BL2 and LAR were shown to have 18% and 29% of patients achieving complete pathological response (pCR), whereas 41% of BL1 patients achieved higher pCR [113].

When TNBC tumours were classified based on integration of copy number alteration and gene expression profile, as stated earlier [114], approximately 25% of all TNBC cases comprised the IntClust 4 group, demonstrating a low level of genomic instability. However, poorly differentiated TNBC presented with the upregulation of TP53 mutations with intermediate genomic instability and showed a worse prognosis in the first five years of disease onset. Furthermore, absent CNAs with increases in tumour-infiltrating lymphocytes (TILs) indicated an improved clinical outcome [100]. Thus, it is evident to show genomic heterogeneity of the TNBC profile [110]. A comparison of the different classifications is illustrated in (**Figure 1-8**)

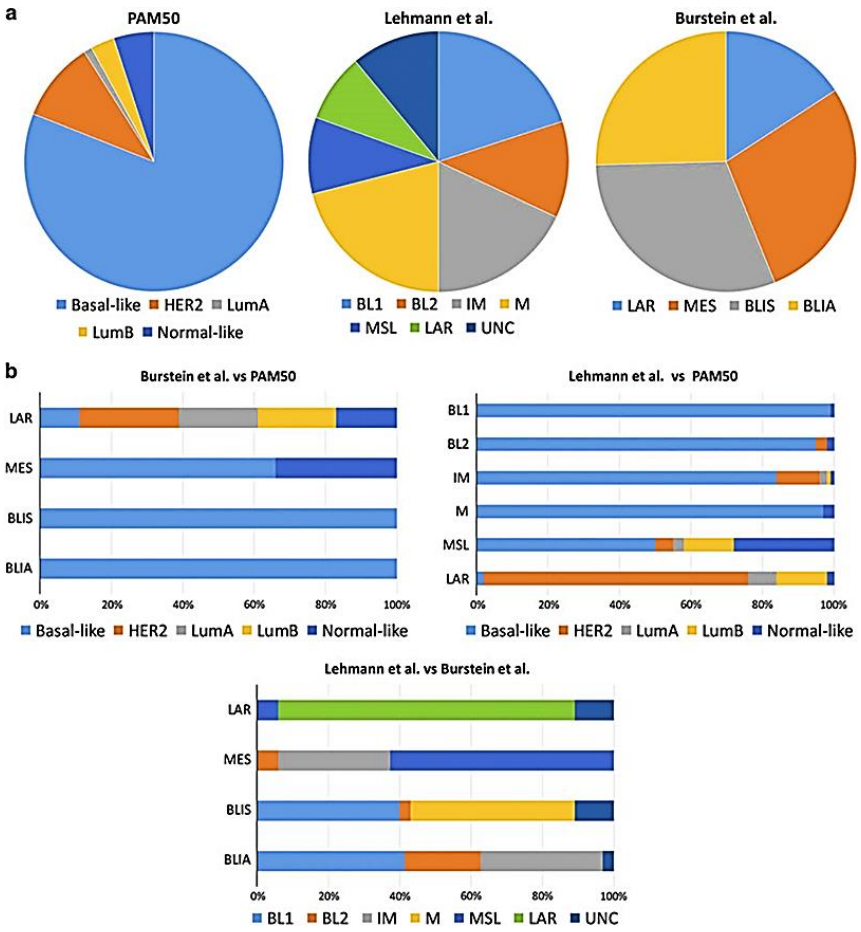


Figure 1-8 :TNBC classification in comparison with PAM50 classification

- Demonstrative graph of the different molecular sub- classification used to stratify TNBC. Taken from Pareja and colleagues [115] (A) The frequency of each subtype in PAM50 classification, Lehman and colleagues [108]. classification, and Burstein and colleagues [116] (B) Comparison between the different sub-classification;
    - Burstein and colleagues [116] versus PAM50[113]
    - Lehman and colleagues [108] versus PAM50[113]
    - Burstein and colleagues [116] versus Lehman and colleagues [108]

-

1.8.4 Gene Expression Analysis of TNBCs by Next-Generation Sequencing

Refined analysis of TNBC subtypes at the molecular level has been possible following the introduction of NGS technology. The aim of such studies is the identification of novel mutations in genes or pathways to enable the development of personalised therapies for improved patient outcomes [32]. One challenge is that TNBC is known to contain high driver mutation frequency [117], with mutations affecting the *TP53, PIK3CA*, and *GATA3* genes frequently detected among BC subtypes [118]. However, the incidence of these gene mutations between different BC subtypes varies. For instance, the *TP53* frequency in TNBC is 80%, while luminal A BC has a 12% mutation load [118]. The most frequently mutated gene in TNBC is *PIK3CA* [118] (**Figure 1-9).**

An expanded 358-gene NGS panel has identified new genetic variants in TNBC [119]. The JAX-CTP assay [119] was able to capture actionable variants, including single nucleotide alterations, small insertion and deletions (≤ 50 bp), and copy number alterations. Tumours with basal-like histologic features harbouring a prominent stromal lymphatic response were enriched in this study. TNBCs with at least one somatic mutation were detected in 90% of cases (18/20, 2.6 mutations per sample). The most common abnormality detected in this study was the amplification of the MYC oncogene, which was detected in 75% of TNBC samples. In addition, six additional amplified genes were detected, including *BCL2, CCDN, FGFRE, EGFR, NOTCH3*, and *ID2. TP53, AURKA*, and *KDR* gene mutations were detected in 30% of TNBC cases, and somatic mutations were observed in *PALB2, MET, PIK3CA, JAK3, FGFR, RET*, and *RB1* genes. Interestingly, TNBC with apocrine morphology had mutations in the *PIK3CA* gene in one of the two cases analysed, which is similar to the Wiseman and colleagues study findings [120].

Several studies showed similar findings for gene mutations and additional alterations from different TNBC cohorts [121–124]. The loss of *INPP4B* and *PTPN12* in up to 56% and 23% of TNBC cases, respectively, were described by Gwennier and colleagues [121]. A novel rearrangement between the *MERTK* and *TMEM87B* genes resulted in upregulation of the *MERTK* intraocular kinase and was identified using segmental transcript analysis [124]. Likewise, *FFRE-TACC3* gene fusion was identified in TNBC analysed cases, similar to identification detected previously in glioblastoma [124].

Persons and colleagues reported on a prospective study utilising a liquid biopsy method to study the mutation profile in TNBC [125]. The study compared tissue sections or biopsies to circulating cell-free DNA. Their NGS study used an assay for 236 cancer-associated genes to test isolated tissue sample DNA and an NGS assay using 27 genes to test cell-free DNA extracted from plasma. The NGS study revealed 70% concordance between patients who had adequate tissue and cell-free DNA [125].

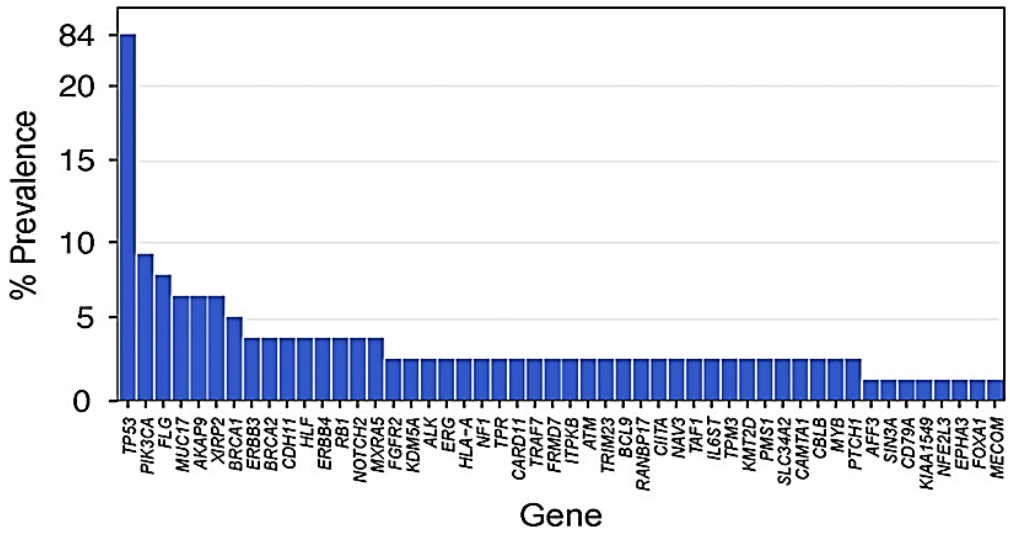


Figure 1-9: Prevalence of gene mutation affecting TNBC genome profile

- Demonstrative graph for the percentage of somatic mutation affecting TNBC-BRCA genomic profile. Taken from Pareja and colleagues [115].

1.8.4.1 Copy Number Aberrations in TNBC

Recurrent copy number aberrations (CNAs) typically involve genes that favour neoplastic development. Therefore, stratification of TNBC genomic profiles for CNA is desired. A study by Fidalgo and colleagues identified the most recurrent CNAs in 16 TNBC patient samples from lymph nodes to assess the mechanism of tumour progression [126]. Their findings suggested a connection between the somatic acquisition of the 19p13.11 susceptibility locus identified in the *BRCA1* gene as a modifier of BC [127] and gain in TNBC with incidence of metastasis to the lymph node [126]. Stevens and colleagues revealed that SNP variation in 19p13.11 is also related to TNBC tumour growth [128].

1.8.4.2 Somatically Acquired Mutations in TNBC

Loss of *BRCA1* or *BRCA2* function during TNBC tumorigenesis occurs in 10% of TNBC cases [129,130]. TNBC tumours with *BRCA1* mutations possess similar molecular profiles to basal-like BC, including high tumour grade, negative expression of ER, PR, and HER2, high expression of Ki-67, and high mutation rates of *TP53* [129,131]. TNBC tumours with BRCA mutations demonstrated overexpression of growth factors, *EGFR* and *C-MYC* amplification, and loss of *RAD51*-focus formations [132]. Increased expression of additional genes involved in DNA repair pathways, such as ataxia telangiectasia mutated (*ATM*) and *TP53*, are characteristically reported in high-grade BC tumours [129,131]. Remarkably, TNBC patients with BRCA1 mutations demonstrated a better response to platinum agents when combined with taxane/anthracycline-based neoadjuvant chemotherapy. Those patients with a TNBC diagnosis with BRCA mutations are referred to as having ‘BRCAness’ [132]. Poly-ADP ribose polymerase (PARP) inhibitors are currently being evaluated in clinical trials. PARP is a nuclear enzyme that interacts with intracellular pathways regulating DNA repair and cell survival through poly-ADP-ribosylation [133,134]. Continuous cell replication is permitted when the designed

molecular agent inhibits the activity of this enzyme. Double-stranded DNA breaks are repaired by *BRCA1/2* through homologous recombination, and *BRCA1/2* mutated TNBC tumour cells are incapable of repairing the damage. TNBC tumour cells with a *BRCA1/2* deficiency are more susceptible to PARP inhibitors and may present favourable responses when the synthetic lethality of cells is induced by PARP inhibitors [133,135].

1.8.5 Morphological Heterogeneity of TNBC

The histological, genetic, transcriptional, and clinical heterogeneity of TNBC represents a major clinical challenge (Figure 1-10). While many TNBC tumours are composed of high-grade lesions from IDC [136], a multitude of rare histologic special subtypes of BC also share the TNBC phenotype. High-grade lesions with a TNBC phenotype include breast carcinoma with apocrine differentiations, metaplastic breast carcinoma (MBC), and BC with medullary features [137]. Low-grade TNBCs, despite their rarity, comprise a diversity of lesions ranging from tumours with absent or indefinite metastatic potential to well-characterised special types of invasive tumours [138]. To distinguish low-grade TNBC tumours, two subsets have been suggested. The first subset of tumours has micro glandular adenosis (MGA), atypical microgranular adenosis (AMGA), and acinic cell carcinoma (ACC). Regardless of their low-grade morphologic presentation and better outcomes, they have a complex genomic landscape common to the TNBC phenotype. Another rare subset is tumours with salivary gland features of the breast, which are recognized by salivary gland-like morphologic presentation and are usually underpinned by specific genetic alterations [139]. Furthermore, solid papillary carcinomas with reverse polarity (SPCRP) and low-grade variants of metaplastic breast carcinoma (MBCs) are considered rare histological variants of low-grade TNBC subtypes [137].

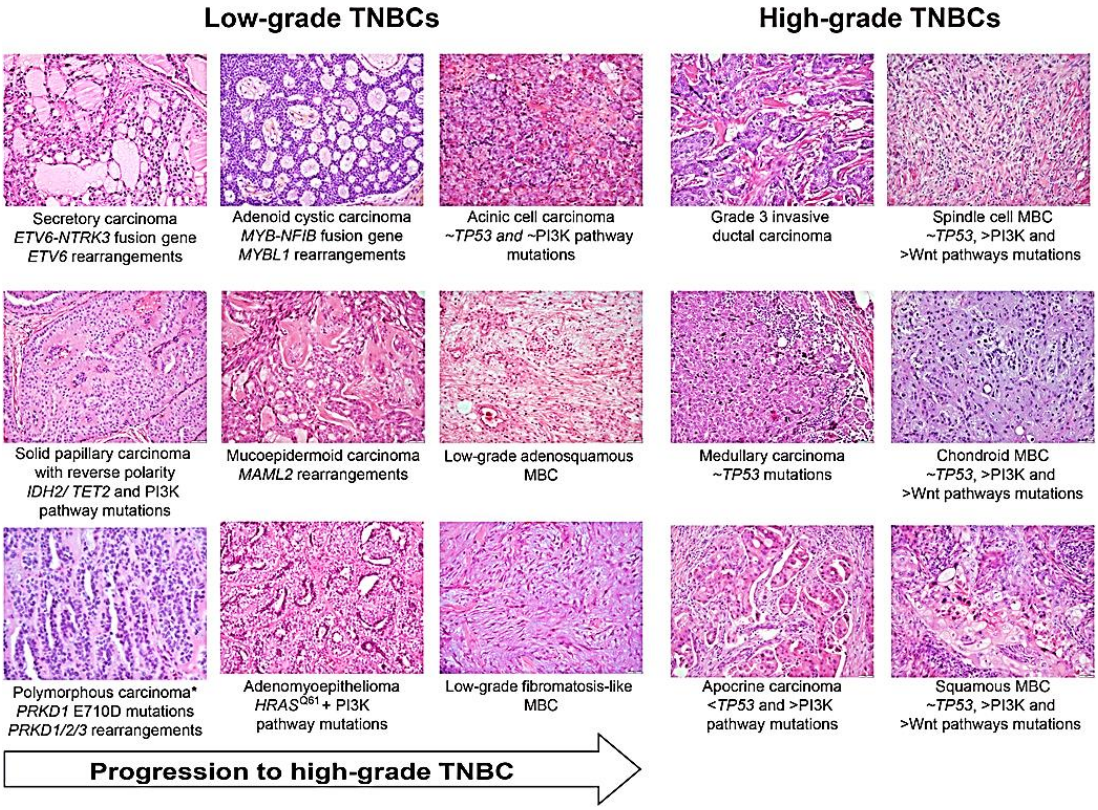


Figure 1-10: Spectrum of TNBC histological subtypes

- Demonstrative spectrum of micrographs of the low- and high-grade histological variants comprises TNBC. Taken from Gayer and colleagues [137].
  MBC; Metaplastic carcinoma of the breast

1.8.5.1 Low-Grade TNBC

In contrast to ER+ breast tumours, the identity of precursor lesions of TNBC remains unclear. The percentage of triple-negative (TN) DCIS is lower than their invasive counterpart, and invasive TNBC generally lacks a DCIS component [140]. The latter observation might be due to the rapid progression of TN DCIS towards invasive disease or that TNBC having a distinct precursor lesions than DCIS [140]. Additionally, studies on BC models showed that most cases of low-grade invasive carcinoma arise from low-grade DCIS, high-grade invasive carcinomas arise from high-grade DCIS, and the genetic profiles of DCIS and the invasive components are slightly different [141]. This is in contrast to carcinomas arising from other organs, like the colon and cervix, where the stepwise

carcinogenesis model is observed [142]. Taken together, TN DCIS might not represent the whole spectrum of TNBC precursor lesions, raising the potential that invasive TNBC can arise de novo.

However, identification of a new low-grade subgroup as a precursor lesion that subsequently becomes invasive high-grade TNBC phenotype tumours based on histologic, immunophenotypic, and genetic similarity was documented [137]. MGA is a major source of TNBC, as it was reported that MGA is a non-obligatory precursor for TNBC [143,144]. However, comparing the low prevalence of MGA indicated that not all TNBC originate from the aforementioned histological subtype [140]. Most TNBCs that arise in a background of MGA are of low-grade morphology [137].

Remarkably, the TCGA BRCA study suggested that there are no significant differences in mutation profiles between the rare MGA, AMGA, and ACC TNBC histological subtypes and more common TNBCs [139]. For example, *BRCA1* germline mutations are observed in MGA, AMGA, and ACC in addition to the common somatic genomic landscape between the entities. *BRCA1* biallelic inactivation by a combination of somatic mutations, germline mutations, and/or somatic loss of heterozygosity has also been observed in three ACC cases [145,146]. In addition, a study by Guenerini-Rocco and colleagues reported a case of MGA that was shown to have a somatic mutation coupled with the loss of heterozygosity in *BRCA1* and *TP53* [147]. These observations, together with data indicating that somatic alteration of *BRCA1* and *TP53* in mice may give rise to well-differentiated tumours [148], support the argument that concurrent loss of *BRCA1* and *TP53* function may lead to the development of high- and low-grade TNBC tumours [137]. Therefore, these observations indicate the existence of low-grade TNBC lesions that may eventually give rise to conventional high-grade TNBC [139]. As opposed to salivary gland-like tumours, specific somatic genetic mutations have yet to be defined in these lesions [145,147].

1.8.5.2 Salivary Gland-Like Tumours of the Breast

Primary TNBC tumours can mimic neoplasms of the salivary glands in histological and genetic differentiation. These tumours lack recurrent TP53 mutations, have lower CNAs, and harbour identified specific genetic mutations. Those specific genetic mutations reinforce some of these tumours regardless of origin site. This indicates that a genetic/phenotypic correlation is very high. Adenoid cystic carcinoma (ADCC) and secretory carcinoma are prototypical examples of this group, and both present with *MYB-NFIB* and *ETV6-NTRK3* fusion genes [149]. Additional salivary gland-like tumours that are possibly strengthened by the same oncogenic drivers that occur in the breast, but are not yet extensively studied, in comparison to other salivary gland counterparts, are polymorphous carcinoma, mucoepidermoid carcinoma, and adenomyoepithiolma [137].

1.8.5.3 High-Grade Histological Types of TNBC

1.8.5.3.1 Invasive Ductal Carcinoma Not Otherwise Specified

Invasive ductal carcinoma not otherwise specified (IDC-NST) constitutes the majority of TNBC cases [136]. Histologically, the tumours present with pushing borders, a high degree of necrosis, striking tumour-infiltrating lymphocytes, an increased mitotic index, and outstanding nuclear pleomorphism [150]. The Integrated Mutation Profiling of Actionable Cancer Target platform is an NGS technique that targets all coding regions of 220 frequent cancer-related genes. Weisman and colleagues reported on 22 IDC-NST cases, identifying an average of 3.41 non-synonymous mutations per case (range: 0–10) [120]. This was statistically significant in comparison with the TCGA cohort of TNBC (range: 0–7; mean = 2.37). *TP53*, *MLL2*, and *PIK3CA* were the most frequently mutated genes in the study, while other infrequent mutations were detected. Statistical significance included mutations in *JAK1, TBK1*, and *TGFBR2*. CNA analysis revealed a high frequency of amplification of *MYC, MCL1*, and *DDR2* observed in the analysis of IDC-NST genome expression [120].

1.8.5.3.2 Infiltrating Lobular Carcinoma

Infiltrating lobular carcinoma (ILC) tumours are described as having either dissociated cells from each other or forming single files or targetoid patterns around uninvolved ducts and are thought to originate from in situ lobular neoplasia. While this type accounts for 10% of BC [151], the ILC-TNBC phenotype is rare, as it accounts for less than 5% [152]. A recent study of 2,952 ILC tumours showed that only 44 (1.49%) cases were of the TNBC molecular phenotype [153]. Interestingly, an analysis of these 44 TNBC tumours with ILC features using the PAM50 genomic profiling technique showed that 48% were classified as luminal A, 3% were luminal B, 32% were HER2-enriched, and only 16% were basal-like [153].

1.8.5.3.3 Metaplastic Breast Carcinoma

Tumours presenting with squamous and/or mesenchymal differentiation comprise MBCs [154]. The tumour spectrum includes a majority of high grade, with conspicuous nuclear pleomorphism and increased mitotic activity. Chemotherapy resistance that leads to a poor outcome is mostly associated with this histological subtype [155]. The morphologic heterogeneity of MBC tumours is associated with inter- and intratumoural transcriptomic and genetic heterogeneity [156]. Different histological components within individual MBC cases may show distinct patterns of CNAs regardless of whether they are clonally related [157]. Therefore, MBC histological subtypes (e.g., spindle cell, squamous, matrix producing, and those with chondroid or osseous differentiation) are underpinned by distinct gene expression profiles. MBCs with spindle cells are of the claudin low intrinsic subtype, whereas the basal-like subtypes are frequently composed of squamous and chondroid histological presentations of MBCs [156]. Applying the TNBC molecular classification by Lehman and colleagues [110,157] , spindle cell and chondroid MBCs were often found in the mesenchymal subtype, while squamous MBCs encompass multiple classifications

[110,157]. MBCs show comparative genomic complexity and a mutation rate of *TP53*, as observed in TNBC [158]. Genes associated with the activation of the PIK3 and WNT pathways were found to be enriched and mutated in TN MBCs [158],[159]. Mutations in *PIK3CA* (29%), *PIK3R1* (11%), and *PTEN* (11%) predominantly affect the PIK3 pathway. TN MBCs have mutated genes in the WNT pathway that include *FAT1* (11%) and *AXIN1* (6%). Although β-catenin expression is consistent with Wnt pathway activation in a subset of MBCs, mutations in the *CTTNB1* gene are mostly infrequent to absent [158–160]. Weisman and colleagues studied one single case of TN MBC genome profile by NGS [161]. Their analysis indicated that the average number of non-synonymous mutations discovered in TN MBC cases was three, which was no different in comparison to cases of TN MBCs from the Cancer Genome Atlas. However, genetic alterations were detected in *TP53, KDM6A*, and *MLL3* [161].

1.8.5.3.4 Carcinoma of the Breast with Medullary Features

Distinct tumour borders with syncytial growth patterns and brisk tumour-infiltrating lymphocytes are the histological characteristics of what was previously known as medullary carcinoma of the breast [162]. However, in the current classification of the World Health organisation (WHO) for BC tumours, BC with medullary features is currently classified under the NST subtype [163].

Remarkably, most of what has been previously known as typical medullary BCs have TN characteristics [164] with a rate of TP53 mutations similar to that observed in the TN-IDC-NST subtype [165]. TN medullary carcinomas are associated with favourable outcomes despite possessing increased mitotic activity and concerning cytological features [162]. Although a TN-atypical medullary carcinoma may not achieve all diagnostic criteria, it has a similar favourable outcome to TN medullary carcinoma of the breast [166]. Accurate distinction between atypical and typical medullary carcinoma is challenging and reflected

in poor inter-observer agreement [162]. Therefore, TNBC cases with medullary features are treated according to a common TNBC treatment protocol. The brisk lymphocytic infiltration of these tumours may be the result of a good prognosis for patients diagnosed with these histological subtypes [167,168].

1.8.5.3.5 Carcinoma of the Breast with Apocrine Features

Apocrine differentiation BC is diagnosed when the tumour epithelium presents with large cells with prominent eosinophilic flocculent cytoplasm, markedly defined cell boards, and large nucleoli with prominent macro nucleoli [169]. Surprisingly, TNBC with apocrine differentiation is most frequently diagnosed in older patients. However, these tumours mostly have favourable prognoses, as the tumours are rarely grade 3 at diagnosis [170]. NGS transcriptomic analysis of TNBC with apocrine differentiation cases by Wiseman and colleagues revealed an average of 4.5 non-synonymous mutations per case [120]. Their transcriptomic analysis demonstrated that the most common mutated genes were *PIK3CA* and *NF1*. Their remarkable discovery was that cases with apocrine differentiation have fewer frequent *TP53* mutations, and therefore, this tumour subtype has a lower level of genetic instability in comparison to other histologic sub-classifications of TNBC [120].

1.8.6 TNBC Heterogeneity: Consequences for Treatment Decisions and Future Targeted Therapeutic Approaches

A combination of treatments, including surgery, radiation, and chemotherapy, are the current management options for TNBC patients. Combination therapy of an anthracycline, such as doxorubicin, in combination with taxane (generally paclitaxel), whether in an adjuvant or neoadjuvant setting, is the most common therapeutic regimen [171]. TNBC patients present with higher response rates to neoadjuvant therapy when compared with patients harbouring ER+ tumours [172]. However, TNBC patients with residual diseases have a poor prognosis and shorter recurrence-free survival compared to those with other

BC subtypes carrying residual diseases. The high degree of tumour heterogeneity and the lack of identification of common targetable genomic mutations in TNBC have hampered the development of precision therapeutics for TNBC [116,172,173].

As noted earlier, the goal of precision medicine in TNBC is to identify therapeutically targetable novel tumour molecular features to improve patient outcomes. One promising example is the use of antibody-drug conjugation, whereby a tumour-selective/specific monoclonal antibody is linked to an effector molecule to enable more precise drug delivery. Glembatumumb vedotin, for instance, is a combination of the cytotoxic drug monomethyl auristatin and an antibody with gpNMB glycoprotein, which is often overexpressed in TNBC tumour cells [174,175]. The second novel antibody drug conjugate used is sacituzumb govitecan, which combines the active metabolite of irinotecan, SN-38, with an anti-torp-2 monoclonal antibody specific to a glycoprotein expressed in TNBC [176].

As previously discussed, 10% of TNBC patients have the BRCA1/BRCA2 mutation in which tumour cells lack the homologous recombination DNA repair mechanism [129]. Most BL1 and BL2 TNBC patients harbour BRCA mutations [130]. Therefore, PARP inhibitors that target the DNA repair pathway might demonstrate efficacy in those TNBC subtypes. Monotherapies, such as veliparib and olaparib, are PARP inhibitors currently in clinical trials for TNBC, which have shown significant improvement in results in the overall response rates when combined with cisplatin [176].

Frequent expression of EGFR has been shown in basal-like BC. A recent study indicated that, while *EGFR* mutations are a rare event, a high *EGFR* copy number was an independent prognostic factor for poor survival in TNBC patients [177]. This can be important when selecting patients for anti-EGFR targeted therapy, specifically TNBC patients from the Lehman BL2 subtype [110].

Recent transcriptomic studies indicate that TNBC tumours with the IM subtype presenting with high levels of tumour-infiltrating CD8 T cells may exhibit chemotherapy sensitivity and improved prognosis, while TNBC tumours exhibiting macrophages have relatively worse outcomes [178]. Programmed cell death ligand 1 (PD-L1) is more commonly expressed in TNBC than in non-TNBC tumours and is responsible for regulating the immune checkpoint. PD-L1 expression differs according to cellular location: 64% in tumour cell membranes, 80% in the cytoplasm, and 93% in the stroma [179]. Interestingly, expression of PD-L1 in the cytoplasm and stroma is associated with improved outcomes compared to PD-L1 expression in the tumour cell membrane [179]. Pembrolizaumb, an anti-PD-L1 antibody, was recently tested in clinical trials and was found to improve the survival rate by 18.5% in metastatic TNBC, strengthening the potential of immune checkpoint therapies against TNBC [180].

AR-targeted therapy for the treatment of TNBC patients expressing AR is currently being investigated in a number of clinical trials [181]. The spectrum of trials includes AR antagonists (bicalutamide and enzalutamide), small molecular inhibitors of AR:HSP90 function (ganetspib), biosynthesis inhibitors (abiraterone acetate), and AR agonists (4-OH-testosterone) [181]. There has been emerging evidence that when PI3K inhibitors are combined with AR antagonists, they exhibit higher efficacy than that of monotherapy in AR-positive TNBC patients [182]. If AR therapy is confirmed to be effective, then AR expression evaluation will be a common routine clinical practice for TNBC patients. Identifying patients with the LAR-TNBC subtype by gene expression profiling is potentially useful for treatment management [32] (**Figure 1-11**)

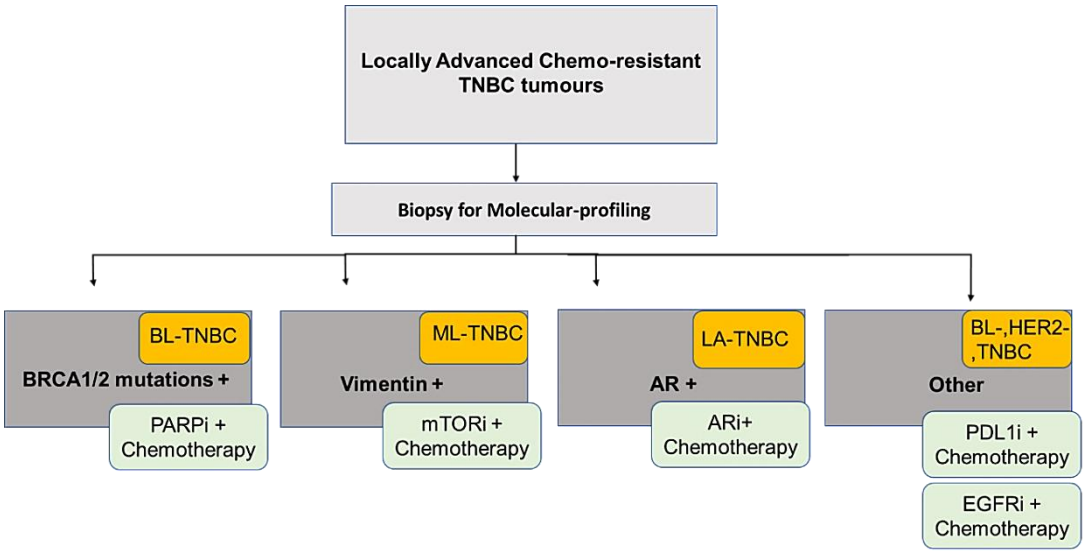


Figure 1-11: Spectrum of current clinical trials for TNBC treatment

Demonstrative graph illustrating the current clinical trials for TNBC treatment based on the sub classification previously discussed, nevertheless, the new targeted therapies currently in development. Taken from Le Du and colleagues [130].

Abbreviations: AR, androgen receptor; ARi, androgen receptor inhibitor; BL, basal-like; EGFR, epidermal growth factor receptor; EGFRi, epidermal growth factor receptor inhibitor; I, immunomodulary; LA, luminal/apocrine; ML, mesenchymal-like; mTORi, mTOR inhibitor; PDL-1i, PDL-1 inhibitor.

1.9 Hypothesis and Aims of the Work

We hypothesised that the prognosis of TNBC is variable and reflects several phenotypic and genotypic characteristics of the tumours. To test this hypothesis, this project analysed both transcriptomic and proteomic profiles of a large, well-described TNBC cohort and related the findings to key clinico-pathologic parameters and identified novel potential mechanisms of metastases.

This study aimed to:

- Collect and catalogue a large series of TNBC cases diagnosed in Nottingham over a 20-year period with associated clinical outcome data to construct tissue microarrays (TMAs) from the Nottingham TNBC cohort
- Utilise high-throughput RNA sequencing transcriptomic analyses of the Nottingham TNBC cohorts to identify differentially expressed genes that predict early metastases.
- Identify a novel prognostic biomarker signature to improve prediction of TNBC outcomes.
- Understand the mechanistic roles and clinicopathological and prognostic value of periplakin (*PPL*) in TNBC.
- Study the mechanisms and consequences of loss of E-cadherin expression as a mediator of metastasis in TNBC as compared to other BC subtypes.

Chapter 2 : Materials and Methods

2.1 Cohort Identification and Data Collection

This study used a retrospective well-characterised series of primary invasive (Stage I–III, excluding T3 and T4 tumours) TNBC samples (n = 393) obtained from patients who presented at Nottingham University Hospitals between 1987 and 2006 (**Figure 2-1**) (Table 2-1). The retrospective cohort clinicopathological data, including patient age at diagnosis, histological phenotype, tumour size, tumour grade, nodal stage, lymphovascular invasion (LVI), and Nottingham Prognostic Index (NPI), were collected from patients' medical records. The mean patient age was 51 years (range: 26–71) and tumour sizes in diameter at the time of presentation ranged from 0.25 to 8.00 cm (1.5–2.8 cm within the interquartile), with a mean tumour size of 2.2 cm. All patients were managed in a standard manner, where all patients underwent a mastectomy or wide local excision, as decided by disease characteristics or patient choice. Adjuvant therapy was based on the NPI and ER status. Prior to 2000, not all ER-negative BC patients were offered chemotherapy. HER2 status was neither assessed in routine practice nor did it influence treatment decisions during that time. Survival data were collected using a prospective method and included BCSS, defined as the interval from the date of primary treatment to the time of death due to BC, in addition to any disease-free interval (DFI), identified as the interval from the date of primary treatment to the first loco-regional recurrence. Additionally, distant metastasis-free survival (DMFS) was defined as the duration from the date of primary diagnosis to the appearance of a distant metastasis. All these parameters were measured in months. Long-term follow-up and treatment data were available and collected from the Nottingham database system. The Nottingham TNBC patient cohort (N = 393) showed that 44% of the patients received classic cyclophosphamide, methotrexate, and 5-fluorouracil (CMF) adjuvant chemotherapy, 29% of the patients received 5-flourouracil, epirubicin, cyclophosphamide (FEC) adjuvant chemotherapy, and 20% received doxorubicin and cyclophosphamide (AC)

adjuvant chemotherapy. Furthermore, 71% of the Nottingham TNBC patient cohort received radiotherapy (RT) treatment. More importantly, none of the Nottingham-TNBC cohort received neoadjuvant therapy.

A subset of the Nottingham TNBC patient cohort (N = 126) was used for NGS analysis and was referred to in this study as the 'Nottingham transcriptome TNBC cohort'. Within this cohort, six patients had normal tissue samples harboured from TDLU areas, one patient had five samples, seven patients had four samples, and four patients had two samples. Selection of this subset cohort was based upon the availability of formalin-fixed paraffin-embedded (FFPE) tissue blocks with a sufficient tumour burden evaluated by the Nottingham Breast Cancer Research Group head pathologist. The Nottingham transcriptome TNBC cohort comprised a well-characterised series who presented at Nottingham University Hospitals between 1987 and 1998. The mean age of the patients was 48 years (range: 27–69). Tumours size ranged in diameter at the time of presentation, with a mean tumour size of 2.41 cm (range: 0.25–8.0 cm). The Nottingham transcriptome TNBC cohort showed that 51% of the patients received CMF adjuvant chemotherapy and 66% received radiotherapy (RT).

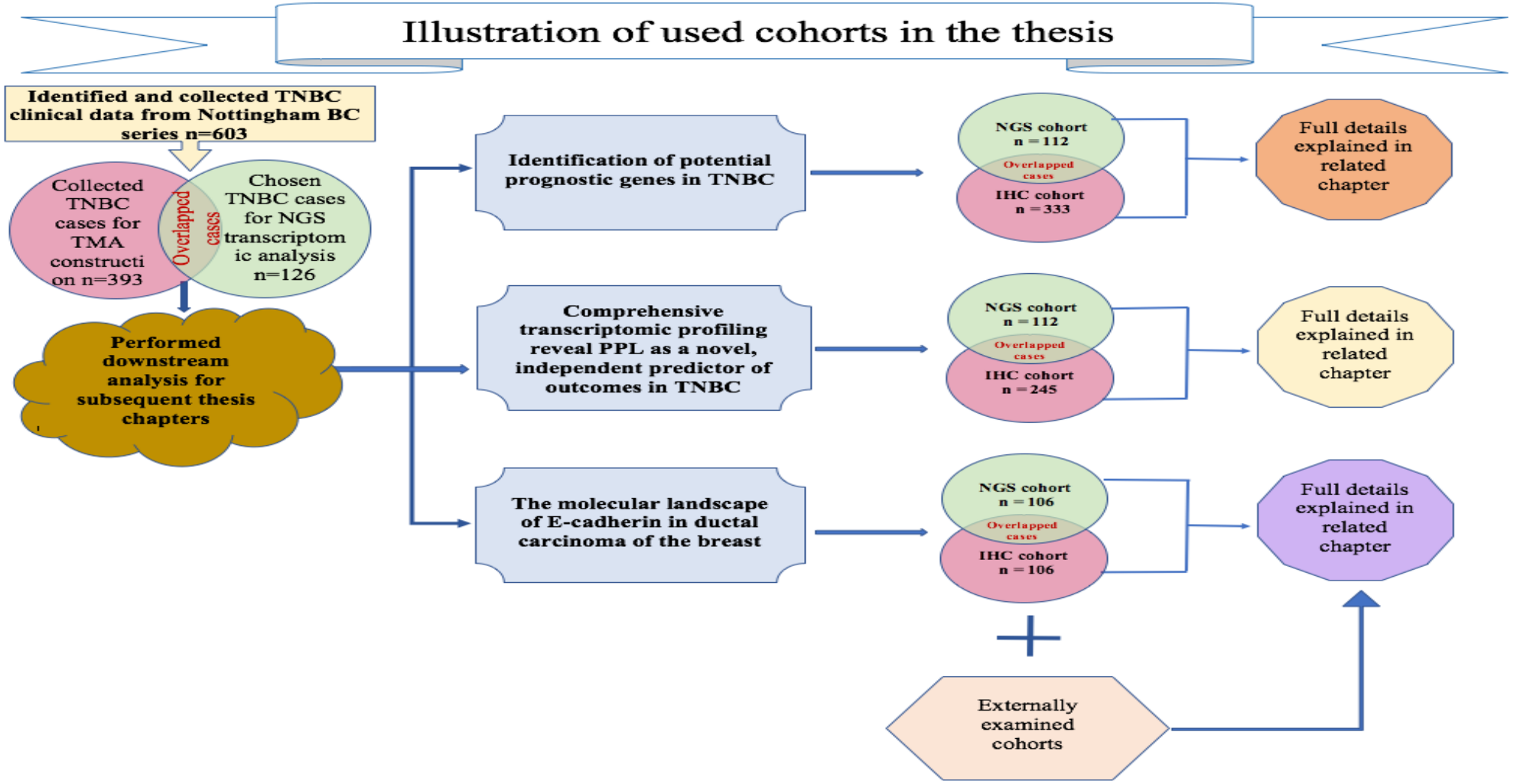


Figure 2-1: Descriptive explanation of used cohorts in the thesis

Table 2-1: Patients cohort

Clinicopathological Parameter	NGS cohort (n=126) No (%)	IHC Cohort (n=393) No (%)	Clinicopathological Parameter	NGS cohort (n=112) No (%)	IHC Cohort (n=393) No (%)
Age (years)			**Distant Metastasis**		
< 50	70 (56)	183 (47)	Yes	51 (41)	115 (30)
≥ 50	56 (44)	210 (53)	No	74 (59)	276 (70)
Tumour Size (cm)			**Recurrence**		
< 2	36 (28)	167 (43)	Yes	53 (46)	133 (34)
≥ 2	90 (72)	225 (57)	No	68 (54)	256 (66)
Grade			**Breast Cancer Specific Survival**		
1	1 (1)	5 (1)	Alive	70 (55)	251 (64)
2	6 (5)	31 (8)	Died of BC	47 (38)	112 (29)
3	118 (94)	356 (91)	Died of other causes	9 (7)	29 (8)
Nodal Stage			**Nottingham Prognostic Index**		
1	76 (61)	264 (67)	Good Prognosis	4 (3)	20 (5)
2	31 (27)	89 (23)	Moderate Prognosis	84 (67)	287 (73)
3	18 (12)	38 (10)	Poor Prognosis	38 (30)	84 (22)
Tumour histological subtype			**Chemotherapy**		
Ductal No special type (NST) & medullary subtype	121 (95)	376 (96)	Yes	63 (51)	240 (61)
			No	49 (40)	123 (31)
Lobular	2 (2)	4 (1)	No treatment information	7 (10)	30 (8)
Miscellaneous and special subtypes	3 (3)	13 (3)	**Radiotherapy**		
Vascular Invasion			Yes	83 (66)	277 (70)
Yes	48 (38)	107 (27)	No	31 (25)	97 (25)
No	78 (62)	286 (67)	No treatment information	12 (9)	19 (5)

2.2 Next-Generation Sequencing (NGS)

2.2.1 RNA Extraction, Library Construction, and NGS Analysis

Representative FFPE blocks of cases were retrieved, and invasive tumour cells were macro-dissected from unstained multiple tumour blocks of each case where tumour cellularity (i.e., tumour burden) was at least 50% of the tissue section area (**Figure 2-2**) illustrates the experimental workflow. Block identification and RNA extraction were conducted by Dr Mohammad Aleskendany prior of the author joining the lab. In brief, haematoxylin and eosin (H&E) sections of tumour blocks were microscopically assessed for the invasive tumour burden and to guide tumour macro-dissection. The chosen FFPE blocks were sent to the USA to carry out RNA extraction and sequencing. Four 10-µm unstained tissue sections were used per case. Macro-dissected tissues were deparaffinised, rehydrated, and centrifuged to remove excess ethanol. RNA was extracted using the Omega Mag-Bind XP formalin-fixed paraffin-embedded RNA isolation kit (M2595-01, Omega) and Kingfisher Flex magnetic particle separator (Thermo Fisher Scientific), as per the manufacturer's instructions. RNA was measured with a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific). First-strand cDNA synthesis was performed on approximately 100 ng RNA at 25°C for 10 min, 42°C for 15 min, and 70°C for 15 min using random hexamers and ProtoScript II Reverse Transcriptase (New England BioLabs, Ipswich, MA). Second-strand synthesis and RNA sequencing libraries were prepared using the Illumina TruSeq RNA access library kit (RS-301-2002, Illumina) and sequenced on an Illumina HiSeq 2500 using the PE75 run chemistry. The targeted read count was 60M total reads per sample. Library preparation and high throughput sequencing was performed at the Emory Integrated Genomics Core Facility (Emory University, Atlanta, USA). To elaborate, the RNA templates were removed, and a second replacement strand was generated by incorporating dUTP to generate double-stranded cDNA. The blunt-ended cDNA was cleaned up from the

second-strand reaction mix with Mag-Bind RxnPure Plus magnetic beads (Omega Biotek, Norcross, GA). The 3' ends of the cDNA were adenylated, followed by the ligation of indexing adaptors. PCR (15 cycles of 98°C for 10 sec, 60°C for 30 sec, and 72°C for 30 sec) was performed to selectively enrich DNA fragments with adapter molecules on both ends. The library was qualified using an Agilent 2200 Tape Station and quantified using the QuantiFluor dsDNA System (Promega, Madison, WI). A four-plex pool of libraries was made by combining 200 ng of each DNA library. The pooled DNA libraries were mixed with capture probes to target regions of interest. The hybridisation was performed by 18 cycles of 1 min incubation, starting at 94°C, and then decreasing by 2°C per cycle. Streptavidin-coated magnetic beads were used to capture probes hybridised to the target regions. The enriched libraries were eluted from the beads and subjected to a second round of hybridisation. The enriched libraries were amplified by 10 cycles of PCR (98°C for 10 sec, 60°C for 30 sec, and 72°C for 30 sec), followed by Rxn Pure magnetic bead clean up. The final libraries were validated using Agilent High Sensitivity D1000 Screen Tape on an Agilent 2200 Tape station instrument. The size distribution of the library ranged from approximately 200 to 1000 bp. Libraries were normalised, pooled, and subjected to clustering, and pair-read sequencing was performed for 75 cycles on a HiSeq2500 instrument (Illumina, Inc. San Diego, CA, USA), according to the manufacturer's instructions. Raw FastQ sequence read files were quality assessed and adapter processed using the trim galore wrapper for Fastqc and Cutadapt reads with phred scores > 30 retained [183]. The resultant quality trimmed reads were aligned to the hg38 (GRCh38.83) build of the human genome using the STAR aligner [184]. Transcript abundance quantification was performed using HTSeq [185]. Normalisation of genes was determined using transcripts per kilobase million (TPM) [186]. Bioinformatic analysis was conducted with Professor Nigel Mongan using secured Linux clusters located at the University of Nottingham School of

Veterinary Medicine and Science, NUH Trust City Hospital site and Biodiscovery Institute. Only one tumour sample per patient was included in downstream analyses by random selection. Normal samples were excluded from all downstream analyses.

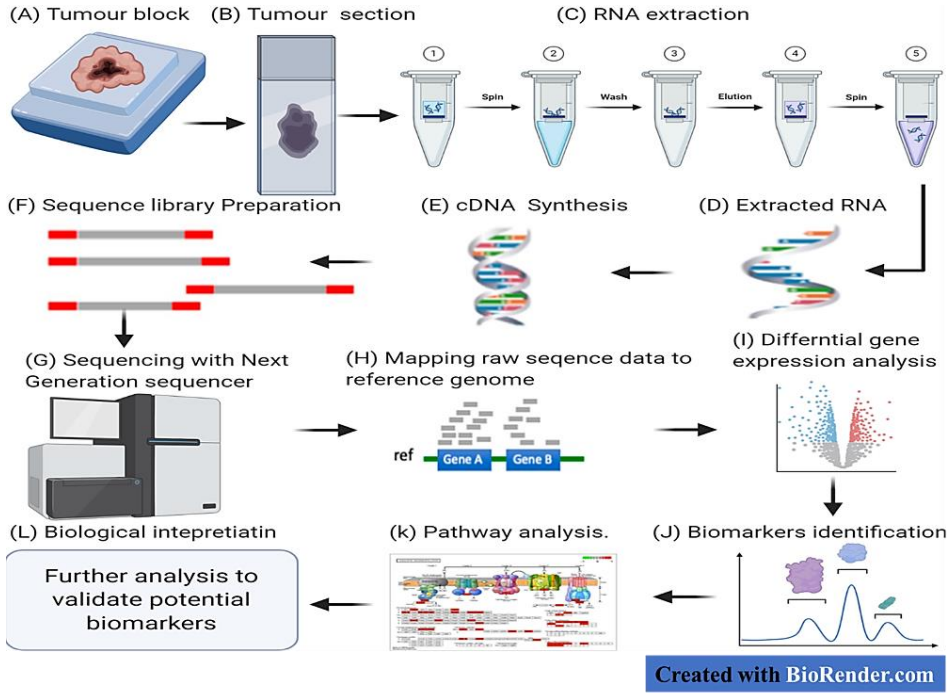


Figure 2-2 :Next generation sequencing flowchart

- This figure illustrate the workflow of RNA-seq and data analysis used in this study: (A) chosen tumour block, (B) four 10µm section were prepared for RNA extraction, (C) Sample preparation steps, (D) Extracted RNA strands, (E) CDNA synthesis, (F) Library preparation and Cutadapt using Trim galore, (G) Next Generation Sequencing reads sequenced with the Illumina Hiseq 2500, (H) transcripts were aligned to the hg38 (GRCh38.83) build of the human genome using the STAR aligner, (I) Differential gene expression analysis using Edge R, (J) Identification of biomarkers by statistical approaches (Edge R or Artificial neural network), (K) Pathway analysis using Webgestalt for Gene Ontology (GO) analysis and Gene Set Enrichment Analysis (GSEA), (L) Biological interpretation and validation of findings using public domain data and Immunohistochemistry.

2.2.2 Differential gene expression (DGE) analysis.

Data gleaned from RNA seq technology are valuable for identifying prognostic candidates. TNBC research can potentially use this information to promote the clinical management of TNBC patients. To statistically quantify potential differences in the gene expression between tested groups, DGE analysis was performed using the RobiNA implementation of EdgeR software [187]. DGE was filtered according to fold change (> ± 2). The threshold for statistical significance was set as $q \leq 0.05$.

2.2.3 Pathway Analysis

To identify canonical pathways that regulate the DGE, gene set enrichment analysis (GSEA) and the Webgestalt online portal (http://www.webgestalt.org/option.php) were used. Pathway analysis was based on DGE originated by the RobiNA analysis, and only those hits with a fold change ≥ ± 2 and significant q values ≤ 0.05 were included [188].

2.3 Tissue Microarray (TMA) Construction

Having the capacity to manage several hundred samples in a single slide and being able to apply techniques such as immunohistochemistry (IHC), TMA enable high-throughput analysis of biomarkers expression [189]. TMA is an economical technique, saving on tissue, reagents, and time, making it quicker and easier to perform IHC and other staining protocols and to assess the results [190]. Furthermore, it improves laboratory standardisation, as more samples are treated consistently [191]. Undertaking TMA involves some careful preparation: the study cohort to be arrayed must be obtained, the receiving array blocks need to be prepared, the array grids designed, the individual array blocks built, and the array blocks readied for sectioning [191]. The most important parts of building a TMA are preparing the donor blocks and devising a suitable TMA design, which are time-consuming processes. TMA construction was performed by the author and a

member of the Nottingham Breast Cancer Research Group pathologists (Dr. Michael Toss).

(**Figure** 2-3) illustrates the experimental workflow.

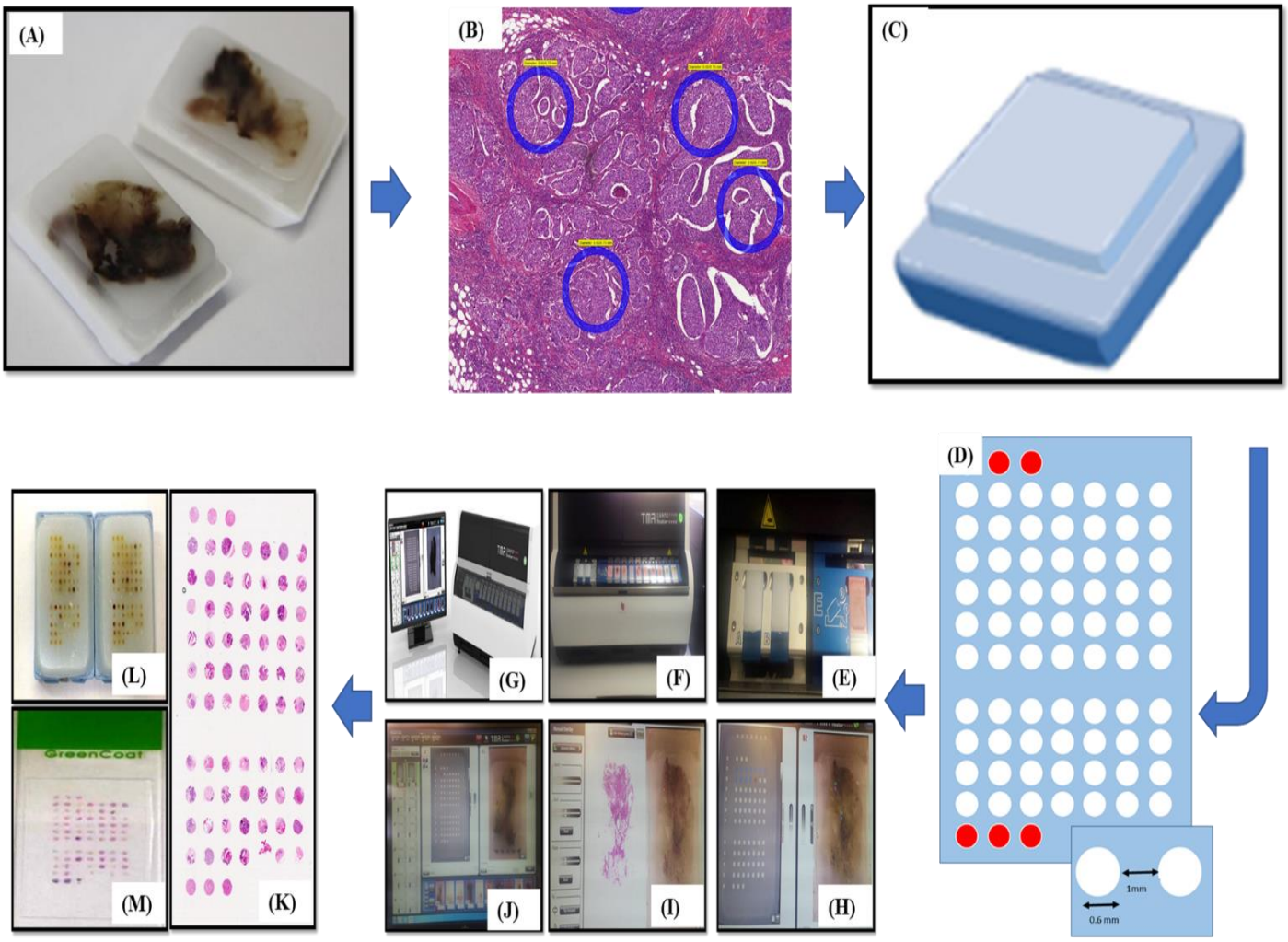


Figure 2-3:TMA construction Protocol

- This figure illustrated the steps taken during TMA construction; (A)Selected tumour block, (B) Selection of 1 mm cores within abundant tumour tissue, (C) Tissue cassettes block containing the paraffin, (D) Layout of 1mm core diameters recipient block, E) TMA machine and screen, F) Loaded machine with donor and recipient blocks, J) Blank recipient blocks loaded in TMA machine, K) Screen showing images of donor block (right one in the screen and recipient block (left one) mapped with selected layout , L) Matching the scanned images with the relevant donor block, M) Marked suitable sites for coring in donor block, N) Constructed TMA blocks, O) H&E stained sections and P) sections are scanned with high resolution Nanozoomer scanner (x20).

2.3.1 Tissue Handling, Processing, and FFPE Block Retrieval from Nottingham Health Science Biobank (NHSB)

To identify suitable cases of TNBC that could be used to build the TMA, slides from various data storage pathology systems were selected and reviewed. An FFPE tumour block (donor) that was characteristic of the specimens from each patient was obtained from the tissue block archives. Those blocks that were ≥ 4 mm in depth bearing viable tumour tissue with an adequate burden of tumour and that had not been sectioned frequently for earlier research were preferred. Once selections were made, specific request applications were filled and signed to comply with NHSB regulations prior to any experiment.

Once the applications were approved and FFPE blocks were retrieved, a clean full-face 4-μm thick section was cut from each block and then stained with H&E. To confirm the presence of tumour and distinguish which areas would be most appropriate for the TMA sampling, the pathologist (Dr. Michael Toss) reviewed each slide using light microscopy. Blocks that were < 4 mm thick or contained an insufficient quantity of or no tumour were rejected in favour of another archive block for the same patient. To avoid the potential risk of exhausting a patient's diagnostic material, where there were no alternative thicker tumour-rich blocks for a particular patient, that patient was eliminated from the study. In total, 120 blocks were used to build the TNBC TMA. The remaining TNBC cases had been arrayed earlier (N = 273). Despite a difference in the number of cases included in the TMA and the number

of cases identified originally, exclusion was random. Thus, no bias was introduced during data analysis.

2.3.2 TMA Construction Protocol for TNBC Cases

From each case, the most suitable FFPE block was selected. Suitability was determined by the abundance of TNBC and the adequacy of the thickness of the tissue for the purpose of arraying. The TMA core diameters were set to 1 mm.

2.3.3 Digital Image Marking

All scanned H&E images were marked for TMA cores using Panoramic Viewer Software v. 1.15.4 (3D HISTECH®, Budapest, Hungary). A maximum number of suitable cores for each image was marked.

2.3.4 Recipient Block Construction

Building of the TMA was achieved using ordinary paraffin wax-based blocks that had been prepared by pouring paraffin into blocks (5–10 mm deep) and then extruding the tissue array cores into the block. To prepare recipient blocks, plain molten paraffin wax (56°–58°C) was poured into a mould. This was then topped with a histopathological tissue cassette [191]. The blocks were cooled on ice until the wax had hardened, and then the quality of the wax was assessed for air bubbles, fault lines, grooves, or stress marks. The wax coverage of the cassette was also checked to ensure that it covered completely but did not have excess wax at the edges or back [192]. A microtome was used to trim the block, ensuring the surface was smooth [192].

2.3.5 TMA Layout Design

TMA Grand Master 2.4-UG-EN software (3D HISTECH®, Budapest, Hungary) was used to design the block layouts. New layouts were devised and given new names. Block sizes were standardised to 24 mm × 35 mm. For each block, three replications were prepared, and each was given a unique version number.

Each array contained 120 samples arranged in three groups (subarrays), with a 1-mm space between each core. The first group contained the 53 cores (6 rows, 10 columns) from the upper half of the block. The second group of the block contained 40 cores (4 rows, 10 columns), while the third group of the block was made up of 23 cores (3 rows, 10 columns). Cores in each subarray were arranged at a centred distance, and the subarrays were separated by a single gap for orientation and macroscopic identification of the TMA block. Positive internal experimental controls were introduced by adding orientation cores from normal kidney, liver, and tonsil tissues outside of the geometric margin of the array on the top and bottom left-hand sides.

Details of the donor block and the TMA position of each core were carefully documented. This enabled the creation of an array map grid for each array block. Each grid was given an individual identifier arranged to correspond with its position in the array block.

2.3.6 Coring Process

Each time, the automated TMA was loaded with 60 donor blocks and two recipient blocks (only one recipient block was loaded during construction cases with only one available core). Donor blocks were given pseudo-anonymised labels, and recipient

blocks were accordingly labelled TNBC-NGS1, TNBC-NGS2, etc. Before the donor-block cores were automatically sampled and embedded into recipient blocks, the software was used to match and overlay the scanned and marked images with their corresponding donor blocks.

2.3.7 TMA Block Annealing and Sectioning

After the TMA blocks were created, they were placed face-up in an oven for 3 h at 37°C to anneal them. This process helps the tissue cores adhere to the walls of the holes in the array block and makes the wax flexible for handling. A clean glass microscope slide was placed on the face surface of the block, then gentle, even pressure was applied, evening out the level of all the cores on the array. Levelling is important to maximise the quantity of sections containing all the samples in the array. After being heated, the blocks were left at room temperature for 3 h to cool. The annealing cycle was replicated in triplicate, and then the blocks were stored overnight at 4°C, ready for sectioning. To anneal agarose-based blocks, they were placed face-up in a metal and mould then placed on a 60°C surface to allow the paraffin wax around the agarose gel to melt without losing the cores. The blocks were allowed to cool and were kept overnight at 4°C.

Using a standard microtome sectioning technique, 4-μm sections were cut from TMA blocks and transferred onto permanent positively charged glass slides (Surgipath, X-tra™ Adhesive, Peterborough, UK). For the purpose of ensuring the quality of the histological samples and the TMA construction, the first section of each block was stained with H&E.

2.4 **Optimisation of Antibodies Used in the Study**

2.4.1 Cell Lines and Culture Media

While cell lysates of TNBC cell lines are optimal to use, the designated cohort contained only TNBC cases, such as MDA-MB-231, MDA-MB-436, MDA-MB-468, and MDA-MB-157. However, different human cell lysates were also used **(**Table 2-2). These included ER-positive (ER+) MCF7, HEK293 (an embryonic kidney cell line), HeLa epithelial cells derived from a human cervical carcinoma patient, Jurkat (a T lymphocyte cell line from a patient with acute T cell leukaemia, which expresses interleukin-2 and CD3), MCF10A (a non-tumourigenic mammary epithelial cell line with characteristics of luminal ductal cells with no myoepithelial features), and SKBR3 (obtained from a patient who underwent radiation). All cell lines were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA) and incubated using a 37°C incubator with a humidified atmosphere of 5% CO_2.

Table 2-2: Cell Lines Used in this Study

Cell Line Type and Origin		Media
TNBC cell lines	MDA-MB-231	RPMI + 10% FBS
	MDA-MB-436	DMEM/F12 with L-glutamine and HEPES + 10% FBS
	MDA-MB-157	RPMI + 10% FBS
	MDA-MB-468	DMEM+ 10% FBS
ER positive cells	MCF7	RPMI + 10% FBS
HER2 positive BC cells	SKBR3	McCOYS 5A + 10 % FBS
Non-tumourigenic BC cell line	MCF-10A	MEMB+ 100ng/ml Cholera toxin
Cervical carcinoma cell line	HeLa	DMEM + 10% FBS
Leukaemia cell line	Jurkat	RPMI + 10% FBS
Embryonic kidney cell line	HEK293	DMEM + 10% FBS

- FBS: fetal bovine serum

2.4.2 Cell Passage Procedure

Cell lines were monitored every two to four days, and once the cells reached 70–80% confluence, the expended cell culture medium was removed from the culture flask and discarded. Dulbecco's phosphate buffered saline (PBS) was used (approximately 2 ml/10 cm^2 culture surface area) to wash the cells. To avoid disturbing the cell layers, a wash solution was carefully added to the side of the vessel that was opposite the attached layer of cells. The vessel was rocked back and forth a number of times to ensure coverage of the culture flask, and then the wash solution was discarded. Five ml

of pre-warmed trypsin dissociation reagent (0.25% phenol red, 100 ml; 15050065, Thermo Fisher Scientific) was added to the side of the flask, covering the cell layer (approximately 0.5 m/10 cm^2). To ensure the cell layer was completely covered, the flask was rocked gently and then incubated for 5 min or as required for the cell line. Using the microscope, cells were checked to see if they had detached. The incubation time was increased, and checks were performed every two minutes if less than 90% of the cells had detached. When 90% or more of the cells had detached, the flask was tilted to allow the cells to drain. Equivalent volumes (twice the volume used for the dissociation reagent (10 ml)) of pre-warmed complete growth medium were added to the flasks. The medium was pipetted over the cell layer surface several times. Then, the cells were transferred to Falcon tubes and centrifuged for 5 to 10 min at 1,000 rpm. Note that the centrifuge speed and time varied based on the cell type. The cell pellet was resuspended in a minimal volume of pre-warmed complete growth medium, and a 10-µl sample was removed for counting. A haemocytometer, a cell counter, and 10 µl trypan blue stain were used to calculate the total number and percentage of viable cells using the following equation: number of cells/ml = mean cell count x dilution factor (10) x 10^4. The cell suspensions were diluted to the seeding density recommended for the cell lines, and appropriate volumes were pipetted into fresh cell culture flasks along with 20 ml of the required growth media.

2.4.2.1 Preparation of Cell Lysates

Eppendorf were appropriately labelled for each cell line, including cell line type, passage number, and date of collection. The cell culture medium was removed from the flasks with caution to avoid disturbing the adherent cells. These cells were washed in

phosphate-buffered saline (PBS) and incubated for 5 min in an incubator at 37°C. Two ml trypsin was then added to cover the cells as previously described. The samples were incubated at 37°C for approximately 5 min. Then the cells were resuspended in cell culture media and transferred to Falcon tubes and centrifuged for 5–10 min at 1,000 rpm, followed by disposing of the old cell culture medium and replacing it with fresh cell culture medium. To ensure that at least 75% of the cells were viable, the cell count procedure was performed to ensure a concentration of 1×10^6 cells per ml. Once cells were determined to be viable and suitable for cell lysate preparation, lysis of the cells was achieved by adding RIPA buffer (25 mM Tris, 150 mM NaCl, 1% Nonidet P-40, 1% sodium deoxycholate, and 0.1% sodium dodecyl sulfate (SDS) to a phosphatase and protease inhibitor cocktail (Pierce, Thermo Fisher Scientific, UK). The cell lysates were then incubated on ice for 10 min, with shaking every two minutes. To remove cell debris, the cell lysates were centrifuged for 20 min at 4°C and 13,000 rpm, and the supernatant was stored at -20°C until needed.

2.4.2.2 Cryopreservation of Cell Lines

Cryovials were labelled, and cells were harvested as described previously. Once > 70% cell viability was achieved, the cells were ready for freezing. Freezing medium was prepared (medium + 10% DMSO). The supernatant was removed, and the freezing medium was added to the required cell density (1×10^6 cells per ml), then 1-ml aliquots were transferred into the cryovials in a -80°C freezer for approximately 24 h. The cryovials were then transferred into liquid nitrogen for long-term storage.

2.5 Specificity of Antibodies by Western Blot

2.5.1 Background Principles

To determine the binding specificity of each of the antibodies that had been bought to its matching antigenic sequence, western blotting (WB) was performed. The binding specificity of each antibody used in this study for the various BC cell lysates was determined qualitatively (Table 2-3). Sometimes, WB produces multiple bands that indicate non-specific binding has taken place, or specific binding has taken place, but it involves a modified protein with two or more epitopes that the same primary antibody has recognised. The isolated protein is described as a ‘blot’ to reflect its transfer from a separation gel medium onto the surface of a non-reactive membrane. Initially, this technique was introduced for the purpose of transferring proteins from polyacrylamide gels to nitrocellulose membranes, as described by Towbin and colleagues [9].

Table 2-3: Western Blotting Evaluation of Antibodies Used in the Study

Antibody	Supplier	Species	Clone	WB Dilution and Incubation	Expected Molecular Weight
E-cadherin	Invitrogen	mouse monoclonal	4A2C7	1/100 O/N	100 KDa
ACSM4	Invitrogen	rabbit polyclonal	PA5-62082 and NBP2-14696PEP	not applicable	65 KDa
SPDYC	Novous Biological	rabbit polyclonal	NBP1-80832 and NBP1-80832PEP	not applicable	33 KDa
PPL	Abcam	rabbit polyclonal	Ab72422	1/500 O/N	200 KDa
Ki67	DACO	mouse monoclonal	MIB-1	not applicable	35 KDa
PIK3CD	Invitrogen	rabbit polyclonal	PA5-15250	1/3000 O/N	119 KDa
ABRAXAS1	Sigma Aldrich	rabbit polyclonal	HPA037654	1/500 R/T	56 KDa
RAD51C	Novous Biological	mouse monoclonal	2H11/6	1/500 O/N	37 KDa
β-actin	Sigma-Aldrich	mouse monoclonal	AC-15	1/5000	42 KDa
β-actin	Sigma-Aldrich	rabbit monoclonal	RM112	1/2000	42 KDa

- O/N: overnight; R/T: room temperature

2.5.2 Polyacrylamide Gel Electrophoresis and Protein Transfer

All experimental tools and reagents required for WB were prepared. To isolate and transfer the proteins, the XCell SureLock™ Mini-Cell and the XCell II™ Blot Module (Invitrogen, Thermo Fisher Scientific, UK) were assembled. Precast NuPAGE™ 4%–12% Bis-Tris 10-well polyacrylamide gels were obtained (Invitrogen, Thermo Fisher Scientific, UK).

The working solution was prepared by adding 50 ml of NuPAGE™ MOPS SDS Running Buffer 20X (Invitrogen, Thermo Fisher Scientific, UK) to 950 ml of distilled water. To prepare the transfer buffer for one gel, 50 ml of NuPAGE™ Transfer Buffer 20X (Invitrogen, Thermo Fisher Scientific, UK) was mixed with 100 ml of 10% methanol, 849 ml of distilled water, and 1 ml of NuPAGE™ Antioxidant (Invitrogen, Thermo Fisher Scientific, UK). The volumes of the other solution stocks were the same. To prepare the wash solution, 1 ml of TWEEN-20 was dissolved in 1 l of PBS. To prepare the blocking solution, 2.5 g of Marvel's fat-free dried milk was dissolved in 50 ml of wash solution. An alternative blocking solution of 1% bovine serum albumin (BSA) was used for some antibodies. All primary antibodies were prepared according to their previously optimised conditions according to the manufacturer's instructions or by modulating the recommended dilutions to obtain optimal results (e.g., in cases of recommended 1:1,000 dilutions, the author tried 1:1,500, 1:2,000, and 1:500). Because β-actin protein is expressed endogenously in all lysates, it is regarded as a normal control; therefore, a primary anti-β-actin antibody produced in mice was used (catalogue number: A5441, Sigma-Aldrich; Clone AC-15; Sigma, UK). The dilution

ratio for the primary anti-β-actin antibody to the blocking solution (either milk or BSA) was 1:5,000.

To detect the primary antibodies of the proteins of interest, IRDye® 680RD donkey (polyclonal) anti-rabbit IgG (LI-COR Biotechnology, Cambridge, UK) and IRDye® 800CW donkey (polyclonal) anti-mouse IgG (LI-COR Biotechnology, Cambridge, UK) antibodies were used. The fluorescent secondary antibodies were diluted in the blocking solution at a ratio of 1:15,000.

2.5.3 Sample Preparation

Lysates from each cell line and MagicMark™ XP western protein standard (Invitrogen, Thermo Fisher Scientific, UK) or Novex™ sharp pre-stained protein standard (Invitrogen, Thermo Fisher Scientific, UK) were thawed on crushed ice. Then, 13 µl of each lysate was mixed with 5 µl of NuPAGE™ LDS 4X sample buffer (Invitrogen, Thermo Fisher Scientific, UK) and 2 µl of NuPAGE™ 10X sample reducing agent (Invitrogen, Thermo Fisher Scientific, UK). All solution mixtures were mixed in chemical-resistant polypropylene microcentrifuge tubes. To mix the contents of the tubes thoroughly, they were spun gently using a laboratory vortex mixer. The protein mixtures were digested on a heating dry block at 100°C for 5 min. All of the samples were prepared in duplicate, and each 20-µl sample was sufficient to fill one well.

2.5.4 Loading Samples and Separation of Proteins Mixture

A precast NuPAGE Bis-Tris 4–12% protein gel cassette (Invitrogen, Thermo Fisher Scientific, UK) was inserted into the XCell SureLock™ Mini-Cell (Invitrogen, Thermo Fisher Scientific, UK) running tank. The running buffer was poured slowly into the running tank until the marked line was reached. When the integrity of the running was

confirmed and no leaks were detected, 1 ml of NuPAGE antioxidant reagent (Invitrogen, Thermo Fisher Scientific, UK) was added to the running buffer to protect the protein from possible oxidation. Then, either 8 µl of Novex^{TM} sharp pre-stained protein standard or 8 µl of MagicMark^{TM} XP (Invitrogen, Thermo Fisher Scientific, UK) were loaded onto the first well, followed by 20 µl of sample mix added to consecutive wells. Any empty wells were filled with 10 µl of loading buffer, and the gel was left to run at 150 V for 60 min.

2.5.5 Transferring Protein Bands onto Nitrocellulose Membrane/PVDF

The second blotting module was assembled according to the manufacturer's instructions. The blotting module was submerged in the transfer buffer in the running tank, and then the unit was run at 30 V for 60 min. To ensure the efficiency of the transfer, Ponceau stain was used followed by blocking using either 5% Marvel milk or 1% BSA for 60 min at room temperature. Then, the membrane was incubated with primary antibody and β-actin with milk or BSA at 4°C overnight in a cold room. After overnight incubation, the membrane was washed in 1X tris-buffered saline (TBS)-TWEEN for 5 × 3 min with continuous rocking. Secondary antibodies were used at the dilution of 1/15,000 in blocking solution (using 5% Marvel milk). To avoid direct exposure to light, the membrane was covered with foil and incubated for 60 min with continuous rocking, followed by three further five-minute rinses in washing buffer.

2.5.6 Visualisation of Protein Bands

The infrared dyes IRDye® 680RD secondary antibody with an emission wavelength of 694 nm and IRDye® 800CW secondary antibody with an emission wavelength of 795 nm were used to capture images of the immunodetected protein bands. Digital images of the membranes capturing red and green fluorescent bands were taken using the Odyssey Fc Imager (model 2800, LI-COR Biotechnology, Cambridge, UK). Where there was an overlap between the molecular weight of the β-actin and the targeted antigen of the biomarker, anti-rabbit and anti-mouse secondary antibody images were created digitally. Yellow bands indicated overlapping proteins with molecular weights of approximately 42 kDa. All images were labelled and stored on a personal hard drive.

2.5.7 Immunising Peptide Blocking (IPB)

An IPB protocol was undertaken to identify non-specific binding of primary antibodies to proteins that were not their specific antigens in the WB gels, i.e., visual bands. The detection of β-actin bands on the same WB gels acted as housekeeping, confirming whether there were any faults with the WB procedure itself. The primary antibody was neutralised by incubating it with an excessive concentration of its recombinant antigen solution, which blocked the antigen-binding site and its fragment antigen-binding (Fab) region. The neutralised primary antibody and a peptide blocking solution (0.25 μg/ml) were incubated on ice in a labelled 2-ml Eppendorf tube for 30 min. This protocol was duplicated with non-blocked primary antibody and recombinant antigen solution. Two successive tissue sections from the same histological sample were prepared according to routine histological procedures. Following the IHC protocol, two stained tissue

sections were obtained. An absolute absence of positive staining on the blocked primary AB confirmed the protein specificity.

2.6 Immunohistochemistry (IHC)

2.6.1 Background Principles

IHC, which was first described in the 1940s, is an important technique for clinically managing BC patients. It provides a range of information about molecular subtypes, the distribution and position of proteins in sections of tumour tissues, and the detection of validated prognostic biomarkers. The contribution that IHC makes to the reliability of pathology reports is key for many neoplastic diseases, including invasive BC [193]. A random number (10–15) of full-face sections were IHC stained to explore the heterogeneity of the expressed target prior to TMA staining and full exploration of designated targets with TNBC outcomes using 4-μm-thick TMA sections of invasive TNBC (Table 2-4).

The NovolinkTM Max Polymer Detection System (RE7280-K, Leica Biosystems, UK) was used with mouse immunoglobulin G (IgG) and rabbit IgG. Each staining run for the different antibodies, including negative and positive controls. Primary antibody was omitted in the negative control, leaving tissue samples incubating in antibody diluent only. Positive controls were provided by tissues incubated with known reactivity to each antibody as described in the supplier's data sheet.

Table 2-4: Immunohistochemistry Evaluation of Antibodies Used in the Study

Antibody	IHC Dilution and Incubation	Pre-treatment	Expression Localisation	Control Tissue
E-cadherin	1/25 O/N	Microwave, citrate pH 6, 20 min	Membranous	BC tissue
ACSM4	1/100 O/N and IPB 1/50	Microwave, citrate pH 6, 20 min	Cytoplasm	Testis tissue
SPDYC	1/50 O/N and IPB 1/20	Microwave, citrate pH 6, 20 min	Cytoplasm	Kidney tissue
PPL	1/50 O/N	Microwave, citrate pH 6, 20 min	Cytoplasm	BC tissue
Ki67	1/100 1 h, R/T	Microwave, citrate pH 6, 20 min	Nucleus	Tonsil tissue
PIK3CD	1/100 1 h, R/T	Microwave, citrate pH 6, 20 min	Cytoplasm and associated fibroblasts	BC tissue
ABRAXAS1	1/50 O/N	Microwave, Citrate PH 6, 20 minutes	Cytoplasm and nucleus	BC tissue
RAD51C	1/50 O/N	Microwave, Citrate PH 6, 20 minutes	Cytoplasm and nucleus	BC tissue

- O/N: overnight; IPB: immunising peptide blocking; R/T: room temperature.

2.6.2 Deparaffinisation and Rehydration of Tissue Sections

All tissue sections were loaded onto Leica's X-tra® slides and labelled. To remove the wax from the TMA sections, the samples were placed on a hot plate (Leica HI1220 flattening table, Leica Biosystems, UK) for 10 min at 60°C. The slides were loaded into the Leica AXL slide rack (Leica Biosystems, UK) for programme 1. The order of programme 1 is as follows: fresh xylene (Fisher Scientific, UK) (2 × 5 min), 100% IMS, (2 × 3 min), 70% IMS (2 × 3 min), and distilled water (5 min).

2.6.3 Heat-Induced Epitope Retrieval: Citrate Buffer, pH 6

Methods using aldehyde or formalin to fix biopsy samples in paraffin can cause transposition or changes to the antigenic site in FFPE tissues. The positivity of IHC staining is reduced by poor epitope exposure. This study used optimum conditions to re-expose the antigenic sites of FFPE tissues. Therefore, various heat-induced epitope retrieval (HIER) methods were tested under different conditions. The commonly used citrate buffer antigen retrieval protocol was determined to be the best antigen retrieval method.

2.6.4 Preparation of Sodium Citrate Buffer, pH 6

To prepare the sodium citrate buffer, 2.94 g of trisodium citrate dihydrate was dissolved in 1 L of distilled water. To achieve an acid solution with pH 6, a few drops of 1 normal HCl were added. This was checked using a laboratory pH meter that was calibrated daily. The solution was stored in a refrigerator at 4°C. The rack from step 2.6.2 was immersed in a plastic Leica AXL vessel filled with 500 ml of citrate buffer. The vessel was microwaved at maximum power for 20 min. Then, the vessel and slides were cooled with flowing water, which displaced the citrate buffer.

2.6.5 Slide Assembly onto Shandon™ Sequenza Coverplates™

To minimise the consumption of antibodies and reagents provided in the Novolink Polymer Detection System kit, Shandon coverplates™ (Thermo Fisher Scientific, UK) were used. The specific matching Shandon Sequenza immunostaining slide rack was also used (Ted Pella, USA). Tissue section slides from the Leica vessel (2.6.4) were covered with a Shandon cover plate applied at an angle of 45°. This process was performed in a water tank to avoid the presence of air bubbles. Sample slides were loaded into slots in the Shandon Sequenza immunostaining slide rack. To rinse the slides, the Sequenza reservoir was half filled with TBS pH 7.6. The TBS flow speed and duration were recorded for each slide. Slides were reloaded if the Sequenza reservoirs drained out in less than five minutes. The sections were kept hydrated at all times with TBS.

2.6.6 Blocking of Endogenous Peroxidases

To block endogenous peroxidases that remain active even in formalin-fixed, paraffin-embedded samples (the presence of which is evidenced by the tissue turning brown), 100 µl of Novolink peroxidase blocking reagent containing 3–4% (v/v) hydrogen peroxide (H_2O_2), was applied to each slide. All slides were left for 5 min for the blocking reaction to take effect, and then they were washed 2 × 5 min with TBS.

2.6.7 Blocking Non-Specific Reacting Proteins

To block stains arising from non-specific protein-protein binding (e.g., mistaken peptide sequences that used to generated the antibody to antigenic sites), 100 µl of Novolink protein blocking reagent, 0.4% casein in PBS with stabilisers, surfactant, and 0.2% Bronidox L broad-spectrum anti-bacterial and anti-fungal preservative were

applied using a micropipette. Tissue slides were incubated for 5 min before being washed with TBS as described previously.

2.6.8 Optimisation of Antibodies Used for IHC

Staining conditions for each antibody were optimised according to the manufacturer's instructions. In addition to the recommended 1:100 dilution, additional dilutions of 1:50, 1:25, 1:150, and 1:200 were also tested. Primary antibodies were diluted in Bond^{TM} primary antibody diluent (catalogue number: AR9352, Leica Biosystems, UK). Antigen retrieval solution step was used through microwave pre-treatment in citrate buffer at pH 6. The incubation duration of the primary antibodies followed the manufacturer's recommendations, which were altered depending on the staining results.

The most accurate evaluation of a cell's protein expression status is achieved with specific antibody staining that isolates immunoreactivity to a specific cellular location (membrane, cytoplasm, or nucleus) with little or no background staining. This was obtained by applying serial dilutions of the antibody to consecutive sections of BC tissue.

2.6.9 Applying Post-Primary Antibody

Post-primary antibody (rabbit anti-mouse IgG < 10 μg/ml in 10% (v/v) animal serum in TBS (0.09%) with ProClin™ 950 as the preservative, 100 μl) was administered to all the TBS-washed slides from the previous step. The slides were incubated for 30 min and then washed for 2 × 5 min with TBS.

2.6.10 Applying Novolink™ Polymer

A total of 100 µl of Novolink™ anti-rabbit Poly-HRP-IgG was loaded onto each slide and left for 30 min, followed by 2 × 5 min washes with TBS.

2.6.11 Blotting with 3,3'-diaminobenzidine (DAB)

DAB peroxidase substrate solution was prepared by mixing DAB chromogen with Novolink™ DAB substrate buffer polymer at a ratio of 1:20. The solution was made fresh and kept in the dark until required. The solution (100 µl) was applied for 5 min, and then the excess chromogen was rinsed for 2 × 5 min with TBS.

2.6.12 Counterstaining with Haematoxylin and Tissue Dehydration

To help with the semi-quantification of IHC-stained proteins, a counterstaining step was performed to produce clear images of the nuclei. To each section, 100 µl of haematoxylin reagent was applied and left for 6 min to react. The slides were submerged in a water tank, and the cover plates were removed. Surplus haematoxylin stain was removed. The slides were put back into the Leica AXL system, and the dehydration programme was activated. The dehydration step included washing the slides 3 × 2 min in baths of ethanol (EtOH), followed by 2 × 5 min washes in xylene. The stained tissues were then mounted in DPX.

2.6.13 Assessment of Immunoreactivity Prior to Immunohistochemical Scoring

To assess the expression in normal breast tissue and evaluate the degree of heterogeneity of expression in tumour tissue, full-face BC tissue sections from 15 cases representative of different BC molecular subtypes, histological grades, and stages were subjected to IHC staining and interpretation before IHC staining and scoring was performed on TMAs. Of the different biomarkers studied, only those that were evenly distributed across the tissues (visually determined by the pattern of staining) progressed for use in the TMAs. All candidate biomarkers studied using TMA cores had been previously validated to check the expression of tumour markers when used in large series [194].

2.6.14 Whole Slide Imaging

The intensity of IHC and H&E stains degrades with lengthy storage, direct exposure to light, or damage to the slide itself. This impacts the quality of the sample. To help preserve the information held within FFPE specimens, histology samples can now be digitalised. This confers several benefits, including needing less physical storage space, the ability to be edited and easily shared, and being readily accessible.

2.7 H-Score

The total area of each TMA core was examined to establish that > 15% was occupied by invasive tumour cells. Those cores that met this criterion were given a ‘histo’ score (H-score), which is a semi-quantitative method of evaluating the intensity of staining [[195]]. Scanned images of TMA cores from different slides were randomly selected and scored using a scale of 0–3 (absent–strong), and the percentage of cells expressing the different scores was recorded. The H-scores were calculated as $1 \times$ (% of 1+ cells) + $2 \times$ (% of 2+ cells) + $3 \times$ (% of 3+ cells).

2.8 **Determination of Cut-Off Points**

A number of methods were used to determine the cut-off values for the dichotomisation of biomarkers, which were individually assessed. These included 1) using cut-offs from previous studies that assessed large amounts of data; 2) depending upon the distribution of data (i.e., whether normally distributed or not), the mean or median was used, respectively; and 3) using X-Tile software (v. 3.6.1, 2003–2005, Yale University, USA http://x-tile.software.informer.com), a non-commercial research tool [196]. The software separates the total patient cohort randomly into two equally sized training and validation sets by producing separate lists of 'censored' and 'uncensored' events, ranked by patients' follow-up time. Uncensored events describe those patients who developed distant metastases or died because of TNBC. Censored patients did not develop distant metastases, were alive, or were lost during follow-up of this study. To determine the optimal cut-off points, the brightest pixel on the X-tile plot diagram was identified. Statistical significance was tested by validating the obtained cut-off points using SPSS with corrected *p* values and relative risk against DMFS (Table 2-5).

Table 2-5: Expression localisation and cut-off details of the stained antibodies.

Antibody	Expression Localisation	Chosen Cut-Off Values
E-cadherin	**Membranous**	**Lower quartile, high expression (H-score ≥ 85)**
ACSM4	**Cytoplasm**	**X-Tile, high expression (H-score ≥ 120)**
SPDYC	**Cytoplasm**	**X-Tile, high expression (H-score ≥ 90)**
PPL	**Cytoplasm**	**X-Tile, high expression (H-score > 0)**
Ki67	**Nucleus**	**High expression (≥ 10 %)** [197]
PIK3CD	**Cytoplasm and associated fibroblasts**	**X-Tile, high cytoplasmic expression (H-score ≥ 130), high fibroblast expression (H-score ≥ 85)**
ABRAXAS1	**Cytoplasm and nucleus**	**X-Tile, high cytoplasmic expression (H-score ≥ 15), high nuclear expression (H-score ≥ 40)**
RAD51C	**Cytoplasm and nucleus**	**X-Tile, high cytoplasmic expression (H-score ≥ 35), high nuclear expression**

2.9 Statistical Analysis

2.9.1 Introduction

To test the significance of the results, several statistical approaches were used. Nominal variables, including categorical and ordinal variables, continuous scale data, and independent or dependent variables exemplify this variety. Multiple statistical analyses were performed using the Statistical Package for Social Sciences (SPSS) software version 24 (SPSS Inc., Chicago, IL, USA). For all statistical tests, the level of significance was set as $p < 0.05$ (two-tailed).

2.9.2 Intra-Class Correlation Coefficient (ICC)

To increase the reliability of the H-scoring method, an ICC test was conducted after at least 30% of the whole cohort was re-evaluated by an additional scorer, who was a pathologist. All TMA slide scores were evaluated blindly, and no collaboration took place between the author and the pathologist. ICC correlation was considered satisfactory if the threshold of H-score agreement was higher than 75%. If there were discrepancies between slide scores, the slides were re-scored, and the discrepancies were discussed. All scorers were blind to the patients' pathological or outcome data.

2.9.3 Chi-Squared Test (Goodness of Fit)

To identify statistical relationships between two or more nominal independent variables, the chi-squared non-parametric statistical test was used. A chi-squared test is used to evaluate the difference of rates of distribution for a selected and categorised biomarker of interest against valid equivalent data belonging to the other nominal variable(s), patient age, tumour size, tumour grade, lymph node (LN) stage, LVI,

chemotherapy, or other previously stained biomarkers, mathematically. Output data were presented in cross-tabulation matrices of rows and columns.

2.9.4 Mann-Whitney U test (Median Analysis)

Univariate analysis of selected biomarkers was undertaken using the Mann-Whitney-Wilcoxon non-parametric test. This enabled statistical associations between the nominal data and the distributed median of the continuous data to be identified.

2.9.5 Student's t-Test and One-Way Analysis of Variance (ANOVA; Mean Analysis)

The statistical difference between the means of two normally distributed variables was evaluated using the Student's t-test and one-way ANOVA.

2.9.6 Spearman Rank Test (Model/Line of Best Fit)

To determine the strength of the linear association between two unnormalised continuous variables, such as gene/protein expression, the Spearman correlation coefficient (r) was used. The value of r ranges from +1 (strong association in the same direction) to -1 (strong inverse correlation), with 0 indicating that the two variables have no association. Correlations with values of -1.0 to -0.5 or 0.5 to 1.0 were considered strong, -0.5 to -0.3) or 0.3 to 0.5 were considered moderate, -0.3 to -0.1 or 0.1 to 0.3 were considered weak, and -0.1 to 0.1) were considered as having no correlation.

2.9.7 Kaplan-Meier Survival Analysis and Log-Rank test

To estimate the statistical survival distribution of two populations of BC patients for censored observations, the log-rank test was used. The associated estimations of TNBC patients' survival were plotted onto Kaplan-Meier (KM) curves. To indicate that a patient had developed distant metastases or had died from TNBC, the survival status was recorded as '1' in SPSS.

2.9.8 Proportional Hazards Model (Cox Regression Model)

The prognostic independence of a differentially expressed gene or protein was measured against standardised prognostic factors in Cox's proportional hazard model. This evaluated the presence of confounders and prognostic or predictive independence of the studied biomarker from standard prognostic/predictive factors, such as tumour grade, LN stage, and tumour size.

2.9.9 Calculation of Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value

To measure the specificity, positive predictive value (PPV), negative predictive value (NPV), and classification accuracy of identified prognostic biomarkers both at the mRNA and protein levels, cases were analysed by the incidence of distant metastases. Sensitivity was calculated based on the percentage of recorded histopathologically proven BC tumours interpreted as positive for distant metastases, defined as (number of true positive distant metastasis cases harbouring high expression of the identified biomarker) × 100/(number of true positive distant metastasis cases harbouring high expression of the identified biomarker + number of false negative distant metastasis cases harbouring low expression of the identified biomarker), where true positive were

the cases correctly recorded as distant metastasis positive and false negative were cases diagnosed as positive for distant metastasis but harboured low expressions of the identified biomarker on both mRNA and protein levels. Specificity was calculated based on the percentage of recorded histopathologically proven BC tumours interpreted as negative for distant metastases, defined as (number of true negative distant metastasis cases harbouring low expression of the identified biomarker) × 100/(number of true negative distant metastasis cases harbouring low expression of the identified biomarker + number of false positive distant metastasis cases harbouring high expression of the identified biomarker), where a true negative was a case correctly recorded as a distant metastasis negative, and a false positive was a case recorded as a distant metastasis negative but which harboured high expressions of the identified biomarker at both the mRNA and protein levels [198].

To evaluate the precision rate of this prognostic index in predicting distant metastasis, the PPV (the proportion of subjects with distant metastasis who are correctly diagnosed; PPV = true positive/(true positive + false positive)) and the NPV (proportion of subjects without distant metastasis who are correctly diagnosed; NPV = true negative/(true negative + false negative)) were calculated [199]. Finally, to investigate the accuracy of the overall ability of the expressions of the identified biomarker to correctly classify cases as high- or low-risk for TNBC metastatic disease, defined as (number of true positive distant metastasis cases harbouring high expressions of the identified biomarker + true negative distant metastasis cases harbouring low expression of the identified biomarker) × 100/(total number of patients who were computed in the analysis) were calculated [200].

2.9.10 Validation of the Findings

Data was sought from online databases for the purposes of comparing and validating the study findings. The prognostic value of the prognostic genes identified was evaluated using the Breast Cancer Gene Expression Miner v4. 0 (Bc-GenExMiner) database, which includes RNA sequence expression data from 4,713 BC patients, including 254 TNBC patients [201]. The Genotype 2 outcome tool (http://www.g-2-o.com) was also integrated to identify prognostic genes. This web-based server uses NGS and microarray data of 6,697 BC patients, including 612 TNBC patients, and also provides outcome data. To create the transcriptomic fingerprint for mutational status, computed receiver operator characteristic (ROC) values obtained from the Cancer Genome Atlas RNA sequence and NGS mutation data were used. The average expression of significant genes was designated as a metagene for a given genotype. Using the gene chip data, multivariate Cox regression, and KM survival analysis, associations were drawn from the expression of the metagenes and patient outcomes [35]. Prognostic genes that had been identified were also computed on the KM plotter dataset to test the prognostic value using 3,951 BC patients, including 255 TNBC patients with available outcome data [202]. The best performing threshold against outcome was used to categorise the cases into high- and low-risk as generated by the KM Plotter (determined by the public domain database).

An ROC curve was created using the ROC plotter data portal to establish the predictive value of the expression of the prognostic genes identified in relation to the effectiveness of adjuvant therapy in TNBC tumours and relapse-free survival [203]. Gene selection and the optimal cut-offs were determined using the Jetset best probe selection [204].

The expression of identified prognostic genes of TNBC patients who do or do not respond to CMF, FAC, FEC, anthracycline, lxabepilone, and taxane chemotherapy interventions was compared to determine potential associations between the genes and chemotherapeutic responses.

2.9.11 The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)

Validated clinicopathological and transcriptomic data of a well-characterised cohort of TNBC patients (N = 313) was obtained from the METABRIC-TNBC [114]. These data were used to evaluate mRNA expression and changes to copy number in the TMA results. The mean tumour size at the time of presentation was 3 cm (range: 1–18 cm) for patients in the METABRIC-TNBC series, and the mean age was 59 years (range: 26–96). DNA and RNA were isolated from fresh frozen samples for the purpose of profiling their transcription, which was achieved using an Affymetrix SNP 6.0 Illumina Total Prep RNA Amplification Kit and Illumina Human HT-12 v3 Expression Bead Chips (Ambion, Warrington, UK). Gene segments and the Šidák correction were used to determine changes to copy number [[205]], whereas gene expression data were pre-processed and normalised as described previously [114]. In this cohort, patients with ER-negative tumours and LN-positive status received adjuvant chemotherapy (Table 2-6).

Table 2-6: METABRIC-TNBC patients' cohort

Clinicopathological Parameter	METABRIC cohort (n=313) No (%)	Clinicopathological Parameter	METABRIC cohort (n=313) No (%)
Age (years)		**Tumour Histological Subtype**	
< 50	**112 (36)**	**Ductal no special type (NST) and medullary subtype**	**282 (90)**
≥ 50	**201 (64)**	**Lobular**	**15 (5)**
Tumour Size (cm)		**Miscellaneous and special subtypes**	**16 (5)**
< 2	**98 (32)**	**Vascular Invasion**	
≥ 2	**211 (68)**	**Yes**	**96 (38)**
Tumour Grade		**No**	**157 (62)**
1	**4 (1)**	**Distant Metastasis**	
2	**40 (13)**	**Yes**	**Not available**
3	**266 (86)**	**No**	**Not available**
Nodal Stage		**Breast Cancer Specific Survival**	
1	**95 (30)**	**Alive**	**166 (53)**
2	**176 (57)**	**Died of BC**	**101 (32)**
3	**40 (13)**	**Died of other causes**	**46 (15)**
Nottingham Prognostic Index		**Chemotherapy**	
Good Prognosis	**18 (6)**	**Yes**	**167 (53)**
Moderate Prognosis	**136 (45)**	**No**	**146 (47)**
Poor Prognosis	**151 (50)**	**No treatment**	**0 (0)**
PAM50 classification		**Radiotherapy**	
Luminal A	**3 (1)**	**Yes**	**222 (71)**
Luminal B	**0 (0)**	**No**	**50 (26)**
HER2 enriched	**38 (12)**	**No treatment**	**41 (13)**
Basal Like	**249 (80)**		
Normal Like	**23 (7)**		

2.9.12 In-House Biomarker Consortium Data

Data on a broad range of biomarkers of well-established clinical and biological relevance to BC were accessible and retained on a web-based interface (Distiller; Slidepath Ltd, Dublin, Ireland). These data included androgen receptor (AR) [206], EMT-related biomarkers (E-cadherin, TGFBeta1, TWIST2, and ZEB) [207–209], DNA repair families (BRCA1, BRCA2, GammaH2ax, and RAD51) [210–212], proliferation biomarkers (Ki67) [213], and tumour suppressors and oncogene biomarkers (TP53, PTEN, ATM) [98,208,214,215]. Except for Ki67 being evaluated on full-face sections, the aforementioned markers used in this study have been previously stained and evaluated on TMA. The author re-scored E-cadherin, Ki67, and AR markers expressed in the TNBC subset of cases. The details are summarised in related chapters.

Chapter 3 : Identification of potential prognostic genes in TNBC

This chapter has been published in Modern Pathology (2019) under the title " A novel prognostic two-gene signature for triple negative breast cancer" with the corresponding DOI: https://doi.org/10.1038/s41379-020-0563-7[216]. The Copyright approval for re-use of these material is provided in (appendix 7.2)

3.1 Background

TNBC comprises 15%-30% of BC, and presents considerable challenges with regard to clinical management due to lack of targeted therapies [217,218]. Moreover, TNBC often has an unfavorable prognosis with increased probability of early metastasis, disease recurrence, and shorter overall survival [219,220]. Although TNBC generally displays aggressive behaviour, patient outcomes can vary considerably. Around 23% of early-diagnosed TNBC patients remain disease free for more than five years while death within five years of diagnosis is inevitable for almost all metastatic TNBC patients [136,150,221]. Therefore, the complexity, molecular variability, and unpredictability of TNBC behavior warrants further investigation [222]. The biological heterogeneity of TNBCs has provided an impetus to develop tools for prognostic stratification, however, there are inconsistent results owing to a small cohort of patients, gene expression datasets obtained from different gene expression platforms and the use of microarray versus reverse transcriptase polymerase chain reaction (RT-PCR), which also makes head-to-head comparison challenging [223,224].

3.2 Available Prognostic gene signatures

Various multigene prognostic tests are available for ER-positive tumours, including Oncotype DX, IHC4, MammaPrint, Genomic Grade index and Prosigna, for patient risk stratification and to guide therapy choice, whereas in ER-negative tumours, and specifically TNBC tumours with a higher proliferation rate, these multigene signatures provide no clinical value [225]. Lehmann et al, used gene expression profiles to classify TNBCs into six molecular subtypes: Basal-like 1 and 2, Mesenchymal, Mesenchymal Stem-like, Immunomodulatory, and Luminal Androgen Receptor [113]. Burstein et al proposed an alternative gene expression classification for TNBC categorising the tumour into four TNBC molecular subtypes: Luminal Androgen Receptor, Mesenchymal, Basal like immune suppressed, and Basal like immune activated [116]. However, distant metastasis-free survival (DMFS) analysis showed poor prognosis for TNBCs regardless of their molecular profile subtype [226]. Therefore, there is an urgent unmet need for clinically validated prognostic markers that can predict outcomes for TNBC patients [226].

3.3 Advancement on transcriptomic technologies

Unbiased omics technologies, including NGS, are expected to lead a paradigm shift for precision medicine from a pathological microscopy-based diagnosis to gene signature-based diagnosis, prognosis, and treatment approaches [227]. NGS enables transcriptomic profiling of TNBC and identification of genomic alterations such as copy number alteration, insertions, deletions and mutations; consequently, studies

exploring inter-tumour heterogeneity in different types of tumours are now possible [40,228]. For successful NGS analysis, clinical samples must be maintained in conditions that would allow for DNA and RNA preservation and subsequent extraction. At present, most clinical samples are processed and archived as FFPE tissue samples in which the DNA and RNA necessary for NGS analysis is often fragmented [229]. However, FFPE tissue samples, if processed and stored properly, have been shown to preserve sufficient DNA and RNA material for extraction for NGS analysis [230].

3.4 Hypothesis and Aim

In this chapter, it was hypothesised that the Identification of a prognostic gene signature for TNBC tumours could significantly impact the survival of patients with these difficult to treat cancers until novel targeted therapies become available. Therefore, the present chapter employs NGS transcriptomic analysis and aims to identify a molecular prognostic signature predicting risk for poor outcomes in TNBC.

3.5 Materials and Methods

3.5.1 Study Cohort

This chapter was carried out as a collaborative work with Dr Graham Ball (School of Biomedical and Natural Sciences, the Nottingham Trent University), who performed the Artificial Neural Network analysis (ANN) to identify prognostic genes associated with TNBC clinical outcome. The experimental work including basic statistical and pathway analysis, IHC staining and evaluation, and development of probabilistic algorithm score were calculated by the author.

This study was conducted on a retrospective well-characterised series of primary invasive TNBC tumours samples obtained from patients presented to Nottingham City Hospital, UK between 1987 to 2006 prepared as TMA (n=393). The NGS cohort (n=112) was referred as "transcriptome cohort" and used a discovery cohort, whereas the "IHC cohort" (n=333) was used to validate the identified biomarkers at the protein level using IHC. Patients outcome data together with clinicopathological characteristics were collected from patients' medical records (Chapter 2). Distribution of clinicopathological parameters between NGS transcriptomic cohort and the TNBC IHC cohort were tested using Chi-square test (Table 3-1).

Table 3-1: Patients cohorts

Clinicopathological Parameter	Transcriptomic cohort (n=112) No (%)	IHC Cohort (n=333) No (%)	P-value	Clinicopathological Parameter	Transcriptomic cohort (n=112) No (%)	IHC Cohort (n=333) No (%)	P-value
Age (years)				Vascular Invasion			
< 50	62 (56)	154 (46)	0.025	Yes	42 (38)	89 (27)	0.005
≥ 50	50 (44)	179 (54)		No	70 (62)	244 (73)	
Tumour Size (cm)				Distant Metastasis			
< 2	32 (28)	144 (43)	<	Yes	46 (41)	91 (28)	0.001
≥ 2	80 (72)	189 (57)	0.001	No	64 (59)	241 (72)	
Grade				Breast Cancer Specific Survival			
1	1 (1)	4 (1)		Alive	61 (55)	216 (65)	
2	6 (6)	24 (7)	0.469	Died of BC	42 (38)	89 (27)	0.033
3	104 (93)	305 (92)		Died of other causes	9 (7)	27 (8)	
Nodal Stage				Chemotherapy			
1	68 (61)	223 (67)		Yes	57 (51)	257 (65)	
2	31 (27)	77 (23)	0.270	No	44(40)	123 (31)	0.005
3	13 (12)	32 (10)		No treatment	11 (10)	14 (4)	
Nottingham				Radiotherapy			
Prognostic Index	3 (3)	15 (4)		Yes	74 (66)	278 (71)	
Good Prognosis	76 (68)	245 (74)	0.037	No	28 (25)	97 (25)	0.668
Moderate Prognosis	33 (29)	72 (22)		No treatment	10 (9)	19 (4)	
Poor Prognosis							

3.5.2 Transcriptomic and Pathway Analysis

Artificial Neural Network (ANN) database mining approach was used to build a classifier using the NGS transcriptomic cohort and identify genes associated with disease outcomes (DMFS and BCSS). In ANN, learning rates and momentum were set at 0.1 and 0.5, respectively [231]. Each tumuor sample had 39,684 corresponding genes. The input codes were "0" if patients showed neither evidence of metastasis (DMFS) nor death from BC (BCSS) within five years, and "1" if metastasis or death due to BC was evident in the first five years after diagnosis. Although BCSS is the ultimate endpoint of cancer outcome, DMFS was chosen as an end point based on the high likelihood of TNBC patients being diagnosed with distant metastases within five years of diagnosis [136]. Prior to ANN testing, a Monte-Carlo cross validation procedure was applied to avoid data over-fitting and false discovery. Documentation of such approach has proven to outperform the commonly used leave-one-out cross validation [232]. The input data were randomly divided into three subsets; 60% for training, 20% for validation to ensure model performance during the training process, and 20% for blind testing of the original model [233]. Genes identification by the forward stepwise approach using ANN was performed as described previously [234]. Based upon the distribution of performance on the aforementioned model, ANN generated two panels of genes, representing the top 1% of the RNA sequence matrices that significantly predicted DMFS and BCSS, respectively. Genes common to both the DMFS and BCSS panels were identified using the Venny 2.0 online tool [235]. Receiver operating characteristics (ROC) curves were generated to assess the predictive value of the

differentially expressed gene panel presenting the sensitivity and specificity of the tested model (Figure 3-1).

Pathway analysis was conducted to identify differentially regulated canonical pathways using the overrepresentation enrichment analysis (ORA) (Chapter 2). The pathway analysis was based on the top 200 ranked genes predicting DMFS and BCSS. The reference gene list was set to the "genome_protein_coding". The ratio of observed versus expected number of genes in the category was recorded for each significant category using the enrichment ratio (R) scores using Panther pathway database [188].

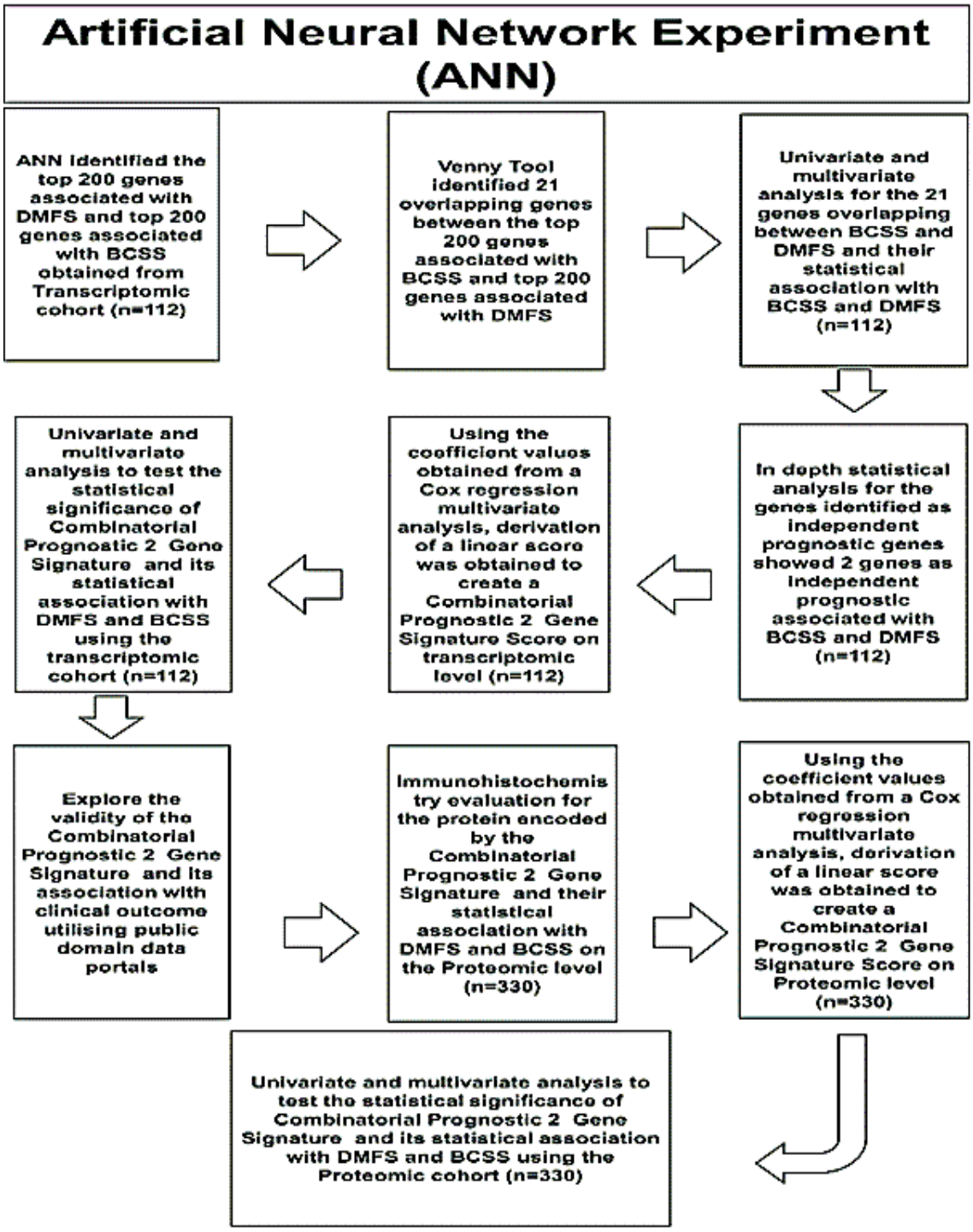


Figure 3-1:Flowchart of Artificial Neural Network analysis

- This figure illustrates the analytical and experimental methodology of the steps used in the conducted experiment utilising the mRNA cohort of 112 TNBC cases to identify genes associated with clinical outcome.

3.5.3 Prognostic Gene Signature Score

In compliance with the Reporting Recommendations for Tumour Marker Prognostic Studies criteria (REMARK), the associations between the expression of genes in our 21-gene panel, common to both the DMFS and BCSS gene prediction panels identified by ANN, and DMFS or BCSS were evaluated both individually, as well as after adjusting for standard prognostic variables [236,237]. Thus, DMFS and BCSS probabilities were individually computed on our gene panel using Kaplan-Meier testing model. Additionally, multivariate Cox regression analysis was used to calculate the estimate effect size [i.e., Hazard ratio (HR), along with 95% confidence interval (CI)] of the genes that were statistically significant in univariate Kaplan-Meier testing model for both DMFS and BCSS, which included the genes and standard prognostic variables, regardless of the statistical significance of standard prognostic variables in univariate analysis. The genes which showed significant prognostic impact independently in multivariate Cox regression analysis were further examined in a combined multivariate Cox regression analysis to identify a signature with a minimum number of genes that showed the most significant association with DMFS and BCSS.

3.5.4 External Validation of Transcriptomic Data

For independent validation of the results, the prognostic value of the two-gene signature predictors of DMFS and BCSS were examined using public domain data (Chapter 2).

3.5.5 Immunohistochemistry and western blotting analyses

Assessment of the protein expression of the identified two-gene prognostic signature was performed using rabbit anti-SPDYC (NBP1-80832, lot # R36476, Novous Biological, UK) and rabbit anti-ACSM4 (PA5-62082, lot # R59771, Thermofisher, UK) (Figure 3-2). Anti SPDYC was neutralized with recombinant human protein antigen (NBP1-80832PEP, lot # PR03260, Novous biological, UK) following the same staining protocol previously described (**Chapter 2**). Of note, the final count of cores eligible for scoring varied between antibodies after elimination of uninformative TMA cores (cores were excluded at random either for loss during TMA construction or antigen retrieval process, tissue folding during processing or cores representing < 15% of tumour cells). In addition, 25% of TMA sections were scored blindly by the author and two other researchers (Dr. Sara Raafat for SPDYC & Dr Michael Toss for ACSM4) to assess inter-observer concordance (ICC). The ICC for SPDYC and ACSM4 were 0.77 and 0.80, respectively, indicating substantial concordance between scorers. Moreover, the discordant cases were re-scored by the author and a second observer and a consensus score was agreed upon and assigned.

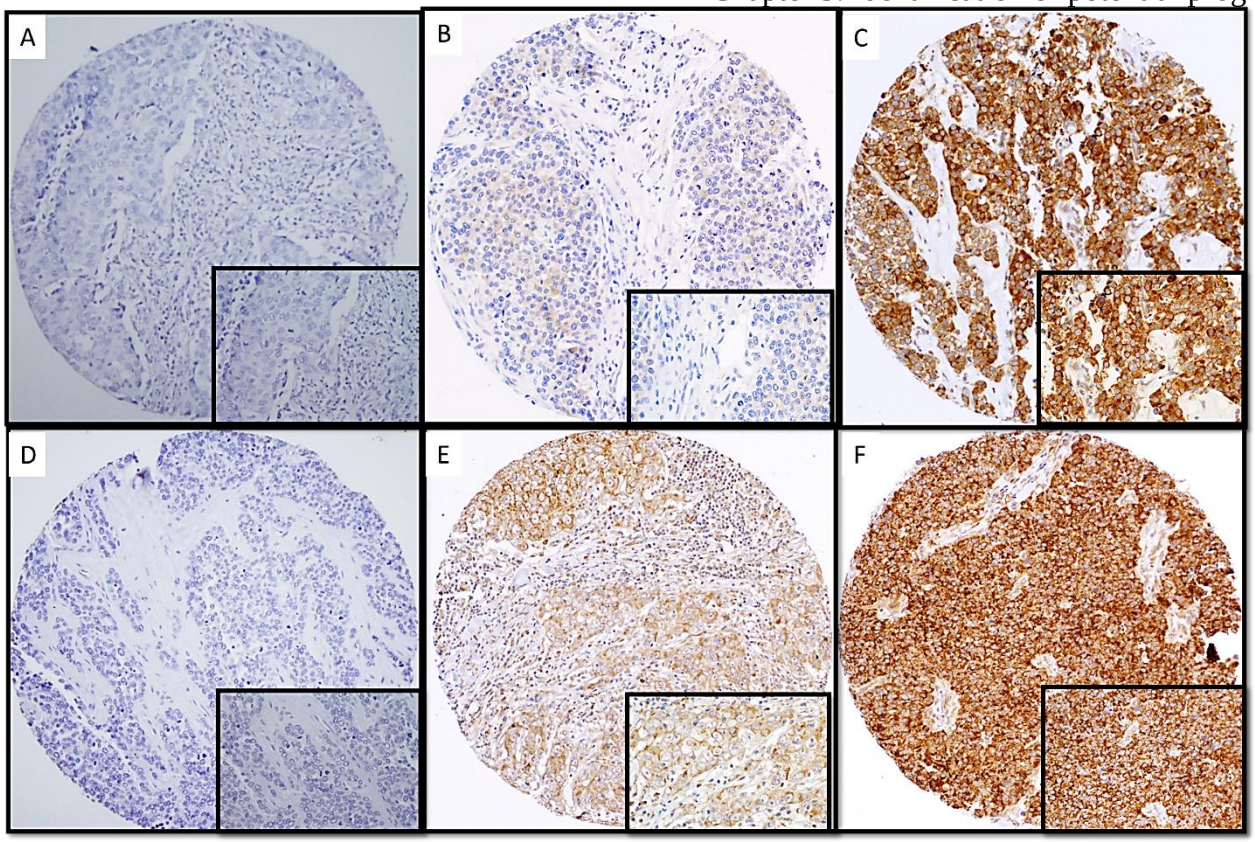


Figure 3-2:Immunohistochemical expression of SPDYC and ACSM4 in FFPE TMA cores from TNBC samples (IHC cohort, n=333) Immunohistochemical protein expression of SPDYC and ACSM4 in TNBC TMA cores (power 10×, inset high power 40×). A) SPDYC neutralised expression, B) SPDYC weak expression, C) SPDYC strong expression, D) ACSM4 neutralized expression, E) ACSM4 moderate expression, F) ACSM4 Strong expression.

3.5.6 Statistical Analysis

IBM SPSS 24.0 (Chicago, IL, USA) software was used for statistical analysis. For dichotomisation of mRNA expression and protein expression levels of different genes, the X-tile bioinformatics version 3.6.1 (Yale University, USA) was utilised with DMFS as an endpoint. Cox proportional hazard models were used for multivariate analysis model adjusting for patients age, tumour grade, nodal stage, tumour size, and LVI status as covariates to adjust for potential confounding influence of these variables on associations between the tested genes and the outcomes of interest. Spearman's Rho test was used to evaluate correlations between continuous variables of the transcriptomic and protein expression data. A *p*-value of <0.05 was deemed significant.

3.6 RESULTS

3.6.1 Gene Selection

To build a classifier panel for outcome prediction in TNBC, ANN analysis of the RNA-sequence matrices data of the transcriptomic cohort was performed and genes were ranked based on relationships between their expression and clinical outcomes in terms of DMFS and BCSS. The top ranked genes predicting DMFS (DMFS genes panel) and those predicting BCSS (BCSS genes panel) were investigated to determine the most statistically enriched pathways, (**Table 3-2 & Appendix for full details**).

Using the Venny tool, we identified a total of 21 genes that were common to both the DMFS and BCSS ANN panels. The 21-gene panel predicted patients' DMFS and BCSS with 92% sensitivity and 94% specificity (Figure 3-3). The probability of finding a

gene by random chance in the top 200 was 0.03, whereas the probability of randomly finding the 21 genes collectively was $6.2x10^{-33}$, (**Figure 3-4**).

Univariate Kaplan–Meier survival analysis showed that elevated expression of some genes was significantly associated with shorter DMFS and BCSS, whereas elevated expressions of other genes showed statistically significant association with longer DMFS and BCSS (**Table 3-3 & Figure 3-5**). Multivariate Cox regression analysis models incorporating patient's age, tumuor grade, nodal stage, tumour size, and LVI status revealed that eight of the 21 genes were independent predictors of DMFS and BCSS, (**Table 3-4**).

Table 3-2:Pathway analysis of genes strongly predictive of both Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112)

Description	Gene Set	Gene Set Size	Overlap	P-value	Enrichment Ratio	Gene Symbol	Gene Name
JAK/STAT signaling pathway	P00038	15	1	0.130	7.1937	STAT1	signal transducer and activator of transcription 1
p38 MAPK pathway	P05918	34	2	0.038	6.3473	MAP3K4	mitogen-activated protein kinase 4
						MAP2K6	mitogen-activated protein kinase 6
Ras Pathway	P04393	70	3	0.025	4.6245	STAT1	signal transducer and activator of transcription 1
						MAP2K6	mitogen-activated protein kinase 6
						MAP3K4	mitogen-activated protein kinase 4
EGF receptor signaling pathv	P00018	115	4	0.019	3.7532	STAT1	signal transducer and activator of transcription 1
						MAP3K4	mitogen-activated protein kinase 4
						MAP2K6	mitogen-activated protein kinase 6
						EGFR	epidermal growth factor receptor
Cadherin signaling pathway	P00012	153	4	0.048	2.821	PCDHGC5	protocadherin gamma subfamily C, 5
						PCDHB1	protocadherin beta 1
						FZD5	frizzled class receptor 5
						EGFR	epidermal growth factor receptor
PDGF signaling pathway	P00047	125	3	0.105	2.5897	MAP3K4	mitogen-activated protein kinase 4
						STAT1	signal transducer and activator of transcription 1
						ETV3	ETS variant 3

- The top 200 ranked genes predicting DMFS (DMFS genes panel) and those predicting BCSS (BCSS genes panel) were investigated to determine the most statistically enriched pathways in our DMFS and BCSS gene lists by conducting Panther enrichment pathway analysis using Webgestalt. Since those aforementioned biologically important pathways showed statistically significant enrichment, they merited more in-depth evaluation, nevertheless, they reinforced the power of our discovered classifier panels in predicting TNBC outcome.

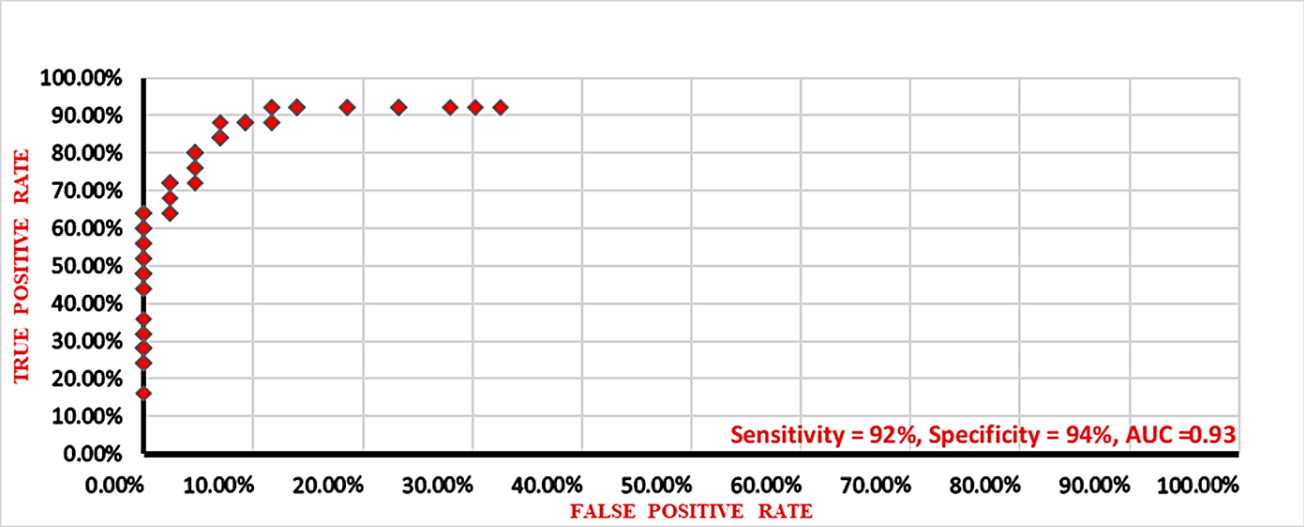


Figure 3-3:Receiver Operative Characteristic curve depicting the sensitivity and specificity of the predicted 21 gene panel associated with both **Breast cancer-specific survival (BCSS) and Distant metastasis-free survival (DMFS) (Transcriptomic cohort, n=112).**

- Area Under the Curve (AUC)

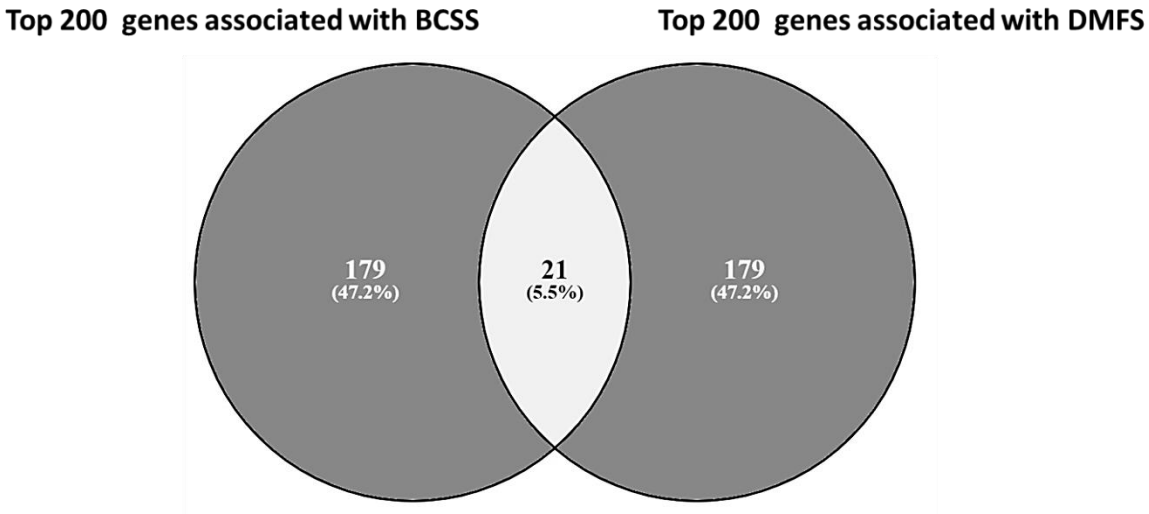
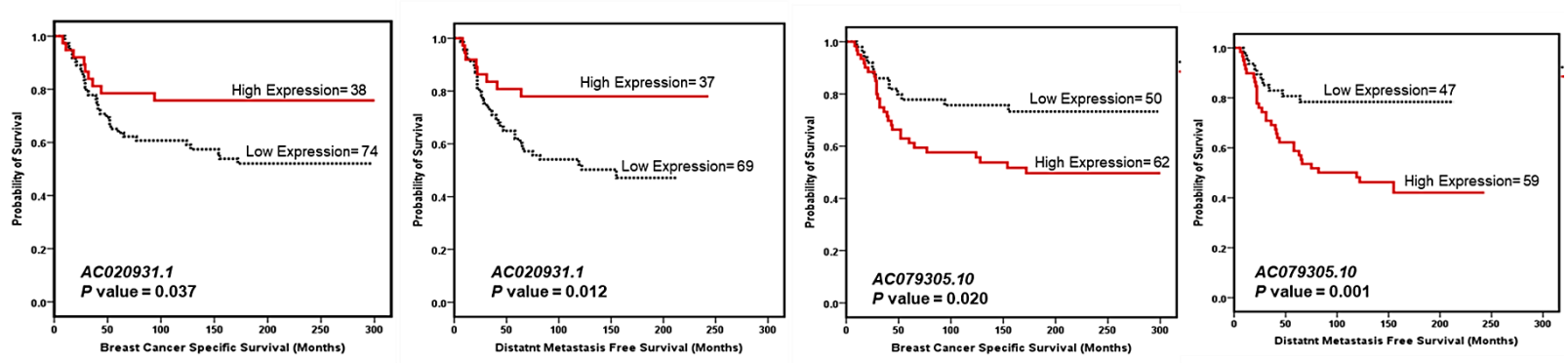


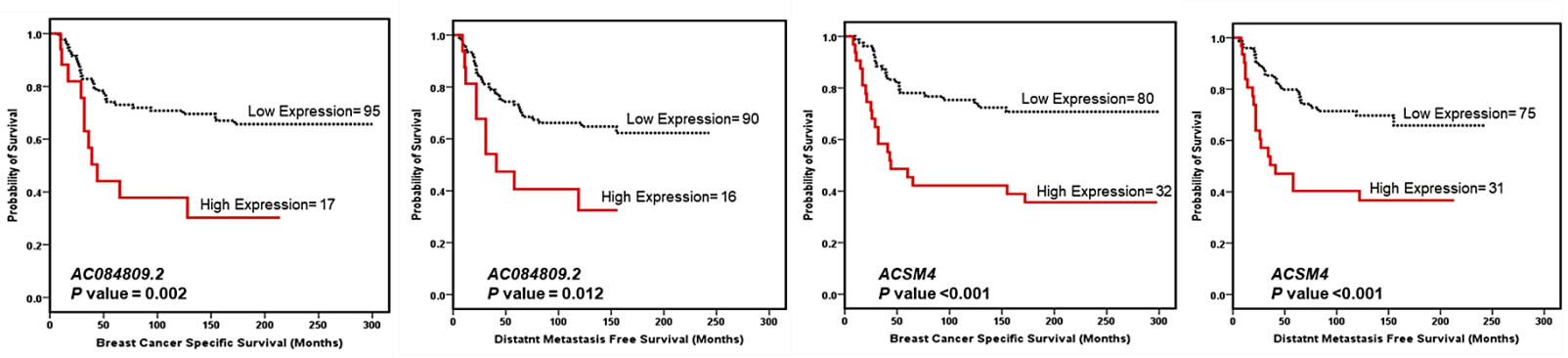
Figure 3-4:The output of Venny diagram tool showing the overlapping genes between the top 200 genes associated with Breast Cancer Specific Survival (BCSS) and the top 200 genes associated with Distant Metastasis Free Survival (DMFS) (Transcriptomic cohort, n=112)

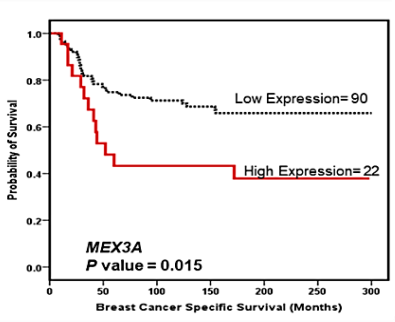
Table 3-3:Univariate analyses for the 21 genes identified by Venny tool that potentially predictive of both Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112)

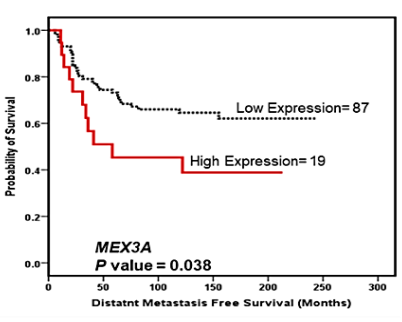
Gene ID	**chromosomal location**	**Breast Cancer specific Survival**			**Distant Metastasis Free Survival**		
		Frequency		**P-value**	**Frequency**		**P-value**
		Below-cut-point Expression (N)	Above cut-point Expression (N)		Below-cut-point Expression (N)	Above cut-point Expression (N)	
AC020931.1	19p13.2	74	38	**0.037**	69	37	**0.012**
AC079305.10	2q31.2	50	62	**0.02**	47	59	**0.001**
AC084809.2	17q11.2	95	17	**0.002**	90	16	**0.012**
ACSM4	12p13.31	80	32	**< 0.001**	75	31	**< 0.001**
MEX3A	1q22	90	22	**0.015**	87	19	**0.038**
NDUFA4L2	12q13.3	72	40	**0.036**	66	40	**0.010**
PAXBP1-AS1	21q22.11	84	28	**0.033**	81	25	**0.020**
BICC1	10q21.1	72	40	0.160	68	38	**0.048**
CCDC54	3q13.12	71	24	0.330	68	38	0.115
DCTN1-AS1	2p13.1	91	21	**0.047**	85	21	**0.006**
DOCK10	2q36.2	91	21	**0.020**	86	20	**0.037**
GTF3C6	6q21	72	40	0.242	68	38	0.076
HARBI1	11p11.2	94	18	0.113	90	16	**0.017**
PPL	16p13.3	94	18	**0.002**	89	17	**< 0.001**
RP1129H23.5	1q22	82	30	**0.002**	77	29	**0.001**
RP11409C19.2	8q13.1	72	40	0.180	70	36	0.068
RPS10P18	10q26.13	95	17	**0.002**	89	17	**0.003**
RPS3AP47	15q15.2	91	21	0.588	86	20	0.509
SNORD99	1p35.3	90	22	0.099	84	22	0.075
SPDYC	11q13.1	90	22	**0.014**	86	20	**0.018**
SRP72P2	6q25.3	91	21	0.381	20	86	0.175
Total		**112**			**106**		

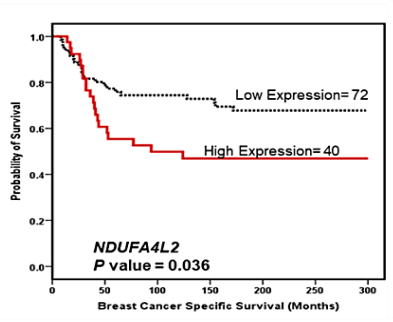
- Significant P-values are bolded

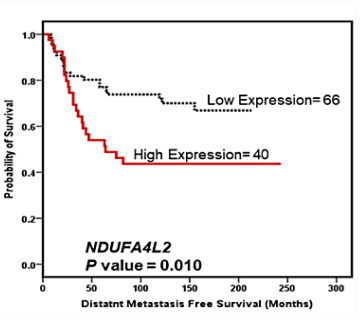

High Expression= 38
Low Expression= 74
AC020931.1
P value = 0.037
Probability of Survival
Breast Cancer Specific Survival (Months)
High Expression= 37
Low Expression= 69
AC020931.1
P value = 0.012
Distatnt Metastasis Free Survival (Months)
Low Expression= 50
High Expression= 62
AC079305.10
P value = 0.020
Breast Cancer Specific Survival (Months)
Low Expression= 47
High Expression= 59
AC079305.10
P value = 0.001
Distatnt Metastasis Free Survival (Months)

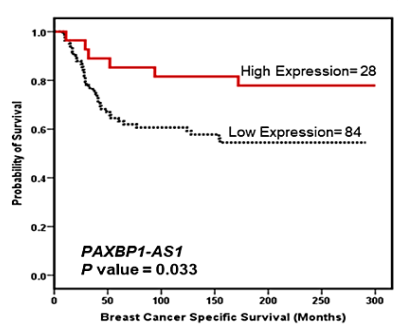

Low Expression= 95
High Expression= 17
AC084809.2
P value = 0.002
Probability of Survival
Breast Cancer Specific Survival (Months)
Low Expression= 90
High Expression= 16
AC084809.2
P value = 0.012
Distatnt Metastasis Free Survival (Months)
Low Expression= 80
High Expression= 32
ACSM4
P value <0.001
Breast Cancer Specific Survival (Months)
Low Expression= 75
High Expression= 31
ACSM4
P value <0.001
Distatnt Metastasis Free Survival (Months)

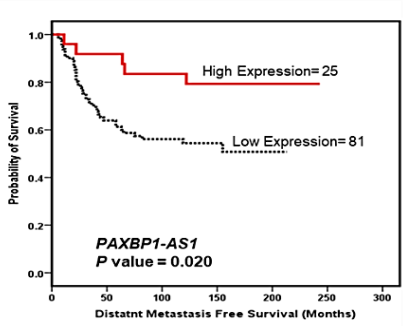

Probability of Survival
Low Expression= 90
High Expression= 22
MEX3A
P value = 0.015
Breast Cancer Specific Survival (Months)

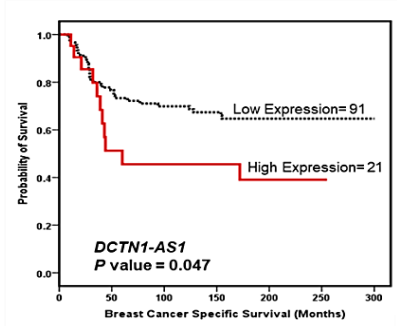

Probability of Survival
Low Expression= 87
High Expression= 19
MEX3A
P value = 0.038
Distatnt Metastasis Free Survival (Months)

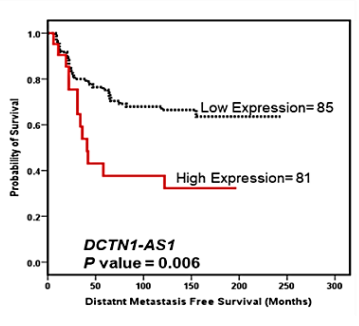

Probability of Survival
Low Expression= 72
High Expression= 40
NDUFA4L2
P value = 0.036
Breast Cancer Specific Survival (Months)

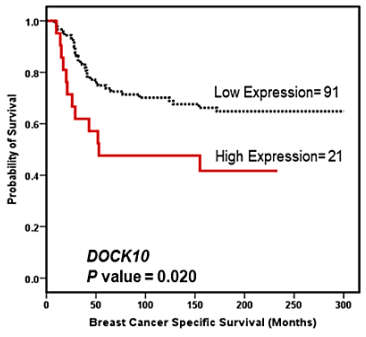

Probability of Survival
Low Expression= 66
High Expression= 40
NDUFA4L2
P value = 0.010
Distatnt Metastasis Free Survival (Months)

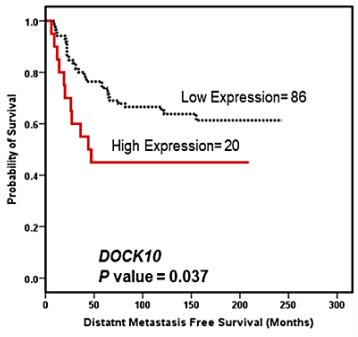

Probability of Survival
High Expression= 28
Low Expression= 84
PAXBP1-AS1
P value = 0.033
Breast Cancer Specific Survival (Months)

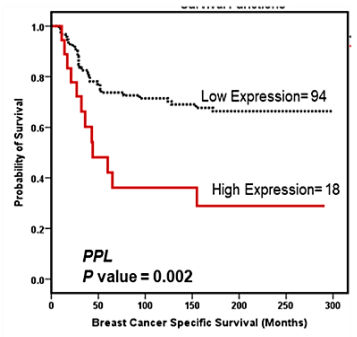

Probability of Survival
High Expression= 25
Low Expression= 81
PAXBP1-AS1
P value = 0.020
Distatnt Metastasis Free Survival (Months)

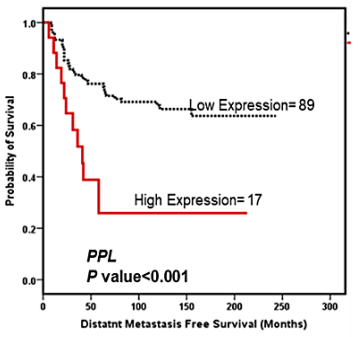

Probability of Survival
Low Expression= 91
High Expression= 21
DCTN1-AS1
P value = 0.047
Breast Cancer Specific Survival (Months)

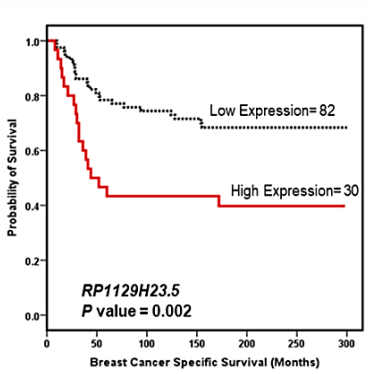

Probability of Survival
Low Expression= 85
High Expression= 81
DCTN1-AS1
P value = 0.006
Distatnt Metastasis Free Survival (Months)

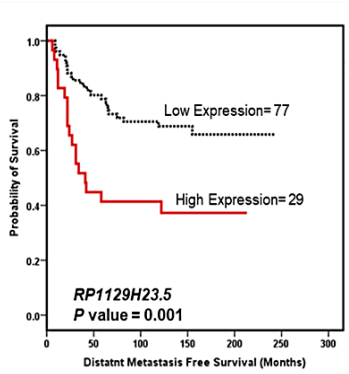

Low Expression= 91
High Expression= 21
DOCK10
P value = 0.020
Probability of Survival
Breast Cancer Specific Survival (Months)

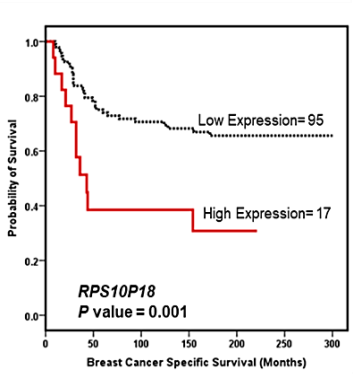

Low Expression= 86
High Expression= 20
DOCK10
P value = 0.037
Probability of Survival
Distatnt Metastasis Free Survival (Months)

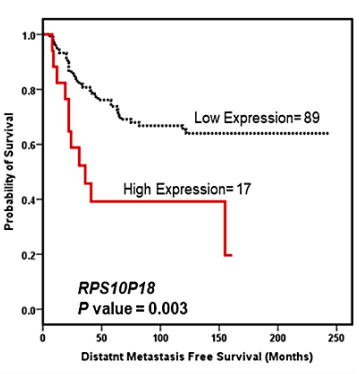

Low Expression= 94
High Expression= 18
PPL
P value = 0.002
Probability of Survival
Breast Cancer Specific Survival (Months)


Low Expression= 89
High Expression= 17
PPL
P value<0.001
Probability of Survival
Distatnt Metastasis Free Survival (Months)


Low Expression= 82
High Expression= 30
RP1129H23.5
P value = 0.002
Probability of Survival
Breast Cancer Specific Survival (Months)


Low Expression= 77
High Expression= 29
RP1129H23.5
P value = 0.001
Probability of Survival
Distatnt Metastasis Free Survival (Months)


Low Expression= 95
High Expression= 17
RPS10P18
P value = 0.001
Probability of Survival
Breast Cancer Specific Survival (Months)


Low Expression= 89
High Expression= 17
RPS10P18
P value = 0.003
Probability of Survival
Distatnt Metastasis Free Survival (Months)

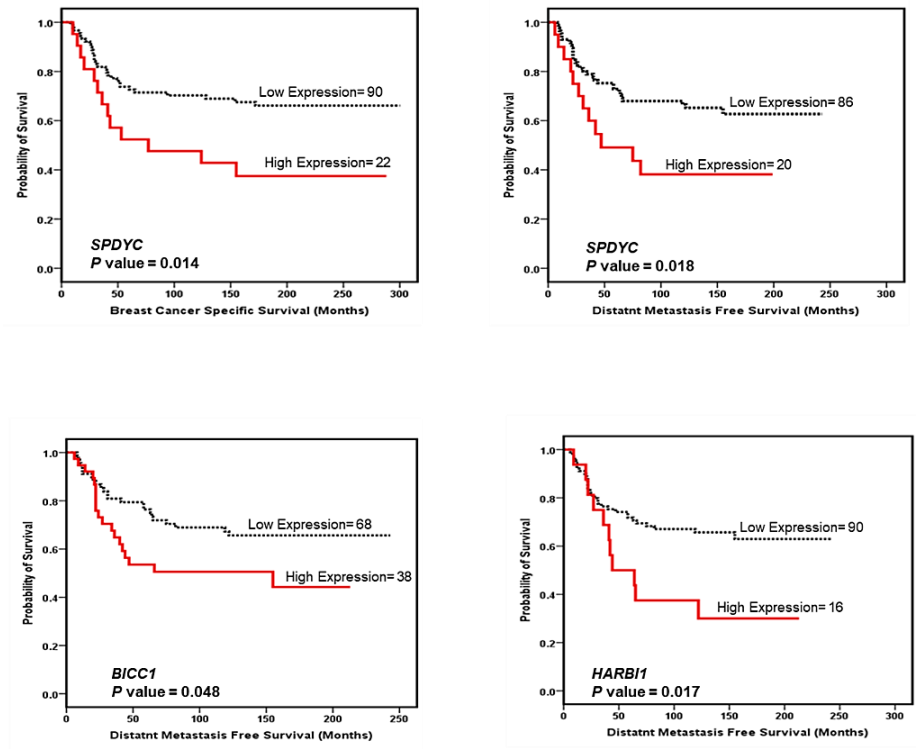


Figure 3-5:Univariate Kaplan-Meier survival analyses of the transcripts identified by Artificial Neural Network analysis to be strongly predictive of Breast Cancer-Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) using RNA seq matrices

Table 3-4:Multivariate Cox regression analysis for genes potentially associated with both Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112)

Covariates	Breast Cancer-Specific Survival				Distant Metastasis-Free Survival			
	P-value	Hazard Ratio	95% CI		P-value	Hazard Ratio	95% CI	
			Lower	Upper			Lower	Upper
Age	0.061	1.836	0.971	3.472	**0.047**	1.92	1.009	3.653
Grade	0.452	1.712	0.421	6.954	0.441	1.743	0.424	7.157
Nodal Stage	0.116	1.582	0.893	2.801	0.173	1.48	0.842	2.6
Tumour Size	0.258	0.675	0.342	1.333	0.137	0.601	0.307	1.176
Vascular Invasion	0.454	1.365	0.604	3.083	0.55	1.277	0.573	2.842
AC020931.1	0.103	0.528	0.245	1.137	**0.043**	0.437	0.196	0.976
Age	0.113	1.682	0.884	3.21	0.101	1.727	0.899	3.315
Grade	0.362	1.914	0.475	7.714	0.341	1.986	0.484	8.15
Nodal Stage	0.102	1.595	0.911	2.793	0.148	1.5	0.866	2.595
Tumour Size	0.351	0.724	0.367	1.427	0.179	0.635	0.327	1.232
Vascular Invasion	0.388	1.413	0.645	3.097	0.416	1.374	0.639	2.953
AC079305.10	**0.046**	1.983	1.012	3.887	**0.004**	2.871	1.394	5.912
Age	0.225	1.512	0.775	2.951	0.184	1.583	0.804	3.117
Grade	0.23	2.364	0.58	9.629	0.219	2.4	0.594	9.7
Nodal Stage	0.077	1.693	0.945	3.034	0.094	1.631	0.92	2.89
Tumour Size	0.239	0.663	0.335	1.314	0.117	0.583	0.296	1.146
Vascular Invasion	0.668	1.203	0.517	2.795	0.723	1.16	0.511	2.637
AC084809.2	**0.047**	2.184	1.01	4.721	0.131	1.848	0.832	4.104
Age	**0.04**	1.969	1.03	3.763	**0.048**	1.919	1.005	3.664
Grade	0.176	2.436	0.671	8.841	0.174	2.466	0.672	9.049
Nodal Stage	**0.015**	2.135	1.161	3.927	**0.027**	1.957	1.08	3.545
Tumour Size	0.165	0.612	0.306	1.224	0.097	0.56	0.282	1.11

Vascular Invasion	0.839	0.914	0.382	2.183	0.796	0.893	0.381	2.096
ACSM4	**<0.001**	3.749	1.964	7.158	**0.001**	3.154	1.645	6.046
Age	0.306	1.421	0.725	2.784	0.198	1.557	0.794	3.053
Grade	0.423	1.762	0.44	7.049	0.318	2.029	0.507	8.129
Nodal Stage	**0.043**	1.808	1.018	3.211	0.07	1.689	0.958	2.977
Tumour Size	0.332	0.715	0.363	1.408	0.145	0.607	0.311	1.188
Vascular Invasion	0.437	1.366	0.623	2.994	0.577	1.247	0.574	2.7111
MEX3A	**0.032**	2.208	1.072	4.546	0.102	1.857	0.884	3.904
Age	0.082	1.751	0.932	3.293	0.077	1.769	0.94	3.329
Grade	0.33	1.937	0.512	7.328	0.32	1.954	0.521	7.33
Nodal Stage	0.141	1.54	0.867	2.738	0.191	1.455	0.83	2.552
Tumour Size	0.384	0.734	0.366	1.472	0.267	0.678	0.341	1.347
Vascular Invasion	0.45	1.358	0.614	3.004	0.523	1.289	0.591	2.811
NDUFA4L2	0.181	1.569	0.811	3.033	0.068	1.846	0.955	3.569
Age	**0.039**	1.96	1.033	3.72	**0.041**	1.949	1.028	3.695
Grade	0.275	2.098	0.555	7.924	0.255	2.169	0.572	8.217
Nodal Stage	0.164	1.524	0.842	2.756	0.196	1.469	0.82	2.63
Tumour Size	0.199	0.639	0.332	1.266	0.099	0.566	0.288	1.114
Vascular Invasion	0.44	1.397	0.598	3.264	0.495	1.339	0.579	3.097
PAXBP1-AS1	**0.049**	0.411	0.17	0.996	**0.036**	0.362	0.14	0.936
Age	0.028	2.081	1.083	3.999	**0.019**	2.185	1.135	4.208
Grade	0.229	2.36	0.583	9.551	0.188	2.581	0.63	10.579
Nodal Stage	**0.035**	1.904	1.047	3.465	**0.034**	1.92	1.052	3.506
Tumour Size	0.26	0.675	0.341	1.337	0.111	0.577	0.294	1.134
Vascular Invasion	0.48	1.346	0.591	3.066	0.51	1.315	0.582	2.97
DCTN1-AS1	**0.003**	3.061	1.452	6.454	**<0.001**	3.874	1.873	8.013
Age	**0.044**	1.92	1.019	3.618	0.05	1.881	0.999	3.543
Grade	0.214	2.293	0.619	8.497	0.203	2.364	0.629	8.885
Nodal Stage	0.272	1.399	0.768	2.55	0.236	1.426	0.793	2.564

Tumour Size	0.204	0.646	0.328	1.269	0.104	0.575	0.295	1.121
Vascular Invasion	0.313	1.573	0.652	3.794	0.462	1.376	0.588	3.217
DOCK10	**0.031**	2.215	1.076	4.56	0.075	1.939	0.935	4.025
Age	0.097	1.74	0.905	3.344	0.11	1.712	0.885	3.313
Grade	0.276	2.19	0.535	8.971	0.241	2.333	0.565	9.626
Nodal Stage	0.198	1.459	0.821	2.595	0.259	1.385	0.787	2.437
Tumour Size	0.301	0.692	0.344	1.39	1.77	0.619	0.309	1.242
Vascular Invasion	0.225	1.687	0.725	3.928	0.266	1.595	0.7	3.637
PPL	**0.004**	2.961	1.424	6.157	**0.004**	2.966	1.414	6.22
Age	0.064	1.845	0.964	3.532	**0.043**	1.977	1.022	3.825
Grade	0.347	1.956	0.483	7.914	0.27	2.212	0.539	9.079
Nodal Stage	0.14	1.538	0.869	2.722	0.207	1.44	0.817	2.538
Tumour Size	0.215	0.645	0.323	1.293.	0.072	0.531	0.267	1.058
Vascular Invasion	0.279	1.581	0.689	3.627	0.336	1.492	0.66	3.374
RP1129H23.5	**0.001**	2.832	1.5	5.346	**0.001**	2.975	1.569	5.641
Age	**0.024**	2.119	1.102	4.078	0.027	2.088	1.085	4.016
Grade	0.36	1.903	0.48	7.541	0.309	2.053	0.513	8.207
Nodal Stage	0.048	1.685	1.005	2.826	0.087	1.566	0.938	2.617
Tumour Size	0.189	0.663	0.32	1.252	0.088	0.556	0.283	1.092
Vascular Invasion	0.181	1.663	0.789	3.503	0.24	1.558	0.755	3.263
RPS10P18	< **0.001**	3.941	1.852	8.389	**0.001**	3.399	1.611	7.171
Age	**0.046**	1.907	1.012	3.596	0.05	1.886	1	3.557
Grade	0.209	2.279	0.631	8.232	0.19	2.378	0.65	8.695
Nodal Stage	0.125	1.561	0.884	2.756	0.129	1.545	0.881	2.71
Tumour Size	0.14	0.596	0.3	1.185	0.066	0.529	0.269	1.042
Vascular Invasion	0.247	1.666	0.702	3.953	0.383	1.452	0.628	3.356
SPDYC	**0.003**	2.892	1.453	5.758	**0.006**	2.646	1.317	5.317

- Significant P-values are bolded CI: Confidence interval

3.6.2 Prognostic Two-Gene Signature

The prognostic gene signature was identified after statistically distilling the eight genes in a multivariate Cox regression analysis to identify a signature with a minimum number of genes that show most significant association with BCSS and DMFS. The analysis revealed two genes *ACSM4* and *SPDYC* that most significantly and independently predicted both DMFS and BCSS (*ACSM4;* DMFS: p=0.015, 95% CI=1.21-6.13, HR=2.72 : BCSS: p=0.004, 95% CI=1.44-6.83, HR=3.14), and (*SPDYC* ; DMFS: p=0.012, 95% CI=1.23-5.45, HR=2.59 :BCSS: p=0.016, 95% CI=1.18-5.09, HR=2.45) (Table 3-5).

Of note, when we examined the association between the mRNA expression of *ACSM4* and *SPDYC,* we observed no linear association (r = 0.036, p=0.710). However, to investigate the prognostic value of the two-gene signature, a linear prognostic score was generated using the sum of the product of normalised expression levels of these two genes and their respective regression coefficients, as follows:

The prognostic two-gene signature score ∑= (*ACSM4* normalized expression * *ACSM4* expression β-value) + (*SPDYC* normalized expression * *SPDYC* expression β-value) (Table 3-6).

Using X-tile cut-off generator, patients with higher mRNA expression score of the prognostic two-gene signature had worse outcome in terms of shorter DMFS and BCSS when compared with those with lower mRNA expression score (Figure 3-6). Cox regression analysis confirmed that the prognostic two-gene signature harbors significant prognostic value in terms of predicting shorter DMFS and BCSS independent of patient age, tumour grade, nodal stage, tumour size, and LVI status (Table 3-7)

Table 3-5:Combined multivariate Cox regression analysis of genes potentially associated with both Breast Cancer-Specific Survival and Distant Metastasis-Free Survival (Transcriptomic Cohort, n=112)

Covariates	Breast Cancer-Specific Survival				Distant Metastasis-Free Survival			
	P-value	Hazard Ratio	95% CI		P-value	Hazard Ratio	95% CI	
			Lower	Upper			Lower	Upper
Age	**0.043**	2.053	1.023	4.119	**0.046**	2.033	1.013	4.079
Grade	0.203	2.331	0.633	8.582	0.208	2.433	0.610	9.703
Nodal Stage	0.181	1.523	0.822	2.819	0.208	1.483	0.803	2.739
Tumour Size	0.202	0.616	0.293	1.296	0.153	0.588	0.283	1.219
Vascular Invasion	0.215	1.813	0.708	4.642	0.253	1.697	0.685	4.206
ACSM4	**0.004**	3.143	1.445	6.834	**0.015**	2.727	1.213	6.135
SPDYC	**0.016**	2.455	1.183	5.096	**0.012**	2.594	1.234	5.450
DCTN1-AS1	0.954	1.032	0.358	2.976	0.304	1.721	0.611	4.849
PPL	0.883	1.079	0.395	2.947	0.926	0.950	0.317	2.841
RP1129H23.5	0.565	1.268	0.566	2.840	0.724	1.158	0.513	2.612
RPS10P18	0.093	2.134	0.882	5.160	0.251	1.716	0.682	4.316
AC079305.10	0.078	2.027	0.924	4.444	0.007	3.014	1.348	6.740
PAXBP1-AS1	0.083	0.432	0.168	1.114	0.051	0.366	0.133	1.005

- Significant P-values are bolded CI: Confidence interval

Table 3-6:Multivariate Cox regression analysis to generate a prognostic score for the two gene signature predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112)

Covariates	Breast Cancer specific Survival					Distant Metastasis Free Survival				
	(B) value	P-value	Hazard Ratio	95% CI		(B) value	P-value	Hazard Ratio	95% CI	
				Lower	Upper				Lower	Upper
ACSM4	1.111	**< 0.001**	3.038	1.653	5.585	1.065	**0.001**	2.900	1.570	5.358
SPDYC	0.745	**0.026**	2.016	1.092	4.063	0.833	**0.016**	2.300	1.166	4.535

- Significant P-values are bolded CI: Confidence interval

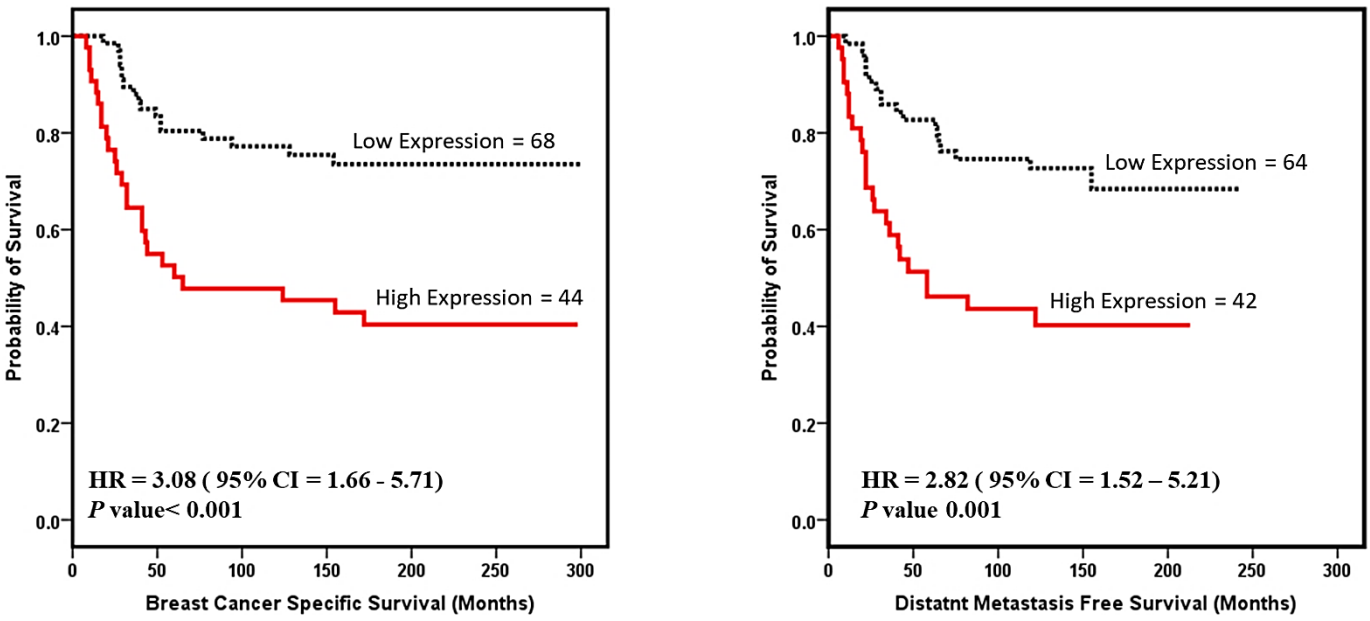


Figure 3-6:Univariate Kaplan-Meier survival analysis of the prognostic two gene signature predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic cohort, n=112)

- Significant P-values are bolded HR: Hazard ratio

Table 3-7:Multivariate Cox regression analysis for the two gene signature predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic Cohort, n=112)

Covariates	Breast Cancer-Specific Survival				Distant Metastasis-Free Survival			
	P-value	Hazard Ratio	95% CI		P-Value	Hazard Ratio	95% CI	
			Lower	Upper			Lower	Upper
Age	**0.011**	2.343	1.217	4.513	**0.015**	2.240	1.169	4.294
Grade	0.182	2.402	0.663	8.705	0.175	2.454	0.671	8.981
Nodal Stage	**0.033**	1.907	1.053	3.454	**0.043**	1.836	1.020	3.307
Tumour Size	0.085	0.540	0.267	1.089	**0.043**	0.490	0.246	0.979
Vascular Invasion	0.694	1.188	1.053	2.805	0.852	1.084	0.466	2.519
two gene signature at mRNA level	< **0.001**	3.891	2.041	7.416	< **0.001**	3.371	1.780	6.384

- Significant P-values are bolded CI: Confidence interval

3.6.3 External Validation of Genomic Findings

Using the Bc-GenExMiner tool to analyze publicly available RNA-sequencing data, we observed that higher expression of *SPDYC* was significantly associated with worse prognosis in the whole/unselective cohorts of BC (n=4308, p<0.0001) [201]. Validating genes expressions on the restricted TNBC cohort (n=254), revealed a similar trend of poor prognosis (p=0.006) [201]. Moreover, the integration of our proposed prognostic two-gene signature in the public domain Genotype 2 outcome, using the median of each gene expression in the whole/unselective cohorts of BC (n=4029), indicated that higher expression of *ACSM4* and *SPDYC* were associated with worse prognosis (both p<0.001). More importantly in the context of this study, the prognostic value of the two-gene signature (*ACSM4 and SPDYC*) was significantly associated with poorer outcome when examined in the TNBC subtype cohort alone (n=612, p<0.001) [35] (Figure 3-7).

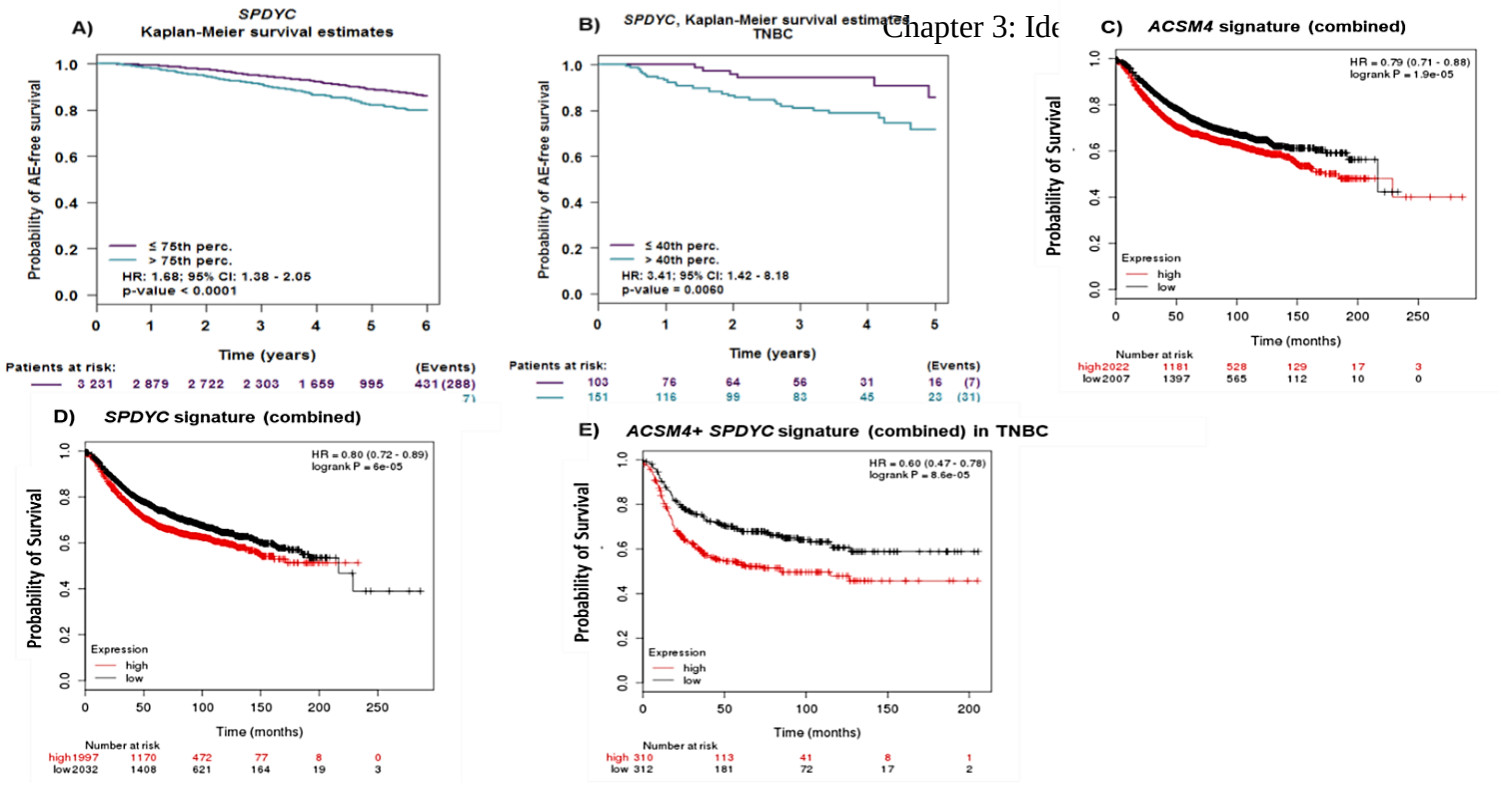


Figure 3-7:Univariate Kaplan Meier survival analysis of our proposed combinatorial two gene signature predicting overall Survival (public domain datasets)

- A) In the Breast Cancer Gene-Expression Miner data portal, high *SPDYC* mRNA expression confers a poor prognosis in the whole (i.e. unselected cohorts) of BC patients (n=4308, p value<0.0001). B) In the Breast Cancer Gene-Expression Miner data portal, high *SPDYC* mRNA expression confers poor prognosis in the TNBC patients (n=254, p value=0.006). C) In the Genotype 2 outcome public portal, high *ACSM4* mRNA expression confers a poor prognosis outcome in the whole (i.e. unselected cohorts) of BC patients (n=4029, p value<0.0001). D) In the Genotype 2 outcome public portal, high *SPDYC* mRNA expression confers a poor prognosis outcome in the whole (i.e. unselected cohorts) of BC patients (n=4029, p value<0.0001). E) In the Genotype 2 outcome public portal, high *SPDYC & ACSM4* mRNA expression confers a poor prognosis outcome in TNBC patients (n=612, p value<0.0001).

3.6.4 Immunohistochemistry of the Prognostic Two-Gene Signature

The morphological assessment of the tissue samples revealed cytoplasmic expression for both proteins; ACSM4 (H-score range 5-295) and SPDYC (H-score range 5-290).

Univariate survival analysis revealed that higher expression of ACSM4 and SPDYC was significantly associated with patients' poor outcomes (DMFS; $p<0.001$, BCSS; $p=0.009$ for ACSM4) and (DMFS and BCSS, both $p=0.004$ for SPDYC) (Figure 3-8), which is concordant with the findings obtained from transcriptomic data.

Multivariate Cox regression analysis showed that SPDYC protein expression was an independent prognostic factor regardless of patient age, tumour grade, nodal stage, tumour size, and LVI status for DMFS ($p=0.015$, 95% CI =1.17 - 4.74, HR=2.365) and BCSS ($p=0.015$, 95% CI =1.18- 4.78, HR=2.377). Likewise, multivariate Cox regression analysis showed that ACSM4 protein expression was a significant independent prognostic factor for DMFS ($p=0.002$, 95% CI=1.35- 3.89, HR= 2.267), but not in BCSS ($p=0.057$, 95% CI=0.98- 2.93 , HR= 1.698) (Table 3-8&Table 3-9).

In a combined multivariate Cox regression analysis, SPDYC protein expression was an independent prognostic factor that predicted shorter DMFS and BCSS (DMFS: $p=0.03$, 95% CI=1.07-5.86, HR=2.50: BCSS: $p=0.03$, 95% CI=1.08-5.96 HR=2.54), regardless of patient age, tumour grade, nodal stage, tumour size, and LVI status. ACSM4 protein expression also was observed to be an independent prognostic factor, associated with shorter DMFS ($p=0.003$, 95% CI =1.01-3.20, HR=1.83), regardless of patient age, tumour grade, nodal stage, tumour size, and LVI status, but not with BCSS ($p=0.27$, 95% CI=0.76-2.56 , HR=1.40) (Table 3-10). Correspondingly, we observed a significant positive linear association between ACSM4 and SPDYC protein expression ($r=0.29$, $p<0.001$), signifying that these proteins might be synergistically driving TNBC disease progression (Figure 3-9). Furthermore, using only cases that were informative for both biomarkers, a linear

prognostic score was generated using Cox proportional hazard analysis to test whether dual expression of SPDYC and ACSM4 proteins was associated with worse outcome. The equation generated used the sum of the product of the quantitative H-score and their respective regression coefficient as follows:

Protein expression prognostic score: $\sum$= (ACSM4 H-score * ACSM4 H-score β value) + (SPDYC H-score * SPDYC H-score β value) (Table 3-11).

This protein expression prognostic score was then dichotomised using X-tile software to determine the optimal score to classify patients into high and low risk groups using DMFS as an end point. In the 257 investigated cases, the scores ranged from 15.43-365.05 with high protein expression risk scores (score > 170) observed in 159/257 (62%) cases.

When testing the association between the prognostic score and outcome, univariate analysis demonstrated that cases with higher protein expression score had a significantly shorter DMFS (p=0.02) but not BCSS (p=0.06) (Figure 3-10). Multivariate Cox regression analysis model demonstrated that protein expression prognostic score was an independent prognostic factor for DMFS (p=0.03, 95% CI=1.04- 3.32 , HR=1.83) independent of patient age, tumour grade, nodal stage, tumour size, and LVI status, but not for BCSS (p=0.07, 95% CI=0.94-2.96, HR=1.83) (Table 3-12).

Finally, when we stratified our cohort based on chemotherapy treatment, the 10-year DMFS of patients who were not offered chemotherapy (n=83) and showed low expression of ACSM4 was 84% compared to 44% of those with high expression and the difference was statistically significant (p=0.005). However, those with low expression of SPDYC had 83% 10-year DMFS compared to 70% in those with high expression but the difference was not statistically significant (p=0.209). Similarly, with the prognostic two gene signature, the 10-year DMFS of patients with low expression was 84% compared to 69% of those with high expression (p=0.309).

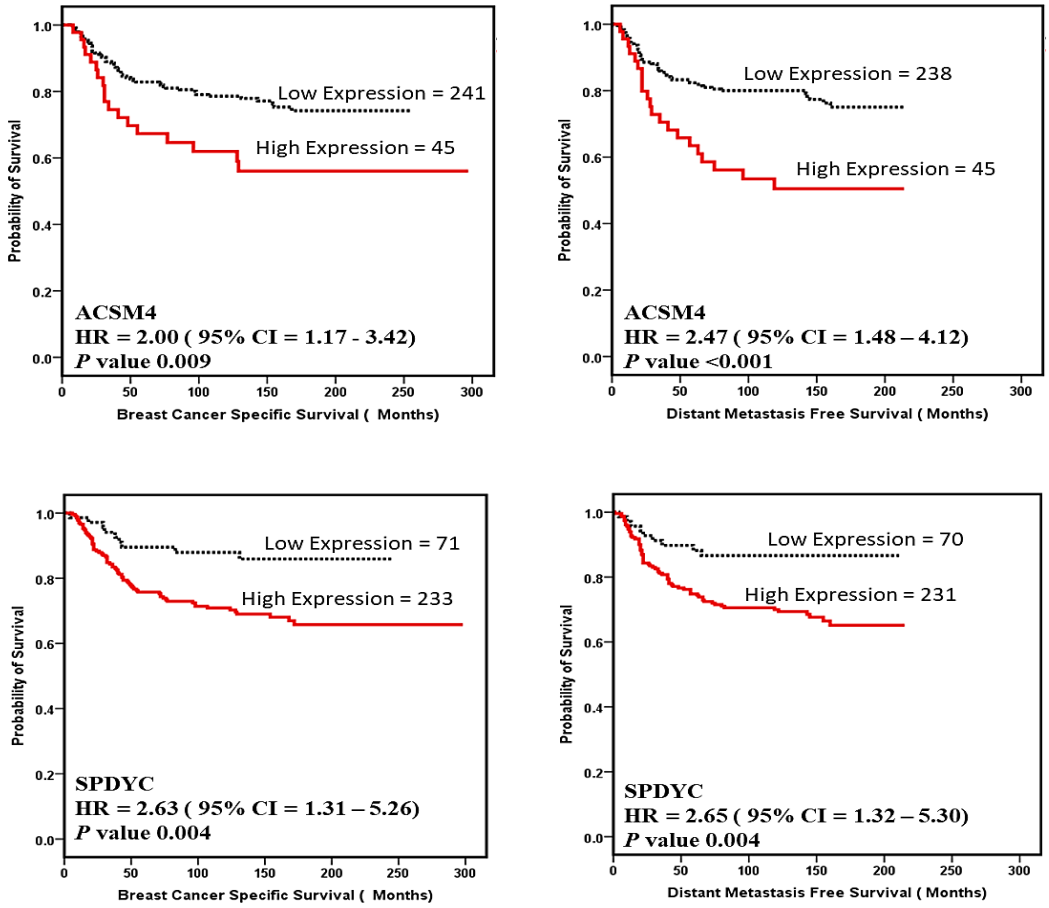


Figure 3-8:Univariate Kaplan Meier survival analysis for ACSM4 and SPDYC protein expression for association with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333)

- Significant P-values are bolded HR: Hazard ratio

Table 3-8:Multivariate Cox regression analysis for Individual Potential proteins associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333)

Covariates	Breast Cancer specific Survival				Distant Metastasis Free Survival			
	P-value	Hazard Ratio	95% CI		P-value	Hazard Ratio	95% CI	
			Lower	Upper			Lower	Upper
Age	**0.011**	1.941	1.165	3.234	**0.030**	1.743	1.055	2.879
Grade	0.494	1.267	0.643	2.498	0.385	1.369	0.674	2.782
Nodal Stage	**0.001**	1.794	1.261	1.552	**0.027**	1.513	1.049	2.180
Tumour Size	0.706	0.908	0.551	1.497	0.619	1.136	0.786	1.879
Vascular Invasion	0.062	1.697	0.975	2.953	0.178	1.466	0.840	2.558
ACSM4	0.057	1.698	0.983	2.933	**0.002**	2.267	1.350	3.809

- Significant P-values are bolded CI: Confidence interval

Table 3-9:Multivariate Cox regression analysis for Individual Potential proteins associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333)

Covariates	Breast Cancer specific Survival				Distant Metastasis Free Survival			
	P-value	Hazard Ratio	95% CI		P-value	Hazard Ratio	95% CI	
			Lower	Upper			Lower	Upper
Age	**0.026**	1.711	1.067	2.745	**0.042**	1.618	1.016	2.576
Grade	0.500	1.292	0.614	2.722	0.422	1.387	0.624	3.079
Nodal Stage	**0.002**	1.757	1.240	2.489	**0.009**	1.600	1.123	2.280
Tumour Size	0.467	0.839	0.522	1.348	0.922	1.024	0.638	1.644
Vascular Invasion	**0.034**	1.783	1.041	3.044	0.086	1.596	0.935	2.724
SPDYC	**0.015**	2.377	1.181	4.783	**0.015**	2.365	1.178	4.748

- Significant P-values are bolded CI: Confidence interval

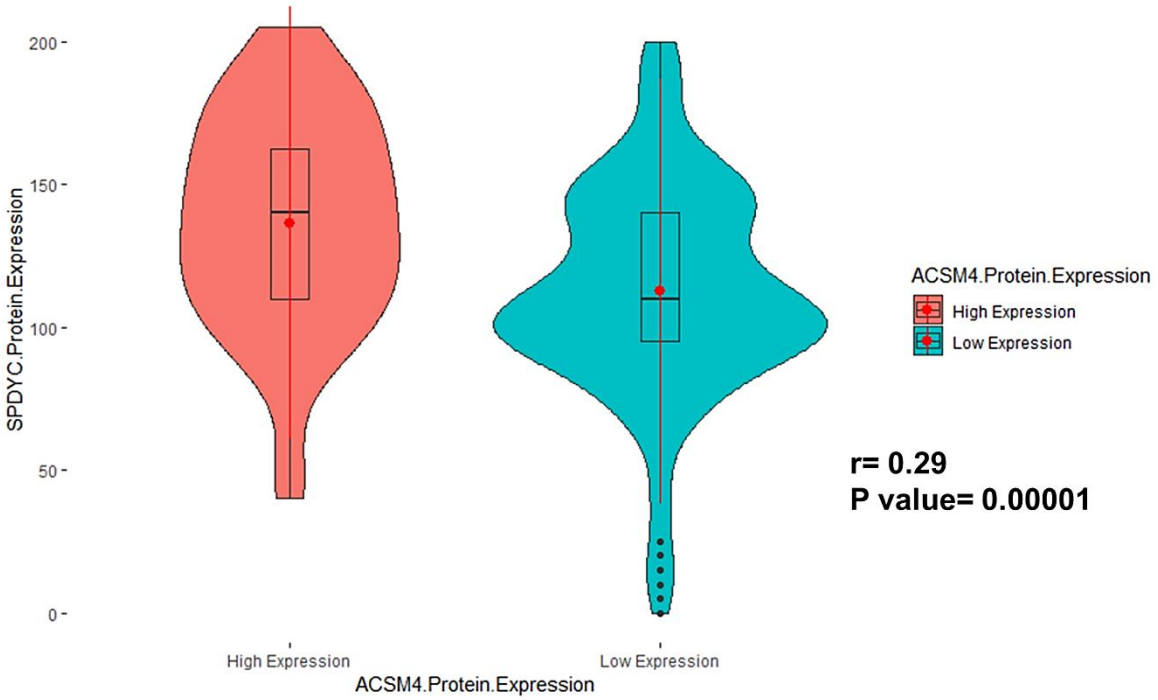


Figure 3-9:Violin plots demonstrating a positive correlation between protein expressions of SPDYC and ACSM4 (Correlation Coefficient, r=0.29, P=0.00001) (IHC cohort, n=333)

Table 3-10:Combined multivariate Cox regression analysis for potential proteins associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333)

Covariates	Breast Cancer-specific Survival				Distant Metastasis-Free Survival			
	P-value	Hazard Ratio	95% CI		P-value	Hazard Ratio	95% CI	
			Lower	Upper			Lower	Upper
Age	**0.049**	1.733	1.003	2.995	0.105	1.548	0.913	2.624
Grade	0.754	1.128	0.530	2.393	0.641	1.210	0.542	2.700
Nodal Stage	0.012	1.640	1.114	2.414	0.051	1.477	0.988	2.185
Tumour Size	0.282	0.741	0.430	1.279	0.942	0.980	0.573	1.678
Vascular Invasion	0.022	2.018	1.108	3.676	0.077	1.702	0.945	3.067
ACSM4	0.274	1.402	0.766	2.568	**0.036**	1.826	1.014	3.204
SPDYC	**0.031**	2.545	1.086	5.960	**0.034**	2.508	1.072	5.869

- Significant P-values are bolded CI: Confidence interval

Table 3-11:Multivariate Cox regression analysis to build prognostic index for the two gene signature predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333)

Covariates	Breast Cancer specific Survival					Distant Metastasis Free Survival				
	(B) value	P-value	Hazard Ratio	95% CI		(B) value	P-value	Hazard Ratio	95% CI	
				Lower	Upper				Lower	Upper
ACSM4	0.46	0.129	1.587	0.874	2.883	0.69	**0.017**	1.972	1.130	3.440
SPDYC	1.05	**0.014**	2.869	1.232	6.678	1.05	**0.015**	2.859	1.229	6.651

- Significant P-values are bolded CI: Confidence interval

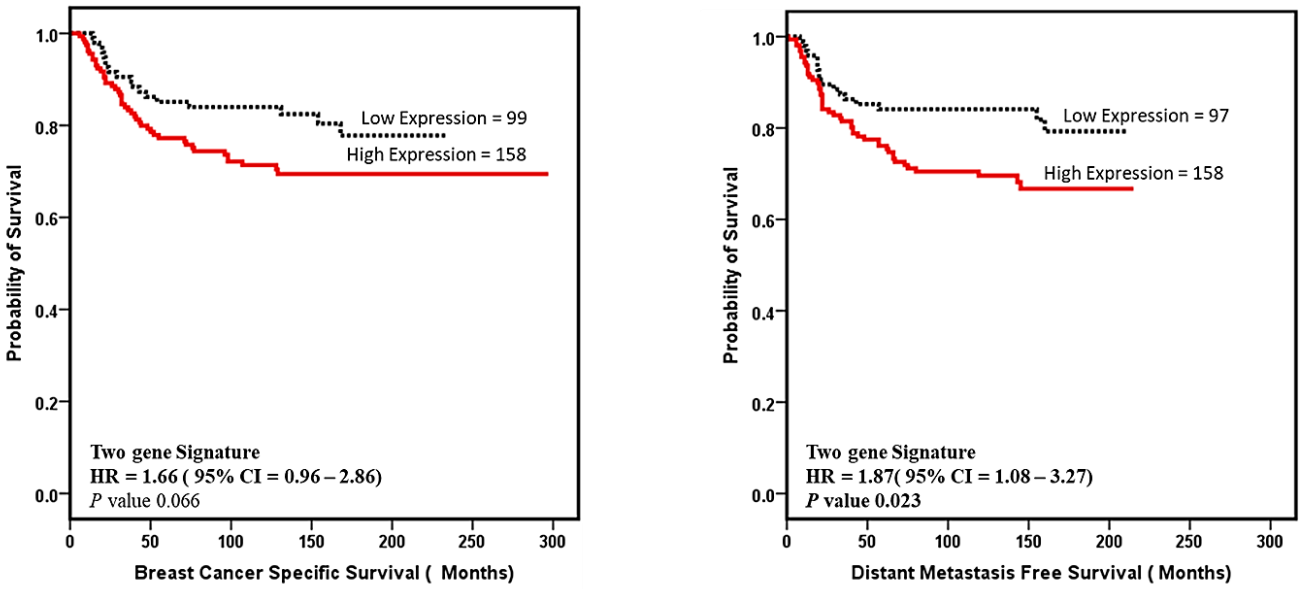


Figure 3-10:Univariate Kaplan Meier survival analysis of the protein expression of the two gene signature score predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333)

- Significant P-values are bolded HR: Hazard rati0

Table 3-12: Multivariate Cox regression analysis for the protein expression prognostic score predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC Cohort, n=333)

Covariates	Breast Cancer specific Survival				Distant Metastasis Free Survival			
	P-value	Hazard Ratio	95.0% CI		P-value	Hazard Ratio	95.0% CI	
			Lower	Upper			Lower	Upper
Age	**0.037**	1.782	1.037	3.064	0.080	01.597	0.946	2.698
Grade	0.734	1.134	0.549	2.343	0.607	1.227	0.563	2.675
Nodal Stage	**0.009**	1.677	1.139	2.468	**0.040**	1.509	1.019	2.235
Tumour Size	0.237	0.721	0.419	1.240	0.796	0.932	0.545	1.592
Vascular Invasion	**0.018**	2.041	1.130	3.684	**0.047**	1.802	1.009	3.219
Two gene signature at protein level	0.077	1.637	0.946	2.960	**0.034**	1.867	1.049	3.323

- Significant P-values are bolded CI: Confidence interval

3.6.5 Testing the performance of the prognostic two-gene at the transcriptomic and protein Levels

The prognostic signature at the mRNA level captured 58% sensitivity, 69% specificity, 54% positive predictive value (PPV), 72% negative predictive value (NPV), and 64% accuracy in dichotomizing distant metastasis outcome of TNBC patients. In comparison, the prognostic signature at the protein level showed 73% sensitivity, 42% specificity, 30% PPV, 82% NPV, and 50% accuracy in dichotomizing distant metastasis outcome of TNBC patients (**Table 3-13**).

Table 3-13:Distribution of cases within transcriptomic and IHC cohorts to test the prognostic index with reference to clinical evidence of distant metastasis.

	Distant Metastasis			
	Expression level	**Yes**	**No**	**Total**
Prognostic index of the two gene signature at mRNA Level	**Above cut-point Expression**	**25 TP**	**21 FP**	**46**
	Below-cut-point Expression	**18 FN**	**46 TN**	**64**
Total		**43**	**67**	**110**
	Distant Metastasis			
	Expression level	**Yes**	**No**	**Total**
Prognostic index of the two gene signature at protein level	**Above cut-point Expression**	**48 TP**	**111 FP**	**159**
	Below-cut-point Expression	**18 FN**	**80 TN**	**98**
Total		**66**	**191**	**257**

TP = true positive, FP = false positive, TN = true negative, and FN = false negative. TP = true positive, FP = false positive, TN = true negative, and FN = false negative. Sensitivity = TP * 100/(TP+FN). Specificity = TN * 100/(TN +FP). PPV, the proportion of subjects with distant metastasis who are correctly diagnosed; PPV = TP* 100/(TP + FP). NPV, proportion of subjects without distant metastasis who are correctly diagnosed; NPV= TN * 100 / (TN + FN). Accuracy = TP + TN * 100/Total number of cases.

3.7 Discussion

Molecular classification of BC provides opportunities for enhanced personalised therapy [238]. In TNBC, conventional prognostic factors such as age, tumour size, tumour grade, and lymph node status have limited risk-predictive influence as these tumours are mostly of higher grade with increased chances of recurrence and metastasis [107]. Therefore, deciphering genomic profiles of TNBC using advanced techniques is an unmet need. Moreover, the utilisation of ANN to mine the transcriptomic profile of TNBC in order to identify genes associated with clinical outcome is a promising approach to stratify patients for risk prediction [239].

In the current study, a discovery phase and two validation phases were implemented. The in-house transcriptomic TNBC cohort was used for the discovery phase for ANN analysis. Whereas the protein expression and publicly available external transcriptomic BC data were used for the validation phases of findings. More importantly, regardless of the statistical differences in the distribution of clinicopathological parameters between transcriptomic and IHC cohorts, our gene signature showed statistical association with outcome both at transcriptomic and protein expression level. Our study supports the utility of applying ANN to integrate distinct clinical and molecular data to find novel prognostic biomarkers associated with TNBC poor outcome.

Our study employed ANN for the analysis of our transcriptomic cohort to discover novel prognostic genes associated with outcome in TNBC. ANN is a powerful tool for the analysis of complex data, overcoming high background noise, and thus identifying the influence of many interacting factors [240]. ANN analysis, unlike conventional statistical approaches such as hierarchical clustering, linear regression, and principal component analysis, is not limited by linear functionality; thus, identification of biological relationships

between biomarkers and clinical outcomes is improved [232]. Furthermore, unlike conventional statistical techniques used in the medical diagnostic and prognostic approaches, ANN can produce greater accuracy model than its counterparts [241]. Therefore, it is highly suitable for the identification of potential key genes driving TNBC outcomes. ANN modelling uses a supervised learning approach, a multi-layer perception architecture with a sigmoid transfer function, where weights are updated by a back propagation algorithm [242].

In this study, ANN analysis identified the top ranked genes predicting DMFS and BCSS. We then employed a web-based tool to identify the signaling pathways significantly enriched in the significant top ranked gene panels. For instance, TNBC patients frequently harbour higher expression of EGFR; however, studies have failed to establish significant benefit from EGFR-targeted therapies or tyrosine kinase inhibitors, suggesting the need to therapeutically target other pathways in these tumours [110,243]. Moreover, the significance and over-activation of pathways such as; P38 MAPK , the PDGF, and the RAS pathways in BC metastatic sites and their association with DMFS and BCSS in TNBC have been previously documented [244–246]. Additionally, the 21 gene panel generated by ANN analysis that was strongly associated with both DMFS and BCSS in TNBC included several novel and potentially targetable biomarkers in TNBC outcome. For instance, higher expression of *DOCK10* (also known as dedicator of cytokeratin-10/ZIZ3) [247], has been previously identified as an indicator of poor prognosis in TNBC patients and as a predictor of distant metastasis [248]. In our transcriptomic cohort, *DOCK10* emerged as a significant prognostic marker of BCSS and DMFS however, it was not significantly prognostic in multivariate Cox regression analysis. We also found that high expression of *BICC1*, an RNA binding protein, a negative regulator of the WNT signaling pathway with potential

involvement in regulating gene expression during embryonic development [249], was associated with DMFS but not with BCSS; thus, it was not included in the final signature.

In our study, we distilled the initial 21 gene panel down to eight genes that when tested individually for their prognostic value, were significantly associated with both DMFS and BCSS using univariate and multivariate analysis after adjusting for the potentially confounding variables. These genes are implicated in pro-oncogenic pathways in BC. *PPL* is a part of the cornified envelop in keratinocytes and desmosomes with intermediate filaments. *PPL* can act in the PKB/AKT-mediated signaling pathway [250]. In TNBC, silencing *PPL* decreased cell migration and invasion [251]. *SPDYC* is a member of the speedy/Ringo cyclin-dependent kinase (CDK) family with known functions in cell cycle transitions and progression [252]. *SPDYC* plays an important role in activating both *CDK1* and *CDK2* expression [253]. CDK2 high expression has been previously described to be associated with shorter survival in metastatic melanoma cases and endocrine resistance in SKBR3-HER2 positive BC cell lines [254],[255]. Furthermore, down regulation of *CDK1* has been found to increase synthetic lethality of TNBC cell lines if accompanied with *c-Myc* over expression [256]. However, *SPDYC* role in BC is still undefined [252]. *ACSM4* encodes a protein with known functions in the conjugation of carboxylic acids and in fatty acid beta oxidation. Interestingly, upregulation of metabolic pathways has been found to interact with cellular transcriptomic and proteomics of both CD4 and CD8 T cells in HIV disease [257]. Although *ACSM4* has been shown to have a role in AIDS progression, there are no reports with its role in BC [258],[259]. We have previously reported a strong correlation between tumour infiltrating lymphocytes (TILs) and TNBC outcome [260]. However, our current analysis did not identify known inflammation and immune response related genes associated with outcome in the TNBC 21 gene panel. Future studies should therefore seek to identify novel mechanisms contributing to aberrant inflammatory and

immune response pathways involved in TILs in TNBC. Furthermore, genes such as *AC020931.1, DCTN1-AS1, RP11-29H23.5, PAXBP1-AS1,* and *RPS10P18* require further investigation to decipher their role and function in BC progression.

The original hypothesis underpinning this study was that a gene expression signature would more accurately predict both DMFS and BCSS in TNBC than a single gene. Multivariate Cox regression analysis enabled us to further filter the set of eight genes to a prognostic two-gene signature (*ACSM4* and *SPDYC*) showing strong association with both DMFS and BCSS. We tested whether immunohistochemical assessment of the protein expression of the ACSM4 and SPDYC genes could be used to predict patient outcomes. Our study confirmed that protein expression had independent prognostic significance in TNBCs and showed strong statistical association with worse outcomes (i.e., shorter DMFS and BCSS). These genes when combined in a linear score, successfully stratified TNBC patients into high- and low-risk subgroups; in the former group, which is at a higher risk of developing distant metastasis, could benefit from greater vigilance and more aggressive treatment regimens. We have validated our ANN investigation and RNA-sequencing results by studying protein expression which showed that a prognostic score derived from the immunohistochemical evaluation of the two biomarkers could significantly predict distant metastasis, and thus support personalised prognostic evaluation and guiding treatment choices to improve disease outcomes.

In this study, the prognostic value of the two-gene signature at the mRNA level yielded 58% sensitivity, and 64% accuracy in dichotomising distant metastasis outcome of TNBC patients. By contrast, at the protein level, our proposed two-gene signature demonstrated 73% sensitivity, and 50% accuracy in dichotomising distant metastasis outcome of TNBC patients. Our proposed two-gene signature showed promising accuracy and sensitivity results in predicting the risk of distant metastasis in TNBC patients, which is even more

important as presently TNBC patients solely rely on chemotherapy treatment. Moreover, those patients who are deemed at high risk of distant metastasis may benefit from the stratification for an improved treatment decision.

Furthermore, our proposed two-gene signature is only based on two genes (*ACSM4* and *SPDYC*), unlike other commercially available prognostic assays including those designed for ER-positive tumours [261]. Our prognostic gene signature may be amenable to the development of affordable molecular tests based on quantitative RT-PCR as the sensitivity, specificity, and accuracy of our two-gene signature is proved to be much stronger at the mRNA level. The prognostic gene signature might be suitable for use in routine clinical practice because the proposed two-gene signature has prognostic value in dichotomising TNBC patients and may provide important information for treatment decisions.

The mainstay of TNBC treatment is cytotoxic chemotherapy [262]. However, chemotherapy decision for metastatic TNBC patients are given based on a combination of aspects relates to the disease and patient physical characteristics (i.e., tumour burden, patient age, co-morbidities, prior treatments received in the adjuvant setting, and patient preference) [263]. Despite the interesting finding of this study and the significant difference in the survival of patients who were not offered chemotherapy based on the expression of ACSM4 (with worse outcome of patients with over expression), the 10-year DMFS of patients with low expression (84%) may not justify recommendation for omission of chemotherapy in those patients. However, to make such a recommendation, a clinical trial utilising a sufficiently large number of TNBC patients may be warranted to determine whether TNBC patients with low ACSM4 expression can avoid chemotherapy without worse outcome.

3.8 Limitation of the study

A challenge of applying the NGS technique to deciphering the molecular characteristics of TNBC tumours includes access to the technology and the integrity of tumour samples to guarantee sufficient tumour RNA extraction [30]. Variation in sample quality and preparation may negatively influence the outputs of NGS analysis and therefore must be carefully controlled. In addition, NGS analysis must consider intrinsic tumour heterogeneity between patients. Samples used in this study were processed in a strictly standardized procedure implemented in Nottingham University Hospitals with immediate sample fixation following surgery, with standard protocols optimised to preserve tissue architecture, subcellular details and importantly the integrity of biologic materials including proteins, DNA, and RNA. Nonetheless, our retrospective study was limited to a single center using an in-house transcriptomic and protein expression cohort for this investigation. Likewise, Within the proteomic cohort, a significant negative linear correlation was observed between *SPDYC* mRNA (assessed by NGS) and protein level (r= -0.29, p=0.008, 80/112) in contrast to *ACSM4*, which showed a trend towards positive linear association (r=0.08, p=0.52, 65/112). Various factors could have contributed to the inverse correlation between the transcriptomic and proteomic levels of SPDYC. For instance, the improved quantification of the mRNA isoform and differential exon usage for the gene underlying the protein, the depth used in sequencing the cases, and potential considerable difference of proteins in vivo half-lives as compared to the mRNA and vice versa [264]. However, the public domain data used in this study supports the value of both *ACMS4* and *SPDYC* high expression conferring poor prognosis for BC patients, especially those diagnosed with TNBC molecular subtype. Hence, further external validation is strongly recommended.

3.9 Conclusion

Personalised medicine seeks to stratify BC patients to ensure optimal treatment and thus, improved patient outcomes. Our study has identified a two-gene signature that stratifies TNBC patients into high and low risk groups for developing distant metastasis, which can potentially guide clinical decision-making. The robust methods used herein to identify our prognostic gene signature followed by validation of the findings at the protein expression level, suggest that this promising two-gene signature provides avenues for further *in vitro* functional investigation and for new drug development for TNBC patients who are in dire need of effective therapeutic options.

Chapter 4 : Comprehensive Transcriptomic Profiling Reveals PPL as a Novel, Independent Predictor of Outcomes in TNBC

This chapter is currently under review in Clinical Cancer research under the title "Mechanistic and clinical evidence supports a key role for Periplakin (PPL) as a novel independent predictor of outcome in Triple Negative Breast Cancer. (appendix 7.3)

4.1 **Background**

The recent advancement in omics technologies, including microarray and other high throughput sequencing platforms such as NGS have shown to provide an effective tool to decode key molecular mechanisms of tumourigenesis by accelerating the discovery of promising biomarkers for cancer diagnosis, treatment and prognosis [265]. The transcriptomic profiling of TNBC by NGS has enabled the identification of genomic alterations such as copy number variations, insertions, deletions, and mutations. As a result, studies exploring both inter- and intra tumour heterogeneity in different tumour types is now possible [40,228]. Through bioinformatic approaches using ANN analysis to mine an in-house transcriptomic TNBC cohort, we have previously identified the cell adhesion molecule periplakin (PPL*)* as a strong predictor of poor outcome in TNBC (Chapter 3)

4.2 **Role of Periplakin in Cancer**

The human *PPL* gene, located on chromosome 16p13.3, encodes the periplakin protein and is a member of the large multi-domain plakin family responsible for retaining epithelial cell adhesion through various functions including envoplakin, plectin, and bullous pemphigoid antigen 1[266–268]. PPL, the smallest member of the plakin family, has an estimated 195 kDa molecular mass, contains a central rod dimerisation domain where its C-terminal region is involved in intermediate filament binding, and a N-terminal domain that has been shown to connect with cortical actin [269]. PPL is recognised as a key component of the cornified envelope precursor and is targeted to cell membranes and desmosomes in keratinocytes [269]. PPL forms a heterodimeric complex with envoplakin, which

collectively make up a key antigen in autoimmune diseases [250]. Previous studies have revealed that PPL acts as a localisation signal in the oncogenic serine/threonine protein kinase AKT/protein kinase B (PKB)-mediated signalling pathway in human cancer cell lines [270]. Activation of the AKT molecule has been shown to induce phosphorylation of regulators of the mammalian target of rapamycin (mTOR) pathway, leading to pathway activation, which enhances protein synthesis, tumour cell growth, and chemoresistance in BC [271].

An integrated bioinformatic approach to screen the transcriptome of TNBC samples and brain-metastatic cancer samples identified upregulated PPL expression as one of the top differentially expressed genes associated with TNBC metastasis to the brain. Additionally, PPL upregulation induced by Cyclin A2 expression in endometrial cancer cells showed association with aggressive tumourigenesis behaviour as well as cisplatin chemotherapy resistance [251,272]. In contrary, PPL downregulation has been linked with tumourigenesis and progression of multiple cancer types such as: bladder, colon, oesophageal squamous cell carcinoma, Prostate [273–277]. However, data pertaining PPL role in TNBC and its correlation with outcome is lacking.

4.3 Hypothesis and Aims of the study

These studies propose promising roles for PPL as a key regulator of various biological processes through its extraordinarily versatile interactome. Therefore, we hypothesised that *PPL* plays an important role in promoting TNBC progression and metastasis, as well as it can potentially act as a novel therapeutic target in TNBC. This study is aimed at exploring the prognostic potentials and possible biological interactions associated with PPL expression in TNBC at transcriptomic, proteomic, and mechanistic functions to investigate its potential to be a candidate for TNBC target therapy.

4.4 Material and Methods

4.4.1 Study cohort

This chapter was carried out as collaborative work with Prof Nigel Monga, Dr. Jennie Jeyapalan, Daisy Haigh, Corinne Woodcock, Jennifer Lothion-Roy, and Anna Harris, who performed the siRNA knockdown experiments for RNA-sequencing of examined cell lines, qRT-PCR confirmation of knockdown, and assisted with bioinformatic analysis interpretation due to COVID19 restrictions in lab access from March 2020. The experimental work including mechanistic studies, bioinformatic and basic statistical analysis, IHC staining and evaluation were completed by the author.

This study was conducted on multiple well-characterised cohorts using different statistical and molecular techniques (Figure 4-1). The TNBC NGS (n=112) and TNBC METABRIC cohorts (n=313) were referred as “transcriptome cohort” and used a discovery cohort, whereas the “IHC cohort” (n=245) was used to validate the prognostic value of PPL expression at the protein level using IHC. Patients outcome data together with clinicopathological characteristics were collected from patients’ medical records (Chapter 2).

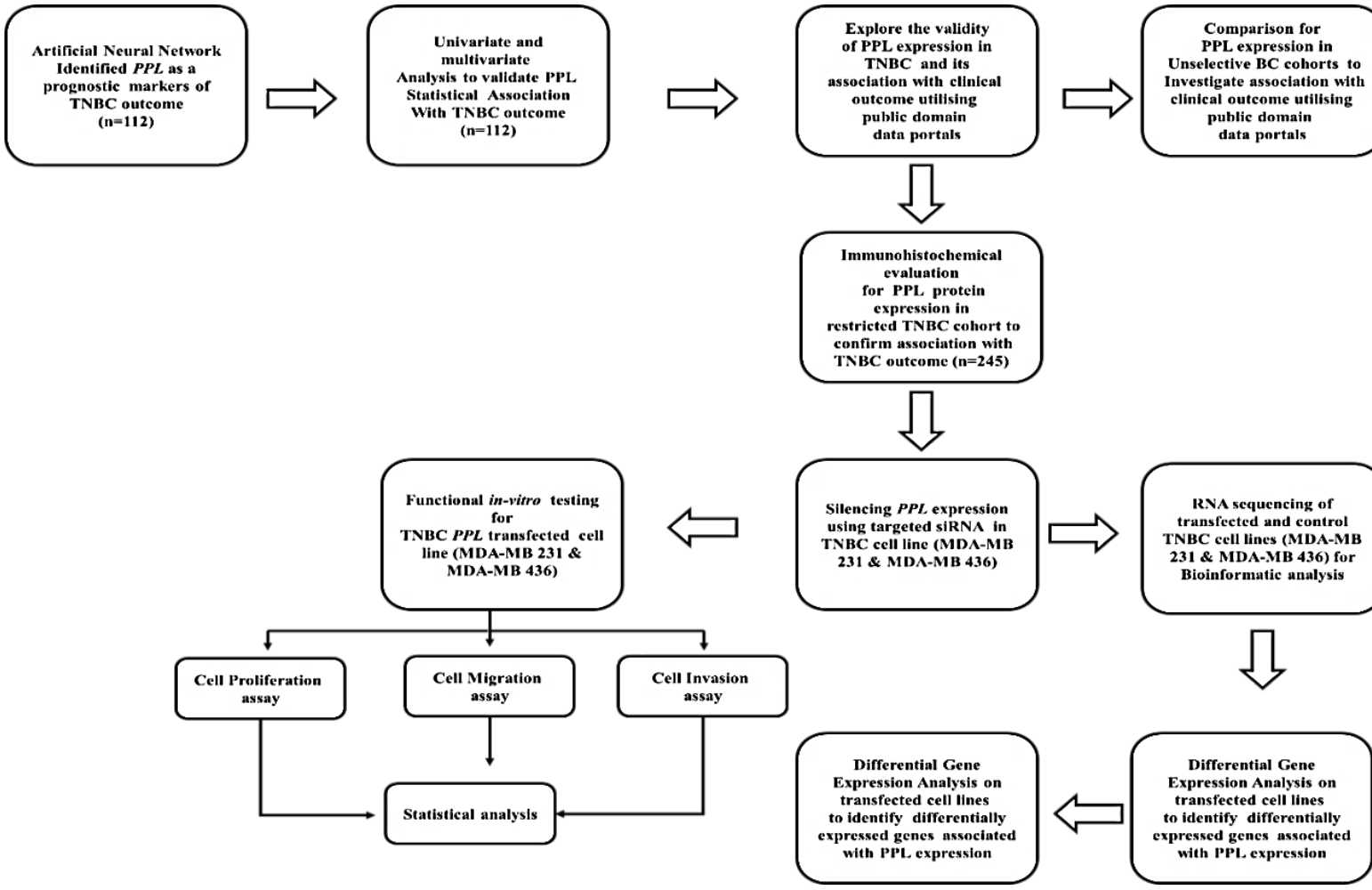


Figure 4-1:Flowchart of Analytical and Experimental Methodology used to explore the prognostic value of PPL expression in TNBC

4.4.2 External Validation of Transcriptomic Data

For independent validation of the results, the prognostic value of the of *PPL* as a predictor of DMFS and BCSS were examined using public domain data (Chapter 2).

4.4.3 Immunohistochemistry and western blotting analyses

Specificity of the rabbit polyclonal anti-PPL antibody (Ab72422, lot number GR2028009-13, Abcam, UK) was validated by western blotting prior to immunohistochemistry staining (IHC) using MCF7 cell lysates (obtained from American Type Culture Collection, Rockville, MD, USA) at a 1:500 dilution overnight at 4°C. Mouse β-Actin (A5441, Sigma-Aldrich; Clone AC-15; Sigma, UK) at 1:5000 was used as a house-keeping protein for overnight incubation at 4°C. Bands were detected using fluorescent secondary antibodies and blot visualisation was performed as previously described (Chapter 2) (Figure 4-2).

PPL mRNA expression obtained from the TNBC-NGS and METABRIC-TNBC cohorts were correlated with PPL protein expression in the same cases where available (n=13). Staining was scored blindly by the author and another researcher (Dr Michael Toss) for 25% of cores to assess inter-observer concordance (ICC). Inter-observer agreement was determined, and the ICC was 0.80, indicating an excellent concordance between scorers. Discordant cases were re-scored by both observers and a final score was agreed.

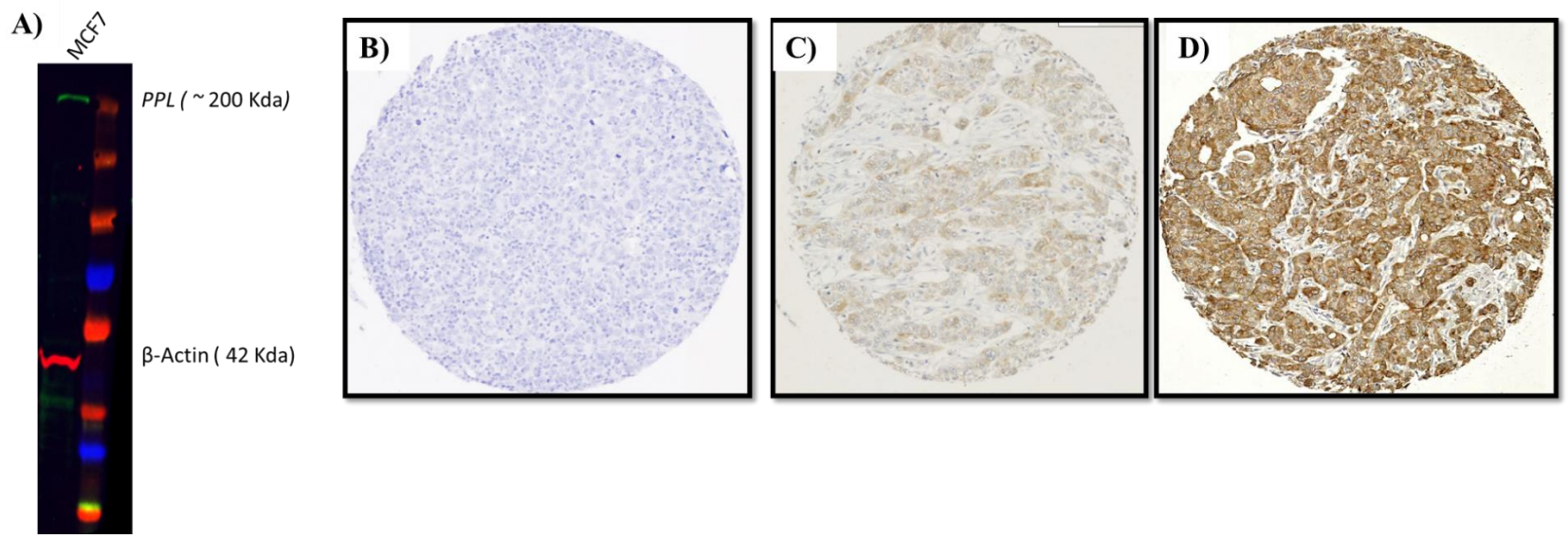


Figure 4-2:Western blot and immunohistochemical expression of PPL in TNBC

- Western blotting results for PPL protein expression in MCF7 cancer cell lines using rabbit polyclonal antibody against human PPL. Green and red bands represent PPL and Beta-Actin (housekeeping) respectively. Immunohistochemical protein expression of PPL in TNBC TMA cores (power 20×). B) PPL negative expression, C) PPL weak expression, D) PPL high expression respectively

4.4.4 Evaluation of the mechanistic and phenotypic functions of PPL in TNBC cell lines

All BC cell lines used in this study were purchased from American Type Culture Collection (ATCC, Manassas, VA USA) and cultured as per ATCC recommendations as previously described (Chapter 2). Cell lines were used within 5 to 13 passages from the ATCC original stocks and authenticated by Short Tandem Repeat (STR) profiling or variant detection within transcriptomic data within the last three years. Cell lines were routinely tested for mycoplasma contamination (CUL001B; R&D Systems) prior to experiments. Antibiotics were not added to any of the experiments.

4.4.5 siRNA transfection

siRNAs were used to functionally deplete *PPL* (*siPPL*) in TNBC cell lines. We first compared the efficacy of three independent siRNA (IDs: S10916, S10917, and S10918) relative to a non-targeting scrambled (ID4390843) control siRNA (*siSCR*) (all purchased from Ambion, ThermoFisher Scientific, UK) (

Table 4-1). All siRNA targeting *PPL* showed similar knockdown (Figure 4-3) as previously reported by Long and colleagues [278]. Therefore, we prioritised only one siRNA (S10917) for subsequent functional and mechanistic studies in MDA-MB-231 and MDA-MB-436 (Figure 4-3). Briefly, 5×10^4 cells were seeded per well in a 24-well plate and transfected by the reverse transfection method with Lipofectamine™ RNAiMAX Transfection Ragent (13778150; ThermoFisher Scientific, UK) as per the manufacturer's protocol. Cells were incubated for five days then lysed and collected in RIPA buffer (89900; ThermoFisher Scientific, UK) and efficiency of knockdown was measured by both western blotting and quantitative real-time PCR.

Table 4-1:The siRNA oligonucleotides used in this study

siRNA oligo	Sequence (5′ to 3′)	Target region
s10916	UGCUUCUAACUCAUAGUCCtt	exon 19, 20
s10917	UAUCUCAUCUAGCAGGUUCtt	exon 18, 19
s10918	UCACGAACAUGUUCCAGUCaa	exon 22

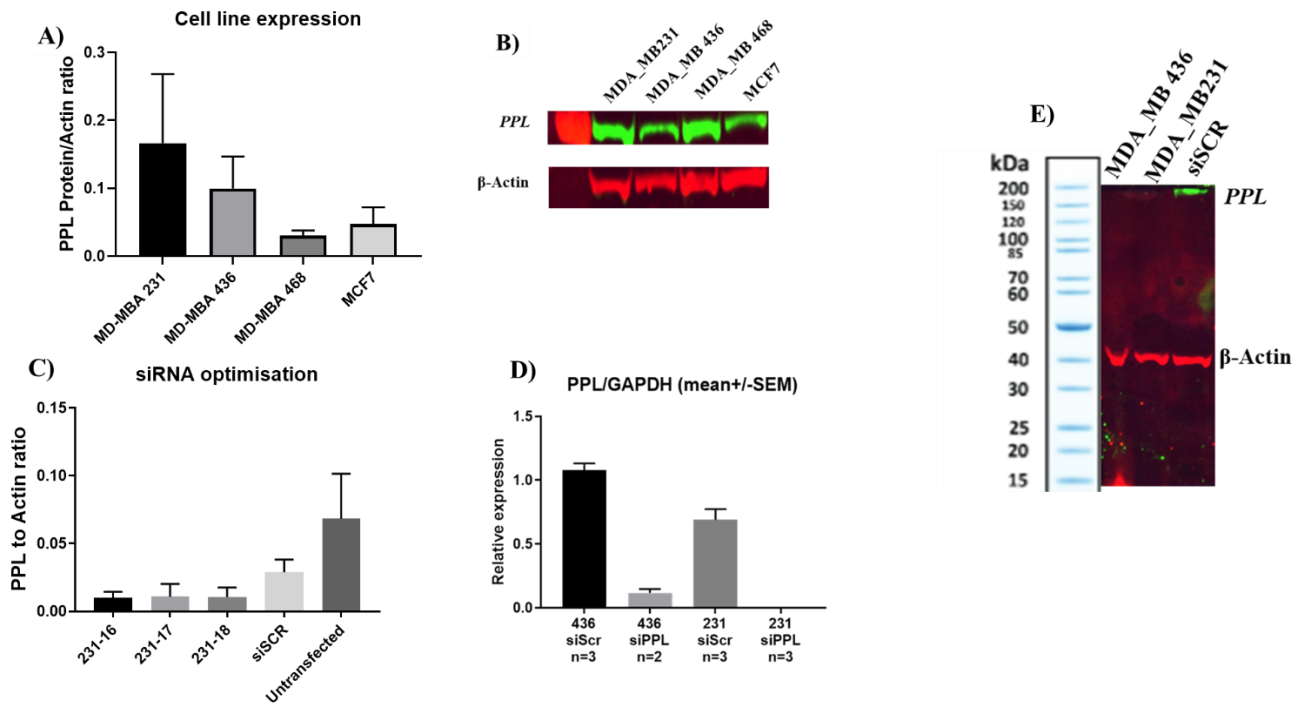


Figure 4-3:Differential expression of PPL protein in breast cancer cell lines. PPL protein levels were quantified by densitometry and normalised to b-actin / GAPDH levels

- PPL at protein level highly expressed in MB-MDA 231 and MB-MDA 436 TNBC cell lines , while low expression was noticed in MB-MDA 468 and MCF7 BC cell lines. B) Three Silencer Select Pre-Designed functional siRNAs (S18, S17 and S16) targeting PPL and a siRNA non-targeting negative control siRNA were used in addition to un transfected cells. B) quantitative RT-PCR analysis to confirm efficacy of PPL depletion in TNBC cell lines. D) Western blotting analysis to confirm efficacy of PPL depletion in TNBC cell lines.

4.4.6 Proliferation Assay

The effect of *PPL* knockdown (KD) on proliferation was assessed by the 3–(4,5–dimethylthiazol-2–yl) -2,5–diphenyltetrazolium (MTS) assay (Promega, (G3580); Cell Titre 96 Aqueous One Solution Cell Proliferation Assay). In brief, the *siSCR* and *siPPL* cells were seeded at 4000 cells/well in 96-well plate at 37 °C in a 5% CO_2 incubator. Proliferation was measured at 24, 48, 72 and 96 hours by adding MTS reagent to the wells as per the manufacturer's protocol. The plates were incubated in for 1 hour and the absorbance of each well was measured using a Synergy™ 4 (BioTek Instruments, Winooski, VT, USA) at 490 nm. All experiments were carried out in triplicate.

4.4.7 Migration and Invasion assays

Cell migration was determined using the Radius™ 24-well from Cell Biolabs (CBA-125; San Diego, CA). In this assay cells were seeded on Radius cell migration plates and allowed to form monolayers before removing the gels. After 24 hours, the central gel layer was removed following the manufacturer's instructions and cell migration was checked by photographing the plates at 24 and 48 hours using a Zeiss LSM-510 inverted microscope (Zeiss, Germany).

We used the CytoSelect 24-Well Cell Invasion Assay (Basement Membrane, Colorimetric) from Cell Biolabs (CBA-110 San Diego, CA) to assess cell invasion. Cells were incubated overnight in serum-free medium before detaching and seeding in matrix-coated trans-wells. The bottom chamber contained media supplemented with 10% FBS. After 24 hours incubation, cells remaining in the top chamber were removed with moistened cotton swabs and treated with cell stain solution from the kit for 10 minutes followed by the adding the extraction solution as per the manufacturer's instructions and calculated OD at 560 nm.

4.4.8 BC cell line RNA-seq and bioinformatic analyses

Optimal *PPL* knockdown was achieved after 96 hours siRNA treatment at which point RNA was isolated using a GenElute RNA extraction kit (Sigma-Aldrich, UK) with on column DNase digestion. RNA quality was assessed by a nanodrop spectrophotometer and Agilent 2100 Bioanalyzer with RIN values >8.8 reported for all samples. RNA sequencing libraries were prepared and 150bp paired end sequencing completed using an Illumina Novoseq (Novogene Co., Ltd, Cambridge, UK). The targeted read count was >20M total reads per sample. The resultant raw reads in FASTQ format were quality assessed and adapter trimmed using the TrimGalore wrapper for FastQC and Cutadapt reads with phred scores >30 retained [183]. The retained quality trimmed reads were aligned to the HG38 (GRCh38.83) build of the human genome using the STAR aligner and gene expression quantified using FeatureCounts [184,279].

4.4.9 Differential Gene Expression Analysis

To identify differential gene expression (DGE) associated with *PPL* KD in BC cell lines, statistical analysis was performed using the Bioconductor package DESeq2 written for the R statistical programming software [280]. The DESeq2 presumes that the RNA-seq counts are allocated according to negative binomial distributions. Therefore, DESeq2 uses comprehensive linear modelling to assess individual null hypotheses of zero log2 fold changes between high and low groups (i.e. no differential expression) for each gene and it uses both an independent-filtering method and the Benjamin and Hochberg method to enhance the power and control of the false discovery rate (FDR) [281]. Thus, the Identified differentially expressed genes (DEGs) were filtered based on fold change $>\pm 2$ and FDR<0.05.

To confirm the genotype of the used cell lines, quality filtered RNA-seq FASTQ files were aligned to the reference human genome using BWA [282].The resultant SAM file was converted to BAM format, sorted, with duplicates removed and indexed using SAMtools [283], read groups assigned using Picard and variants identified using the GATK HaplotypeCaller [284] to create specific variant call files (VCF). The resultant VCF files were then compared to genotype profiles for MDA-MB-231 (1998455) and MDA-MB-436 (2164998) in the COSMIC database to confirm the presence of cell line specific unique variants [285]. Single nucleotide variants (SNVs) characteristic to each cell line as reported in the COSMIC data base were confirmed (MDA-MB-231 ten SNVs and MDA-MB-436 one SNV) (Appendix).

4.4.10 Pathway Analysis

To perform Pathway analysis, the online publicly-available web-based gene set analysis tool Webgestalt (http://www.webgestalt.org/option.php) was used to identify differentially regulated canonical pathways using the gene set enrichment analysis (GSEA). The pathway analysis was based on the previously identified DEGs associated with PPL KD using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database [188].

4.4.11 RNA isolation and quantitative real-time PCR (qRT-PCR) analysis

To validate *PPL* mRNA KD on BC cell lines, RNA extraction was performed using GenElute RNA extraction kits (Sigma Aldrich,UK) and cDNA synthesised using Qscript cDNA synthesis kits (Quanta bio, USA). For mRNA expression analysis, Taqman^{TM} assays (Applied Biosystems, Germany) for *PPL* (Hs01011417_m1) and *GAPDH* (Hs03929097_g1) were used. The qRT-PCR reactions were performed (LC480, Roche), at 95 ^{o}C for 10 mins for 1 cycle; 95 ^{o}C for 10 secs and 60 ^{o}C for 30 secs for 40 cycles. Cq values were then converted to relative fold change using the Pfaffl method (Pfaffl., 2001).

PPL expression values normalised to *GAPDH* and calibrated to *siSCR* to obtain relative fold change expression.

4.4.12 Statistical Analysis

IBM SPSS 24.0 (Chicago, IL, USA) software was used for statistical analysis. For dichotomisation of mRNA expression and protein expression levels of PPL, the X-tile bioinformatics version 3.6.1 (Yale University, USA) was utilised with DMFS (when this information was available) and substituted by BCSS for the METABRIC-TNBC cohort as an endpoint. Cox proportional hazard models were used for multivariate analyses adjusting for tumour stage, tumour grade, patient age, nodal stage, LVI and chemotherapy status as covariates to adjust for potential confounding influence of these variables on associations between PPL and the outcomes of interest. Correlation between *PPL* mRNA and protein expression in the tumour cells were analysed using Spearman's Rho correlation coefficient test. The Chi-square test was used for the categorical variables' correlation. Differences between two groups were assessed using t-test (normalised data) or Mann-Whitney test (non-normal distribution). Data is represented as the mean ± standard deviation (SD) of at least two independent experiments performed in triplicate using GRAPH PAD PRISM v.7 for data presentation. A p-value of <0.05 was deemed significant. To examine the performance of PPL expression to predict distant metastasis in TNBC, the specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy probabilities were calculated both at the mRNA and protein levels as previously described (Chapter 2) . Briefly, the probabilities were calculated based on performance of PPL expression predicting histopathologically proven TNBC tumours interpreted as positive for distant metastasis.

4.5 RESULTS

4.5.1 Clinical significance of PPL expression in TNBC at both transcriptomic and protein level

In the TNBC NGS cohort, upregulated *PPL* mRNA expression was significantly associated with shorter DMFS and BCSS (both $p<0.01$) (Figure 4-4). In a multivariate Cox regression analysis, *PPL* mRNA expression was also an independent poor prognostic factor for DMFS and BCSS regardless of patient age, grade, nodal stage, tumour size, LVI status, and chemotherapy (both; $p <0.01$) (Table 4-2).

In the METABRIC TNBC cohort, upregulated *PPL* mRNA expression was also significantly associated with shorter BCSS ($p= 0.04$) (Figure 4-4). In a multivariate Cox regression analysis, upregulated *PPL* mRNA expression was again an independent poor prognostic factor for BCSS regardless of patient age, grade, nodal stage, tumour size, LVI status, and chemotherapy ($p= 0.019$) (Table 4-2).

Furthermore, *PPL* copy number alteration and *PPL* mRNA showed a significant association ($p= 0.014$) (Table 4-3)*; PPL* copy number gain was noted in 45% of TNBC cases, whereas only 14% of cases showed copy number loss. Given this association of upregulated *PPL* expression and poor TNBC patient outcome at the transcriptomic level, we next evaluated if PPL protein expression could also prognosticate TNBC tumours. The morphological assessment of TNBC tissue samples showed cytoplasmic expression of PPL (H-score range 0-200). High PPL protein expression was observed in 176 out of 245 cases (59%, H score >0). There was a positive linear association between the mRNA and protein expression of PPL (Spearman's coefficient 0.66; $p=0.013$).

At the protein level, High PPL protein expression was significantly associated with poor DMFS and BCSS (both $p<0.01$) (Figure 4-5) . In a Multivariate Cox regression analysis, high PPL expression was a predictor of shorter DMFS and BCSS, independent of patient age, grade, nodal stage, tumour size, LVI status, and chemotherapy (both $p<0.01$) (Table 4-4).

Table 4-2:Multivariate Cox regression analysis for PPL expression associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic cohorts)

Covariates	TNBC NGS cohort (n=112)								TNBC METABRIC cohort (n=313)			
	Breast Cancer specific Survival				Distant Metastasis Free Survival				Breast Cancer specific Survival			
	P-value	Hazard Ratio	95% CI		P-value	Hazard Ratio	95% CI		P-value	Hazard Ratio	95% CI	
			Lower	Upper			Lower	Upper			Lower	Upper
Age	0.281	1.587	0.686	3.672	0.243	1.614	0.722	3.607	0.423	0.821	0.508	1.329
Tumour Size	0.585	0.801	0.362	1.773	0.480	0.757	0.349	1.639	0.632	0.871	0.494	1.535
Grade	0.389	1.883	0.446	7.949	0.277	2.227	0.526	9.440	0.370	1.416	0.662	3.029
Vascular Invasion	0.144	1.980	0.792	4.946	0.152	1.876	0.792	4.442	0.016	1.895	1.125	3.191
Nodal Stage	0.408	1.300	0.698	2.421	0.522	1.216	0.668	2.215	0.025	1.674	1.066	2.629
Chemotherapy	0.745	0.872	0.383	1.989	0.475	0.749	0.339	1.656	0.537	1.211	0.660	2.220
PPL	**0.002**	3.605	1.610	8.075	**0.004**	3.293	1.466	7.349	0.019	1.821	1.104	3.002

- Significant P-values are bolded CI: Confidence interval

Table 4-3:PPL copy number alteration (CNA) in TNBC-METABRIC cohort (n=313)

CNA region for *PPL*	**CNA**	**Cases with reduced *PPL* mRNA expression (%)**	**Cases with high *PPL* mRNA Expression (%)**	**Total (%)**	**p-value**
Chromosome 16p13.3	Loss	6 (86)	1 (14)	7 (100)	**0.014**
	Gain	12 (55)	10 (45)	22 (100)	
	Neutral	229 (81)	55 (19)	284 (100)	
	Total	247 (79)	66 (21	313 (100)	

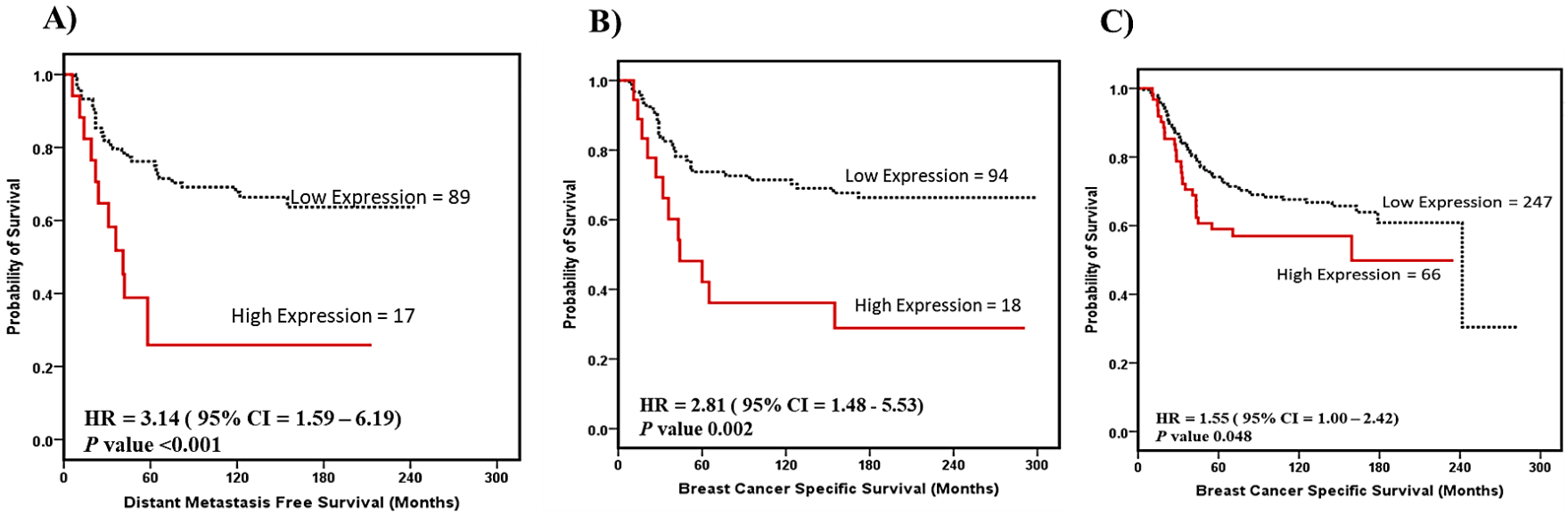


Figure 4-4:Univariate Kaplan-Meier survival analysis of PPL expression predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (Transcriptomic cohorts)

- A) High PPL mRNA expression confers a poor prognosis association with distant metastasis free survival in TNBC-NGS patients (n=106, p value <0.001). B) high PPL mRNA expression confers a poor prognosis association with Breast cancer specific survival in TNBC-NGS patients (n=112, p value 0.002). C) high PPL mRNA expression confers a poor prognosis association with Breast cancer specific survival in TNBC-METABRIC patients (n=313, p value 0.048).

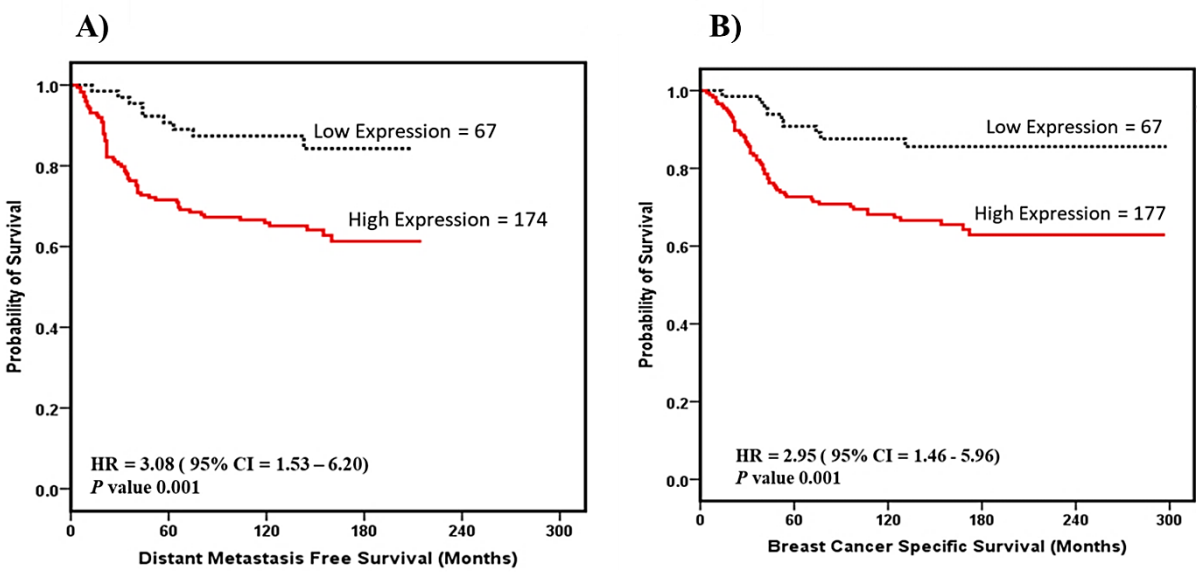


Figure 4-5:Univariate Kaplan-Meier survival analysis of PPL expression predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC cohort).

- A) High PPL protein expression confers a poor prognosis association with distant metastasis free survival in TNBC patients (n=241, p value 0.001).B) High PPL mRNA expression confers a poor prognosis association with Breast cancer specific survival in TNBC patients (n=244, p value 0.002).

Table 4-4:Multivariate Cox regression analysis for PPL expression associated with Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) (IHC cohort, n=245)

Covariates	Breast Cancer specific Survival				Distant Metastasis Free Survival			
	P-value	Hazard Ratio	95% CI		P-value	Hazard Ratio	95% CI	
			Lower	Upper			Lower	Upper
Age	0.009	2.220	1.218	4.044	0.020	1.986	1.116	3.536
Tumour Size	0.039	0.567	0.330	0.972	0.670	0.889	0.519	1.524
Grade	0.149	1.875	0.798	4.404	0.070	2.433	0.930	6.366
Vascular Invasion	0.016	2.090	1.146	3.813	0.042	1.836	1.023	3.295
Nodal Stage	0.003	1.779	1.210	2.617	0.021	1.547	1.072	2.309
Chemotherapy	0.284	0.731	0.411	1.298	0.093	0.620	0.356	1.082
PPL	**0.006**	2.910	1.360	6.226	**0.005**	2.800	1.366	5.738

- Significant P-values are bolded CI: Confidence interval

4.5.2 PPL expression prognostic value in TNBC chemotherapy treated patients

In the TNBC NGS cohort, our analysis showed that patients with upregulated mRNA expression of *PPL* were significantly associated with shorter BCSS and DMFS (both $p<0.01$) (Figure 4-6). Similarly, the METABRIC TNBC cohort showed comparable association with shorter BCSS ($p=0.01$) (Figure 4-6). Likewise, our analysis showed that patients with upregulated PPL protein expression had both significantly shorter BCSS and DMFS (both $p<0.01$) (Figure 4-7).

PPL upregulation in TNBC was significantly associated with poor clinical outcome, leading us to hypothesise that PPL expression may have a role in chemotherapy resistance. Therefore, the ability of PPL expression to predict response to chemotherapy was investigated by analysing its association with clinical outcome in a subgroup of TNBC patients who received chemotherapy alone.

Investigation of the level *PPL* expression in patients within the NGS cohort who were not offered chemotherapy (n=44) found a statistically significant difference in the 5-year DMFS. In patients with low expression of *PPL,* DMFS was 72% compared to 24% in patients with high *PPL* expression ($p=0.003$). At the PPL protein level, our results found that the 5-year DMFS of patients with low expression of PPL was 94% compared to 63% of those with high expression and the difference was statistically significant ($p=0.002$, n=123).

To further investigate our findings, we utilised the ROC plotter data portal to predict the prognostic value of PPL expression in TNBC patients'response to chemotherapy. Our investigation reflected an observational trend that PPL expression was higher in those patients who did not respond to chemotherapy ($p=0.058$). While our results did not reach statistical significance which might be due to the small number of analysed cases, it is suggesting that PPL could- be used as predictive marker for TNBC (Figure 4-8).

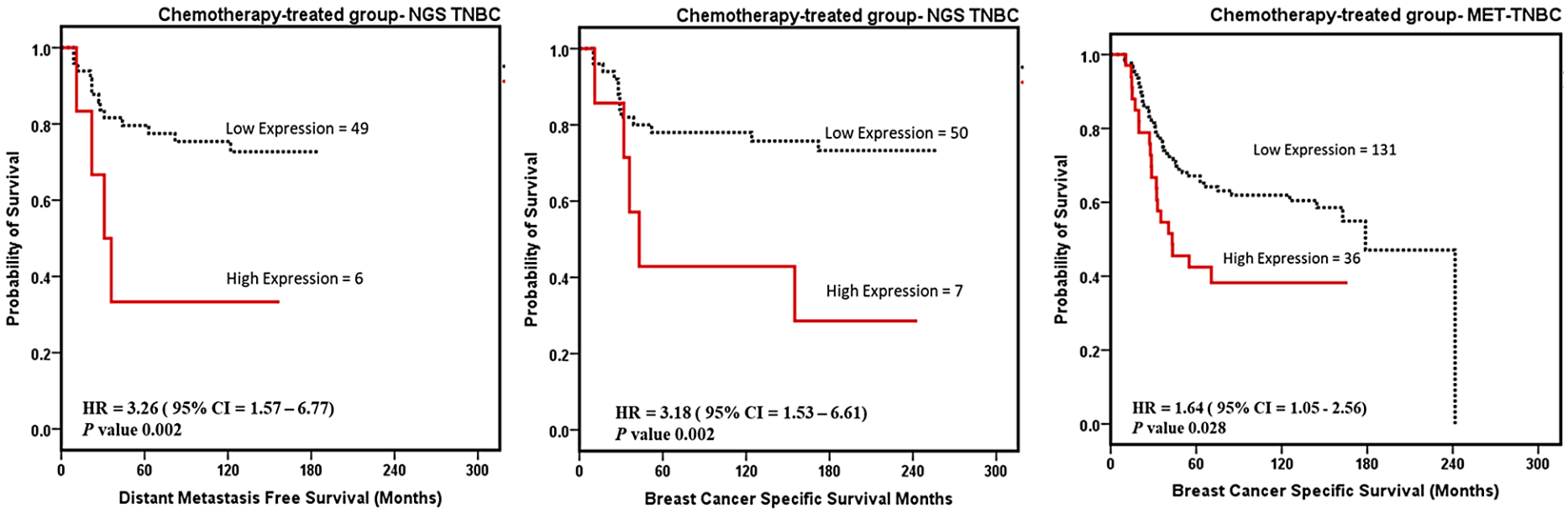


Figure 4-6:Univariate Kaplan-Meier survival analysis of PPL expression predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) in TNBC patients who received chemotherapy treatment only (Transcriptomic cohorts)

- A) High PPL mRNA expression confers a poor prognosis association with distant metastasis-free survival in TNBC-NGS patients who received chemotherapy treatment only (n=55, p value 0.002). B) high PPL mRNA expression confers a poor prognosis association with breast cancer specific survival in TNBC-NGS patients who received chemotherapy treatment only (n=57, p value 0.002). C) high PPL mRNA expression confers a poor prognosis association with breast cancer specific survival in TNBC-METABRIC patients who received chemotherapy treatment only (n=167, p value 0.028).

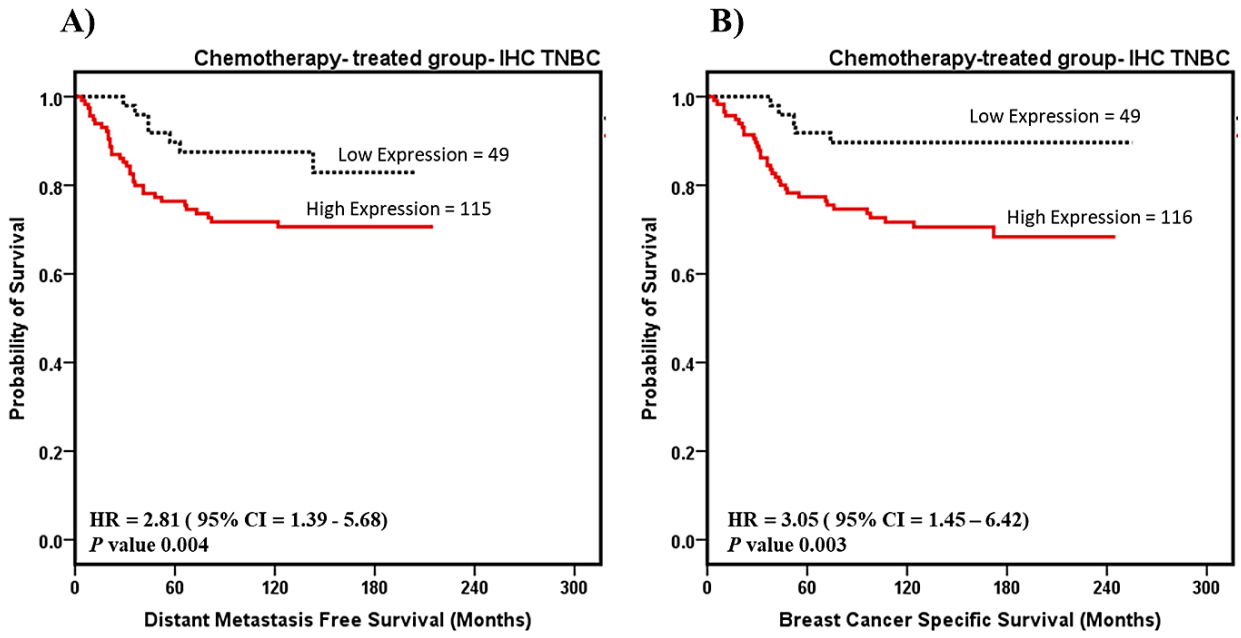


Figure 4-7:Univariate Kaplan-Meier survival analysis of PPL expression predicting Breast Cancer Specific Survival (BCSS) and Distant Metastasis-Free Survival (DMFS) in TNBC patients who received chemotherapy treatment only (IHC cohort)

- High PPL protein expression confers a poor prognosis association with distant metastasis free survival in TNBC patients who received chemotherapy treatment only (n=164, p value 0.004). B) High PPL protein expression confers a poor prognosis association with Breast cancer specific survival in TNBC patients who received chemotherapy treatment only (n=165, p value 0.003).

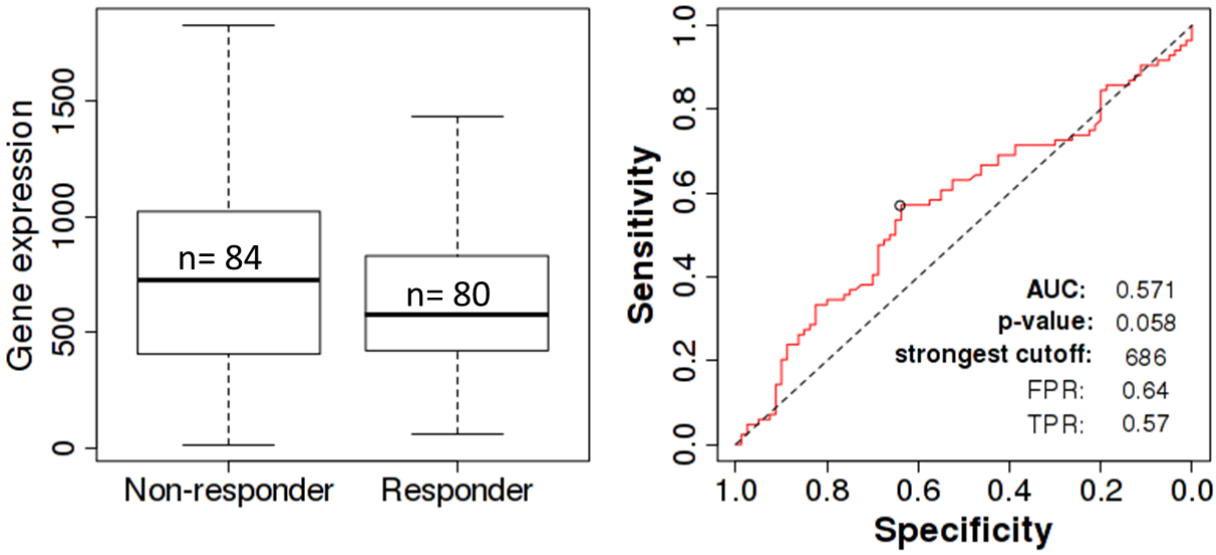


Figure 4-8:Analysis of PPL expression association with chemotherapy response in TNBC restrictive cohort using public domain data

- Using transcriptomic data of 164 TNBC patients with available treatment and response data, potential associations between PPL and patient responses to chemotherapy were evaluated. The PPL mRNA expression levels in responder and non-responder patients, measured as Recurrence free survival after five years of chemotherapy completion, are shown for several chemotherapy regimens.

4.5.3 External Validation of PPL expression in TNBC cohorts

Utilising the Bc-GenExMiner tool to analyse publicly available RNA-sequencing data, we noticed that higher expression of *PPL* was significantly associated with worse prognosis in the basal like subtype of BC (n=761, p< 0.0062). Furthermore, the restricted TNBC cohort (n=254) showed a comparable association with poor prognosis (p= 0.036). Additionally, examination of the median expression of *PPL* in the public domain Genotype 2 outcome TNBC subtype cohort (n=612) found that high expression of *PPL* was similarly associated with poor outcome, however statistical significance was not achieved (p=0.13). Nonetheless, when investigating the prognostic value of *PPL* expression via the KM plotter domain utilising a restrictive TNBC cohort (n=255), high *PPL* expression was predictive of shorter recurrence free survival (*p*= 0.050). A similar trend was observed with DMFS but was not statistically significant (p=0.06) (Figure 4-9).

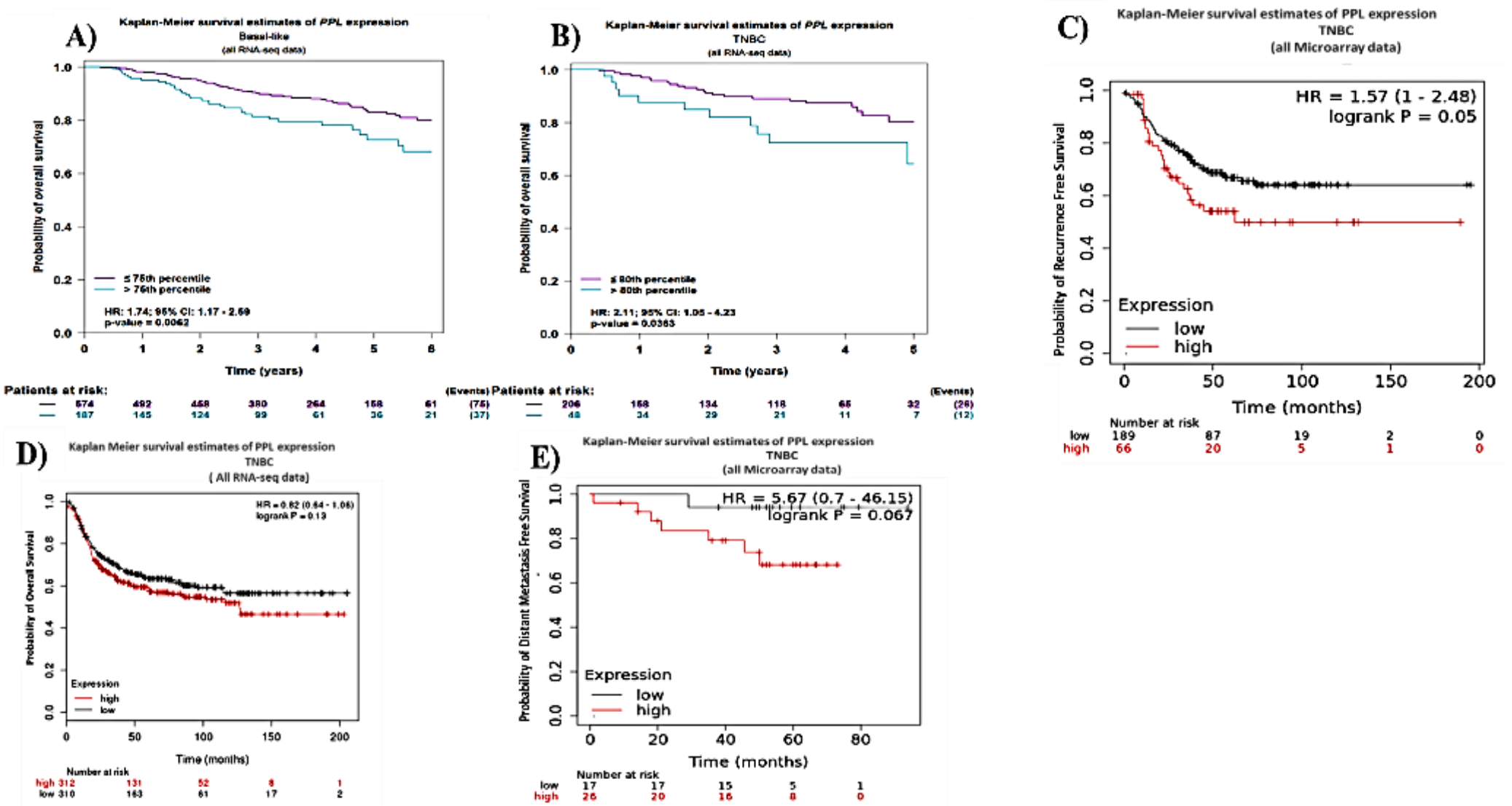


Figure 4-9:Univariate Kaplan-Meier survival analysis of PPL expression predicting patients outcome using public domain data

- To validate our findings in TNBC patients, we utilised the Breast Cancer Gene-Expression Miner v4.0 (bc-GenExMiner v4.0) datasets which includes 761 Basal like breast cancer patients, including 254 TNBC patients, the Genotype 2 outcome public portal, A genome-wide approach to link genotype to clinical outcome by utilising next generation sequencing and gene chip data of 6,697 breast cancer patients including 612 TNBC patients. And the KM-plotter datasets which includes 3951 breast cancer patients including 255 TNBC patients with available outcome data. A) In the Breast Cancer Gene-Expression Miner data portal, high PPL mRNA expression confers a poor prognosis association with overall survival in the Basal like molecular subtype of Breast cancer patients (n=761, p value 0.0062).B) In the Breast Cancer Gene-Expression Miner data portal, high PPL mRNA expression confers a poor prognosis association with overall survival in TNBC patients (n=254, p value 0.036). C) In the KM plotter public portal, high PPL mRNA expression confers a poor prognosis association with recurrence free survival in the TNBC patients (n=255, p value 0.05). D) In the Genotype 2 outcome public portal, high PPL mRNA expression confers a poor prognosis association with overall survival in the TNBC patients (n=612, p value 0.13). E) In the KM plotter public portal, high PPL mRNA expression confers a poor prognosis association with distant metastasis free survival in the TNBC patients (n=43, p value 0.067).

4.5.4 Testing the performance of PPL expression to predict distant metastasis in TNBC cohorts at the transcriptomic and protein level

In the TNBC-NGS cohort, upregulated *PPL* mRNA expression showed 72% sensitivity, 64% specificity, 28% positive predictive value (PPV), 92% negative predictive value (NPV), and 65% accuracy in dichotomising distant metastasis outcome of TNBC patients. Similarly, upregulated PPL protein expression captured 36% sensitivity, 87% specificity, 88% PPV, 34% NPV, and 50% accuracy in dichotomising distant metastasis outcome of TNBC patients in the TNBC-IHC cohort. (Table 4-5).

Table 4-5:Distribution of cases within NGS transcriptomic and IHC cohorts to test the prognostic index of PPL expression with reference to clinical evidence of distant metastasis.

	Distant Metastasis			
Prognostic index of PPL expression at mRNA Level within the TNBC NGS cohort	**Expression level**	**Yes**	**No**	**Total**
	Above cut-point Expression	**13 TP**	**33 FP**	**46**
	Below-cut-point Expression	**5 FN**	**59 TN**	**64**
Total		**18**	**92**	**110**
	Distant Metastasis			
Prognostic index of PPL expression at protein Level within the TNBC NGS cohort	**Expression level**	**Yes**	**No**	**Total**
	Above cut-point Expression	**64 TP**	**9 FP**	**73**
	Below-cut-point Expression	**113 FN**	**58 TN**	**171**
Total		**177**	**67**	**244**

- TP = true positive, FP = false positive, TN = true negative, and FN = false negative. Sensitivity = TP * 100/(TP+FN). Specificity = TN * 100/(TN +FP). PPV, the proportion of subjects with distant metastasis who are correctly diagnosed; PPV = TP* 100/(TP + FP). NPV, proportion of subjects without distant metastasis who are correctly diagnosed; NPV= TN * 100 / (TN + FN). Accuracy = TP + TN * 100/Total number of cases.

4.5.5 PPL gene silencing reduces proliferation, cell migration and invasion

qRT-PCR and western blot were used to confirm siRNA knockdown of *PPL* in TNBC cells relative to *siSCR* control treated cells. Using the MTS proliferation assay, PPL knockdown reduced proliferation of both MDA-MB-231 and MDA-MB-436 TNBC cells at 48 and 72hrs ($p < 0.01$) (Figure 4-10) siRNA-mediated knockdown of *PPL* also significantly reduced the migration rate of both TNBC cell lines (both $p<0.01$) but only significantly reduced the invasion rate of MDA-MB-231 ($p=0.03$) (Figure 4-10). These results indicate that *PPL* may significantly contribute to proliferation, migration and in certain context invasion in TNBC cells.

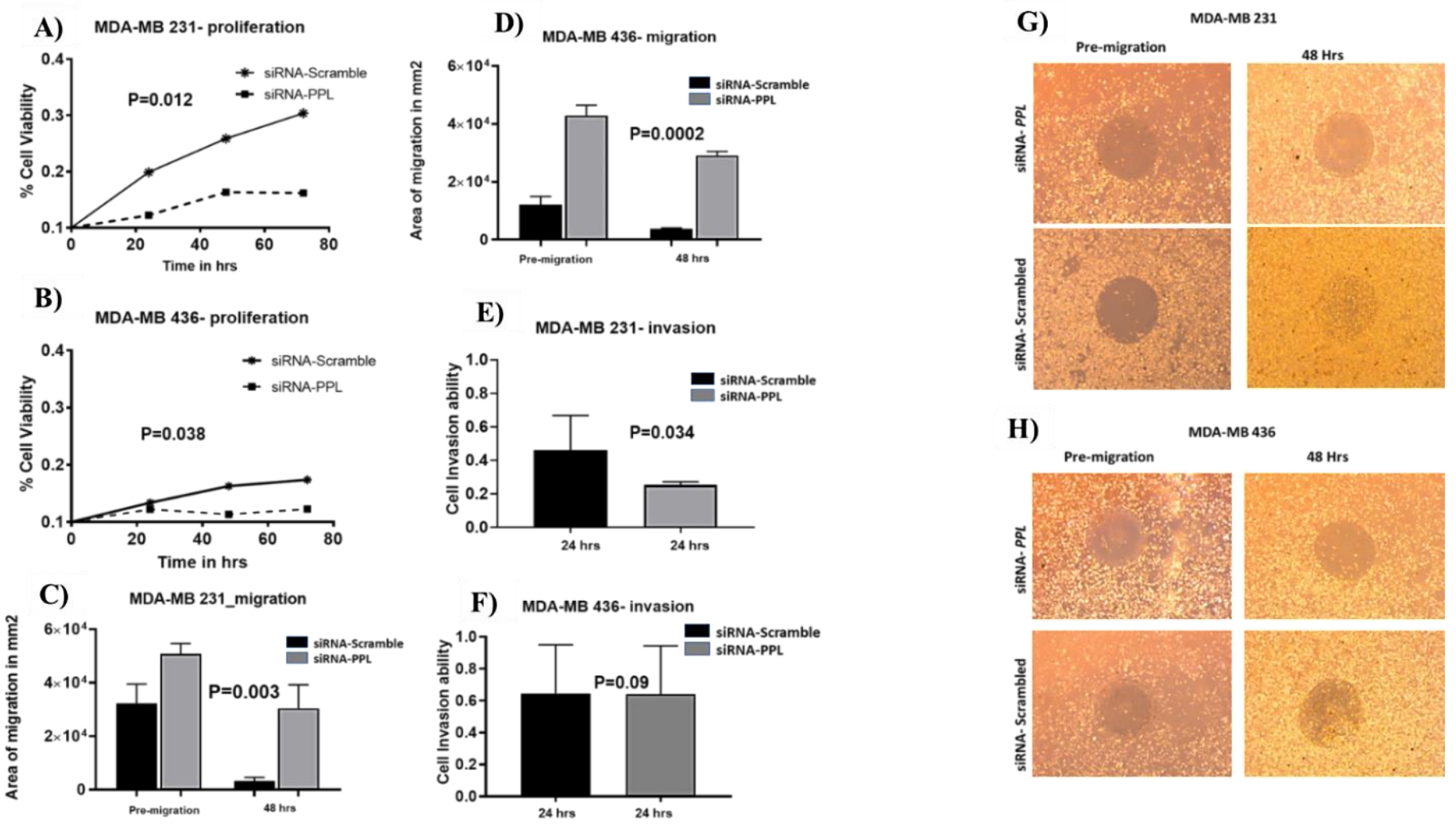


Figure 4-10:Mechanistic evaluation of PPL expression

- Cell proliferation was significantly reduced after PPL depletion in MDA-MB-231(p value 0.012) as detected by MTT assay. B) Cell proliferation was significantly reduced after PPL depletion in MDA-MB-436 (p value 0.038) as detected by MTT assay. C) PPL depletion showed the migration rate was significantly reduced in the MDA-MB-231 transfected cells compared to the control groups (p value 0.003) ass detected by migration assay. D) PPL depletion showed the migration rate was significantly reduced in the MDA-MB-436 transfected cells compared to the control groups (p value 0.0002) ass detected by migration assay. E) PPL depletion showed the number of invaded cells was significantly reduced in the MDA-MB-231 transfected cells compared to the control groups (p value 0.034) ass detected by Trans-well assay. F) PPL depletion showed the number of invaded cells was slightly reduced in the MDA-MB-436 transfected cells compared to the control groups (p value 0.09) ass detected by Trans-well assay. (G & H) Representative images of the migration assay in the MDA-MB 231 and MDA-MB-436 TNBC cells, the number of migrated cells was significantly increased in the control cells after 48 hours compared with the PPL depleted group.

4.5.6 Clinical significance of PPL expression in unselective BC cohorts at transcriptomic level

In contrast to the findings associated with *PPL* expression in TNBC cases, high *PPL* mRNA expression within the entire METABRIC cohort was significantly associated with prolonged BCSS (n=1975, p=0.014). Furthermore, when stratifying the METABRIC cohort based on hormonal subclasses, high *PPL* expression was predictive of prolonged BCSS (ER positive tumours, n=1871, p<0.001; PR positive tumours, n=1037, p<0.001; HER2 negative tumours, n=1774, p=0.002) (Figure 4-11). In a multivariate Cox regression analysis, high *PPL* expression was a predictor of prolonged BCSS independent of patient age, tumour size, HER2 expression, and ER expression (p=0.04) (

Table 4-6). Furthermore, the association between high *PPL* mRNA expression and excellent patient outcome in entire and unselected cohorts of BC cases was verified by the public Genotype 2 outcome domain (n=4047; p<0.001) and via the KM plotter public domain (recurrence free survival: n=3951; p<0.001, overall survival: n=1402; p=0.005). A similar trend was observed with DMFS, but this did not achieve statistical significance (n=1746; p=0.08). |Similarly, no statistically significant association between *PPL* expression and survival was found in the Breast Cancer Gene-Expression Miner domain (n=4307; p=0.08) (Figure 4-11).

Finally, the ROC plotter was used to explore the prognostic value of *PPL* expression and its association with adjuvant chemotherapy response. It was found that *PPL* expression can be used as a predictive biomarker for good response to endocrine therapy in ER positive tumours, in contrast to its role in TNBC tumours (Figure 4-12).

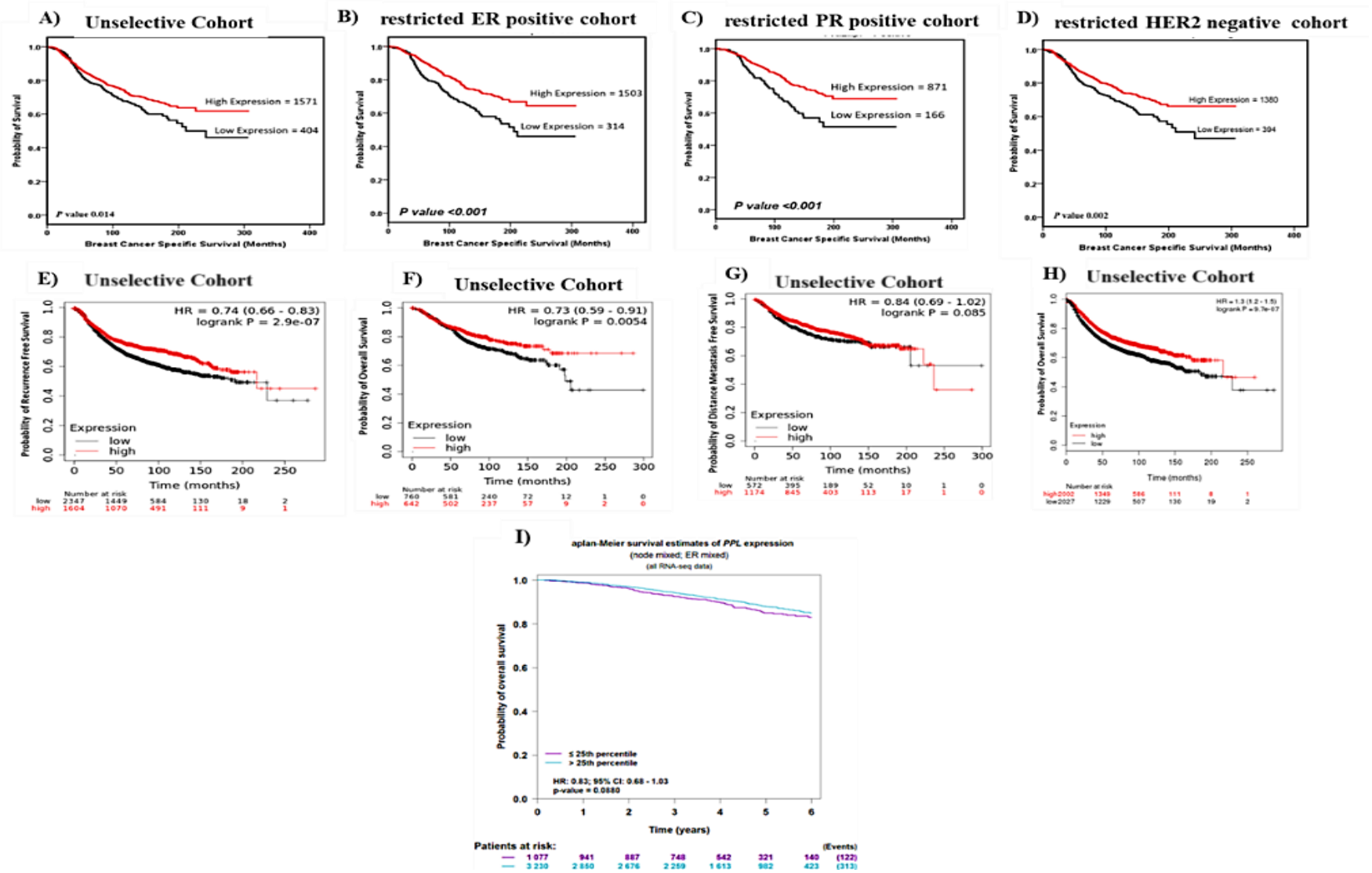


Figure 4-11:Univariate Kaplan-Meier survival analysis of PPL expression association with Breast cancer outcome using public domain data

- To explore our findings in unselected cohorts of breast cancer patients , we utilised the METABRIC dataset which includes 1980 breast cancer patients, the KM-plotter datasets which includes 3951 breast cancer patients, the Breast Cancer Gene-Expression Miner v4.0 (bc-GenExMiner v4.0) datasets which includes 5861 breast cancer patients & Genotype 2 outcome public portal, A genome-wide approach to link genotype to clinical outcome by utilising next generation sequencing and gene chip data of 6,697 breast cancer patients A) In the METABRIC cohort, high PPL mRNA expression confers a good prognosis in the whole (i.e. unselected cohorts) of Breast cancer patients (n=1975, p value0.014). B) In the METABRIC cohort, high PPL mRNA expression confers a good prognosis in the ER positive tumours of Breast cancer patients (n=1817, p value < 0.001). C) In the METABRIC cohort, high PPL mRNA expression confers a good prognosis in the PR positive tumours of Breast cancer patients (n=1037, p value < 0.001). D) In the METABRIC cohort, high PPL mRNA expression confers a good prognosis in the HER2 negative tumours of Breast cancer patients (n=1774, p value 0.002). E) In the KM plotter public portal, high PPL mRNA expression confers a good prognosis with recurrence free survival in the whole (i.e. unselected cohorts) of Breast cancer patients (n=3951, p value < 0.001) F) In the KM plotter public portal, high PPL mRNA expression confers a good prognosis with overall survival in the whole (i.e. unselected cohorts) of Breast cancer patients (n=1402, p value 0.005). G) In the KM plotter public portal, high PPL mRNA expression confers a good prognosis with distant metastasis free survival in the whole (i.e. unselected cohorts) of Breast cancer patients (n=1746, p value 0.08) H) In the Genotype 2 outcome public portal, high PPL mRNA expression confers a good prognosis in the whole (i.e. unselected cohorts) of Breast cancer patients (n=4047, p value <0.001). I) In the Breast Cancer Gene-Expression Miner data portal, high PPL mRNA expression confers a good prognosis with distant metastasis free survival in the whole (i.e. unselected cohorts) of Breast cancer patients (n=4307, p value 0.08)

Table 4-6:Multivariate Cox regression analysis for PPL expression associated with Breast Cancer Specific Survival (BCSS) (METABRIC cohort)

Covariates	Breast Cancer specific Survival			
	P-value	Hazard Ratio	95% CI	
			Lower	Upper
Age	0.399	0.911	0.733	1.132
Tumour Size	**>0.001**	1.985	1.594	2.472
ER status	**>0.001**	0.634	0.515	0.780
HER2 status	**>0.001**	1.734	1.369	2.196
PPL	**0.041**	0.807	0.657	0.992

- Significant P-values are bolded CI: Confidence interval

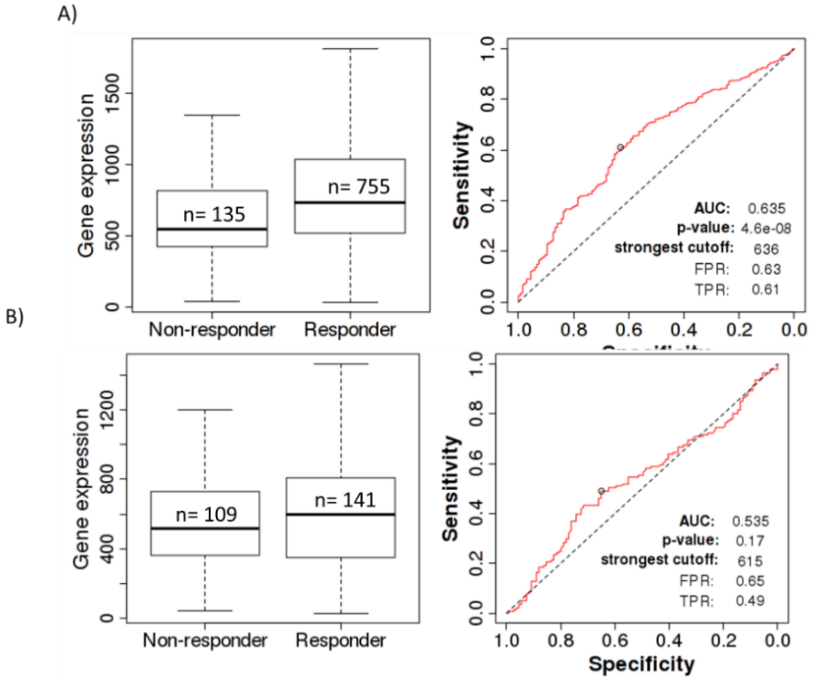


Figure 4-12:Investigation of PPL expression association with endocrine and Chemotherapy response using public domain data

- Using transcriptomic data of 885 ER positive patients with available treatment and response data, potential associations between PPL and patient responses to endocrine therapy were evaluated. The PPL mRNA expression levels in responder and non-responder patients, measured as Recurrence free survival after five years of endocrine therapy completion, are shown for several endocrine therapy regimens. B) Using transcriptomic data of 250 ER positive patients with available treatment and response data, potential associations between PPL and patient responses to chemotherapy were evaluated. The PPL mRNA expression levels in responder and non-responder patients, measured as Recurrence free survival after five years of chemotherapy completion, are shown for several chemotherapy regimens.

4.5.7 Changes in TNBC transcriptome upon PPL knockdown

To understand the molecular mechanism by which PPL expression promotes TNBC proliferation, invasion, metastasis, and chemotherapy resistance in TNBC, we performed RNA-sequencing on MDA-MB-231 and MDA-MB-436 cell lines treated with *siSCR* and *siPPL*. The differential gene expression (DGE) analysis on the scrambled MDA-MB-231 cell line vs *PPL*-depleted MDA-MB-231 cells revealed 1055 differentially expressed genes (DEG) comprising 615 upregulated and 440 downregulated genes (Figure 4-13). Interestingly, the 440 downregulated genes with PPL knockdown showed significant enrichment with cell adhesion molecules pathway, TNF signalling pathway, ECM-receptor interaction, MAPK signalling pathway, IL-17 signalling pathway, PI3K-Akt signalling pathway, and the EGFR tyrosine kinase inhibitor resistance pathway. (Table 4-7). For MDA-MB-436, our analysis identified 965 DEG comprising 415 upregulated and 550 downregulated DEG. Similarly, the 550 downregulated DEG seen in the *PPL*-depleted cells showed significant enrichment with the cytokine-cytokine receptor interaction pathway, TNF signalling pathway, IL-17 signalling pathway, and chemokine signalling pathway (Table 4-8). We investigated the overlapping DEG in both TNBC cell lines, and our analysis identified 90 significantly upregulated and 76 significantly downregulated genes (Figure 4-13). Pathway analysis of the overlapped 76 downregulated genes identified significant enrichment with the transcriptional misregulation in cancer pathway, focal adhesion pathway, PI3K-Akt signalling pathway, RAP1 signalling pathway, EGFR tyrosine kinase inhibitor resistance pathway, RAS signalling pathway, and MAPK signalling pathway (Table 4-9). Gene ontology (GO) analysis of the aforementioned 76 downregulated genes revealed a variety of biological processes including regulation of endothelial cell chemotaxis (GO:2001026), whereas other genes were involved in other related biological processes (Table 4-10& Figure 4-14). Additionally, when performing GO

analysis for molecular functions, our analysis showed a series of important molecular functions involving vascular endothelial growth factor binding (GO:0038085) and integrin binding (GO:0005178) (Table 4-11& Figure 4-15).

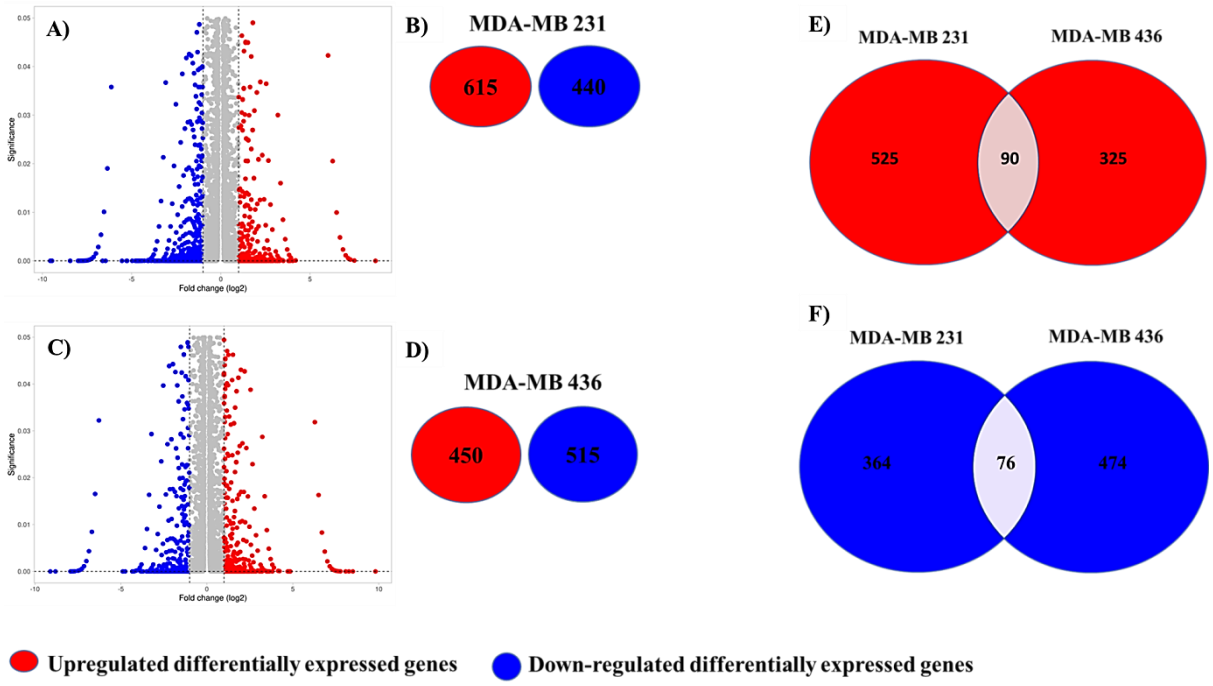


Figure 4-13:Differential gene expression analysis on scrambled TNBC cell lines vs siRNA depletion of PPL TNBC cells.

- Differential gene expression (DGE) analysis of scrambled TNBC cell lines vs siRNA depletion of *PPL* TNBC cells. A) Volcano plot obtained from DGE analysis comparing the scrambled MDA-MB-231 cell line vs siRNA depletion of *PPL* cells. B) Venn diagram of obtained from differentially expressed genes (DEG) comparing the scrambled MDA-MB-231 cell line vs siRNA depletion of *PPL* cells. C) Volcano plot obtained from DEG analysis comparing the scrambled MDA-MB-436 cell line vs siRNA depletion of PPL cells. D) Venn diagram of obtained from DEG comparing the scrambled MDA-MB-436 cell line vs siRNA depletion of *PPL* cells. E) Venn diagram of obtained DEG from comparing upregulated genes originated scrambled TNBC cell lines vs siRNA depletion of *PPL* cells. F) Venn diagram of obtained DEG from comparing downregulated genes originated scrambled TNBC cell lines vs siRNA depletion of *PPL* cells.

Table 4-7:The top regulated pathway as defined by both differential gene expression and pathway analysis on siRNAs depleted endogenous PPL MDA-MB-231 TNBC cell line using the over-representation analysis (ORA) for identification of gene ontologies.

Gene Set	Description	Size	Expect	Enrichment Ratio	P Value
hsa04640	Hematopoietic cell lineage	97	3.3247	5.4141	3.49E-09
hsa05323	Rheumatoid arthritis	90	3.0848	5.511	7.27E-09
hsa05150	Staphylococcus aureus infection	56	1.9194	6.252	2.96E-07
hsa04672	Intestinal immune network for IgA production	49	1.6795	6.5497	5.68E-07
hsa05332	Graft-versus-host disease	41	1.4053	7.116	8.24E-07
hsa04940	Type I diabetes mellitus	43	1.4738	6.7851	1.33E-06
hsa05321	Inflammatory bowel disease (IBD)	65	2.2279	5.3863	1.62E-06
hsa04612	Antigen processing and presentation	77	2.6392	4.9258	1.72E-06
hsa05330	Allograft rejection	38	1.3025	6.9101	3.87E-06
hsa05164	Influenza A	171	5.861	3.2418	5.46E-06
hsa05310	Asthma	31	1.0625	7.5292	6.79E-06
hsa04514	Cell adhesion molecules (CAMs)	144	4.9356	3.4444	7.64E-06
hsa04610	Complement and coagulation cascades	79	2.7077	4.4318	1.35E-05
hsa04668	TNF signalling pathway	110	3.7703	3.7133	2.1E-05
hsa05140	Leishmaniasis	74	2.5364	4.3369	3.81E-05
hsa04145	Phagosome	152	5.2098	3.0711	5.96E-05
hsa04512	ECM-receptor interaction	82	2.8106	3.9138	0.0001
hsa05410	Hypertrophic cardiomyopathy (HCM)	83	2.8448	3.8667	0.000112
hsa05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	72	2.4678	4.0522	0.000154
hsa05416	Viral myocarditis	59	2.0222	4.4505	0.000159
hsa05320	Autoimmune thyroid disease	53	1.8166	4.4039	0.000397
hsa05152	Tuberculosis	179	6.1352	2.6079	0.000404
hsa04010	MAPK signalling pathway	295	10.111	2.1758	0.000471
hsa05165	Human papillomavirus infection	339	11.619	2.0655	0.000557
hsa05168	Herpes simplex infection	185	6.3409	2.5233	0.000583
hsa05414	Dilated cardiomyopathy (DCM)	90	3.0848	3.2418	0.00096

hsa05166	**Human T-cell leukemia virus 1 infection**	**255**	**8.7401**	**2.1739**	**0.001157**
hsa04510	**Focal adhesion**	**199**	**6.8207**	**2.3458**	**0.001281**
hsa05145	**Toxoplasmosis**	**113**	**3.8731**	**2.8401**	**0.001631**
hsa05134	**Legionellosis**	**55**	**1.8851**	**3.7133**	**0.002545**
hsa05418	**Fluid shear stress and atherosclerosis**	**139**	**4.7642**	**2.5188**	**0.002838**
hsa00770	**Pantothenate and CoA biosynthesis**	**19**	**0.65123**	**6.1423**	**0.003479**
hsa05169	**Epstein-Barr virus infection**	**201**	**6.8893**	**2.1773**	**0.003709**
hsa04657	**IL-17 signalling pathway**	**93**	**3.1876**	**2.8235**	**0.004423**
hsa04151	**PI3K-Akt signalling pathway**	**354**	**12.133**	**1.8132**	**0.004834**
hsa01521	**EGFR tyrosine kinase inhibitor resistance**	**79**	**2.7077**	**2.9545**	**0.005438**

- The category size is calculated based on the number of overlapping genes between the annotated genes in the category and the reference gene list for the "ORA" method. The category expect is the number of categories expected from set cover" indicates the number of the expected reduced sets of the weighted set cover algorithm for redundancy reduction in the report. The category ratio is the enrichment ratio (for ORA) as it indicates the gene ontology sets with FDR < 0.05. The category p value is indicative of the weighted set cover and maximum coverage called size-constrained weighted set cover where weights are assigned to gene sets with smaller enrichment p values.

Table 4-8:The top regulated pathway as defined by both differential gene expression and pathway analysis on siRNAs depleted endogenous PPL on MDA-MB-436 TNBC cell line using the over-representation analysis (ORA) for identification of gene ontologies.

Gene Set	Description	Size	Expect	Enrichment Ratio	P Value
hsa04621	NOD-like receptor signalling pathway	168	3.284	3.6541	0.000102
hsa04060	Cytokine-cytokine receptor interaction	294	5.747	2.7841	0.000187
hsa05132	Salmonella infection	86	1.6811	4.7588	0.000258
hsa04668	TNF signalling pathway	110	2.1502	4.1856	0.000282
hsa05323	Rheumatoid arthritis	90	1.7593	4.5473	0.000353
hsa04657	IL-17 signalling pathway	93	1.8179	4.4006	0.000441
hsa05146	Amoebiasis	96	1.8766	4.2631	0.000547
hsa04640	Hematopoietic cell lineage	97	1.8961	4.2192	0.000586
hsa05142	Chagas disease (American trypanosomiasis)	102	1.9938	4.0124	0.000819
hsa04062	Chemokine signaling pathway	189	3.6945	2.9774	0.00114

- The category size is calculated based on the number of overlapping genes between the annotated genes in the category and the reference gene list for the "ORA" method. The category expect is the number of categories expected from set cover" indicates the number of the expected reduced sets of the weighted set cover algorithm for redundancy reduction in the report. The category ratio is the enrichment ratio (for ORA) as it indicates the gene ontology sets with FDR < 0.05. The category p value is indicative of the weighted set cover and maximum coverage called size-constrained weighted set cover where weights are assigned to gene sets with smaller enrichment p values.

Table 4-9:The top regulated pathway as defined by both differential gene expression and pathway analysis on the 76 overlapping down regulated genes resulted the comparison of scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies.

Gene Set	Description	Size	Expect	Enrichment Ratio	P Value
hsa05202	Transcriptional misregulation in cancer	186	0.51162	7.8183	0.00146
hsa04510	Focal adhesion	199	0.54738	7.3075	0.00187
hsa04270	Vascular smooth muscle contraction	121	0.33283	9.0136	0.00417
hsa04151	PI3K-Akt signalling pathway	354	0.97373	4.1079	0.0144
hsa05218	Melanoma	72	0.19805	10.099	0.01638
hsa04015	Rap1 signalling pathway	206	0.56663	5.2944	0.01791
hsa01521	EGFR tyrosine kinase inhibitor resistance	79	0.2173	9.2038	0.01951
hsa04014	Ras signalling pathway	232	0.63815	4.7011	0.02448
hsa00604	Glycosphingolipid biosynthesis	15	0.04126	24.237	0.04051
hsa04010	MAPK signalling pathway	295	0.81144	3.6971	0.04521

- The category size is calculated based on the number of overlapping genes between the annotated genes in the category and the reference gene list for the "ORA" method. The category expect is the number of categories expected from set cover" indicates the number of the expected reduced sets of the weighted set cover algorithm for redundancy reduction in the report. The category ratio is the enrichment ratio (for ORA) as it indicates the gene ontology sets with FDR < 0.05. The category p value is indicative of the weighted set cover and maximum coverage called size-constrained weighted set cover where weights are assigned to gene sets with smaller enrichment p values.

Table 4-10: The Biological process category by the Gene Ontology Consortium shows the 76 core common genes that are significantly downregulated among scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies.

Gene Set	Description	Size	Overlap	Expect	Enrichment Ratio	P Value	Gene symbol
GO:2001026	regulation of endothelial cell chemotaxis	21	3	0.06301	47.611	3.3E-05	MET, FGF1, KDR
GO:0035767	endothelial cell chemotaxis	29	3	0.08701	34.477	8.8E-05	MET, FGF1, KDR
GO:0051639	actin filament network formation	8	2	0.024	83.32	0.00024	PLS3, LCP1
GO:0003158	endothelium development	117	4	0.35106	11.394	0.00041	PTN, MET, FGF1, KDR
GO:0051130	positive regulation of cellular component organization	1203	10	3.6096	3.0474	0.00072	PTN,MET,KDR, LCP1,MMP3, TAPT1,ZNF804A, NEGR1,PPM1E,CDH4
GO:0031346	positive regulation of cell projection organization	365	6	1.0952	5.4786	0.00075	PTN,TAPT1,ZNF804A, NEGR1,CDH4, RELN
GO:2001028	positive regulation of endothelial cell chemotaxis	15	2	0.04501	44.437	0.0009	MET,KDR
GO:0051017	actin filament bundle assembly	148	4	0.44407	9.0076	0.001	PLS3,MET, LCP1,PPM1E
GO:0048012	hepatocyte growth factor receptor signaling pathway	16	2	0.04801	41.66	0.00103	MET,ESM1
GO:0061572	actin filament bundle organization	151	4	0.45307	8.8286	0.00108	PLS3,MET,LCP1, PPM1E

Table 4-11: The Molecular function category by the Gene Ontology Consortium shows the 76 core common genes that are significantly downregulated among scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies.

Gene Set	Description	Size	Overlap	Expect	Enrichment Ratio	P Value	Gene symbol
GO:0038085	vascular endothelial growth factor binding	5	2	0.013499	148.16	7.09E-05	PTN,KDR
GO:0005178	integrin binding	127	4	0.34287	11.666	0.000376	FGF1,KDR,LCP1, ECM1
GO:0019838	growth factor binding	138	4	0.37257	10.736	0.000515	PTN,KDR, IGFBPL1, ESM1,
GO:0050839	cell adhesion molecule binding	478	6	1.2905	4.6494	0.0017	FGF1,CALD1,KDR,LCP1,ESM1,CDH4
GO:0005520	insulin-like growth factor binding	28	2	0.075594	26.457	0.002576	IGFBPL1,ESM1
GO:0004864	protein phosphatase inhibitor activity	39	2	0.10529	18.995	0.004956	PTN,BOD1
GO:0003779	actin binding	419	5	1.1312	4.42	0.005239	PLS3, CALD1, CNN1, LCP1, TAGLN
GO:0019212	phosphatase inhibitor activity	42	2	0.11339	17.638	0.00573	PTN, BOD1
GO:0008201	heparin binding	160	3	0.43197	6.945	0.009171	PTN,FGF1, RSPO2
GO:0005326	neurotransmitter transporter activity	54	2	0.14579	13.719	0.00933	SLC36A1, CPLX1

- The category size is calculated based on the number of overlapping genes between the annotated genes in the category and the reference gene list for the "ORA" method. The category expect is the number of categories expected from set cover" indicates the number of the expected reduced sets of the weighted set cover algorithm for redundancy reduction in the report. The category ratio is the enrichment ratio (for ORA) as it indicates the gene ontology sets with FDR < 0.05. The category p value is indicative of the weighted set cover and maximum coverage called size-constrained weighted set cover where weights are assigned to gene sets with smaller enrichment p values.

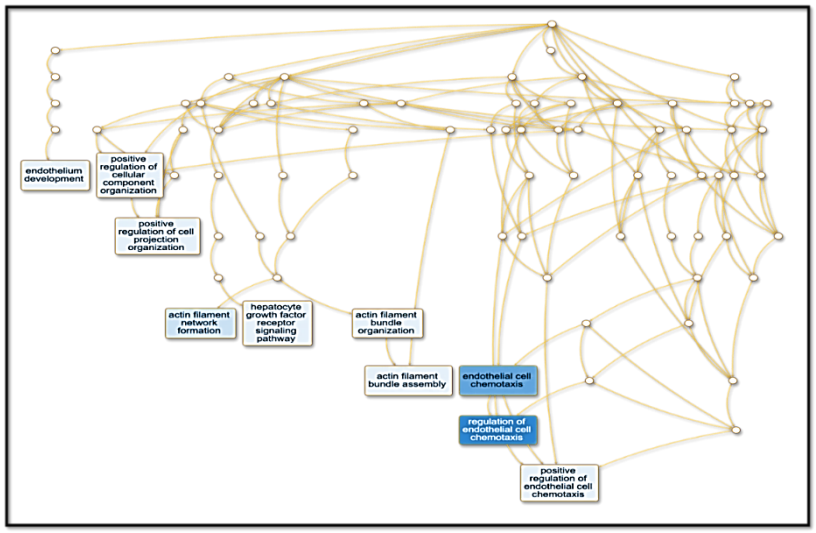


Figure 4-14:The Biological process category by the Gene Ontology Consortium shows the 76 core common genes that are significantly downregulated among scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies.

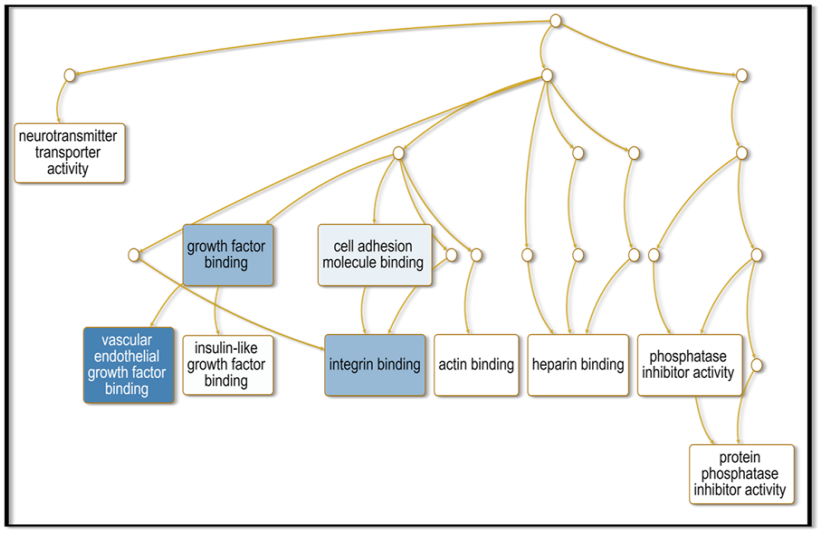


Figure 4-15:The Molecular function category by the Gene Ontology Consortium shows the 76 core common genes that are significantly downregulated among scrambled against siRNA depletion of PPL in TNBC cell lines using the over-representation analysis (ORA) for identification of gene ontologies.

4.6 Discussion

The high propensity of developing distant metastasis and limited treatment choices for TNBC remain a clinical challenge for oncologists as ~46% of TNBC patients go on to develop distant metastasis within the first three years of disease diagnosis [286,287]. TNBC patients do not benefit from hormonal or trastuzumab-based therapy due to the loss of target receptors [172] and, as previously reported, TNBC patients usually acquire chemotherapy resistance, thus efficacy of this regime continues to be poor [288]. Hence, identification of novel and prognostic biomarkers associated with TNBC outcome and response to chemotherapy are of urgent need. In this study, we investigated the significance of our previously identified *PPL* gene as one of the key genes associated with clinical outcome in TNBC [216]. We explored the predictive importance of PPL expression in TNBC to provide robust evidence that upregulated PPL expression is prognostic of patient outcome and could be beneficial in predicting the early onset of distant metastasis in TNBC patients.

DNA amplification is a major feature of oncogenic transformation. Amplification of chromosome 16p13.3, where *PPL* resides, is described as the third most amplified region in BC (57%) [289]. Previous comparative genomic hybridisation assays argued that copy number (CN) amplification within the 16p13.3 amplicons induces the upstream of PI3K pathway which has been linked to poor prognosis in BC [290]. In our study, we have documented that *PPL* CN amplification is strongly associated with upregulated *PPL* mRNA and protein expression. However, *PPL* CN gain only explains a portion of *PPL* upregulated cases, indicating that other downstream pathways may be modulating PPL expression.

In our study, we have provided evidence that upregulated *PPL* expression at both transcriptomic and protein levels is an independent and strong predictor of DMFS and BCSS in TNBC by univariate and multivariate analysis followed by validation in different

platforms. Our data also demonstrated the potential role of upregulated *PPL* expression and its significant association with chemotherapy resistance in TNBC at both transcriptomic and protein level with increased risk of distant metastasis and death among the TNBC chemotherapy-treated patients, but not in ER positive patients. Our results revealed that patients with upregulated *PPL* expression are less likely to benefit from chemotherapy. Additionally, the prognostic value of *PPL* expression in dichotomising distant metastasis probabilities in TNBC patients was calculated. *PPL* expression demonstrated 64% specificity and 65% accuracy in dichotomising distant metastasis outcome of TNBC patients at the mRNA level. Similarly, PPL expression yielded 87 % specificity and 50% accuracy in dichotomising distant metastasis outcome of TNBC patients at the protein level. PPL has displayed promising specific and accurate results as a potential prognostic biomarker both at the transcriptomic and protein level. Therefore, it may facilitate TNBC progression prediction and patient fall at higher risk may benefit from aggressive treatment plans.

In concordance with our findings, previous studies indicated that dysregulation of *PPL* expression causes resistance to erlotinib and other EGFR inhibitors in oral squamous cell carcinoma and is also associated with cisplatin resistance in endometrial cell lines [272,291]. *PPL* has also been reported *to be* linked to the PI3K cascade and found to enhance AKT activity [270]. However, the precise molecular mechanism of PPL expression on chemotherapy resistance in TNBC but not ER positive patients remains unclear and therefore requires further investigation.

We have provided evidence that endogenous PPL significantly promoted cell proliferation, migration and invasion probabilities in TNBC cell lines. In concordance with our findings, *PPL* knockdown has been associated with reduced cell migration and invasion in both pharyngeal squamous cancer and TNBC cells [251,268]. Theoretically, PPL may be linked

with cell proliferation or cell death due to its impact on cellular structure [5,292]. Therefore, distortion of this through *PPL* knockdown may result in interference with cell adhesion to the extra cellular matrix, thus promoting epithelial tumourigenesis [293].

Using a panel of well-characterised TNBC cases, we sought to identify the molecular mechanisms which underlie PPL expression in TNBC that differentiate its prognostic abilities from ER positive tumours. Therefore, DGE analysis and the KEGG pathway were applied to explore the identified DEG. The KEGG pathway analysis showed that overlapped DEG obtained from TNBC cell lines were significantly enriched in the transcriptional misregulation in cancer pathway, focal adhesion pathway, PI3K-Akt signalling pathway, RAP1 signalling pathway, EGFR tyrosine kinase inhibitor resistance pathway, RAS signalling pathway, and MAPK signalling pathway, which are known to modulate treatment responses and aggressive behaviour of TNBC tumours [294–297]. Also, GO annotation analysis revealed our overlapped DEG in TNBC cells with siRNA depletion of *PPL* were significantly enriched in the following GO categories: (GO:0038085, GO:0005178, GO:0019838, GO:0050839, and GO:0005520), which are strongly associated with the biological and aggressive behaviour of TNBC [298,299]. The obtained results, collectively, suggested that PPL expression may play a pivotal role in TNBC tumourigenesis and chemoresistance.

4.7 **Limitations of the study**

We have made a strong effort to perform a comprehensive investigation on PPL expression in TNBC, but several limitations need to be recognised. Firstly, we did not identify any association between PPL expression and well-established tumour prognostic factors such as age, tumour size, tumour grade, and lymph node status, which we guess is due to the relatively small sample size and chosen cut-off points. However, the aforementioned

prognostic factors have been shown to be of limited predictive value due to the typically high-grade nature of TNBC tumours [107]. Therefore, PPL has a strong predictive value for chemoresistance and poor patient outcome regardless of the aforementioned prognostic factors as our data has identified. Thus, evaluation of PPL expression in a larger TNBC cohort is warranted. While the molecular heterogeneity TNBC, reflected in our cell lines, resemble mesenchymal type and harbour stem/progenitor-cell characteristics, were equally effective candidates to resemble early onset of distant metastasis in TNBC patients. While the effects seen phenotypically with loss of proliferation, migration and invasion when PPL knockdown was achieved. However, strong evidence of genomic differences among TNBC cell lines is acknowledged [110,300]. For instance, MDA-MB-231 is recognised to harbour mutations in *BRAF, CDK2A, KRAS, NF2, TP53, PDGFR* but no mutation in the *BRCA1* gene whereas MDA-MB-436 is documented to carry homozygous deleterious mutations in *BRCA1, TP53,* Cell-cycle and DNA damage response genes [110,300,301]. However, our DGE findings represent a fraction of our results affected by *PPL* knockdown. The identification of novel pathways never before recognised to be associated with *PPL* expression as well as known pathways pertaining to DEG lists not previously associated with *PPL* expression present novel targets for PPL-based combination therapies. Indeed, preliminary data and initial experiments targeting some of the identified DEG and pathways in TNBC may have failed due to not investigating the development of PPL inhibitors.

4.8 Conclusion

Our data highlights the potential pro-oncogenic role of PPL in TNBC, and that it is a strong predictor of distant metastasis and poor patient outcome. The association of PPL with well-established tumour characteristics at the *in vitro* level has provided evidence as to its oncogenic role in TNBC. Our study concluded that upregulated PPL in TNBC may activate several oncogenic pathways such as the TNF signalling pathway, MAPK signalling pathway, PI3K-Akt signalling pathway, and the EGFR tyrosine kinase inhibitor resistance pathway, which can induce chemoresistance in TNBC. Our findings suggest that *PPL* expression both at transcriptomic and protein levels could be used as a predictor of outcome and tumour progression in TNBC patients. Further research *in vivo* is needed for validation of PPL expression as a potential therapeutic target in TNBC.

Chapter 5 : The Molecular Landscape of E-Cadherin in Ductal Carcinoma of The Breast

This chapter has been published in Modern Pathology (2019) under the title " The Molecular Mechanisms Underlying Reduced E-cadherin Expression in Invasive Ductal Carcinoma of the Breast: High Throughput Analysis of Large Cohorts" with the corresponding DOI: https://doi.org/10.1038/s41379-019-0209-9 [302]. The Copyright approval for re-use of these material is provided (appendix 7.4)

5.1 Background

Worldwide, BC accounts for 23% of total diagnosed cancer cases, and is the second leading cause of cancer-related death amongst women [2]. In solid tumours, cell-cell dyscohesion is a recognised phenomenon allowing tumuor cells to grow invasively into surrounding tissues [303]. E-cadherin, a calcium-dependent adhesion molecule encoded by the *CDH1* gene located on chromosome 16q22.1 [304], plays an important role in gland formation, cell differentiation, polarity and maintaining the integrity of epithelial cells [305]. Subsequently, decreased expression of E-cadherin, which is frequently seen in breast cancer, may lead to cellular de-differentiation and invasiveness [306,307].

5.2 Prognostic Significance of E-cadherin expression in BC

Reduced/loss of E-cadherin expression is observed in the vast majority of ILC and LCIS tumours together with loss of *CDH1* gene copy number [308–310] or *CDH1* gene mutation[311] in a large proportion of cases suggests a plausible role for E-cadherin as a tumour suppressor gene [310,312]. However, there is limited evidence to support a role for E-cadherin as a tumour suppressor gene in IDC tumours [207]. In fact, DCIS and low grade IDC tumours generally show stronger E-cadherin membrane staining than that seen in the normal breast epithelial cells, denoting increased expression, rather than a loss of expression [313]. Although some studies indicated that a proportion of IDC shows loss/reduced E-cadherin protein expression, these tumours were typically high grade aggressive tumours. Of note, accumulating evidence suggests that high grade IDC are characterised by genomic instability with loss of increasing number of tumour suppressor

genes during the carcinogenesis process that contributes to their aggressive behaviour [207]. In addition, reduced/loss of E-cadherin expression is frequently associated with loss of ER expression, larger tumour size, and with the development of metastasis and recurrence [314–317]. Furthermore, in a recent study by our group, reduced/loss of E-cadherin expression was an indicator of BC poor prognosis and advanced tumour characteristics including; larger tumour size, higher histological grade, early development of distant metastasis, and strong association with ER negativity [318]. Likewise, multiple reports indicated the strong association between reduced/loss of E cadherin expression and worse patient outcome in TNBC patients [319–321].

These findings suggest that E-cadherin loss occurs as a late event in the process of carcinogenesis arising in association with or as a part of genomic instability, rather than as an early neoplastic event as seen in ILC tumours [313,322,323]. However, the reasons for dysregulation of E-cadherin protein expression remain ill-defined [318].

5.3 **Hypothesis and Aims**

In this study, it was hypothesised that reduced/loss of E-cadherin expression in IDC and its molecular implications are different from ILC not only at the morphological levels but also at the molecular levels. Therefore, this study aimed to examine the mechanisms of reduced/loss E-cadherin expression in high grade IDC carcinoma compared to ILC carcinoma and its potential molecular implications. Nonetheless, an improved insight can be obtained from decoding the transcriptome of TNBC molecular subtype by NGS to unravel the reasoning of heterogeneous E-cadherin expression in cases that did not show evidence of EMT.

5.4 Materials and Methods

5.4.1 Study cohorts

This study was conducted on multiple well-characterised cohorts of high grade IDC using different molecular techniques (Figure 5-5-1).First, a well-characterised cohort of primary grade 3 IDC from patients presenting to Nottingham City Hospital between 1989 and 1998 (n=813), and for whom detailed immunohistochemistry evaluation of E-cadherin expression was carried out by a previous member of our group, Dr Abd El-Rehim and colleagues [318]. To complete the TNBC IDC cohort, missing cases (n=43) were stained and evaluated for E-cadherin protein expression by the author.

The mean patient age was 52 years (range 18-71), and tumour size ranged in diameter from 0.1 cm to 5 cm at time of presentation, with a mean tumour size of 2 cm (Table 5-1). To understand the molecular biology of E-cadherin expression, high grade IDC carcinoma (n=883) cases in the Molecular Taxonomy of the METABRIC cohort[114] were used to investigate copy number alterations and *CDH1* mRNA expression. The mean patient age was 59 (range 26-96) and mean tumour size at time of presentation was 3 cm (range from 1 cm to 18 cm). In this cohort, patients with ER positive tumour and /or lymph node negative at time of diagnosis did not receive adjuvant chemotherapy, whereas those with ER negative tumours and lymph node positive status received adjuvant treatment. Nonetheless, an additional TNBC NGS cohort (n=106) was used to investigate E-cadherin reduced/loss expression in this subtype of BC. The mean patient age was 48 (range 27-69) and tumour size ranged in diameter from 1 to 6 cm at time of presentation, with a mean tumour size of 2 cm (Table 5-1).

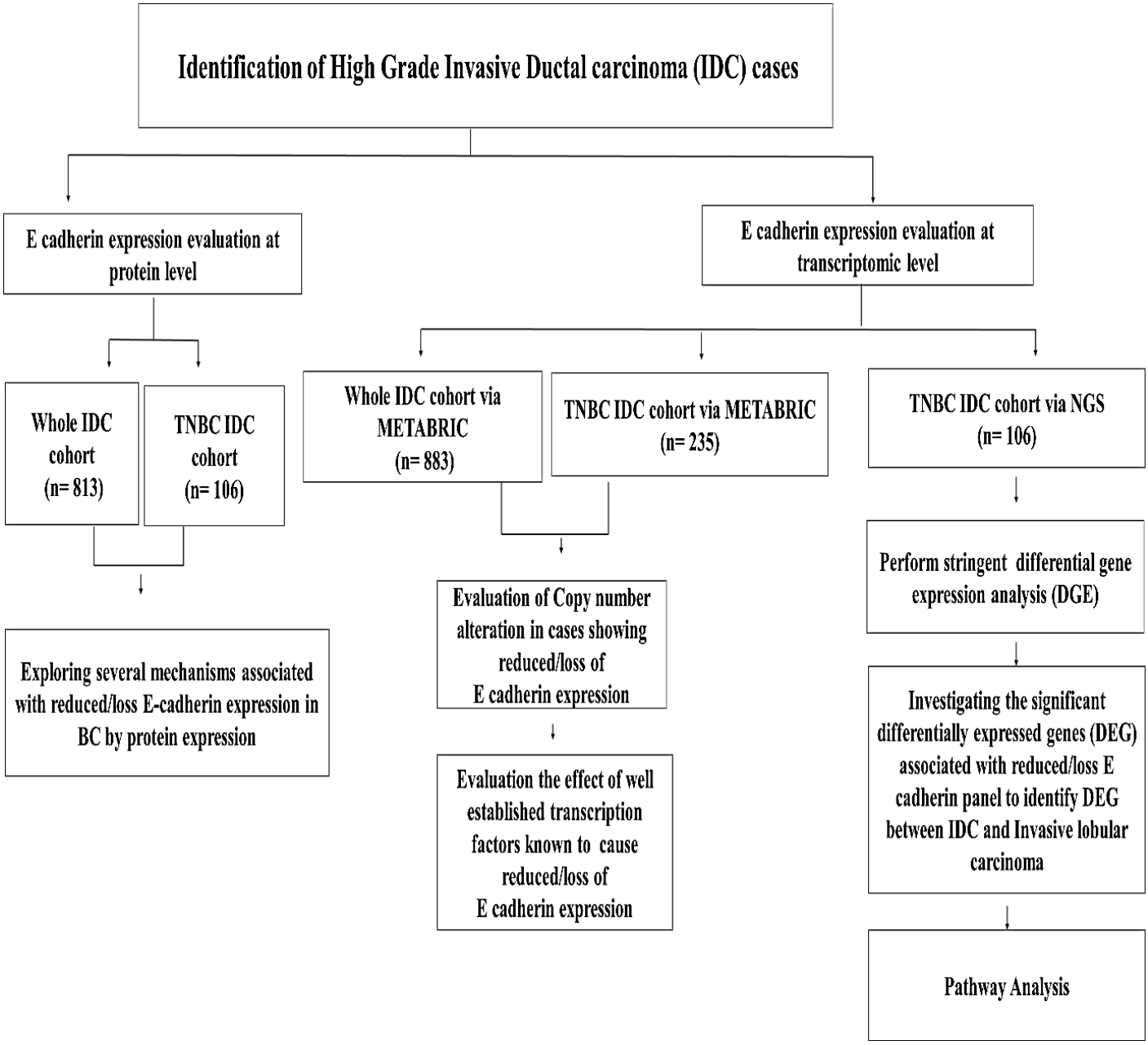


Figure 5-5-1:Flowchart of Analytical and Experimental Methodology of the molecular landscape of E-cadherin in ductal carcinoma of the breast study

Table 5-1:Patients cohort

Clinicopathological Parameter	**Whole IDC protien cohort (n=813) N (%)**	**Whole IDC in METABRIC cohort (n=883) N (%)**	**Triple negative - IDC cohort (n=106) N (%)**
Age (years)			
< 50	326 (40)	266 (30)	59 (56)
≥ 50	487 (60)	617 (70)	47 (44)
Tumour size (cm)			
≤ 2cm	319 (39)	325 (37)	29 (27)
>2cm	494 (61)	551 (63)	77 (73)
Nodal Stage			
1	455 (56)	186 (27)	64 (60)
2	258 (32)	420 (62)	29 (27)
3	100 (12)	73 (11)	13 (13)
Estrogen receptor status			
Positive	452 (56)	372 (42)	0 (0)
Negative	361 (44)	511 (58)	106 (100)
Nottingham prognostic index			
Good prognosis	1 (0.1)	0 (0)	2 (2)
Moderate prognosis	567 (69.9)	705 (80)	72 (68)
Poor prognosis	245 (30)	178 (20)	32 (30)

5.4.2 Immunohistochemistry and western blotting analyses

Mouse monoclonal anti-E-cadherin antibody [Cl;4A2C7, Ref#180223, LOT 954621A, Invitrogen, UK] was used to assess protein expression on immunohistochemically-stained 4mm thick full-face tissue sections after prior validation of the antibody by Western blot using MDA-MB-231 and MDA-MB-157 BC cell lysates as previously mentioned (**Chapter 2**). Sections were incubated for 24 hours with the anti-CDH1 antibody diluted to a concentration of 1:25. Scoring of membranous protein expression was performed using the modified histo-score [195]. This study used the lower quartile from the modified histochemical score value (i.e. 85/0-300) to stratify the cohort into high/normal and reduced/loss E-cadherin expression groups (Figure 5-2). Cases in the METABRIC cohort were stratified using a similar approach for total *CDH1* mRNA expression. Copy number alteration and *CDH1* mRNA expression were correlated with E-cadherin protein expression in the same cases where available (n=131).

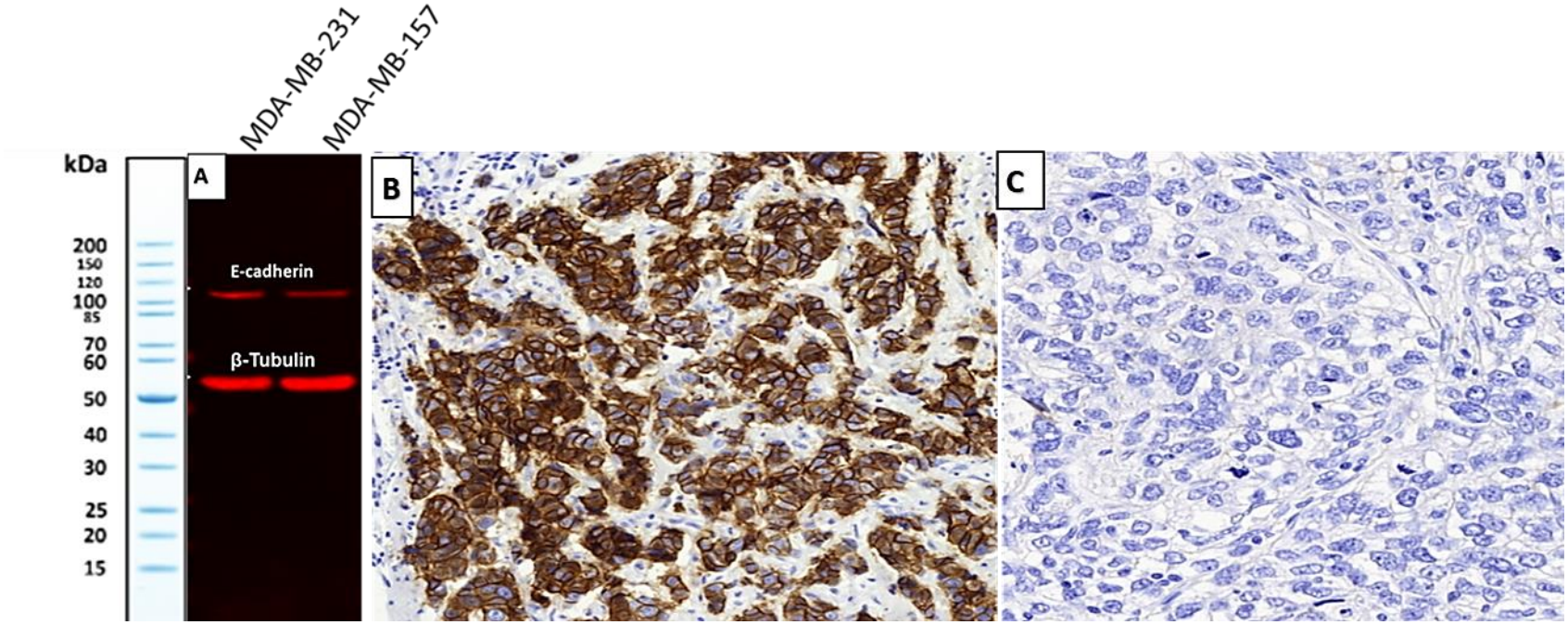


Figure 5-2:Western blotting & Immunohistochemistry

- Specificity of the E-cadherin antibody was confirmed by western blotting using MDA-MB231 and MDA-MB157 cell lines, which indicated a specific band at approximately 100 kDa. β-Tubulin antibody (clone AA2, Ref# T8328, Sigma-Aldrich, UK) was kindly gifted from Dr Reem Ali , member of Dr Madhusudan Srinivasan group. B) Immunohistochemistry for E-cadherin expression in IDC tumours showing strong membranous staining (power 40×). C) Immunohistochemistry for E-cadherin expression in IDC tumours showing complete loss of membranous staining (power 40×)

5.4.3 Statistical analysis

IBM SPSS 24.0 (Chicago, IL, USA) software was used for statistical analysis as explained in (**Chapter 2**). The chi-squared test was used to assess the effect of copy number alteration on reduced/loss of CDH1 mRNA expression. Furthermore, we evaluated copy number alteration of established tumuor suppressor genes in cases that exhibited reduced/loss of *CDH1* mRNA expression and copy number loss via copy number alteration in the METABRIC cohort, to infer genetic instability as the likely driver of the reduced/loss of *CDH1* mRNA expression using chi-squared test. Mann-Whitney test was used to compare the expression of *CDH1* mRNA expression with expression of well-established transcription factors affecting E-cadherin expression [29]. Furthermore, we evaluated expression of a set of genes previously demonstrated to have 93% predictive accuracy in distinguishing ILC from IDC via the prediction analysis for microarrays test [324]. Expression of proteins related to DNA repair, available transcription factors, and proliferation were compared to the expression of the E-cadherin protein using the Mann-Whitney test (**Table 5-2**) (**Chapter 2**). Furthermore, the association of E-cadherin protein expression with that of transcription factors mRNA expression (assessed using NGS - HTSEQ values) was evaluated using the Mann-Whitney test. Two tailed p-value <0.05 was considered as statistically significant. RNA-Seq values were expressed as standard error of means in GRAPH PAD PRISM v.7 for data presentation.

Table 5-2:Biomarkers included in this study to examine the potential association with altered E-cadherin expression in high grade IDC protein cohort (n=813).

Rational of Inclusion		Biomarkers	Staining and evaluation
Transcription Factor	**Dysregulation of these biomarkers has been reported to trigger EMT and suppress E cadherin expression** [29,209]**.**	**TGFBeta1**	**Dr. Mohammed Aleskandarany** [207–209]**,**
		TWIST2	
		ZEB	
DNA repair	**Dysregulation or alteration of expression of these biomarkers may infer to a general cancer genomic instability** [214,325,326]	**BRCA1**	**Dr. Alaa Alshareeda** [210–212]
		BRCA2	
		GammaH2ax	
		RAD51	
Proliferation and other biomarkers family		**Ki67**	**Dr. Mohammed Aleskandarany** [213]
		PTEN	**Dr. Alaa Alshareeda** [98,208,214,215]
		ATM	
		TP53	

5.4.4 Bioinformatics investigation and Pathway analysis

To expand our understanding on the molecular biology underlying reduced/loss of E cadherin expresstion at the protein level, DGE analysis was performed using the Robina implementation of Edge-R statistical tool [187] (**Chapter 2**). The dichotomisation of cases into reduced/loss of E cadherin expression versus high E cadherin expression groups relied

on the lower quartile H-scores (85/300 H-score). Henceforth, DGE was performed between cases showing reduced/loss of E cadherin expression (n=46) against cases harbouring normal/above threshhold E cadherin expression (n=60). Pathway analysis was used to identify differentially regulated canonical pathways generated from DEG panel by Edge R using KEGG pathway database [188] as previously described (**Chapter 2**).

5.5 RESULTS

5.5.1 Evaluation of E-cadherin protein expression in the high-grade IDC protein cohort (n=813)

A total of 217/813 (27%) of high-grade IDC and 46/106 (43%) of TNBC-IDC cancer cases showed reduced/loss membrane expression of E-cadherin. Within the METABRIC cohort, reduced/loss *CDH1* mRNA expression was observed in 208/883 (23%) cases. Furthermore, TNBC-IDC tumours showed reduced/loss of *CDH1* mRNA expression in 90/235 (38%) cases. Reduced/loss *CDH1* mRNA expression cases were observed in 104 cases of the IDC - Basal like tumours (37%), 18 cases of the IDC- HER2 enriched tumours (11%), 40 cases of IDC- luminal A tumours (27%) , 29 cases of IDC luminal B tumours (12%) , and 17 cases of the Normal like tumours (29%) molecular subtypes. In the subset of Nottingham cases that were included in the METABRIC dataset (n=131), there was a positive linear correlation between *CDH1* mRNA and the dichotomised E-cadherin protein expression (r=0.27, p=0.002). Interestingly, reduced/loss E-cadherin protein expression was associated with *GammaH2AX* (p<0.0001), *RAD51* (p=0.050) and PTEN (p=0.003) protein expression in the whole IDC protein cohort (Table 5-3).

Table 5-3:Correlation between level of proteins associated with altered E-cadherin expression in high-grade IDC cohort (n=813)

Identifier		Protein of interest	E-cadherin **expression frequency**		Mean RANK		Z-score	***p*-value**
			Low (%)	High (%)	Low	High		
Protein Level	Transcription Factor	TGFBeta1	143 (27)	371 (73)	257	258	-0.04	0.96
		TWIST2	130 (25)	376 (75)	259	251	-0.60	0.540
		ZEB	126 (26)	352 (74)	240	239	-0.19	0.840
	DNA repair family	BRCA1	178 (26)	491 (74)	336	335	-0.05	0.950
		BRCA2	132 (26)	375 (74)	248	256	-1.19	0.230
		GammaH2ax	120 (23)	387 (77)	221	264	-3.48	**<0.001**
		RAD51	110 (25)	320 (75)	199	221	-1.96	**0.050**
	Proliferation and other markers family	Ki67	174 (25)	501 (75)	329	341	-1.14	0.250
		PTEN	104 (28)	260 (72)	164	189	-2.93	**0.003**
		ATM	143 (28)	358 (72)	245	253	-0.70	0.480
		TP53	214 (26)	590 (74)	408	400	-0.47	0.630

5.5.2 E-cadherin copy number alteration in high-grade IDC tumours

To investigate whether reduced/loss of E-cadherin expression in the IDC cases is due to copy number alteration, we examined copy number alteration and *CDH1* mRNA levels. We observed that 44/208 (21%) of cases showed significant association between loss of *CDH1* copy number and reduced/loss of *CDH1* mRNA expression (p=0.003) (Table 5-4). Only 1 case with copy number loss did not show any association with the transcription factors investigated while the remaining cases showed upregulation of one or more transcription factors (Table 5-5). Interestingly, 77% of tumours presenting with reduced/loss *CDH1* mRNA expression did not show *CDH1* copy number loss, indicating that other mechanisms are implicated. Subsequently, investigating the TNBC IDC cases within METABRIC, only 7/90 (8%) of cases showed copy number to be associated with reduced/loss *CDH1* mRNA expression. However, there was no statistical association between copy number loss and reduced/loss of *CDH1* mRNA expression (p=0.10) (Table 5-6). More importantly, amongst those cases, only 1 (copy number loss) case did not show any association with any transcription factors while the rest of the 6/90 (7%) cases (copy number loss) showed upregulation of one or more transcription factors (Table 5-7). Moreover, 83/90 (92%) of TNBC IDC tumours within the METABRIC cohort, with reduced/loss *CDH1* mRNA expression showed neutral/amplified *CDH1* copy number expression.

In addition, reduced/loss *CDH1* mRNA expression in IDC showed copy number loss of multiple well-established breast cancer tumour suppressor genes located at different chromosome loci; *TP53, ATM, BRCA1, and BRCA2* ($p<0.001$) (Table 5-8).

Table 5-4:CDH1 copy number alteration (CNA) in whole IDC METABRIC cohort (n=883)

CNA region for *CDH1*	**CNA**	***Cases with reduced/loss CDH1*** mRNA ***expression* (%)**	***Cases with normal/above threshold CDH1*** mRNA ***Expression* (%)**	**Total (%)**	***p*-value**
Chromosome 16q22.1	Loss	44 (35)	81 (65)	125 (100)	**0.003**
	Gain	4 (15)	23 (85)	27 (100)	
	Neutral	160 (22)	571 (78)	731 (100)	
	Total	208 (23)	675 (77)	883 (100)	

Table 5-5:Frequency of genes associated with reduced E-cadherin expression in the whole IDC METABRIC cohort (n= 208)

Frequency of Cumulative Number of Upregulation of Transcription Factor suppressor of *CDH1* mRNA expression in cases with reduced/loss of *CDH1* mRNA expression	Frequency (%)
11	1 (1)
10	1 (1)
9	5 (3)
8	19 (12)
7	35 (21)
6	42 (26)
5	38 (23)
4	32 (20)
3	21 (13)
2	8 (5)
1	5 (3)

- Dichotomisation of mRNA level was based on median of the total expression of each gene.

Table 5-6:CDH1 copy number alteration (CNA) in the high grade TNBC IDC cohort within the METABRIC cohort (n=235)

CNA region for *CDH1*	**CNA**	***Cases with reduced/loss CDH1*** mRNA ***expression (%)***	***Cases with normal/above threshold CDH1*** mRNA ***Expression (%)***	**Total *(%)***	***p*-value**
Chromosome 16q22.1	Loss	7 (70)	3 (30)	10 (100)	0.10
	Gain	4 (36)	7 (64)	11 (100)	
	Neutral	79 (37)	135 (63)	214 (100)	
	Total	90 (38)	145 (62)	235 (100)	

- Dichotomisation of mRNA level was based on median of the total expression of each gene.

Table 5-7:Frequency of genes associated with reduced E-cadherin expression in the high grade TNBC IDC cohort within the METABRIC cohort (n=90)

Frequency of cumulative number of upregulation of transcription factors' suppressor of *CDH1* mRNA expression in cases with reduced/loss of *CDH1* mRNA expression	Frequency (%)
10	1 (1)
9	2 (2)
8	10 (11)
7	20 (24)
6	15 (17)
5	14 (16)
4	11 (12)
3	9 (10)
2	3 (3)
1	4 (4)

- Dichotomisation of mRNA level was based on median of the total expression of each gene.

Table 5-8:Copy number alteration (CNA) within other chromosomes in the whole IDC METABRIC cohort with reduced/loss CDH1 mRNA expression. (n=208)

Gene	**Cases with reduced/loss *CDH1* mRNA expression**			***p*-value**
	Loss (%)	**Gain (%)**	**Neutral (%)**	
ATM	26 (12)	2 (1)	180 (87)	**<0.001**
PTEN	17 (8)	8 (4)	178 (88)	0.53
RB1	36 (17)	24 (12)	145 (71)	0.46
TP53	25 (12)	0 (0)	183 (88)	**<0.001**
BRCA1	13 (6)	8 (4)	187 (90)	**<0.001**
BRCA2	12 (6)	15 (7)	180 (87)	**<0.001**

- Dichotomisation of mRNA level was based on median of the total expression of each gene.

5.5.3 Expression of E-cadherin suppressor transcription factors

In cases with reduced/loss E-cadherin expression (n=208) from the METABRIC cohort, upregulated mRNA expression was observed with *ZEB2* (56%), *TWIST2* (54%), *NFKB1* (54%), *ZEB1* (53%), *TWIST1* (52%), *SLUG* (51%), *SNAIL* (50%), *GSK3BETA* (49%), *TGFB1* (47%), *LLGL2 (38%),* and *CRUMBS3* (34%). Only 4% of the cases were affected by 9 or more upregulated transcription factors (Table 5-9). Upregulated expression of *TWIST2*, *ZEB2*, *NFKB1*, *LLGL2* and *CRUMBS3* were significantly associated with reduced/loss of *CDH1* mRNA expression (Table 5-10). In TNBC IDC cases with reduce/loss E-cadherin expression, upregulated mRNA expression was observed with *ZEB2* (63%), *SLUG* (62%), *TWIST2* (59%), *TWIST1* (57%), *ZEB1* (54%), *SNAIL* (52%),

TGFB1 (51%), *GSK3BETA* (50%), *NFKB1* (46%), *LLGL2 (24%)*, and *CRUMBS3* (24%) (Table 5-11). Only 3% of the cases harboured 9 or more upregulated transcription factors (Table 5-7**).** Upregulated expression of *TWIST2*, *TWIST1*, *ZEB2*, *ZEB1*, *SLUG*, *LLGL2* and *CRUMBS3* were significantly associated with reduced/loss of *CDH1* mRNA expression (Table 5-12).

Table 5-9:known Transcriptional factors associated with reduced E-cadherin expression examined in the reduced/loss of CDH1 mRNA expression IDC METABRIC cases (n=208)

Upregulation of transcription factors' suppressor of *CDH1* mRNA expression in cases with reduced/loss of *CDH1* mRNA expression	**Percentage**
ZEB2	**56%**
TWIST2	**54%**
NFKB1	**54%**
ZEB1	**53%**
TWIST1	**52%**
SLUG	**51%**
SNAIL	**50%**
GSK3BETA	**49%**
TGFB1	**47%**
LLGL2	**38%**
CRUMBS3	**34%**

- Dichotomization of mRNA level was based on median of the total expression of each gene.

Table 5-10:Correlation between mRNA levels of the genes associated with altered E-cadherin expression in BC in the whole IDC METABRIC cohort (n=883)

Identifier	**Genome of interest**	**E-cadherin** expression **frequency**		**Mean RANK**		**Z-score**	***p*-value**
		Low (%)	**High (%)**	**Low**	**High**		
Transcription Factor mRNA Level	*TGFB1*	208 (23)	675 (77)	443	442	-0.07	0.940
	TWIST2			481	430	-2.51	**0.012**
	TWIST1			466	434	-1.58	0.110
	ZEB2			489	427	-3.05	**0.002**
	ZEB1			445	441	-0.21	0.830
	SLUG			463	436	-1.33	0.180
	SNAIL			457	437	-0.94	0.340
	NFKB1			489	427	-3.06	**0.002**
	LLGL2			349	470	-6.00	**<0.001**
	GSK3B			439	443	-0.150	0.880
	CRUMBS			335	475	-6.93	**<0.001**

- Dichotomization of mRNA level was based on median of the total expression of each gene

Table 5-11:Genes associated with reduced E-cadherin expression in the TNBC IDC cohort within the METABRIC cohort (n=235)

Upregulation of transcription factor suppressor of *CDH1* mRNA expression in cases with reduced/loss of *CDH1* mRNA expression	**Percentage**
ZEB2	**63%**
SLUG	**62%**
TWIST2	**59%**
TWIST1	**57%**
ZEB1	**54%**
SNAIL	**52%**
TGFB1	**51%**
GSK3BETA	**50%**
NFKB1	**46%**
LLGL2	**24%**
CRUMBS3	**24%**

- Dichotomization of mRNA level was based on median of the total expression of each gene

Table 5-12:Correlation between mRNA level of the genes associated with E-cadherin expression in the high grade TNBC IDC cohort within the METABRIC cohort (n=235)

Identifier	Gene of interest	E-cadherin expression frequency		Mean RANK		Z- score	*p*-value
		Low (%)	H high (%)	Low	High		
Transcription Factor mRNA Level	*TGFB1*	90 (38)	145 (62)	114	120	- 0.74	0.460
	TWIST2			130	110	- 2.16	**0.030**
	TWIST1			132	109	- 2.51	**0.010**
	ZEB2			138	106	- 3.49	**<0.001**
	ZEB1			134	108	- 2.825	**0.005**
	SLUG			129	111	- 2.048	**0.040**
	SNAIL			113	121	- 0.94	0.340
	NFKB1			115	119	- 0.47	0.630
	LLGL2			89	135	- 5.00	**<0.001**
	GSK3B			108	124	-1.71	0.080
	CRUMBS			91	135	- 4.80	**<0.001**

- Dichotomization of mRNA level was based on median of the total expression of each gene

5.5.4 Proteins associated with E-cadherin expression in invasive TNBC-IDC

There was no significant statistical correlation between reduced/ loss of E-cadherin expression with transcription factors, DNA repair family, nor other markers such as ki67, ATM and PTEN on the protein level in TNBC cohort (Table 5-13 **&** Table 5-14**&** Figure 5-3)

Table 5-13:Correlation between level of proteins known to control E-cadherin expression using triple negative -IDC cohort (n=106)

Identifier		Protein of interest	E-cadherin **expression frequency**		Mean RANK		Z-score	***p*-value**
			Low (%)	High (%)	Low	High		
Protein Level	Transcription Factor	TGFBeta1	27 (47)	30 (53)	27	31	-0.87	0.380
		TWIST2	26 (48)	28 (52)	27	31	-0.23	0.810
		ZEB	25 (49)	26 (51)	27	25	-0.79	0.430
	DNA repair family	BRCA1	33 (44)	41(56)	37	38	-0.26	0.790
		BRCA2	26 (45)	31 (55)	26	32	-1.90	0.060
		GammaH2ax	27 (43)	35 (57)	28	34	-1.47	0.140
		RAD51	23 (51)	22 (49)	21	24	-0.85	0.390
	Proliferation and other markers family	Ki67	35 (47)	40 (53)	42	34	-1.61	0.100
		PTEN	21 (47)	19 (53)	22	19	-0.96	0.390
		ATM	28 (49)	29 (51)	29	29	-0.18	0.850
		TP53	39 (41)	45 (59)	43	42	-0.32	0.750

Table 5-14:Correlation between level mRNA expression of other genes known to control E-cadherin expression using triple negative invasive breast carcinoma cohort (n=106)

Identifier		**Gene of interest**	**E-cadherin expression frequency**		**Mean RANK**		***Z*-score**	***p*-value**
			Low (%)	**High** (%)	**Low**	**High**		
mRNA Level	**Transcription Factor family**	*TGFB1*	46 (43)	60 (57)	56	51	-0.88	0.370
		TWIST2			56	51	-1.28	0.200
		TWIST1			53	53	-0.01	0.990
		ZEB2			56	52	-0.74	0.450
		ZEB1			57	51	-0.94	0.340
		SLUG			51	55	-0.69	0.480
		SNAIL			55	53	-0.43	0.660
		NFKB1			57	51	-0.95	0.340
		LLGL2			58	50	-1.33	0.180
		GSK3B			54	53	-0.29	0.760
		CRUMBS			53	54	-0.07	0.930

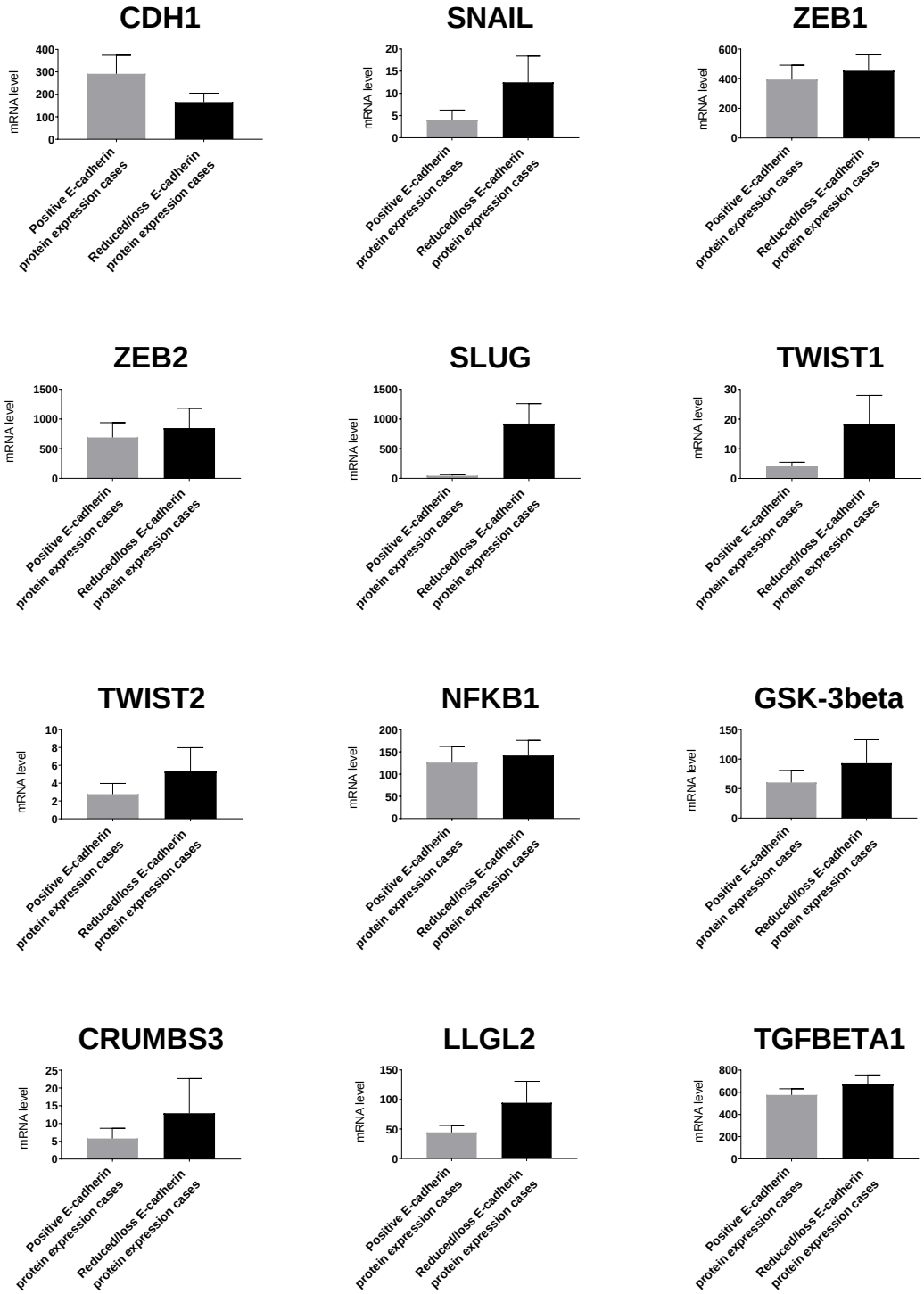


Figure 5-3:Frequency of Transcription factors suppressing E-cadherin expression in Cases showing reduced/loss CDH1 mRNA expression in the triple negative breast cancer cohort

5.5.5 E-cadherin loss and expression of genes deferentially expressed between ILC and IDC within the TNBC IDC cohort.

There was no significant association between reduced/ loss of E-cadherin expression in the high grade TNBC tumours and those genes differentially expressed between ILC and IDC (*Cathepsin B, TPI1, SPRY1, SCYA14, TFAP2B, thrombospondin 4, Osteopontin, HLA-G, CHC1*) [324] (Table 5-15 **&** Figure 5-4).

Table 5-15:Genes differentially expressed between lobular versus ductal breast carcinomas in triple negative breast cancer cohort (n=106)

Identifier	Gene of Interest	E-cadherin expression frequency		Mean RANK		*Z*-score	*p*-value
		Low (%)	High (%)	Low	High		
mRNA Level	*Cathepsin B*	46 (43)	60 (57)	51	55	-0.57	0.560
	TPI1			56	51	-0.87	0.380
	SPRY1			54	53	-0.39	0.690
	SCYA14			54	53	-0.20	0.830
	TFAP2B			52	55	-0.46	0.640
	Thrombospondin 4			53	54	-0.09	0.920
	Osteopontin			54	53	-0.26	0.790
	HLA-G			52	55	-0.51	0.600
	CHC1			53	54	-0.13	0.890

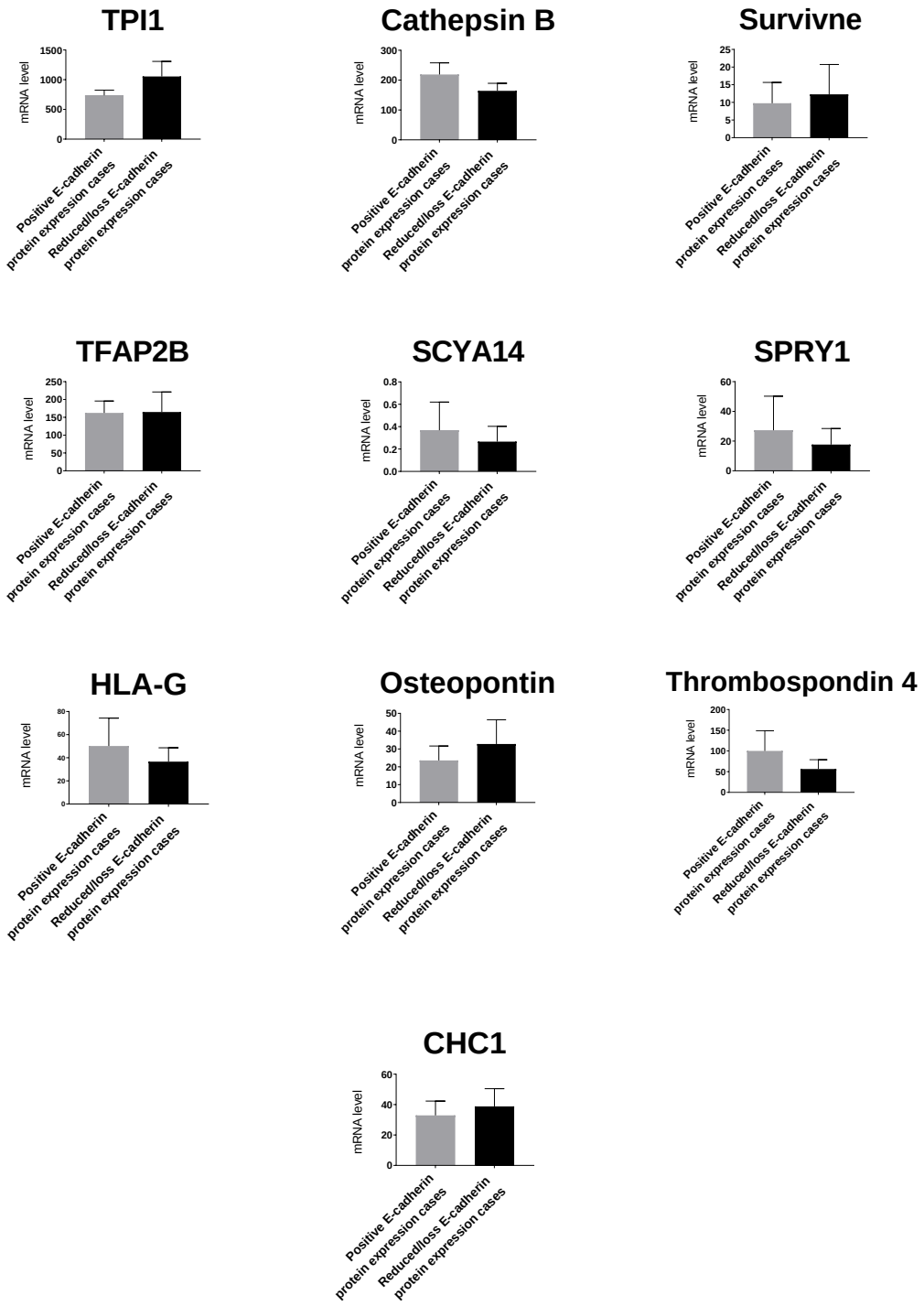


Figure 5-4:Differentially expressed genes between lobular versus ductal breast carcinomas in the triple negative breast cancer cohort

5.5.6 Bioinformatics investigation and pathway analysis

DGE Analysis identified 2143 differentially expressed genes (Benjamin-Hochberg; $p< 0.05$, differentially expressed by >two-fold, false discovery rate <0.05). Triple negative IDC with reduced/loss E-cadherin expression (n=46) showed 849 significantly overexpressed and 1294 downregulated genes. It is noteworthy that dysregulation of genes regulating Wnt signaling pathway, the top predicted master regulator of E-cadherin expression, based on p-value, whose activity could explain protein expression differences were (*FZD2, GNG5, HLTF, WNT2, and CER1*), PIK3-AKT signaling pathway top predicted master regulator controlling E-cadherin expression were (*FGFR2, GNF5, GNGT1, IFNA17, and IGF1*) (**Table 5-16**). Importantly, key genes differentially expressed between ILC and invasive ductal tumours [324] did not show association with E-cadherin reduced/loss of expression in the invasive triple negative ductal carcinoma (**Table 5-17**).

Table 5-16:Pathway analysis results using Web gestalt to identify differentially regulated canonical pathways in the triple negative breast cancer cohort

Master regulator	Gene symbol	Gene name	score	FDR score from GSEA	*p* value from GSEA
PI3K-Akt signalling pathway	*FGFR2*	Fibroblast growth factor receptor 2	0.04	0.0069	**<0.001**
	GNG5	G protein subunit gamma 5	0.03		
	GNGT1	G protein subunit gamma transducin1	0.02		
	IFNA17	Interferon alpha 17	0.03		
	IGF1	Insulin like growth factor	0.03		
Wnt signalling pathway	*FZD2*	frizzled class receptor 2	0.01	0.0024	**<0.001**
	GNG5	G protein subunit gamma 5	0.03		
	HLTF	helicase like transcription factor	0.01		
	WNT2	Wnt family member 2	0.03		
	CER1	Cerberus 1, DAN family BMP antagonist	0.01		

Table 5-17:Genomic study and pathway analysis based on data generated from next generation RNA sequencing in triple negative breast cancer cohort

Identifier	Log Fold Change	False Discovery Rate	*p*-value
FZD2	3.60	0.01	**<0.001**
GNG5	-2.73	0.03	**<0.001**
HLTF	3.39	0.01	**<0.001**
WNT2	-2.47	0.03	**<0.001**
CER1	4.17	0.01	**<0.001**
FGFR2	-3.26	0.04	**<0.001**
GNGT1	-4.49	0.02	**<0.001**
IFNA17	-6.90	0.03	**<0.001**
IGF1	-2.99	0.03	**<0.001**

- Negative (-) correlation with CDH1 genomic and protein expression (downregulated when CDH1 is low/negative cases). Positive (+) correlation with CDH1 genomic and protein expression (upregulated when CDH1 is low/negative cases). Significant *p* values are in bold

5.6 Discussion

Reduced/loss of E-cadherin expression is recognised as part of the main molecular events driving loss of cell-cell adhesion and thus facilitating cancer invasion and metastasis [327]. reports have suggested that E-cadherin can serve as a phenotypic marker to distinguish between ILC and non-invasive lobular tumours [324]. Mechanisms seeding reduced/loss of E-cadherin expression comprise *CDH1* gene mutation [311], truncating mutation [328], promoter hypermethylation [329] and transcriptional inactivation [330]. Reduced/loss of E-cadherin expression is observed in 84% of ILC [310]. Several studies have shown that approximately 38% of high-grade IDC tumours show reduced/ loss of E-cadherin expression, and this phenomenon has been linked to aggressive tumour behaviour. Interestingly, *CDH1* gene mutations were not identified in this subgroup [312,323,331].

One of the recognised mechanisms leading to reduced/loss of E-cadherin expression is loss of heterozygosity at chromosome 16q22.1, where the *CHD1* gene is located [332]. Studies investigating the mechanism underlying reduced/loss of E-cadherin protein expression in ILC cases uncovered loss of wild-type allele due to loss of heterozygosity at 16q22.1 occurring in more than 70% of ILC cases [308,309]. Furthermore, *CDH1* gene mutation and promoter hypermethylation were observed in 20% and 56% of ILC, respectively [308]. Interestingly, co-occurrence of these mechanisms rarely occurs in ILC tumours [333]. Remarkably, mutational inactivation of *CDH1* gene mostly coexists with loss of the wild type allele in ILC [334]. Since reduced/loss of E-cadherin expression in ILC tumours is predominantly caused by loss of heterozygosity, it has been suggested that copy number loss of the *CDH1* gene can be utilised to discriminate between IDC and ILC tumours when it is difficult to differentiate them based on histological evaluation [335]. Our investigation revealed that copy number loss occurred in only 21% of IDC tumours displaying reduced/loss of E-cadherin expression. Therefore, other mechanisms must underlie the

downregulation of E-cadherin in the majority of cases. Other mechanisms of E-cadherin reduced/loss of expression without copy number loss include DNA hypermethylation, a mechanism that may induce the *CDH1* reduced/loss of mRNA expression detected in 60% of metastatic IDC [336].

Loss of *CDH1* gene at 16q22.1 in ILC is one of the main genet events and is observed early in the process of carcinogenesis in lobular carcinomas. We hypothesized that reduced/loss of E-cadherin expression in a subset of IDC tumours might be the result of genomic instability and occurs as a late event during the process of cancer progression. Our results demonstrate that loss of *CDH1* copy number is associated with copy number alteration of multiple well-established BC tumour suppressor genes located at different chromosomes;copy number loss of *ATM* (11q22.3), *PTEN (*10q23.31), *RB1* (13q14.2), *TP53* (17p13.1), *BRCA1* (17q21.31), and *BRCA2* (13q13.1) tumour suppressor genes. Moreover, DNA damage response pathways which are crucial for detecting DNA lesions and arresting the cell cycle until the DNA is repaired or inducing cell death if cells sustain irreparable DNA damage [337], play key roles in preventing genetic instability and tumourigenesis [338]. Investigation of correlations between reduced/loss of E-cadherin expression and expression of biomarkers related to DNA damage response pathways in BC revealed negative correlation between reduced/loss of E-cadherin protein expression and GammaH2AX, RAD51 and PTEN expression suggesting that reduced/loss of E-cadherin expression is associated with impaired DNA damage response and likely, genomic instability. Taken together, these results support our hypothesis that reduced/loss of E-cadherin expression in invasive IDC is associated with genomic instability.

Reduced/loss of E-cadherin expression can also be caused by overexpression of its associated transcription factors [29,339–341]. Our results showed a negative correlation between reduced/loss of *CDH1* mRNA expression and the mRNA expression of

transcription factors known to suppress E-cadherin expression and cause disruption of cell-cell adhesion[342]-[343]; in fact, 76% of cases harbouring reduced/loss of *CDH1* mRNA show upregulation of one or more of these transcriptional repressors.

Remarkably, other key players in EMT such as TGFBeta1, SNAIL and SLUG did not show any correlation with E-cadherin reduced/loss of mRNA expression. These observations suggest that reduced/loss of E-cadherin expression is not merely a surrogate for EMT but represents a readout of other pathways controlling E-cadherin expression at membranes level.

Of note, reduced/loss of E-cadherin protein expression occurs in up to 50% of TNBC ductal IDC tumours, which may contribute to increased lymph node metastasis, and poor patient outcomes [320]. We observed a negative correlation between reduced/loss of *CDH1* mRNA expression and the mRNA expression of multiple transcription factors known to suppress E-cadherin expression in our TNBC cohort. On the contrary, when we investigated the same genotype within the cohort tested by NGS, none of these transcription factors showed statistically significant associations with E-cadherin expression. It is possible that different molecular mechanisms regulate E-cadherin expression, although we cannot exclude the possibility that our cohort is small to such associations.

More importantly, genes differentially expressed between IDC and ILC tumours as identified by Waldman and colleagues [324] and could represent the effect of E-cadherin loss in ILC compared to IDC tumours showed no statistically significant difference, when tested on mRNA level, in TNBC cases showing reduced/loss of E-cadherin expression compared to these tumour with normal E cadherin expression. This may indicate not only that more complex molecular mechanisms are responsible for E-cadherin reduced/loss of protein expression in these cases but also E-cadherin loss in IDC does not produce the same effects in ILC tumours. This may also be supported by the lack of morphological features

and metastatic behaviour characteristic of ILC in IDC tumours lacking E-cadherin expression.

In this study, DEG investigating differences between cases with reduced/loss of E-cadherin expression and cases with normal/high expression showed dysregulation of genes regulating *PIK3-AKT* signaling pathway. Our analysis exposed a negative correlation between the genes regulating this pathway and reduced/loss of E-cadherin protein expression suggesting that overexpression of those indicators may promote signaling via the *PIK3-AKT* pathway and thus negatively regulate E-cadherin expression. Receptors such as insulin like growth factor receptor 1, can induce the activity of *Akt* pathway [344]. Our results are in agreement with reports indicating activation of *PIK3-AKT* represses E-cadherin expression and stimulates cell migration [345]. Nonetheless, dysregulation of genes regulating Wnt signaling pathway was also present in our results. Mutation or deregulation of gene expression of the canonical Wnt pathway is implicated in cancer [346–348].

5.7 **Limitation of the study**

Our study limitation relates to comparing gene expression data obtained from microarrays, as used in the METABRIC cohort comprising different molecular subtypes of invasive ductal tumours, and the RNA sequencing dataset available for our triple negative breast cancers only. We have chosen TNBC to study E-cadherin protein expression in IDC cases as up to 50% of this molecular subtype show reduced/loss of E-cadherin protein expression [320,349]. On the contrary, studies have shown that reduced/loss of E-cadherin expression occurs in 23% and 27% of luminal and HER2 enriched subtypes, respectively [349]. RNA-Seq approaches cover multiple aspects of the transcriptome without any prior knowledge, allowing to identify novel transcripts, splice junctions and noncoding RNAs [350]. We

acknowledge that comparison between these two different approaches may or may not provide the same results due to intrinsic differences in assay design [351]. For instance, next generation sequencing may have different lower limits of detection or may encompass different genomic regions [350]. More importantly, IDC cases used in the METABRIC cohort comprise different molecular subtypes, while the RNA sequencing data was acquired for a TNBC cohort, which also may play a role in our study. Therefore, further validation of our findings is warranted.

5.8 **Conclusion**

Reduced/loss E-cadherin expression in IDC is a complex biological phenomenon, which, according to the findings of this study, appears to be a part of the genomic instability process occurring late in the process of carcinogenesis rather than an initial neoplastic event and results in different effects to those produced in invasive lobular carcinomas. Using the high throughput next generation sequencing, we have unravelled potential novel regulators controlling different signalling pathways that regulate E-cadherin protein expression in IDC. These regulators warrant further investigation and validation using different platforms.

Chapter 6 : General Discussion

This thesis has identified novel molecular features and biological characteristics underpinning the clinical and morphological heterogeneity of BC particularly TNBC. This biological heterogeneity must be considered in the development of improved preventive, diagnostic, prognostic and therapeutic strategies for TNBC with the goal of improving patient well-being and outcomes. This study has deepened the understanding of the distinctive molecular profiles and how these contribute to the wide variety of morphological and molecular entities of TNBC and indeed, inter- and intratumoural heterogeneity [352].

6.1 Background

TNBC constitutes an aggressive BC subtype which is clinically associated with frequent recurrence, distant metastasis and death, especially in the first five years of follow-up [136,150,221]. The prognoses for patients with TNBCs, as determined by genomic prognostic signatures, are generally poor, and this reflects the tendency for advanced clinical stages and histological grades by the time a diagnosis is made [353–355]. The treatment of TNBCs is constrained by the absence of targeted therapies available for ER and HER2 expression in BC tumours, and cytotoxic chemotherapy, radiation and surgery are currently the only treatment options available to TNBC patients [107]. As a result, a lot of current research is focused on the identification of effective, stratifying biomarkers and novel therapeutic targets for the diversity of TNBCs observed in the clinic. One method that has been explored is the profiling of gene expression [110,217]. This approach enables the identification of particular TNBC molecular subtypes. To date, the clinical exploitation of gene profiling has been limited by its cost, impracticality for use on a regular basis and poor predictive ability. This is in part because of the limited availability of genomic data from TNBC patients to inform the development of accurate predictive models. Similarly, the absence of this data has contributed to the lack of patient-specific precision medicine approaches to treat TNBC. Thus, there is still a lack of cost-effective, feasible tests that can

discern risk-predictive biomarkers. However, this information is urgently needed to stratify TNBCs accurately, reveal variations in their tumour biology and identify potential pharmaceutical interventions to treat them.

6.2 Hypothesis, Aims and Methods of the Study

In this study, it was hypothesised that TNBC has a variable prognosis which depends on phenotypic and genotypic characteristics that reflect complex mechanisms governed by the dysregulation of distinct oncogenic transcriptional networks which enable TNBC metastasis. Therefore, it is unlikely that a single molecule will be identifiable as a promotor of metastasis or act as a prognostic determinant. Instead, multiple pathways coordinate to promote metastasis in TNBC. Once such pathways are identified, they could be used to better predict the outcomes of TNBC patients.

The aims of this study were to test this hypothesis and to identify potential promotors and prognostic predictors of metastasis that might be able to stratify TNBC patients into high and low risk groups, with the goal of improving risk stratification and therapy recommendations for TNBC patients. This was accomplished through an evaluation of a large and well-characterised cohort of TNBC cases treated in a single institution, with available transcriptomic profiles, accompanied by the utilisation of available public domain data to validate the prognostic value of identified targets, and constructed TMA with clinical data information. A biomarker panel of adverse clinical outcomes in TNBC was developed, incorporating important genetic changes responsible for metastasis. The biomarker panel was chosen based on: 1) an exploration of the transcriptomic data from a large cohort of TNBC samples, using advanced statistical approaches and aiming to identify a molecular combinatorial prognostic signature predicting risk scores for poor outcomes in TNBC; and 2) an integrative exploration of the prognostic potential and possible biological

interactions associated with PPL expression in TNBC in terms of its transcriptomic, proteomic and mechanistic functions, to investigate its potential for TNBC targeted therapy and 3) an examination of the literature to assess the possible underlying mechanisms and the molecular implications of the loss of E-cadherin, a key molecular player known to trigger tumour progression and metastasis in TNBC in comparison to other molecular subtypes of BC.

6.3 Key Findings of the Study

6.3.1 Identification of Potential Prognostic Genes in TNBC

To date, attempts to refine the prognostic stratification of TNBC patients has had limited success as most studies have used small numbers of patients or have examined gene expression datasets obtained from different gene expression platforms, which have contributed to poorly reproducible results [224]. In addition, utilising transcriptional profiles to classify TNBC patients into molecular subtypes, does not alter the early onset of distant metastasis and poor prognosis [226]. Owing to the biological heterogeneity of TNBCs, in this thesis, a new approach is proposed for risk assessment of distant metastasis and BCSS probabilities in TNBC using the transcriptomic data available and the application of several statistical analysis tools to ensure the rigor and accuracy of this approach. Using this approach, a prognostic two gene signature and enriched pathways involving EGFR, P38 MAPK , the PDGF, and the RAS pathways, highlight their strong association with TNBC metastasis and worse patient outcomes [110,243–246].

Interestingly, the prognostic value of the two-gene signature (ACSM4 and SPDYC) yielded 58% sensitivity and 64% accuracy in predicting the distant metastasis outcomes of TNBC patients. In contrast, at the protein level, our proposed two-gene signature demonstrated 73% sensitivity and 50% accuracy in predicting the distant metastasis outcomes of TNBC

patients. Our proposed two-gene signature showed promising accuracy and sensitivity in predicting the risk of distant metastasis in TNBC patients, which is important because presently, TNBC patients rely solely on chemotherapy treatment. Moreover, those patients who are deemed at high risk for distant metastasis by our stratification model may benefit from more aggressive treatment approaches.

Furthermore, our proposed gene signature is only based on two genes (*ACSM4* and *SPDYC*) unlike other commercially available prognostic assays including those designed for ER-positive tumours [261]. Our prognostic gene signature may be amenable to the development of affordable molecular tests based on quantitative RT-PCR as the sensitivity, specificity and accuracy of our two-gene signature was confirmed at the mRNA level. The prognostic gene signature might be suitable for use in routine clinical practice because the proposed two-gene signature has prognostic value in classifying TNBC patients and may provide important information for treatment decisions.

6.3.2 Comprehensive Transcriptomic Profiling Reveals PPL as a Novel, Independent Predictor of Outcomes in TNBC

Next, this study addressed the clinical and mechanistic importance of periplakin (PPL) in TNBC. Transcriptomic analysis of the prognostic value of *PPL* expression in TNBC cohorts was statistically significant in predicting outcomes in TNBC patients. In addition, a high PPL protein expression, defined as any positive staining in tumour epithelial cells, was strongly associated with TNBC outcomes. Hence, the use of multiple transcriptomics cohorts coupled with protein expression and a positive correlation between protein and mRNA expression paved the way for the *in vitro* studies reported here. Previously, PPL has only been associated with poor outcomes in endometrial cancer but not in other types of cancer or in any BC subtypes [272]. The *in vitro* studies reported here provided proof of concept experiments in which silencing PPL expression in two TNBC cell line models

demonstrated reduced proliferative, migratory and invasive capacities. Therefore, we hypothesised that our identified biomarkers are potential therapeutic targets in TNBC.

Post-transcriptional gene silencing using siRNA holds a promising future for the development of new therapeutic targets [356]. The use of siRNA nanoparticles as platforms for targeted therapy encompasses several advantages, such as, but not limited to, silencing any gene in mammalian cells, synthesis and optimisation to attain maximum gene silencing and the availability of several siRNA software that facilitate the optimisation of siRNA length, specificity and nucleotide content constructed from numeral criteria available online [357,358]. Interestingly, preclinical studies on multiple solid tumours showed the benefit of using siRNA nanoparticles in supressing tumour growth, tumour angiogenesis and metastatic capacities and enhancing drug delivery [359–363]. SiRNA therapeutics have been tested in clinical trials in a variety of cancer contexts (NCT03608631, NCT03087591, NCT01591356, NCT00672542 retrieved from clinicaltrials.gov). Collectively, the aforementioned studies helped to pave the way for new approaches to target PPL function in TNBC cells.

Notably, the data in this study is limited, and future investigation using a wide spectrum of TNBC cell lines is needed. However, our existing results highlight the role of PPL expression in TNBC tumour progression. Therefore, PPL may represent a novel target for siRNA targeted therapy in TNBC patients if a robust preclinical platform for drug development was feasible. Before this can be accomplished, further mechanistic studies using mouse models to validate our findings with *PPL* siRNA silencing are vital. Once this is achieved, a strict adherence to published guidelines needs to be followed in order to validate the *PPL* siRNA therapeutic target in TNBC patients [356]. These include providing evidence that designed *PPL* siRNA operates post transcriptionally, illustrating strong siRNA silencing specificity, further validation demonstrating that *PPL* mRNA expression

can be silenced via different siRNA sequencing to exclude non-specific effects [364,365] and selecting the optimal delivery method for siRNA [364–366]. Once these are completed and validation data is obtained, a preclinical development plan encompassing multiple major steps could be implemented as previously described [367].

In summary, the data obtained identified PPL as a novel therapeutic target in TNBC which may offer a new therapeutic approach to overcome treatment resistance.

6.3.3 The Molecular Landscape of E-Cadherin Expression in Ductal Carcinoma of the Breast

As discussed in depth in the first chapter, BC metastasis to distant organs is the main cause of BC-related deaths upon post-mortem examination [368]. Metastasis in BC is orchestrated by specific molecular changes, which involve different proteins and converging pathways.[369] For instance, a key characteristic of metastatic cells is their transition from an epithelial phenotype to being mesenchymal-like (EMT) [28]. During the EMT, tumour cells acquire the capacity to migrate to and invade other tissues [26]. A reduction or loss of E-cadherin expression is recognised as one of the main molecular events driving the loss of cell-cell adhesion and thereby facilitating EMT changes, eventually promoting cancer cell invasion and metastasis [327]. In BC, a reduction or loss of E-cadherin expression is frequently associated with a loss of ER expression, larger tumour sizes, and the development of metastasis and recurrence [314–317]. A reduction or loss of E-cadherin protein expression occurs in 42% of TNBC-IDC cases, exemplifying a leading molecular phenotype showing this phenomenon. This may explain increased lymph node metastasis and therefore, poor patient outcomes [320]. This study utilised a targeted approach to understand the molecular mechanisms underlying the reduction or loss of E-cadherin expression in BC. This approach was combined with multiple high throughput molecular techniques to decipher this complex biological phenomenon. The study results

indicated that the reduction or loss of E-cadherin in IDC is a multifactorial process that is likely related to genomic instability that occurs late in the process of carcinogenesis and influences IDC clinical behaviour. In contrast, an initial neoplastic event occurs in ILC tumours, caused mostly by a loss of heterozygosity at 16q22.1 that occurs in more than 70% of cases [308,309].

6.4 **Concluding remarks**

TNBC heterogeneity results from multiple different convergent phenotypic and genotypic factors that contribute to treatment resistance and the aggressive behaviour of tumours [370]. The use of large cohorts of annotated TNBC in various molecular techniques and bioinformatics approaches provides evidence of the molecular heterogeneity of TNBC and an improved understanding of the disease. We identified several genes and proteins that are strongly predictive of TNBC outcomes including *ACSM4, SPDYC* and *PPL*, which are potential therapeutic targets. Also, the methodological approaches performed in this study can be used to investigate underlying molecular functions and identify genetic panels predicting patients' outcomes in other subtypes of BC as well as in different types of tumours. We provide further insights into the mechanism of the altered adhesion process and the impact of reduced or lost E-cadherin expression in TNBC. Further studies on the therapeutic value of the identified novel genes are warranted.

6.5 **Strengths of the study**

- The study was designed to use a large cohort of TNBC patients. In addition, individual data relating to clinicopathology, tissue markers and outcomes was obtained, confirmed and prospectively maintained.
- By using a high-throughput statistical analysis of TNBC RNA-seq cohorts, a unique TNBC-gene signature was proposed for the purpose of predicting outcomes.

- The study's findings on the discovery cohort and protein expression were validated by combining data from mechanistic studies with in-house TNBC transcriptomic data and public domain data.
- In addition to WB and IPB, antibodies were used with positive and negative controls. In addition, published guidelines for conducting prognostic marker studies were closely followed to ensure that the interpretations of stains and statistical correlations were appropriate [189,237].
- Instead of using arbitrary cut-off levels of positivity for different markers, various statistical approaches were used to identify the optimal cut-off points.
- To evaluate the expression of markers, visual evaluations were made by an experienced pathologist instead of the use of automated image analysis methods. The reason for this was that experienced, manual evaluation is more discriminating than automated methods. The former is able to exercise judgement, but the latter are less able to differentiate between non-malignant nuclei such as the foci of normal, mammary epithelial cells and in-situ lesions trapped in tissue cores.
- Several published studies were derived from this thesis.

6.6 Limitations of the study

- The retrieval, reformatting and validation of the NGS transcriptomic and clinical data took a considerable amount of time.
- Although a TMA composed of TNBC patient specimens (n=393) was conducted, at the outset of this study, it was planned to prepare a TNBC TMA composed of n=603 cases. However, due to national restrictions related to the COVID-19 pandemic, this was not possible.

- Investigating the role of TILS and associated markers in TNBC at the transcriptomic level was not feasible because of time and restrictions related to the COVID-19 pandemic lockdown.
- The IHC data were only assessed in patients from one centre and no multicentre validation was available.
- Prior to studying the biomarker panel, a training subset of full-face sections was performed to determine its suitability for use on TMAs. Studying tissue markers in a large sample size has specific technical limitations, such as core loss. However, there has been excellent consistency between the TMA results and the full-face sections of BC and other cancer tissues. Also, the use of one TMA core to study the expression of tumour markers has been validated by previous research [371,372].
- The processes of qualitative assessment and interpretation of the expression of biomarkers are inherently subjective [373]. However, a standard approach and independent confirmation by additional assessors were used to minimise subjective bias. Inter- and intra-observer agreement in staining interpretation was evaluated and excellent concordance was attained.
- The paraffin blocks used to study the expression of tissue markers in historical samples were relatively old. This may have exhausted or consumed some blocks. Furthermore, as TNBC treatment has evolved quickly, some of the therapeutic interventions used when the patients were treated are no longer routinely employed.

6.7 **Future Plan**

- Further functional studies are required to validate the functions of the proteins identified here. These could include *in vitro* cell line and xenograft studies using multiple TNBC cell lines such as MDA-MB 231, MDA-MB 436, MDA-MB 468 and MDA-MB 157

for knockdown. CRISPR-cas9 mediated targeting of the identified genes (particularly the oncogenes) would be valuable in exploring the therapeutic capacity of the target genes. Animal models can also be prepared to examine identified biomarkers and associations with TNBC tumour progression to understand the underlying mechanisms.

- Assess and characterise the role of TILS and associated markers in TNBC at the transcriptomic and protein levels to decipher the molecular profiles of these tumours and potentially decipher the mechanisms and effects of TILS in TNBC.
- To increase the robustness of this study results, future studies should include multicentre evaluation for the discovered biomarker panel. The impact and clinical utility of the results obtained will increase if the study's conclusions are clinically implemented.
- Study the prognostic or predictive power of the studied biomarker (PPL) and the proposed gene signature (ACSM4 and SPDYC) for metastatic risk stratification in TNBC to validate their utility and their predictive performance.
- Contribute to the literature by producing several publications to understand the molecular heterogeneity of TNBC by deciphering the transcriptomics and proteomics of global TNBC data available online.

References

1 Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;144:646–674.

2 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.

3 Wild CP WESB, ed. World Cancer Report: Cancer Research for Cancer Prevention (PDF) – IARC E-Bookshop. [cited 27 September 2020]. Available from: https://shop.iarc.fr/products/world-cancer-report-cancer-research-for-cancer-prevention-pdf.

4 Beckmann MW, Niederacher D, Schnüren HG, Gusterson BA, Bender HG. Multistep carcinogenesis of breast cancer and tumour heterogeneity. J. Mol. Med. 1997;75:429–439.

5 Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57–70.

6 Sugimura T, Terada R, Yokota J, Hirohashi S, Wakabayashi K. Multiple genetic alterations in human carcinogenesis. In: Environmental Health Perspectives. : National Institute of Environmental Health Sciences, 1992. p. 5–12.

7 Zhu K, Liu Q, Zhou Y, Tao C, Zhao Z, Sun J, et al. Oncogenes and tumor suppressor genes: Comparative genomics and network perspectives. BMC Genomics 2015;16:S8.

8 Hulpiau P, van Roy F. Molecular evolution of the cadherin superfamily. Int. J. Biochem. Cell Biol. 2009;41:349–369.

9 Li B, Shi H, Wang F, Hong D, Lv W, Xie X, et al. Expression of E-, P- and n-cadherin and its clinical significance in cervical squamous cell carcinoma and precancerous

lesions. PLoS One 2016;11. doi:10.1371/journal.pone.0155910.

10 Harris TJC, Tepass U. Adherens junctions: From molecules to morphogenesis. Nat. Rev. Mol. Cell Biol. 2010;11:502–514.

11 Tepass U, Truong K, Godt D, Ikura M, Peifer M. Cadherins in embryonic and neural morphogenesis. Nat. Rev. Mol. Cell Biol. 2000;1:91–100.

12 Guo R, Sakamoto H, Sugiura S, Ogawa M. Endothelial cell motility is compatible with junctional integrity. J Cell Physiol 2007;211:327–335.

13 van Roy F, Berx G. The cell-cell adhesion molecule E-cadherin. , 2008.

14 Halbleib JM, Nelson WJ. Cadherins in development: Cell adhesion, sorting, and tissue morphogenesis. Genes Dev. 2006;20:3199–3214.

15 Vleminckx K, Vakaet L, Mareel M, Fiers W, Van Roy F. Genetic manipulation of E-cadherin expression by epithelial tumor cells reveals an invasion suppressor role. Cell 1991;66:107–119.

16 Jeanes A, Gottardi CJ, Yap AS. Cadherins and cancer: How does cadherin dysfunction promote tumor progression? Oncogene. 2008;27:6920–6929.

17 Hennig G, Lowrick O, Birchmeier W, Behrens J. Mechanisms identified in the transcriptional control of epithelial gene expression. J Biol Chem 1996;271:595–602.

18 Russell H, Pranjol MZI. Transcription factors controlling E-cadherin downregulation in ovarian cancer. Biosci. Horizons. 2018;11:2018.

19 Stemmler MP. Cadherins in development and cancer. Mol. Biosyst. 2008;4:835–850.

20 Yoshiura K, Kanai Y, Ochiai A, Shimoyama Y, Sugimura T, Hirohashi S. Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. Proc Natl Acad Sci U S A 1995;92:7416–7419.

21 Kaur P, Porras TB, Ring A, Carpten JD, Lang JE. Comparison of TCGA and GENIE

genomic datasets for the detection of clinically actionable alterations in breast cancer. Sci Rep 2019;9. doi:10.1038/s41598-018-37574-8.

22 Cancer Genome Atlas Network TCGA, Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, et al. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.

23 Berx G, Becker KF, Höfler H, Van Roy F. Mutations of the human E-cadherin (CDH1) gene. Hum. Mutat. 1998;12:226–237.

24 Valastyan S, Weinberg RA. Tumor metastasis: Molecular insights and evolving paradigms. Cell. 2011;147:275–292.

25 Nicolini A, Giardino R, Carpi A, Ferrari P, Anselmi L, Colosimo S, et al. Metastatic breast cancer: an updating. Biomed Pharmacother 2006;60:548–556.

26 Nieto MA. The snail superfamily of zinc-finger transcription factors. Nat. Rev. Mol. Cell Biol. 2002;3:155–166.

27 Younis LK, El Sakka H, Haque I. The Prognostic Value of E-cadherin Expression in Breast Cancer. Int J Health Sci (Qassim) 2007;1:43–51.

28 Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial-Mesenchymal Transitions in Development and Disease. Cell. 2009;139:871–890.

29 Baranwal S, Alahari SK. Molecular mechanisms controlling E-cadherin expression in breast cancer. Biochem Biophys Res Commun 2009;384:6–11.

30 de Abreu FB, Peterson JD, Amos CI, Wells WA, Tsongalis GJ. Effective quality management practices in routine clinical next-generation sequencing. Clin Chem Lab Med 2016;54:761–771.

31 Lih C-J, Harrington RD, Sims DJ, Harper KN, Bouk CH, Datta V, et al. Analytical Validation of the Next-Generation Sequencing Assay for a Nationwide Signal-Finding

Clinical Trial. J Mol Diagnostics 2017;19:313–327.

32 Marotti JD, De Abreu FB, Wells WA, Tsongalis GJ. Triple Negative Breast Cancer: Next-Generation Sequencing for Target Identification. 2017. doi:10.1016/j.ajpath.2017.05.018.

33 Li W, Qiu T, Guo L, Ying J, Zhou A. NGS-based oncogenic mutations analysis in advanced colorectal cancer patients improves targeted therapy prediction. Pathol Res Pract 2019;215:483–489.

34 Pang KH, Esperto F, Noon AP. Opportunities of next-generation sequencing in non-muscle invasive bladder cancer outcome prediction. Transl. Androl. Urol. 2017;6:1043–1048.

35 Pongor LLL, Kormos M, Hatzis C, Pusztai L, Szabó A, Gyorffy B, et al. A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. Genome Med 2015;7:104.

36 Dong L, Wang W, Li A, Kansal R, Chen Y, Chen H, et al. Clinical Next Generation Sequencing for Precision Medicine in Cancer. Curr Genomics 2015;16:253–263.

37 Ignatiadis M, Singhal SK, Desmedt C, Haibe-Kains B, Criscitiello C, Andre F, et al. Gene Modules and Response to Neoadjuvant Chemotherapy in Breast Cancer Subtypes: A Pooled Analysis. J Clin Oncol 2012;30:1996–2004.

38 Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, et al. Biological Processes Associated with Breast Cancer Clinical Outcome Depend on the Molecular Subtypes. Clin Cancer Res 2008;14:5158–5165.

39 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63.

40 Desmedt C, Voet T, Sotiriou C, Campbell PJ. Next-generation sequencing in breast cancer: first take home messages. Curr Opin Oncol 2012;24:597–604.

41 Narrandes S, Xu W. Gene Expression Detection Assay for Cancer Clinical Use. J Cancer 2018;9:2249–2265.

42 Hassiotou F, Geddes D. Anatomy of the human mammary gland: Current status of knowledge. Clin. Anat. 2013;26:29–48.

43 Medina D. The mammary gland: a unique organ for the study of development and tumorigenesis. J. Mammary Gland Biol. Neoplasia. 1996;1:5–19.

44 Menes TS, Kerlikowske K, Lange J, Jaffer S, Rosenberg R, Miglioretti DL. Subsequent breast cancer risk following diagnosis of atypical ductal hyperplasia on needle biopsy. JAMA Oncol 2017;3:36–41.

45 Maxwell AJ, Clements K, Hilton B, Dodwell DJ, Evans A, Kearins O, et al. Risk factors for the development of invasive cancer in unresected ductal carcinoma in situ. Eur J Surg Oncol 2018;44:429–435.

46 Stingl J, Caldas C. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. Nat. Rev. Cancer. 2007;7:791–799.

47 Sjostrom M, Hartman L, Honeth G, Grabau D, Malmstrom P, Hegardt C, et al. Stem cell biomarker ALDH1A1 in breast cancer shows an association with prognosis and clinicopathological variables that is highly cut-off dependent. J Clin Pathol 2015;68:1012–1019.

48 Zhang S, Li R, Hang X, Yan M, Niu F, Liu L, et al. Can CD44+/CD24- tumor cells be used to determine the extent of breast cancer invasion following neoadjuvant chemotherapy? J Breast Cancer 2011;14:175–180.

49 Braunstein LZ, Taghian AG. Molecular Phenotype, Multigene Assays, and the

Locoregional Management of Breast Cancer. Semin. Radiat. Oncol. 2016;26:9–16.

50 Lopez-Garcia MA, Geyer FC, Lacroix-Triki M, Marchió C, Reis-Filho JS. Breast cancer precursors revisited: Molecular features and progression pathways. Histopathology. 2010;57:171–192.

51 Bray F, Soerjomataram I. The Changing Global Burden of Cancer: Transitions in Human Development and Implications for Cancer Prevention and Control. In: Disease Control Priorities, Third Edition (Volume 3): Cancer. : The World Bank, 2015. p. 23–44.

52 Youlden DR, Cramb SM, Dunn NAM, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: An international comparison of screening, incidence, survival and mortality. Cancer Epidemiol. 2012;36:237–248.

53 Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359–E386.

54 Saggu S, Rehman H, Abbas ZK, Ansari AA. Recent incidence and descriptive epidemiological survey of breast cancer in Saudi Arabia. Saudi Med J 2015;36:1176–1180.

55 Bcheraoui C El, Basulaiman M, Wilson S, Daoud F, Tuffaha M, AlMazroa MA, et al. Breast cancer screening in saudi arabia: Free but almost no takers. PLoS One 2015;10. doi:10.1371/journal.pone.0119051.

56 Gierisch JM, Coeytaux RR, Urrutia RP, Havrilesky LJ, Moorman PG, Lowery WJ, et al. Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: A systematic review. Cancer Epidemiol. Biomarkers Prev. 2013;22:1931–1943.

57 Crandall CJ, Aragaki AK, Cauley JA, McTiernan A, Manson JAE, Anderson G, et al. Breast tenderness and breast cancer risk in the estrogen plus progestin and estrogen-alone women's health initiative clinical trials. Breast Cancer Res Treat 2012;132:275–285.

58 Russo J, Moral R, Balogh GA, Mailo D, Russo IH. The protective role of pregnancy in breast cancer. Breast Cancer Res. 2005;7:131–142.

59 Zhou Y, Chen J, Li Q, Huang W, Lan H, Jiang H. Association between breastfeeding and breast cancer risk: Evidence from a meta-analysis. Breastfeed Med 2015;10:175–182.

60 HA A, KE K, H D, LE K, TH M, NS ES. Epidemiology and prognosis of breast cancer in young women. J Thorac Dis 2013;5 Suppl 1. doi:10.3978/J.ISSN.2072-1439.2013.05.24.

61 Slaoui M, Mouh FZ, Ghanname I, Razine R, El Mzibri M, Amrani M. Outcome of breast cancer in Moroccan young women correlated to clinic-pathological features, risk factors and treatment: A comparative study of 716 cases in a single institution. PLoS One 2016;11. doi:10.1371/journal.pone.0164841.

62 Honrado E, Benítez J, Palacios J. The molecular pathology of hereditary breast cancer: Genetic testing and therapeutic implications. Mod Pathol 2005;18:1305–1320.

63 Larsen MJ, Thomassen M, Gerdes AM, Kruse TA. Hereditary breast cancer: Clinical, Pathological and molecular characteristics. Breast Cancer Basic Clin Res 2014;8:145–155.

64 Shroff RT, Hendifar A, McWilliams RR, Geva R, Epelbaum R, Rolfe L, et al. Rucaparib Monotherapy in Patients With Pancreatic Cancer and a Known Deleterious BRCA Mutation . JCO Precis Oncol 2018;2018:1–15.

65 Wang GH, Zhao CM, Huang Y, Wang W, Zhang S, Wang X. BRCA1 and BRCA2 expression patterns and prognostic significance in digestive system cancers. Hum Pathol 2018;71:135–144.

66 Flom JD, Ferris JS, Tehranifar P, Terry MB. Alcohol intake over the life course and mammographic density. Breast Cancer Res Treat 2009;117:643–651.

67 Liu Y, Nguyen N, Colditz GA. Links between alcohol consumption and breast cancer: A look at the evidence. Women's Heal. 2015;11:65–77.

68 Dossus L, Boutron-Ruault MC, Kaaks R, Gram IT, Vilier A, Fervers B, et al. Active and passive cigarette smoking and breast cancer risk: Results from the EPIC cohort. Int J Cancer 2014;134:1871–1888.

69 Neuhouser ML, Aragaki AK, Prentice RL, Manson JAE, Chlebowski R, Carty CL, et al. Overweight, obesity, and postmenopausal invasive breast cancer risk: A secondary analysis of the women's health initiative randomized clinical trials. JAMA Oncol 2015;1:611–621.

70 McDonnell DP, Park S, Goulet MT, Jasper J, Wardell SE, Chang CY, et al. Obesity, cholesterol metabolism, and breast cancer pathogenesis. Cancer Res. 2014;74:4976–4982.

71 Baglietto L, Krishnan K, Stone J, Apicella C, Southey MC, English DR, et al. Associations of mammographic dense and nondense areas and body mass index with risk of breast cancer. Am J Epidemiol 2014;179:475–483.

72 Rakha EA, El-Sayed ME, Lee AHS, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of nottingham histologic grade in invasive breast carcinoma. J Clin Oncol 2008;26:3153–3158.

73 ELSTON CW, ELLIS IO. pathological prognostic factors in breast cancer. I. The value

of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991;19:403–410.

74 Lee AHS, Ellis IO. The Nottingham prognostic index for invasive carcinoma of the breast. Pathol. Oncol. Res. 2008;14:113–115.

75 Rakha EA, Tan PH, Varga Z, Tse GM, Shaaban AM, Climent F, et al. Prognostic factors in metaplastic carcinoma of the breast: A multi-institutional study. Br J Cancer 2015;112:283–289.

76 Mojarad S, Venturini B, Fulgenzi P, Papaleo R, Brisigotti M, Monti F, et al. Prediction of nodal metastasis and prognosis of breast cancer by ANN-based assessment of tumour size and p53, Ki-67 and steroid receptor expression. Anticancer Res 2013;33:3925–3933.

77 Saadatmand S, Bretveld R, Siesling S, Tilanus-Linthorst MMA. Influence of tumour stage at breast cancer detection on survival in modern times: Population based study in 173 797 patients. BMJ 2015;351. doi:10.1136/bmj.h4901.

78 Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham prognostic index in primary breast cancer. Breast Cancer Res Treat 1992;22:207–219.

79 Kurshumliu F, Gashi-Luci L, Kadare S, Alimehmeti M, Gozalan U. Classification of patients with breast cancer according to Nottingham Prognostic Index highlights significant differences in immunohistochemical marker expression. World J Surg Oncol 2014;12:243.

80 Domínguez MA, Marcos M, Meiriño R, Villafranca E, Dueñas MT, Arias F, et al. Prognostic and predictive factors in early breast cancer. An. Sist. Sanit. Navar. 2001;24:99–110.

81 Blamey RWW, Ellis IOO, Pinder SEE, Lee AHSHS, Macmillan RDD, Morgan

DALAL, et al. Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990–1999. Eur J Cancer 2007;43:1548–1555.

82 Rakha EA, Martin S, Lee AHS, Morgan D, Pharoah PDP, Hodi Z, et al. The prognostic significance of lymphovascular invasion in invasive breast carcinoma. Cancer 2012;118:3670–3680.

83 Stacker SA, Williams SP, Karnezis T, Shayan R, Fox SB, Achen MG. Lymphangiogenesis and lymphatic vessel remodelling in cancer. Nat. Rev. Cancer. 2014;14:159–172.

84 Al-Rohil RN, Milton DR, Nagarajan P, Curry JL, Feldmeyer L, Torres-Cabala CA, et al. Intratumoral and peritumoral lymphovascular invasion detected by D2-40 immunohistochemistry correlates with metastasis in primary cutaneous Merkel cell carcinoma. Hum Pathol 2018;77:98–107.

85 Payne SJL, Bowen RL, Jones JL, Wells CA. Predictive markers in breast cancer - The present. Histopathology. 2008;52:82–90.

86 Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J. Clin. Oncol. 2010;28:2784–2795.

87 Lee ES, Han W, Kim MK, Kim J, Yoo T kyung, Lee MH, et al. Factors associated with late recurrence after completion of 5-year adjuvant tamoxifen in estrogen receptor positive breast cancer. BMC Cancer 2016;16:430.

88 Rakha EA, Reis-Filho JS, Ellis IO. Combinatorial biomarker expression in breast cancer. Breast Cancer Res. Treat. 2010;120:293–308.

89 Cancello G, Montagna E. Treatment of breast cancer in young women: Do we need

more aggressive therapies? J. Thorac. Dis. 2013;5:S47.

90 Figueroa-Magalhães MC, Jelovac D, Connolly RM, Wolff AC. Treatment of HER2-positive breast cancer. Breast. 2014;23:128–136.

91 Ingthorsson S, Andersen K, Hilmarsdottir B, Maelandsmo GM, Magnusson MK, Gudjonsson T. HER2 induced EMT and tumorigenicity in breast epithelial progenitor cells is inhibited by coexpression of EGFR. Oncogene 2016;35:4244–4255.

92 Feigin ME, Muthuswamy SK. Polarity proteins regulate mammalian cell-cell junctions and cancer pathogenesis. Curr. Opin. Cell Biol. 2009;21:694–700.

93 Ahmed S, Sami A, Xiang J. HER2-directed therapy: current treatment options for HER2-positive breast cancer. Breast Cancer 2015;22:101–116.

94 Metzger-Filho O, Tutt A, de Azambuja E, Saini KS, Viale G, Loi S, et al. Dissecting the heterogeneity of triple-negative breast cancer. J Clin Oncol 2012;30:1879–1887.

95 Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–10874.

96 Desmedt C, Sotiriou C, Piccart-Gebhart MJ. Development and validation of gene expression profile signatures in early-stage breast cancer. Cancer Invest. 2009;27:1–10.

97 Reis-Filho JS, Tutt ANJJ. Triple negative tumours: a critical review. Histopathology 2008;52:108–118.

98 Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, et al. Triple-Negative Breast Cancer: Distinguishing between Basal and Nonbasal Subtypes. Clin Cancer Res 2009;15:2302–2310.

99 Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al.

Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol 2007;8:R76.

100 Dawson S-JJ, Rueda OM, Aparicio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. : EMBO J, 2013.

101 Kumar P, Aggarwal R. An overview of triple-negative breast cancer. Arch Gynecol Obstet 2016;293:247–269.

102 Kurebayashi J. Possible treatment strategies for triple-negative breast cancer on the basis of molecular characteristics. In: Breast Cancer. : Breast Cancer, 2009. p. 275–280.

103 Livasy CA. Pathologic Evaluation of Triple-Negative Breast Cancer. In: Triple-Negative Breast Cancer. : Springer International Publishing, 2018. p. 1–22.

104 Schneider BP, Winer EP, Foulkes WD, Garber J, Perou CM, Richardson A, et al. Triple-Negative Breast Cancer: Risk Factors to Potential Targets. Clin Cancer Res 2008;14:8010–8018.

105 Dolle JM, Daling JR, White E, Brinton LA, Doody DR, Porter PL, et al. Risk factors for triple-negative breast cancer in women under the age of 45 years. Cancer Epidemiol Biomarkers Prev 2009;18:1157–1166.

106 Boyle P. Triple-negative breast cancer: Epidemiological considerations and recommendations. Ann Oncol 2012;23. doi:10.1093/annonc/mds187.

107 Rakha EA, Chan S. Metastatic triple-negative breast cancer. Clin Oncol (R Coll Radiol) 2011;23:587–600.

108 Abramson VG, Lehmann BD, Ballinger TJ, Pietenpol JA. Subtyping of triple-negative breast cancer: Implications for therapy. Cancer 2015;121:8–16.

109 Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al.

Immunohistochemical and Clinical Characterization of the Basal-Like Subtype of Invasive Breast Carcinoma. Clin Cancer Res 2004;10:5367–5374.

110 Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121:2750–2767.

111 Bertucci F, Finetti P, Cervera N, Charafe-Jauffret E, Mamessier E, Adélaïde J, et al. Gene Expression Profiling Shows Medullary Breast Cancer Is a Subgroup of Basal Breast Cancers. Cancer Res 2006;66:4636–4644.

112 Wang J, Wei Q, Wang X, Tang S, Liu H, Zhang F, et al. Transition to resistance: An unexpected role of the EMT in cancer chemoresistance. Genes Dis 2016;3:3–6.

113 Lehmann BD, Jovanović B, Chen X, Estrada M V., Johnson KN, Shyr Y, et al. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. PLoS One 2016;11:e0157368.

114 Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346–352.

115 Pareja F, Geyer FC, Marchiò C, Burke KA, Weigelt B, Reis-Filho JS. Triple-negative breast cancer: the importance of molecular and histologic subtyping, and recognition of low-grade variants. npj Breast Cancer 2016;2:16036.

116 Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SAW, et al. Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-Negative Breast Cancer. Clin Cancer Res 2015;21:1688–1698.

117 Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486:395–

399.

118 Cancer Genome Atlas Network T. Comprehensive molecular portraits of human breast tumours. 2012. doi:10.1038/nature11412.

119 Dillon JL, Mockus SM, Ananda G, Spotlow V, Wells WA, Tsongalis GJ, et al. Somatic gene mutation analysis of triple negative breast cancers. The Breast 2016;29:202–207.

120 Weisman PS, Ng CKY, Brogi E, Eisenberg RE, Won HH, Piscuoglio S, et al. Genetic alterations of triple negative breast cancer by targeted next-generation sequencing and correlation with tumor morphology. Mod Pathol 2016;29:476–488.

121 Gewinner C, Wang ZC, Richardson A, Teruya-Feldstein J, Etemadmoghadam D, Bowtell D, et al. Evidence that Inositol Polyphosphate 4-Phosphatase Type II Is a Tumor Suppressor that Inhibits PI3K Signaling. Cancer Cell 2009;16:115–125.

122 Sun T, Aceto N, Meerbrey KL, Kessler JD, Zhou C, Migliaccio I, et al. Activation of Multiple Proto-oncogenic Tyrosine Kinases in Breast Cancer via Loss of the PTPN12 Phosphatase. Cell 2011;144:703–718.

123 Wu M-Q, Hu P, Gao J, Wei W-D, Xiao X-S, Tang H-L, et al. Low expression of tyrosine-protein phosphatase nonreceptor type 12 is associated with lymph node metastasis and poor prognosis in operable triple-negative breast cancer. Asian Pac J Cancer Prev 2013;14:287–292.

124 Shaver TM, Lehmann BD, Beeler JS, Li C-I, Li Z, Jin H, et al. Diverse, Biologically Relevant, and Targetable Gene Rearrangements in Triple-Negative Breast Cancer and Other Malignancies. Cancer Res 2016;76:4850–4860.

125 Parsons HA, Beaver JA, Cimino-Mathews A, Ali SM, Axilbund J, Chu D, et al. Individualized Molecular Analyses Guide Efforts (IMAGE): A Prospective Study of

Molecular Profiling of Tissue and Blood in Metastatic Triple-Negative Breast Cancer. Clin Cancer Res 2017;23:379–386.

126 Fidalgo F, Gonçalves A, Rodrigues T, Fiorini A, Carraro D, Rosenberg C, et al. Genome-wide profiling of copy number alterations in triple-negative breast cancer identifies a region at 19p13 associated with lymph node metastasis. BMC Proc 2013;7:P24.

127 Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor–negative breast cancer in the general population. Nat Genet 2010;42:885–892.

128 Stevens KN, Fredericksen Z, Vachon CM, Wang X, Margolin S, Lindblom A, et al. 19p13.1 Is a Triple-Negative-Specific Breast Cancer Susceptibility Locus. Cancer Res 2012;72:1795–1803.

129 Foulkes WD, Stefansson IM, Chappuis PO, Bégin LR, Goffin JR, Wong N, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003;95:1482–1485.

130 Le Du F, Eckhardt BL, Lim B, Litton JK, Moulder S, Meric-Bernstam F, et al. Is the future of personalized therapy in triple-negative breast cancer based on molecular subtype? Oncotarget 2015;6:12890–12908.

131 Yun MH, Hiom K. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. Nature 2009;459:460–463.

132 Turner N, Tutt A, Ashworth A. Opinion: Hallmarks of “BRCAness” in sporadic cancers. Nat Rev Cancer 2004;4:814–819.

133 Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast

cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol 2016;13:674–690.

134 Helleday T. The underlying mechanism for the PARP and BRCA synthetic lethality: Clearing up the misunderstandings. Mol Oncol 2011;5:387–393.

135 Buisson R, Dion-Côté A-M, Coulombe Y, Launay H, Cai H, Stasiak AZ, et al. Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating homologous recombination. Nat Struct Mol Biol 2010;17:1247–1254.

136 Foulkes WD, Smith IE, Reis-Filho JS. Triple-Negative Breast Cancer. N Engl J Med 2010;363:1938–1948.

137 Geyer FC, Pareja F, Weigelt B, Rakha E, Ellis IO, Schnitt SJ, et al. The Spectrum of Triple-Negative Breast Disease: High and Low Grade Lesions. Am J Pathol 2017;18:124–132.

138 Rakha EA, Badve S, Eusebi V, Reis-Filho JS, Fox SB, Dabbs DJ, et al. Breast lesions of uncertain malignant nature and limited metastatic potential: proposals to improve their recognition and clinical management. Histopathology 2016;68:45–56.

139 Geyer FC, Berman SH, Marchiò C, Burke KA, Guerini-Rocco E, Piscuoglio S, et al. Genetic analysis of microglandular adenosis and acinic cell carcinomas of the breast provides evidence for the existence of a low-grade triple-negative breast neoplasia family. Mod Pathol 2017;30:69–84.

140 Badve S, Dabbs DJ, Schnitt SJ, Baehner FL, Decker T, Eusebi V, et al. Basal-like and triple-negative breast cancers: A critical review with an emphasis on the implications for pathologists and oncologists. Mod. Pathol. 2011;24:157–167.

141 Hwang ES, DeVries S, Chew K Lo, Moore DH, Kerlikowske K, Thor A, et al. Patterns of chromosomal alterations in breast ductal carcinoma in situ. Clin Cancer Res

2004;10:5160–5167.

142 Leong ASY, Sormunen RT, Vinyuvat S, Hamdani RW, Suthipintawong C. Biologic markers in ductal carcinoma in situ and concurrent infiltrating carcinoma: A comparison of eight contemporary grading systems. Am J Clin Pathol 2001;115:709–718.

143 Koenig C, Dadmanesh F, Bratthauer GL, Tavassoli FA. Carcinoma arising in microglandular adenosis: An immunohistochemical analysis of 20 intraepithelial and invasive neoplasms. Int J Surg Pathol 2000;8:303–315.

144 Geyer FC, Kushner YB, Lambros MB, Natrajan R, MacKay A, Tamber N, et al. Microglandular adenosis or microglandular adenoma? A molecular genetic analysis of a case associated with atypia and invasive carcinoma. Histopathology 2009;55:732–743.

145 Guerini-Rocco E, Hodi Z, Piscuoglio S, Ng CK, Rakha EA, Schultheis AM, et al. The repertoire of somatic genetic alterations of acinic cell carcinomas of the breast: an exploratory, hypothesis-generating study. J Pathol 2015;237:166–178.

146 Ripamonti CB, Colombo M, Mondini P, Siranoush M, Peissel B, Bernard L, et al. First description of an acinic cell carcinoma of the breast in a BRCA1 mutation carrier: a case report. BMC Cancer 2013;13:46.

147 Guerini-Rocco E, Piscuoglio S, Ng CK, Geyer FC, De Filippo MR, Eberle CA, et al. Microglandular adenosis associated with triple-negative breast cancer is a neoplastic lesion of triple-negative phenotype harbouring *TP53* somatic mutations. J Pathol 2016;238:677–688.

148 Liu X, Holstege H, van der Gulden H, Treur-Mulder M, Zevenhoven J, Velds A, et al. Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of

human BRCA1-mutated basal-like breast cancer. Proc Natl Acad Sci 2007;104:12111–12116.

149 Persson M, Andren Y, Mark J, Horlings HM, Persson F, Stenman G. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. Proc Natl Acad Sci 2009;106:18740–18744.

150 Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-Negative Breast Cancer: Clinical Features and Patterns of Recurrence. Clin Cancer Res 2007;13:4429–4434.

151 Rakha EA, Ellis IO. Lobular breast carcinoma and its variants. Semin Diagn Pathol 2010;27:49–61.

152 Harbhajanka A, Lamzabi I, Singh RI, Ghai R, Reddy VB, Bitterman P, et al. Correlation of clinicopathologic parameters and immunohistochemical features of triple-negative invasive lobular carcinoma. Appl Immunohistochem Mol Morphol 2014;22. doi:10.1097/PAI.0b013e3182a5a02d.

153 Conforti F, Pala L, Guerini Rocco E, Pagan E, Bagnardi V, Montagna E, et al. Biological and clinical features of early triple-negative invasive lobular carcinomas of the breast. J Clin Oncol 2020;38:e12570–e12570.

154 Weigelt B, Eberle C, Cowell CF, Ng CKY, Reis-Filho JS. Metaplastic breast carcinoma: more than a special type. Nat Rev Cancer 2014;14:147–148.

155 Jung S-Y, Kim HY, Nam B-H, Min SY, Lee SJ, Park C, et al. Worse prognosis of metaplastic breast cancer patients than other patients with triple-negative breast cancer. Breast Cancer Res Treat 2010;120:627–637.

156 Weigelt B, Ng CKY, Shen R, Popova T, Schizas M, Natrajan R, et al. Metastatic breast carcinomas display genomic and transcriptomic heterogeneity. Mod Pathol

2015;28:340–351.

157 Geyer FC, Weigelt B, Natrajan R, Lambros MB, de Biase D, Vatcheva R, et al. Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas. J Pathol 2010;220:562–573.

158 Ng CKY, Piscuoglio S, Geyer FC, Burke KA, Pareja F, Eberle CA, et al. The Landscape of Somatic Genetic Alterations in Metaplastic Breast Carcinomas. Clin Cancer Res 2017;23:3859–3870.

159 Hennessy BT, Gonzalez-Angulo A-M, Stemke-Hale K, Gilcrease MZ, Krishnamurthy S, Lee J-S, et al. Characterization of a Naturally Occurring Breast Cancer Subset Enriched in Epithelial-to-Mesenchymal Transition and Stem Cell Characteristics. Cancer Res 2009;69:4116–4124.

160 Hayes MJ, Thomas D, Emmons A, Giordano TJ, Kleer CG. Genetic Changes of Wnt Pathway Genes Are Common Events in Metaplastic Carcinomas of the Breast. Clin Cancer Res 2008;14:4038–4044.

161 Weisman PS, Ng CKY, Brogi E, Eisenberg RE, Won HH, Piscuoglio S, et al. Genetic Alterations of Triple Negative Breast Cancer By Targeted Next Generation Sequencing And Correlation With Tumor Morphology HHS Public Access. Mod Pathol 2016;29:476–488.

162 Lakhani, S.R., Ellis. I.O., Schnitt, S.J., Tan, P.H., van de Vijver, M. J. (2012). WHO Classification of Tumours of the Breast, Fourth Edition - WHO - OMS - 240. WHO Classification of Tumours of the Breast, Fourth Edition - WHO - OMS - [Internet]. [cited 8 October 2017]. Available from: http://apps.who.int/bookorders/anglais/detart1.jsp?codlan=1&codcol=70&codcch=4004.

163 WHO Classification of Tumours Editorial Board, ed. WHO Classification of Tumours, 5th Edition, Volume 2, 5th ed. , 2019.

164 Xu R, Feiner H, Li P, Yee H, Inghirami G, Delgado Y, et al. Differential amplification and overexpression of HER-2/neu, p53, MIB1, and estrogen receptor/progesterone receptor among medullary carcinoma, atypical medullary carcinoma, and high-grade invasive ductal carcinoma of breast. Arch Pathol Lab Med 2003;127:1458–1464.

165 Vincent-Salomon A, Gruel N, Lucchesi C, MacGrogan G, Dendale R, Sigal-Zafrani B, et al. Identification of typical medullary breast carcinoma as a genomic sub-group of basal-like carcinomas, a heterogeneous new molecular entity. Breast Cancer Res 2007;9:R24.

166 Mateo AM, Pezzi TA, Sundermeyer M, Kelley CA, Klimberg VS, Pezzi CM. Atypical medullary carcinoma of the breast has similar prognostic factors and survival to typical medullary breast carcinoma: 3,976 cases from the National Cancer Data Base. J Surg Oncol 2016;114:533–536.

167 Rakha EA, Aleskandarany M, El-Sayed ME, Blamey RW, Elston CW, Ellis IO, et al. The prognostic significance of inflammation and medullary histological type in invasive carcinoma of the breast. Eur J Cancer 2009;45:1780–1787.

168 Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, et al. Prognostic and Predictive Value of Tumor-Infiltrating Lymphocytes in a Phase III Randomized Adjuvant Breast Cancer Trial in Node-Positive Breast Cancer Comparing the Addition of Docetaxel to Doxorubicin With Doxorubicin-Based Chemotherapy: BIG 02-98. J Clin Oncol 2013;31:860–867.

169 Celis JE, Gromova I, Gromov P, Moreira JMA, Cabezón T, Friis E, et al. Molecular pathology of breast apocrine carcinomas: A protein expression signature specific for benign apocrine metaplasia. FEBS Lett 2006;580:2935–2944.

170 Mills AM, E. Gottlieb C, M. Wendroth S, M. Brenin C, Atkins KA. Pure Apocrine Carcinomas Represent a Clinicopathologically Distinct Androgen Receptor–Positive Subset of Triple-Negative Breast Cancers. Am J Surg Pathol 2016;40:1109–1116.

171 Anders CK, Carey LA. Biology, Metastatic Patterns, and Treatment of Patients with Triple-Negative Breast Cancer. Clin Breast Cancer 2009;9:S73–S81.

172 Ballinger T, Kremer J, Miller K. Triple Negative Breast Cancer—Review of Current and Emerging Therapeutic Strategies. Oncol Hematol Rev 2016;12:89.

173 Schmadeka R, Harmon BE, Singh M. Triple-Negative Breast Carcinoma: Current and Emerging Concepts. Am J Clin Pathol 2014;141:462–477.

174 Bendell J, Saleh M, Rose AAN, Siegel PM, Hart L, Sirpal S, et al. Phase I/II Study of the Antibody-Drug Conjugate Glembatumumab Vedotin in Patients With Locally Advanced or Metastatic Breast Cancer. J Clin Oncol 2014;32:3619–3625.

175 Yardley DA, Weaver R, Melisko ME, Saleh MN, Arena FP, Forero A, et al. EMERGE: A Randomized Phase II Study of the Antibody-Drug Conjugate Glembatumumab Vedotin in Advanced Glycoprotein NMB–Expressing Breast Cancer. J Clin Oncol 2015;33:1609–1619.

176 Rugo HS, Olopade OI, DeMichele A, Yau C, van ’t Veer LJ, Buxton MB, et al. Adaptive Randomization of Veliparib–Carboplatin Treatment in Breast Cancer. N Engl J Med 2016;375:23–34.

177 Park HS, Jang MH, Kim EJ, Kim HJ, Lee HJ, Kim YJ, et al. High EGFR gene copy number predicts poor outcome in triple-negative breast cancer. Mod Pathol 2014;27:1212–1222.

178 Matsumoto H, Koo S, Dent R, Tan PH, Iqbal J. Role of inflammatory infiltrates in triple negative breast cancer: Table 1. J Clin Pathol 2015;68:506–510.

179 Beckers RK, Selinger CI, Vilain R, Madore J, Wilmott JS, Harvey K, et al. Programmed death ligand 1 expression in triple-negative breast cancer is associated with tumour-infiltrating lymphocytes and improved outcome. Histopathology 2016;69:25–34.

180 Nanda R, Chow LQM, Dees EC, Berger R, Gupta S, Geva R, et al. Pembrolizumab in Patients With Advanced Triple-Negative Breast Cancer: Phase Ib KEYNOTE-012 Study. J Clin Oncol 2016;34:2460–2467.

181 Pietri E, Conteduca V, Andreis D, Massa I, Melegari E, Sarti S, et al. Androgen receptor signaling pathways as a target for breast cancer treatment. Endocr Relat Cancer 2016;23:R485–R498.

182 Lehmann BD, Bauer JA, Schafer JM, Pendleton CS, Tang L, Johnson KC, et al. PIK3CA mutations in androgen receptor-positive triple negative breast cancer confer sensitivity to the combination of PI3K and androgen receptor inhibitors. Breast Cancer Res 2014;16:406.

183 Krueger F. Babraham Bioinformatics - Trim Galore! [Internet]. a wrapper tool around Cutadapt FastQC to consistently apply Qual. Adapt. trimming to FastQ files 2015. [cited 17 June 2020]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

184 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.

185 Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166–169.

186 Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 2010;26:493–500.

187 Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res 2012;40:W622–W627.

188 Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res 2017;45:W130–W137.

189 Ilyas M, Grabsch H, Ellis IO, Womack C, Brown R, Berney D, et al. Guidelines and considerations for conducting experiments using tissue microarrays. Histopathology. 2013;62:827–839.

190 Jawhar NMT. Tissue microarray: A rapidly evolving diagnostic and research tool. Ann. Saudi Med. 2009;29:123–127.

191 Dancau AM, Simon R, Mirlacher M, Sauter G. Tissue microarrays. In: Methods in Molecular Biology. : Humana Press Inc., 2016. p. 53–65.

192 Lacombe A, Carafa V, Schneider S, Sticker-Jantscheff M, Tornillo L, Eppenberger-Castori S. Re-Punching Tissue Microarrays Is Possible: Why Can This Be Useful and How to Do It. Microarrays 2015;4:245–254.

193 Coons AH, Creech HJ, Jones RN. Immunological Properties of an Antibody Containing a Fluorescent Group. Proc Soc Exp Biol Med 1941;47:200–202.

194 Torhorst J, Bucher C, Kononen J, Haas P, Zuber M, Köchli OR, et al. Tissue microarrays for rapid linking of molecular changes to clinical endpoints. Am J Pathol 2001;159:2249–2256.

195 McCarty KS, Miller LS, Cox EB, Konrath J, McCarty KS. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. Arch Pathol Lab Med 1985;109:716–721.

196 Camp RL, Dolled-Filhart M, Rimm DL. X-tile: A new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res 2004;10:7252–7259.

197 Aleskandarany MA, Rakha EA, MacMillan RD, Powe DG, Ellis IO, Green AR. MIB1/Ki-67 labelling index can classify grade 2 breast cancer into two clinically distinct subgroups. Breast Cancer Res Treat 2011;127:591–599.

198 Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ 1994;308:1552.

199 Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ 1994;309:102.

200 Bhargava EK, Rathore PK, Raj A, Meher R, Rana K. Diagnostic Efficacy of Computed Tomography in Detecting Cervical Metastases in Clinically N0 Head and Neck Squamous Cell Carcinoma. Indian J Otolaryngol Head Neck Surg 2016;68:25–29.

201 Jézéquel P, Campone M, Gouraud W, Charbonnel C, Leux C, Ricolleau G CL bc-G an easy-to-use online platform for gene prognostic analyses in breast cancer. BCRT 2012; 131: 765-775. bc-GenExMiner [Internet]. [cited 30 November 2017]. Available from: http://bcgenex.centregauducheau.fr/BC-GEM/GEM-Citation.php.

202 Győrffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res. Treat. 2010;123:725–731.

203 Fekete JT, Győrffy B. ROCplot.org: Validating predictive biomarkers of chemotherapy/hormonal therapy/anti-HER2 therapy using transcriptomic data of 3,104 breast cancer patients. Int J Cancer 2019;145:3140–3151.

204 Li Q, Birkbak NJ, Gyorffy B, Szallasi Z, Eklund AC. Jetset: Selecting the optimal microarray probe set to represent a gene. BMC Bioinformatics 2011;12:474.

205 Šidák Z. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. J Am Stat Assoc 1967;62:626–633.

206 Aleskandarany MA, Abduljabbar R, Ashankyty I, Elmouna A, Jerjees D, Ali S, et al. Prognostic significance of androgen receptor expression in invasive breast cancer: transcriptomic and protein expression analysis. Breast Cancer Res Treat 2016;159:215–227.

207 Rakha EA, Teoh TK, Lee AHS, Nolan CC, Ellis IO, Green AR. Further evidence that E-cadherin is not a tumour suppressor gene in invasive ductal carcinoma of the breast: an immunohistochemical study. Histopathology 2013;62:695–701.

208 Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JFR, et al. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer 2005;116:340–350.

209 Aleskandarany MA, Negm OH, Green AR, Ahmed MAH, Nolan CC, Tighe PJ, et al. Epithelial mesenchymal transition in early invasive breast cancer: an immunohistochemical and reverse phase protein array study. Breast Cancer Res Treat 2014;145:339–348.

210 Aleskandarany M, Caracappa D, Nolan CC, Macmillan RD, Ellis IO, Rakha EA, et al. DNA damage response markers are differentially expressed in BRCA-mutated breast cancers. Breast Cancer Res Treat 2015;150:81–90.

211 Albarakati N, Abdel-Fatah TMA, Doherty R, Russell R, Agarwal D, Moseley P, et al. Targeting BRCA1-BER deficient breast cancer by ATM or DNA-PKcs blockade either

alone or in combination with cisplatin for personalized therapy. Mol Oncol 2015;9:204–217.

212 Alshareeda AT, Negm OH, Aleskandarany MA, Green AR, Nolan C, TigHhe PJ, et al. Clinical and biological significance of RAD51 expression in breast cancer: a key DNA damage response protein. Breast Cancer Res Treat 2016;159:41–53.

213 Aleskandarany MA, Green AR, Benhasouna AA, Barros FF, Neal K, Reis-Filho JS, et al. Prognostic value of proliferation assay in the luminal, HER2-positive, and triple-negative biologic classes of breast cancer. Breast Cancer Res 2012;14:R3.

214 Aleskandarany MA, Rakha EA, Ahmed MA, Powe DG, Ellis IO, Green AR. Clinicopathologic and molecular significance of phospho-Akt expression in early invasive breast cancer. Breast Cancer Res Treat 2011;127:407–416.

215 Abdel-Fatah TMA, Arora A, Moseley PM, Perry C, Rakha EA, Green AR, et al. DNA repair prognostic index modelling reveals an essential role for base excision repair in influencing clinical outcomes in ER negative and triple negative breast cancers. Oncotarget 2015;6:21964–21978.

216 Alsaleem MA, Ball G, Toss MS, Raafat S, Aleskandarany M, Joseph C, et al. A novel prognostic two-gene signature for triple negative breast cancer. Mod Pathol 2020. doi:10.1038/s41379-020-0563-7.

217 Ahn SG, Kim SJ, Kim C, Jeong J. Molecular Classification of Triple-Negative Breast Cancer. J Breast Cancer 2016;19:223.

218 Liedtke C, Bernemann C, Kiesel L, Rody A. Genomic profiling in triple-negative breast cancer. Breast Care (Basel) 2013;8:408–413.

219 Khalifeh IM, Albarracin C, Diaz LK, Symmans FW, Edgerton ME, Hwang RF, et al. Clinical, Histopathologic, and Immunohistochemical Features of Microglandular

Adenosis and Transition Into In Situ and Invasive Carcinoma. Am J Surg Pathol 2008;32:544–552.

220 Stead LA, Lash TL, Sobieraj JE, Chi DD, Westrup JL, Charlot M, et al. Triple-negative breast cancers are increased in black women regardless of age or body mass index. Breast Cancer Res 2009;11:R18.

221 Haffty BG, Yang Q, Reiss M, Kearney T, Higgins SA, Weidhaas J, et al. Locoregional Relapse and Distant Metastasis in Conservatively Managed Triple Negative Early-Stage Breast Cancer. J Clin Oncol 2006;24:5652–5657.

222 Yam C, Mani SA, Moulder SL. Targeting the Molecular Subtypes of Triple Negative Breast Cancer: Understanding the Diversity to Progress the Field. Oncologist 2017;22:1086–1093.

223 Alizadeh AA, Ross DT, Perou CM, van de Rijn M. Towards a novel classification of human malignancies based on gene expression patterns. J Pathol 2001;195:41–52.

224 Katagiri T, Yoshimaru T, Matsuo T, Kiyotani K, Miyoshi Y, Tanahashi T, et al. Molecular features of triple negative breast cancer cells by genome-wide gene expression profiling analysis. Int J Oncol 2012;42:478–506.

225 Győrffy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L. Multigene prognostic tests in breast cancer: past, present, future. Breast Cancer Res 2015;17:11.

226 Ménard S. Heterogeneity of triple-negative breast carcinomas. Oncologie 2012;14:28–30.

227 Nagahashi M, Wakai T, Shimada Y, Ichikawa H, Kameyama H, Kobayashi T, et al. Genomic landscape of colorectal cancer in Japan: clinical implications of comprehensive genomic sequencing for precision medicine. Genome Med 2016;8:136.

228 Lips EH, Michaut M, Hoogstraat M, Mulder L, Besselink NJM, Koudijs MJ, et al.

Next generation sequencing of triple negative breast cancer to find predictors for chemotherapy response. Breast Cancer Res 2015;17:134.

229 Endrullat C, Glökler J, Franke P, Frohme M. Standardization and quality management in next-generation sequencing. Appl Transl Genomics 2016;10:2–9.

230 McDonough SJ, Bhagwate A, Sun Z, Wang C, Zschunke M, Gorman JA, et al. Use of FFPE-derived DNA in next generation sequencing: DNA extraction methods. PLoS One 2019;14:e0211400.

231 Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323:533–536.

232 Lancashire LJ, Powe DG, Reis-Filho JS, Rakha E, Lemetre C, Weigelt B, et al. A validated gene expression profile for detecting clinical outcome in breast cancer using artificial neural networks. Breast Cancer Res Treat 2010;120:83–93.

233 Picard RR, Cook RD. Cross-Validation of Regression Models. J Am Stat Assoc 1984;79:575–583.

234 Xu Q-S, Liang Y-Z, Du Y-P. Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration. J Chemom 2004;18:112–120.

235 Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams [Internet]. [cited 20 June 2019]. Available from: http://bioinfogp.cnb.csic.es/tools/venny/.

236 Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): Explanation and Elaboration. PLoS Med 2012;9:e1001216.

237 McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al.

REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer 2005;93:387–391.

238 van de Vijver MJ, He YD, van ’t Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. N Engl J Med 2002;347:1999–2009.

239 Lundin M, Lundin J, Burke HB, Toikkanen S, Pylkkänen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer. Oncology 1999;57:281–286.

240 Ball G, Mian S, Holding F, Allibone RO, Lowe J, Ali S, et al. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. Bioinformatics 2002;18:395–404.

241 Tafeit E, Reibnegger G. Artificial Neural Networks in Laboratory Medicine and Medical Outcome Prediction. Clin Chem Lab Med 1999;37:845–853.

242 Abdel-Fatah TMA, Agarwal D, Liu D-X, Russell R, Rueda OM, Liu K, et al. SPAG5 as a prognostic biomarker and chemotherapy sensitivity predictor in breast cancer: a retrospective, integrated genomic, transcriptomic, and protein analysis. Lancet Oncol 2016;17:1004–1018.

243 Costa R, Shah AN, Santa-Maria CA, Cruz MR, Mahalingam D, Carneiro BA, et al. Targeting Epidermal Growth Factor Receptor in triple negative breast cancer: New discoveries and practical insights for drug development. Cancer Treat Rev 2017;53:111–119.

244 Forte L, Turdo F, Ghirelli C, Aiello P, Casalini P, Iorio MV, et al. The PDGFRβ/ERK1/2 pathway regulates CDCP1 expression in triple-negative breast

cancer. BMC Cancer 2018;18:586.

245 Fan Y, Li M, Ma K, Hu Y, Jing J, Shi Y, et al. Dual-target MDM2/MDMX inhibitor increases the sensitization of doxorubicin and inhibits migration and invasion abilities of triple-negative breast cancer cells through activation of TAB1/TAK1/p38 MAPK pathway. Cancer Biol Ther 2019;20:617–632.

246 Adeyinka A, Nui Y, Cherlet T, Snell L, Watson PH, Murphy LC. Activated mitogen-activated protein kinase expression during human breast tumorigenesis and breast cancer progression. Clin Cancer Res 2002;8:1747–1753.

247 Ruiz-Lafuente N, Alcaraz-García M-J, García-Serna A-M, Sebastián-Ruiz S, Moya-Quiles M-R, García-Alonso A-M, et al. Dock10, a Cdc42 and Rac1 GEF, induces loss of elongation, filopodia, and ruffles in cervical cancer epithelial HeLa cells. Biol Open 2015;4:627–635.

248 Westcott JM, Prechtl AM, Maine EA, Dang TT, Esparza MA, Sun H, et al. An epigenetically distinct breast cancer cell subpopulation promotes collective invasion. J Clin Invest 2015;125:1927–1943.

249 Kraus MR-C, Clauin S, Pfister Y, Di Maïo M, Ulinski T, Constam D, et al. Two mutations in human BICC1 resulting in Wnt pathway hyperactivity associated with cystic renal dysplasia. Hum Mutat 2012;33:86–90.

250 Ruhrberg C, Hajibagheri MA, Parry DA, Watt FM. Periplakin, a novel component of cornified envelopes and desmosomes that belongs to the plakin family and forms complexes with envoplakin. J Cell Biol 1997;139:1835–1849.

251 Choi YK, Woo S-M, Cho S-G, Moon HE, Yun YJ, Kim JW, et al. Brain-metastatic triple-negative breast cancer cells regain growth ability by altering gene expression patterns. Cancer Genomics Proteomics 2013;10:265–275.

252 Cheng A, Solomon MJ. Speedy/Ringo C regulates S and G 2 phase progression in human cells. Cell Cycle 2008;7:3037–3047.

253 Mourón S, De Cárcer G, Seco E, Fernández-Miranda G, Malumbres M, Nebreda AR. RINGO C is required to sustain the spindle-assembly checkpoint. J Cell Sci 2010;123:2586–2595.

254 Bogunovic D, O'Neill DW, Belitskaya-Levy I, Vacic V, Yu Y Lo, Adams S, et al. Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival. Proc Natl Acad Sci U S A 2009;106:20429–20434.

255 Karavasilis V, Reid A, Sinha R, de Bono JS. Cancer drug resistance. In: Cancer Drug Design and Discovery.1st ed. : Elsevier Inc., 2008. p. 405–423.

256 Liu Y, Zhu YH, Mao CQ, Dou S, Shen S, Tan Z Bin, et al. Triple negative breast cancer therapy with CDK1 siRNA delivered by cationic lipid assisted PEG-PLA nanoparticles. J Control Release 2014;192:114–121.

257 Wu JQ, Dwyer DE, Dyer WB, Yang YH, Wang B, Saksena NK. Genome-wide analysis of primary CD4+ and CD8+ T cell transcriptomes shows evidence for a network of enriched pathways associated with HIV disease. Retrovirology 2011;8:18.

258 Guzmán-Fulgencio M, Jiménez JL, Jiménez-Sousa MA, Bellón JM, García-Álvarez M, Soriano V, et al. ACSM4 polymorphisms are associated with rapid AIDS progression in HIV-infected patients. J Acquir Immune Defic Syndr 2014;65:27–32.

259 Hendrickson SL, Lautenberger JA, Chinn LW, Malasky M, Sezgin E, Kingsley LA, et al. Genetic variants in nuclear-encoded mitochondrial genes influence AIDS progression. PLoS One 2010;5:e12862.

260 Althobiti M, Aleskandarany MA, Joseph C, Toss M, Mongan N, Diez-Rodriguez M, et

al. Heterogeneity of tumour-infiltrating lymphocytes in breast cancer and its prognostic significance. Histopathology 2018;73:887–896.

261 Gyanchandani R, Lin Y, Lin H-M, Cooper K, Normolle DP, Brufsky A, et al. Intratumor Heterogeneity Affects Gene Expression Profile Test Prognostic Risk Stratification in Early Breast Cancer. Clin Cancer Res 2016;22:5362–5369.

262 Isakoff SJ. Triple-negative breast cancer: Role of specific chemotherapy agents. Cancer J. 2010;16:53–61.

263 Biganzoli L, Cufer T, Bruning P, Coleman R, Duchateau L, Calvert AH, et al. Doxorubicin and paclitaxel versus doxorubicin and cyclophosphamide as first-line chemotherapy in metastatic breast cancer: The European Organization for Research and Treatment of Cancer 10961 Multicenter Phase III Trial. J Clin Oncol 2002;20:3114–3121.

264 Liu Y, Beyer A, Aebersold R. Leading Edge Review On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 2016;165:535–550.

265 Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat. Clin. Pract. Oncol. 2008;5:588–599.

266 Aho S, McLean WHI, Li K, Uitto J. cDNA cloning, mRNA expression, and chromosomal mapping of human and mouse periplakin genes. Genomics 1998;48:242–247.

267 Sonnenberg A, Liem RKH. Plakins in development and disease. Exp. Cell Res. 2007;313:2189–2203.

268 Tonoike Y, Matsushita K, Tomonaga T, Katada K, Tanaka N, Shimada H, et al. Adhesion molecule periplakin is involved in cellular movement and attachment in pharyngeal squamous cancer cells. BMC Cell Biol 2011;12:41.

269 DiColandrea T, Karashima T, Määttä A, Watt FM. Subcellular distribution of envoplakin and periplakin: Insights into their role as precursors of the epidermal cornified envelope. J Cell Biol 2000;151:573–585.

270 Pieter A, van den Heuvel J, de Vries-Smits AMM, van Weeren PC, Dijkers PF, de Bruyn KMT, et al. Binding of protein kinase B to the plakin family member periplakin. J Cell Sci 2002;115:3957–3966.

271 Nedeljković M, Damjanović A. Mechanisms of Chemotherapy Resistance in Triple-Negative Breast Cancer-How We Can Rise to the Challenge. Cells. 2019;8:957.

272 Suzuki A, Horiuchi A, Ashida T, Miyamoto T, Kashima H, Nikaido T, et al. Cyclin A2 confers cisplatin resistance to endometrial carcinoma cells via up-regulation of an Akt-binding protein, periplakin. J Cell Mol Med 2010;14:2305–2317.

273 Matsumoto K, Ikeda M, Sato Y, Kuruma H, Kamata Y, Nishimori T, et al. Loss of periplakin expression is associated with pathological stage and cancer-specific survival in patients with urothelial carcinoma of the urinary bladder. Biomed Res 2014;35:201–206.

274 Otsubo T, Hagiwara T, Tamura-Nakano M, Sezaki T, Miyake O, Hinohara C, et al. Aberrant DNA hypermethylation reduces the expression of the desmosome-related molecule periplakin in esophageal squamous cell carcinoma. Cancer Med 2015;4:415–425.

275 Nishimori T, Tomonaga T, Matsushita K, Oh-Ishi M, Kodera Y, Maeda T, et al. Proteomic analysis of primary esophageal squamous cell carcinoma reveals downregulation of a cell adhesion protein, periplakin. Proteomics 2006;6:1011–1018.

276 Li X, Zhang G, Wang Y, Elgehama A, Sun Y, Li L, et al. Loss of periplakin expression is associated with the tumorigenesis of colorectal carcinoma. Biomed

Pharmacother 2017;87:366–374.

277 Sun C, Gu Y, Chen G, Du Y. Bioinformatics Analysis of Stromal Molecular Signatures Associated with Breast and Prostate Cancer. J Comput Biol 2019;26:1130–1139.

278 Long HA, Boczonadi V, McInroy L, Goldberg M, Määttä A. Periplakin-dependent re-organisation of keratin cytoskeleton and loss of collective migration in keratin-8-downregulated epithelial sheets. J Cell Sci 2006;119:5147–5159.

279 Yang Liao GKSWS. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features | Bioinformatics | Oxford Academic. Bioinformatics 2014;30:923–930.

280 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014 1512 2014;15:1–21.

281 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B 1995;57:289–300.

282 Heng Li, Richard Durbin. Fast and accurate long-read alignment with Burrows–Wheeler transform. 2010;26:589–595.

283 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–2079.

284 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–1303.

285 John G Tate, Sally Bamford, Harry C Jubb, Zbyslaw Sondka, David M Beare, Nidhi Bindal, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer | Nucleic Acids Research | Oxford Academic. Nucleic Acids Res 2019;47:D941–D947.

286 Zhang FL, Cao JL, Xie HY, Sun R, Yang LF, Shao ZM, et al. Cancer-associated MORC2-mutant M276I regulates an hnRNPM-mediated CD44 splicing switch to promote invasion and metastasis in triple-negative breast cancer. Cancer Res 2018;78:5780–5792.

287 Lin NU, Claus E, Sohl J, Razzak AR, Arnaout A, Winer EP. Sites of distant recurrence and clinical outcomes in patients with metastatic triple-negative breast cancer: High incidence of central nervous system metastases. Cancer 2008;113:2638–2645.

288 Ocana A, Pandiella A. Targeting oncogenic vulnerabilities in triple negative breast cancer: Biological bases and ongoing clinical studies. Oncotarget. 2017;8:22218–22234.

289 Naylor TL, Greshock J, Wang Y, Colligon T, Yu QC, Clemmer V, et al. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. Breast Cancer Res 2005;7:R1186–R1198.

290 Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She QB, Maurer M, et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. Proc Natl Acad Sci U S A 2007;104:7564–7569.

291 Lee HM, Kelly GM, Zainal NS, Yee PS, Fadlullah MZH, Lee BKB, et al. The 4717C > G polymorphism in periplakin modulates sensitivity to EGFR inhibitors. Sci Rep 2019;9. doi:10.1038/s41598-019-38742-0.

292 Perez-Moreno M, Jamora C, Fuchs E. Sticky business: Orchestrating cellular signals at adherens junctions. Cell. 2003;112:535–548.

293 Silva JM, Ezhkova E, Silva J, Heart S, Castillo M, Campos Y, et al. Cyfip1 Is a Putative Invasion Suppressor in Epithelial Cancers. Cell 2009;137:1047–1061.

294 Khan MA, Jain VK, Rizwanullah M, Ahmad J, Jain K. PI3K/AKT/mTOR pathway inhibitors in triple-negative breast cancer: a review on drug discovery and future challenges. Drug Discov. Today. 2019;24:2181–2191.

295 Al-Mahmood S, Sapiezynski J, Garbuzenko OB, Minko T. Metastatic and triple-negative breast cancer: challenges and treatment options. Drug Deliv. Transl. Res. 2018;8:1483–1507.

296 Dankort D, Maslikowski B, Warner N, Kanno N, Kim H, Wang Z, et al. Grb2 and Shc Adapter Proteins Play Distinct Roles in Neu (ErbB-2)-Induced Mammary Tumorigenesis: Implications for Human Breast Cancer. Mol Cell Biol 2001;21:1540–1551.

297 Giltnane JM, Balko JM. Rationale for targeting the Ras/MAPK pathway in triple-negative breast cancer. Discov Med 2014;17:275–283.

298 Huang P, Ouyang DJ, Chang S, Li MY, Li L, Li QY, et al. Chemotherapy-driven increases in the CDKN1A/PTN/PTPRZ1 axis promote chemoresistance by activating the NF-κB pathway in breast cancer cells 11 Medical and Health Sciences 1112 Oncology and Carcinogenesis. Cell Commun Signal 2018;16. doi:10.1186/s12964-018-0304-4.

299 Bhola NE, Jansen VM, Koch JP, Li H, Formisano L, Williams JA, et al. Treatment of triple-negative breast cancer with TORC1/2 inhibitors sustains a drug-resistant and notch-dependent cancer stem cell population. Cancer Res 2016;76:440–452.

300 Chavez KJ, Garimella S V., Lipkowitz S. Triple negative breast cancer cell lines: One tool in the search for better treatment of triple negative breast cancer. Breast Dis 2010;32:35–48.

301 Kao J, Salari K, Bocanegra M, Choi Y La, Girard L, Gandhi J, et al. Molecular

profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One 2009;4. doi:10.1371/journal.pone.0006146.

302 Alsaleem M, Toss MS, Joseph C, Aleskandarany M, Kurozumi S, Alshankyty I, et al. The molecular mechanisms underlying reduced E-cadherin expression in invasive ductal carcinoma of the breast: high throughput analysis of large cohorts. Mod Pathol 2019;32:967–976.

303 Hart IR, Saini A. Biology of tumour metastasis. Lancet (London, England) 1992;339:1453–1457.

304 Berx G, Staes K, van Hengel J, Molemans F, Bussemakers MJ, van Bokhoven A, et al. Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1). Genomics 1995;26:281–289.

305 Paredes J, Figueiredo J, Albergaria A, Oliveira P, Carvalho J, Ribeiro AS, et al. Epithelial E- and P-cadherins: Role and clinical significance in cancer. Biochim Biophys Acta - Rev Cancer 2012;1826:297–311.

306 Behrens J, Mareel MM, Van Roy FM, Birchmeier W. Dissecting tumor cell invasion: epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion. J Cell Biol 1989;108:2435–2447.

307 Frixen UH, Behrens J, Sachs M, Eberle G, Voss B, Warda A, et al. E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. J Cell Biol 1991;113:173–185.

308 Sarrió D, Moreno-Bueno G, Hardisson D, Sánchez-Estévez C, Guo M, Herman JG, et al. Epigenetic and genetic alterations of APC and CDH1 genes in lobular breast cancer: relationships with abnormal E-cadherin and catenin expression and

microsatellite instability. Int J Cancer 2003;106:208–215.

309 Huiping C, Sigurgeirsdottir JR, Jonasson JG, Eiriksdottir G, Johannsdottir JT, Egilsson V, et al. Chromosome alterations and E-cadherin gene mutations in human lobular breast cancer. Br J Cancer 1999;81:1103–1110.

310 de Leeuw WJF, Berx G, Vos CBJ, Peterse JL, Van de Vijver MJ, Litvinov S, et al. Simultaneous loss of E-cadherin and catenins in invasive lobular breast cancer and lobular carcinomain situ. J Pathol 1997;183:404–411.

311 Sgeirsson KSA, Joâ Nasson JG, Tryggvadoâ Ttir L, Â Lafsdoâ Ttir KO, Sigurgeirsdoâ Ttir JR, Ingvarsson S, et al. Altered expression of E-cadherin in breast cancer: patterns, mechanisms and clinical signi®cance. [cited 12 July 2018]. Available from: https://skemman.is/bitstream/1946/21610/1/Asgeirsson2000.pdf.

312 Acs G, Lawton TJ, Rebbeck TR, LiVolsi VA, Zhang PJ. Differential Expression of E-Cadherin in Lobular and Ductal Neoplasms of the Breast and Its Biologic and Diagnostic Implications. Am J Clin Pathol 2001;115:85–98.

313 Jeschke U, Mylonas I, Kuhn C, Shabani N, Kunert-Keil C, Schindlbeck C, et al. Expression of E-Cadherin in Human Ductal Breast Cancer Carcinoma In Situ, Invasive Carcinomas, their Lymph Node Metastases, their Distant Metastases, Carcinomas with Recurrence and in Recurrence. 2007;27:1969–1974.

314 da Silva BB, dos Santos AR, Pires CG, Correa-Lima MA, Pereira-Filho JDD, dos Santos LG, et al. E-cadherin expression in estrogen receptor-positive and negative breast carcinomas of postmenopausal women. Eur J Gynaecol Oncol 2010;31:90–93.

315 Siitonen SM, Kononen JT, Helin HJ, Rantala IS, Holli KA, Isola JJ. Reduced E-cadherin expression is associated with invasiveness and unfavorable prognosis in breast cancer. Am J Clin Pathol 1996;105:394–402.

316 Mahler-Araujo B, Savage K, Parry S, Reis-Filho JS. Reduction of E-cadherin expression is associated with non-lobular breast carcinomas of basal-like and triple negative phenotype. J Clin Pathol 2008;61:615–620.

317 Lipponen P, Saarelainen E, Ji H, Aaltomaa S, Syrjänen K. Expression of E-cadherin (E-CD) as related to other prognostic factors and survival in breast cancer. J Pathol 1994;174:101–109.

318 Rakha EA, Abd El Rehim D, Pinder SE, Lewis SA, Ellis IO. E-cadherin expression in invasive non-lobular carcinoma of the breast and its prognostic significance. Histopathology 2005;46:685–693.

319 Wu Q, Wang J, Liu Y, Gong X. Epithelial cell adhesion molecule and epithelial-mesenchymal transition are associated with vasculogenic mimicry, poor prognosis, and metastasis of triple negative breast cancer. Int J Clin Exp Pathol 2019;12:1678–1689.

320 Kashiwagi S, Yashiro M, Takashima T, Nomura S, Noda S, Kawajiri H, et al. Significance of E-cadherin expression in triple-negative breast cancer. Br J Cancer 2010;103:249–255.

321 Li P, Sun T, Yuan Q, Pan G, Zhang J, Sun D. The expressions of NEDD9 and E-cadherin correlate with metastasis and poor prognosis in triple-negative breast cancer patients. Onco Targets Ther 2016;9:5751–5759.

322 McCart Reed AE, Kutasovic JR, Lakhani SR, Simpson PT. Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics. Breast Cancer Res 2015;17:12.

323 Gillett CE, Miles DW, Ryder K, Skilton D, Liebman RD, Springall RJ, et al. Retention of the expression of E-cadherin and catenins is associated with shorter survival in grade III ductal carcinoma of the breast. J Pathol 2001;193:433–441.

324 Korkola JE, DeVries S, Fridlyand J, Hwang ES, Estep ALH, Chen Y-Y, et al. Differentiation of lobular versus ductal breast carcinomas by expression microarray analysis. Cancer Res 2003;63:7167–7175.

325 Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. Nat Genet 2001;27:247–254.

326 Bukholm IK, Nesland JM, Kåresen R, Jacobsen U, Børresen-Dale AL. Expression of E-cadherin and its relation to the p53 protein status in human breast carcinomas. Virchows Arch 1997;431:317–321.

327 Debies MT, Welch DR. Genetic basis of human breast cancer metastasis. J Mammary Gland Biol Neoplasia 2001;6:441–451.

328 Strathdee G. Epigenetic versus genetic alterations in the inactivation of E-cadherin. Semin Cancer Biol 2002;12:373–379.

329 Graff JR, Herman JG, Lapidus RG, Chopra H, Xu R, Jarrard DF, et al. E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. Cancer Res 1995;55:5195–5199.

330 Batlle E, Sancho E, Francí C, Domínguez D, Monfar M, Baulida J, et al. The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nat Cell Biol 2000;2:84–89.

331 Oka H, Shiozaki H, Kobayashi K, Inoue M, Tahara H, Kobayashi T, et al. Expression of E-cadherin cell adhesion molecules in human breast cancer tissues and its relationship to metastasis. Cancer Res 1993;53:1696–1701.

332 Cleton-Jansen AM, Callen DF, Seshadri R, Goldup S, Mccallum B, Crawford J, et al. Loss of heterozygosity mapping at chromosome arm 16q in 712 breast tumors reveals factors that influence delineation of candidate regions. Cancer Res 2001;61:1171–

1177.

333 Cheng C-W, Wu P-E, Yu J-C, Huang C-S, Yue C-T, Wu C-W, et al. Mechanisms of inactivation of E-cadherin in breast carcinoma: modification of the two-hit hypothesis of tumor suppressor gene. Oncogene 2001;20:3814–3823.

334 Berx G, Cleton-Jansen AM, Strumane K, de Leeuw WJ, Nollet F, van Roy F, et al. E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. Oncogene 1996;13:1919–1925.

335 Li J, Wang K, Li S, Timmermans-Wielenga V, Rank F, Wiuf C, et al. DNA copy number aberrations in breast cancer by array comparative genomic hybridization. Genomics Proteomics Bioinformatics 2009;7:13–24.

336 Caldeira JRF, Prando ÉC, Quevedo FC, Neto FAM, Rainho CA, Rogatto SR. CDH1promoter hypermethylation and E-cadherin protein expression in infiltrating breast cancer. BMC Cancer 2006;6:48.

337 Sengodan SK, K.H. S, Nadhan R, Srinivas P. Regulation of epithelial to mesenchymal transition by BRCA1 in breast cancer. Crit Rev Oncol Hematol 2018;123:74–82.

338 Yasmeen A, Liu W, Dekhil H, Kassab A, Aloyz R, Foulkes WD, et al. BRCA1 mutations contribute to cell motility and invasion by affecting its main regulators. Cell Cycle 2008;7:3781–3783.

339 Becker KF, Atkinson MJ, Reich U, Becker I, Nekarda H, Siewert JR, et al. E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. Cancer Res 1994;54:3845–3852.

340 Berx G, Cleton-Jansen AM, Nollet F, de Leeuw WJ, van de Vijver M, Cornelisse C, et al. E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast

cancers. EMBO J 1995;14:6107–6115.

341 Cleton-Jansen A-M. E-cadherin and loss of heterozygosity at chromosome 16 in breast carcinogenesis: different genetic pathways in ductal and lobular breast cancer? Breast Cancer Res 2002;4:5–8.

342 Spaderna S, Schmalhofer O, Wahlbuhl M, Dimmler A, Bauer K, Sultan A, et al. The Transcriptional Repressor ZEB1 Promotes Metastasis and Loss of Cell Polarity in Cancer. Cancer Res 2008;68:537–544.

343 Aigner K, Dampier B, Descovich L, Mikula M, Sultan A, Schreiber M, et al. The transcription factor ZEB1 (δEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. Oncogene 2007;26:6979–6988.

344 Foulstone E, Prince S, Zaccheo O, Burns J, Harper J, Jacobs C, et al. Insulin-like growth factor ligands, receptors, and binding proteins in cancer. J Pathol 2005;205:145–153.

345 Irie HY, Pearline R V, Grueneberg D, Hsia M, Ravichandran P, Kothari N, et al. Distinct roles of Akt1 and Akt2 in regulating cell migration and epithelial-mesenchymal transition. J Cell Biol 2005;171:1023–1034.

346 Bienz M, Clevers H. Linking colorectal cancer to Wnt signaling. Cell 2000;103:311–320.

347 Prasad CP, Gupta SD, Rath G, Ralhan R. Wnt signaling pathway in invasive ductal carcinoma of the breast: relationship between beta-catenin, dishevelled and cyclin D1 expression. Oncology 2007;73:112–117.

348 Prasad CP, Mirza S, Sharma G, Prashad R, DattaGupta S, Rath G, et al. Epigenetic alterations of CDH1 and APC genes: Relationship with activation of Wnt/β-catenin Pathway in invasive ductal carcinoma of breast. Life Sci 2008;83:318–325.

349 Liu J-B, Feng C-Y, Deng M, Ge D-F, Liu D-C, Mi J-Q, et al. E-cadherin expression phenotypes associated with molecular subtypes in invasive non-lobular breast cancer: evidence from a retrospective study and meta-analysis. World J Surg Oncol 2017;15:139.

350 Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. Briefings Funct Genomics Proteomics 2009;8:174–183.

351 Wolff A, Bayerlová M, Gaedcke J, Kube D, Beißbarth T. A comparative study of RNA-Seq and microarray data analysis on the two examples of rectal-cancer patients and Burkitt Lymphoma cells. PLoS One 2018;13:e0197162.

352 Testa U, Castelli G, Pelosi E. Breast Cancer: A Molecularly Heterogenous Disease Needing Subtype-Specific Treatments. Med Sci 2020;8:18.

353 Santuario-Facio SK, Cardona-Huerta S, Perez-Paramo YX, Trevino V, Hernandez-Cabrera F, Rojas-Martinez A, et al. A New Gene Expression Signature for Triple Negative Breast Cancer Using Frozen Fresh Tissue before Neoadjuvant Chemotherapy. Mol Med 2017;23:101.

354 Vishnoi M, Liu NH, Yin W, Boral D, Scamardo A, Hong D, et al. The identification of a TNBC liver metastasis gene signature by sequential CTC-xenograft modeling. Mol Oncol 2019;13:1913–1926.

355 de Ruijter TC, Veeck J, de Hoon JPJ, van Engeland M, Tjan-Heijnen VC. Characteristics of triple-negative breast cancer. J Cancer Res Clin Oncol 2011;137:183–192.

356 Singh A, Trivedi P, Jain NK. Advances in siRNA delivery in cancer therapy. Artif Cells, Nanomedicine, Biotechnol 2018;46:274–283.

357 Hoelder S, Clarke PA, Workman P. Discovery of small molecule cancer drugs:

Successes, challenges and opportunities. Mol. Oncol. 2012;6:155–176.

358 Fakhr E, Zare F, Teimoori-Toolabi L. Precise and efficient siRNA design: A key point in competent gene silencing. Cancer Gene Ther. 2016;23:73–82.

359 Zhao X, Li F, Li Y, Wang H, Ren H, Chen J, et al. Co-delivery of HIF1α siRNA and gemcitabine via biocompatible lipid-polymer hybrid nanoparticles for effective treatment of pancreatic cancer. Biomaterials 2015;46:13–25.

360 Parvani JG, Gujrati MD, Mack MA, Schiemann WP, Lu ZR. Silencing β3 integrin by targeted ECO/siRNA nanoparticles inhibits EMT and metastasis of triple-negative breast cancer. Cancer Res 2015;75:2316–2325.

361 Su S, Tian Y, Li Y, Ding Y, Ji T, Wu M, et al. “triple-punch” strategy for triple negative breast cancer therapy with minimized drug dosage and improved antitumor efficacy. ACS Nano 2015;9:1367–1378.

362 Zhang X, Wang Q, Qin L, Fu H, Fang Y, Han B, et al. EGF-modified mPEG-PLGA-PLL nanoparticle for delivering doxorubicin combined with Bcl-2 siRNA as a potential treatment strategy for lung cancer. Drug Deliv 2016;23:2936–2945.

363 Liu JY, Chiang T, Liu CH, Chern GG, Lin TT, Gao DY, et al. Delivery of siRNA Using CXCR4-targeted nanoparticles modulates tumor microenvironment and achieves a potent antitumor response in liver cancer. Mol Ther 2015;23:1772–1782.

364 Jackson AL, Bartz SR, Schelter J, Kobayashi S V., Burchard J, Mao M, et al. Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol 2003;21:635–637.

365 Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for RNA interference. Nat Biotechnol 2004;22:326–330.

366 Ozcan G, Ozpolat B, Coleman RL, Sood AK, Lopez-Berestein G. Preclinical and

clinical development of siRNA-based therapeutics. Adv. Drug Deliv. Rev. 2015;87:108–119.

367 Steinmetz KL, Spack EG. The basics of preclinical drug development for neurodegenerative disease indications. BMC Neurol 2009;9:S2.

368 Jurrius P, Green T, Garmo H, Young M, Cariati M, Gillett C, et al. Invasive breast cancer over four decades reveals persisting poor metastatic outcomes in treatment resistant subgroup – the "ATRESS" phenomenon. Breast 2020;50:39–48.

369 Fazilaty H, Gardaneh M, Bahrami T, Salmaninejad A, Behnam B. Crosstalk between breast cancer stem cells and metastatic niche: Emerging molecular metastasis pathway? Tumor Biol. 2013;34:2019–2030.

370 Jitariu AA, Cîmpean AM, Ribatti D, Raica M. Triple negative breast cancer: The kiss of death. Oncotarget. 2017;8:46652–46662.

371 Wang Y, Savage K, Grills C, McCavigan A, James JA, Fennell DA, et al. A TMA de-arraying method for high throughput biomarker discovery in tissue research. PLoS One 2011;6. doi:10.1371/journal.pone.0026007.

372 Hewitt SM. Tissue microarrays as a tool in the discovery and validation of predictive biomarkers. Methods Mol Biol 2012;823:201–214.

373 Oyama T, Ishikawa Y, Hayashi M, Arihiro K, Horiguchi J. The effects of fixation, processing and evaluation criteria on immunohistochemical detection of hormone receptors in breast cancer. Breast Cancer. 2007;14:182–188.

Appendices

7.1 Chapter 1: Copyright approval



29/12/2020 https://marketplace.copyright.com/rs-ui-web/mp/license/2ea48359-7ffb-400b-a9fc-a2aafde256a5/0584df68-27ad-4a99-8903-b5950...

Copyright Clearance Center Marketplace™

Elsevier Science & Technology Journals - License Terms and Conditions

This is a License Agreement between Mansour Alsaleem ("You") and Elsevier Science & Technology Journals ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier Science & Technology Journals, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	28-Dec-2020	Type of Use	Republish in a thesis/dissertation
Order license ID	1086850-1		
ISSN	1097-4172	Publisher	Cell Press; M I T PRESS
		Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Cell	Publication Type	e-Journal
Article Title	Hallmarks of cancer: the next generation.	Start Page	646
		End Page	674
Date	01/01/1974	Issue	5
Language	English	Volume	144
Country	United States of America	URL	http://www.elsevier.com/journals/cell/0092-8674
Rightsholder	Elsevier Science & Technology Journals		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	Distribution	U.K. and Commonwealth (excluding Canada)
Number of charts / graphs / tables / figures requested	1	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the disabled?	No
		Minor editing privileges?	No
Who will republish the content?	Academic institution	Incidental promotional use?	No
Duration of Use	Life of current edition	Currency	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Integrative Molecular and Pathologic Profiling of Triple Negative Breast Cancer	Institution name	University of Nottingham
		Expected presentation date	2021-12-07
Instructor name	Mansour Alsaleem		

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	Mansour Alsaleem

REUSE CONTENT DETAILS

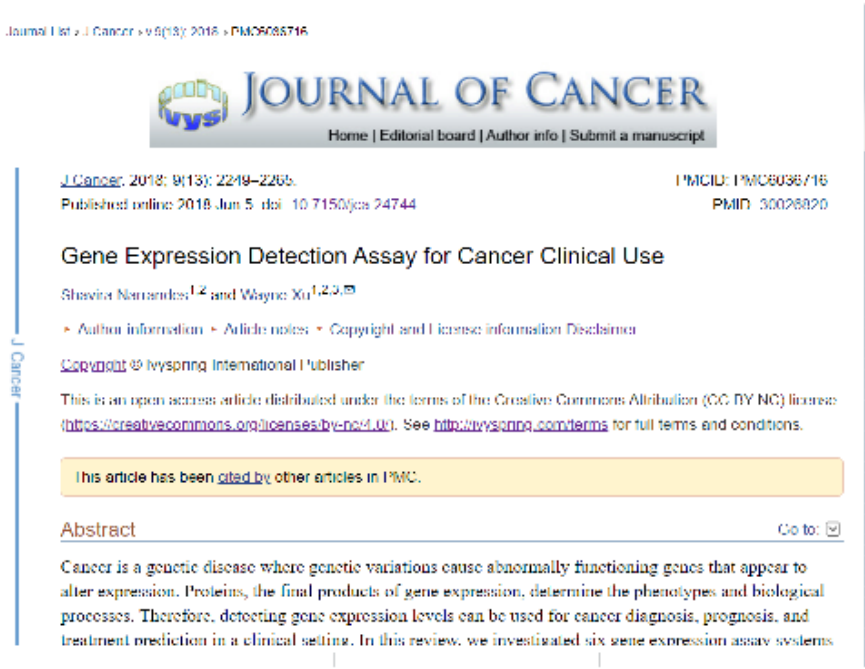
https://marketplace.copyright.com/rs-ui-web/mp/license/2ea48359-7ffb-400b-a9fc-a2aafde256a5/0584df68-27ad-4a99-8903-b5950fae2271 1/4

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jan 04, 2021

This Agreement between Mr. Mansour Alsaleem ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4982150879277
License date	Jan 04, 2021
Licensed Content Publisher	Elsevier
Licensed Content Publication	The American Journal of Pathology
Licensed Content Title	The Spectrum of Triple-Negative Breast Disease High- and Low-Grade Lesions
Licensed Content Author	Felipe C. Geyer,Fresia Pareja,Britta Weigelt,Emad Rakha,Ian O. Ellis,Stuart J. Schnitt,Jorge S. Reis-Filho
Licensed Content Date	Oct 1, 2017
Licensed Content Volume	187
Licensed Content Issue	10
Licensed Content Pages	13
Start Page	2139
End Page	2151
Type of Use	reuse in a thesis/dissertation



Journal List > J Cancer > v.9(13); 2018 > PMC6036716

JOURNAL OF CANCER

Home | Editorial board | Author info | Submit a manuscript

J Cancer. 2018; 9(13): 2249–2265. PMCID: PMC6036716
Published online 2018 Jun 5. doi: 10.7150/jca.24744 PMID: 30026920

Gene Expression Detection Assay for Cancer Clinical Use

Shawna Hernandez[1,2] and Wayne Xu[1,2,3]✉

▸ Author information ▸ Article notes ▾ Copyright and License information Disclaimer

Copyright © Ivyspring International Publisher

This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY NC) license (https://creativecommons.org/licenses/by-nc/4.0/). See http://ivyspring.com/terms for full terms and conditions.

This article has been cited by other articles in PMC.

Abstract

Go to:

Cancer is a genetic disease where genetic variations cause abnormally functioning genes that appear to alter expression. Proteins, the final products of gene expression, determine the phenotypes and biological processes. Therefore, detecting gene expression levels can be used for cancer diagnosis, prognosis, and treatment prediction in a clinical setting. In this review, we investigated six gene expression assay systems



IVYSPRING
INTERNATIONAL PUBLISHER

Home | About | Journals | Author | Contact

License & Copyright

New articles in our journals are distributed under the Creative Commons Attribution License (CC-BY license). The Open Access statement is: "This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). See http://ivyspring.com/terms for full terms and conditions." Authors retain copyright of the article, and grant Ivyspring International Publisher a license to publish, use, reproduce and distribute the article and the material it contains. Therein Ivyspring International Publisher shall not be liable for the world in all languages and in all media (now or hereafter known). This is applicable to articles published after June 10, 2015.

For articles published before June 10, 2015, the Creative Commons Attribution Non-Commercial 4.0 License was used by default and copyright is held by the publisher. Please refer to each individual article's copyright and open access statement.

Permission request

Request to reproduce figures or tables in other publications or journals: There is no need to obtain permission if copyrights of the figures or tables belong to Ivyspring International Publisher. Please acknowledge the original source of publication in our journal with full citation. If the copyrights of the figures or tables belong to other publishers or authors, please contact the original publishers or authors for permissions.

Authors or co-authors: As an author or co-author of the article, you are free to include or reuse any part of the materials in your thesis or other publications that you write, with citation of the original source in our journal. There is no need to obtain permission.

Terms of use

By accessing this web site, downloading, printing, or reading any article published in a journal of Ivyspring International Publisher (Journal), you are deemed that you agree to all of the following terms and conditions:

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Dec 29, 2020

This Agreement between Mr. Mansour Alsaleem ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4977400868468
License date	Dec 27, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	The EMBO Journal
Licensed Content Title	A new genome-driven integrated classification of breast cancer and its implications
Licensed Content Author	Carlos Caldas, Samuel Aparicio, Oscar M Rueda, et al
Licensed Content Date	Feb 8, 2013
Licensed Content Volume	32
Licensed Content Issue	5
Licensed Content Pages	12
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Dec 29, 2020

This Agreement between Mr. Mansour Alsaleem ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4977390920271
License date	Dec 27, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Histopathology
Licensed Content Title	Breast cancer precursors revisited: molecular features and progression pathways
Licensed Content Author	Maria A Lopez-Garcia, Felipe C Geyer, Magali Lacroix-Triki, et al
Licensed Content Date	Aug 19, 2010
Licensed Content Volume	57
Licensed Content Issue	2
Licensed Content Pages	7
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic

Journal List > Oncotarget > v.6(15); 2015 May 30 > PMC4536987

Oncotarget
Open Access Impact Journal

Oncotarget. 2015 May 30; 6(15): 12890–12908.
Published online 2015 May 7. doi: 10.18632/oncotarget.3849

PMCID: PMC4536987
PMID: 25973541

Is the future of personalized therapy in triple-negative breast cancer based on molecular subtype?

Fanny Le Du,[1,2] Bedrich L. Eckhardt,[1] Bora Lim,[1,6] Jennifer K. Litton,[1,3,6] Stacy Moulder,[1,6] Funda Meric-Bernstam,[4,6] Ana M. Gonzalez-Angulo,[1,5,6] and Naoto T. Ueno[1,6]

▸ Author information ▸ Article notes ▾ Copyright and License information Disclaimer

Copyright : © 2015 Le Du et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

7.2 Chapter 3: Identification of potential prognostic genes in TNBC

Authors' list

Mansour A. Alsaleem (MAA) , Graham Ball (GB) , Michael S. Toss (MS), Sara Raafat (SR) , Mohammed Aleskandarany (MA) , Chitra Joseph (CJ), Angela Ogden (AO), Shristi Bhattarai (SB), Padmashree C. G. Rida (PR), Francesca Khani (FK), Melissa Davis (MD), Olivier Elemento (OE), Ritu Aneja (RA), Ian O. Ellis (IE), Andrew Green(AR), Nigel P. Mongan (NPM) & Emad Rakha (ER)

SPRINGER NATURE

A novel prognostic two-gene signature for triple negative breast cancer

Author: Mansour A. Alsaleem et al
Publication: Modern Pathology
Publisher: Springer Nature
Date: May 13, 2020

Copyright © 2020, The Author(s), under exclusive licence to United States & Canadian Academy of Pathology

Author Request

If you are the author of this content (or his/her designated agent) please read the following. If you are not the author of this content, please click the Back button and select no to the question "Are you the Author of this Springer Nature content?".
Ownership of copyright in original research articles remains with the Author, and provided that, when reproducing the contribution or extracts from it or from the Supplementary Information, the Author acknowledges first and reference publication in the Journal, the Author retains the following non-exclusive rights:
To reproduce the contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).
The author and any academic institution, where they work, at the time may reproduce the contribution for the purpose of course teaching.
To reuse figures or tables created by the Author and contained in the Contribution in oral presentations and other works created by them.
To post a copy of the contribution as accepted for publication after peer review (in locked Word processing file, of a PDF version thereof) on the Author's own web site, or the Author's institutional repository, or the Author's funding body's archive, six months after publication of the printed or online edition of the Journal, provided that they also link to the contribution on the publisher's website.
Authors wishing to use the published version of their article for promotional use or on a web site must request in the normal way.

If you require further assistance please read Springer Nature's online author reuse guidelines.

For full paper portion: Authors of original research papers published by Springer Nature are encouraged to submit the author's version of the accepted, peer-reviewed manuscript to their relevant funding body's archive, for release six months after publication. In addition, authors are encouraged to archive their version of the manuscript in their institution's repositories (as well as their personal Web sites), also six months after original publication.

v1.0

BACK

CLOSE WINDOW

Table 1: Genes associated with Distant Metastasis Free Survival (DMFS) identified by Artificial Neural Network analysis (Transcriptomic Cohort, n=112)

Gene Id	initial_alias	description
AANAT	ENSG00000129673	aralkylamine N-acetyltransferase [Source: HGNC Symbol;Acc:HGNC:19]
AC004893.10	ENSG00000238109	ring finger protein 14(RNF14) pseudogene
AC005042.4	ENSG00000204380	PKP4 antisense RNA 1 [Source: HGNC Symbol; Acc:HGNC:52580]
AC007750.5	ENSG00000236841	novel transcript
AC012358.4	ENSG00000227799	pseudogene similar to RIKEN cDNA 2210021J22
AC020931.1	ENSG00000257110	novel transcript
AC064852.4	ENSG00000241409	novel transcript
AC068491.3	ENSG00000223973	platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa (PAFAH1B1) pseudogene
AC079305.10	ENSG00000222043	novel transcript
AC079807.2	ENSG00000233230	novel transcript
AC084809.2	ENSG00000226377	novel transcript
AC093724.2	ENSG00000213222	translocase of outer mitochondrial membrane 40 (TOMM40) pseudogene
AC094019.4	ENSG00000236732	ribosomal protein L21 (RPL21) pseudogene
AC098824.6	ENSG00000232202	istidine-rich domain (CHORD)-containing 1 (CHORDC1) pseudogene
ACSM4	ENSG00000215009	acyl-CoA synthetase medium chain family member 4 [Source: HGNC Symbol;Acc:HGNC:32016]
ACSM5	ENSG00000183558	Acyl-CoA Synthetase Medium Chain Family Member 5
ANO1-AS1	ENSG00000254902	ANO1 antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:40016]
AP001062.7	ENSG00000184441	novel transcript, antisense to C21orf2
AP001189.4	ENSG00000236304	uncharacterized LOC107984360 [Source: NCBI gene;Acc:107984360]
AP004290.1	ENSG00000236583	insulin-like growth factor 2 mRNA binding protein 2 (IGF2BP2) pseudogene
ARMS2	ENSG00000254636	age-related maculopathy susceptibility 2 [Source: HGNC Symbol;Acc:HGNC:32685]

ASIP	ENSG00000101440	agouti signaling protein [Source: HGNC Symbol;Acc:HGNC:745]
ATP6V1G1P4	ENSG00000233346	novel transcript
BANF1P4	ENSG00000223828	barrier to autointegration factor 1 pseudogene 4 [Source: HGNC Symbol;Acc:HGNC:43884]
BICC1	ENSG00000122870	BicC family RNA binding protein 1 [Source: HGNC Symbol;Acc:HGNC:19351]
BRD7P4	ENSG00000218676	bromodomain containing 7 pseudogene 4 [Source: HGNC Symbol;Acc:HGNC:37630]
BRI3P1	ENSG00000225169	brain protein I3 pseudogene 1 [Source: HGNC Symbol;Acc:HGNC:33533]
C14orf1	ENSG00000133935	ergosterol biosynthesis 28 homolog [Source: HGNC Symbol;Acc:HGNC:1187]
C1orf204	ENSG00000188004	small nucleolar RNA host gene 28 [Source: HGNC Symbol;Acc:HGNC:27647]
CBX3P5	ENSG00000257666	chromobox 3 pseudogene 5 [Source: HGNC Symbol;Acc:HGNC:42877]
CCDC54	ENSG00000138483	coiled-coil domain containing 54 [Source: HGNC Symbol;Acc:HGNC:30703]
CCDC6	ENSG00000108091	coiled-coil domain containing 6 [Source: HGNC Symbol;Acc:HGNC:18782]
CEACAM6	ENSG00000086548	carcinoembryonic antigen related cell adhesion molecule 6 [Source: HGNC Symbol;Acc:HGNC:1818]
CEMP1	ENSG00000205923	cementum protein 1 [Source: HGNC Symbol;Acc:HGNC:32553]
CEP164	ENSG00000110274	centrosomal protein 164 [Source: HGNC Symbol;Acc:HGNC:29182]
CEP89	ENSG00000121289	centrosomal protein 89 [Source: HGNC Symbol;Acc:HGNC:25907]
CLUHP2	ENSG00000228947	novel transcript
COL5A1-AS1	ENSG00000204011	COL5A1 antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:31368]
COL5A2	ENSG00000204262	collagen type V alpha 2 chain [Source: HGNC Symbol;Acc:HGNC:2210]
CRYM	ENSG00000103316	crystallin mu [Source: HGNC Symbol;Acc:HGNC:2418]
CTB-33G10.1	ENSG00000243829	ribosomal protein S9 (RPS9) pseudogene
CTB-47B11.3	ENSG00000248544	novel transcript, antisense to CYFIP2
CTSZ	ENSG00000101160	cathepsin Z [Source: HGNC Symbol;Acc:HGNC:2547]
CUBNP3	ENSG00000235690	cubilin pseudogene 3 [Source: HGNC Symbol;Acc:HGNC:44985]
DBNL	ENSG00000136279	drebrin like [Source: HGNC Symbol;Acc:HGNC:2696]

DCC	ENSG00000187323	DCC netrin 1 receptor [Source: HGNC Symbol;Acc:HGNC:2701]
DCTN1-AS1	ENSG00000237737	DCTN1 antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:44151]
DHODH	ENSG00000102967	dihydroorotate dehydrogenase (quinone) [Source: HGNC Symbol;Acc:HGNC:2867]
DNAH3	ENSG00000158486	dynein axonemal heavy chain 3 [Source: HGNC Symbol;Acc:HGNC:2949]
DOCK10	ENSG00000135905	dedicator of cytokinesis 10 [Source: HGNC Symbol;Acc:HGNC:23479]
DSP	ENSG00000096696	desmoplakin [Source:HGNC Symbol;Acc:HGNC:3052]
EIF5B	ENSG00000158417	eukaryotic translation initiation factor 5B [Source: HGNC Symbol;Acc:HGNC:30793]
ERP44	ENSG00000023318	endoplasmic reticulum protein 44 [Source: HGNC Symbol;Acc:HGNC:18311]
EWSAT1	ENSG00000212768	novel transcript
EXOSC3P1	ENSG00000229007	exosome component 3 pseudogene 1 [Source: HGNC Symbol;Acc:HGNC:33989]
FAM138E	ENSG00000248894	novel transcript
FAM196A	ENSG00000188916	inhibitory synaptic factor 2A [Source: HGNC Symbol;Acc:HGNC:33859]
FAM212A	ENSG00000185614	inka box actin regulator 1 [Source: HGNC Symbol;Acc:HGNC:32480]
FGD1	ENSG00000102302	FYVE, RhoGEF and PH domain containing 1 [Source: HGNC Symbol;Acc:HGNC:3663]
FGL2	ENSG00000127951	fibrinogen like 2 [Source: HGNC Symbol;Acc:HGNC:3696]
FOXQ1	ENSG00000164379	forkhead box Q1 [Source: HGNC Symbol;Acc:HGNC:20951]
FUCA2	ENSG00000001036	alpha-L-fucosidase 2 [Source: HGNC Symbol;Acc:HGNC:4008]
GAPDHP70	ENSG00000249489	glyceraldehyde 3 phosphate dehydrogenase pseudogene 70 [Source: HGNC Symbol;Acc:HGNC:4148]
GBP7	ENSG00000213512	guanylate binding protein 7 [Source: HGNC Symbol;Acc:HGNC:29606]
GFOD1-AS1	ENSG00000237786	GFOD1 antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:40956]
GPD1	ENSG00000167588	glycerol-3-phosphate dehydrogenase 1 [Source: HGNC Symbol;Acc:HGNC:4455]
GPR132	ENSG00000183484	G protein-coupled receptor 132 [Source: HGNC Symbol;Acc:HGNC:17482]
GTF3C6	ENSG00000155130	general transcription factor IIIC subunit 6 [Source: HGNC Symbol;Acc:HGNC:20872]

HARBI1	ENSG00000180423	harbinger transposase derived 1 [Source: HGNC Symbol;Acc:HGNC:26522]
HEPHL1	ENSG00000181333	hephaestin like 1 [Source: HGNC Symbol;Acc:HGNC:30477]
HIGD2A	ENSG00000146066	HIG1 hypoxia inducible domain family member 2A [Source: HGNC Symbol;Acc:HGNC:28311]
HIRIP3	ENSG00000149929	HIRA interacting protein 3 [Source: HGNC Symbol;Acc:HGNC:4917]
HLA-F-AS1	ENSG00000214922	HLA-F antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:26645]
HNRNPA1P30	ENSG00000233780	heterogeneous nuclear ribonucleoprotein A1 pseudogene 30 [Source: HGNC Symbol;Acc:HGNC:39548]
HNRNPA1P44	ENSG00000249271	heterogeneous nuclear ribonucleoprotein A1 pseudogene 44 [Source: HGNC Symbol;Acc:HGNC:48774]
HSP90B1	ENSG00000166598	heat shock protein 90 beta family member 1 [Source: HGNC Symbol;Acc:HGNC:12028]
IGFL1	ENSG00000188293	IGF like family member 1 [Source: HGNC Symbol;Acc:HGNC:24093]
IGHG3	ENSG00000211897	immunoglobulin heavy constant gamma 3 (G3m marker) [Source: HGNC Symbol;Acc:HGNC:5527]
IPO9-AS1	ENSG00000231871	IPO9 antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:40892]
IQCC	ENSG00000160051	IQ motif containing C [Source: HGNC Symbol;Acc:HGNC:25545]
KIF11	ENSG00000138160	kinesin family member 11 [Source: HGNC Symbol;Acc:HGNC:6388]
KRT18P57	ENSG00000215867	keratin 18 pseudogene 57 [Source: HGNC Symbol;Acc:HGNC:48884]
LEF1-AS1	ENSG00000232021	LEF1 antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:40339]
LILRA6	ENSG00000244482	leukocyte immunoglobulin like receptor A6 [Source: HGNC Symbol;Acc:HGNC:15495]
LOXL4	ENSG00000138131	lysyl oxidase like 4 [Source: HGNC Symbol;Acc:HGNC:17171]
LRRC45	ENSG00000169683	leucine rich repeat containing 45 [Source: HGNC Symbol;Acc:HGNC:28302]
MAP3K14	ENSG00000006062	mitogen-activated protein kinase kinase kinase 14 [Source: HGNC Symbol;Acc:HGNC:6853]
MARCKSL1	ENSG00000175130	MARCKS like 1 [Source: HGNC Symbol;Acc:HGNC:7142]
MARS	ENSG00000166986	methionyl-tRNA synthetase [Source: HGNC Symbol;Acc:HGNC:6898]
METTL1	ENSG00000037897	methyltransferase like 1 [Source: HGNC Symbol;Acc:HGNC:7030]

MEX3A	ENSG00000254726	mex-3 RNA binding family member A [Source: HGNC Symbol;Acc:HGNC:33482]
MOGAT1	ENSG00000124003	monoacylglycerol O-acyltransferase 1 [Source: HGNC Symbol;Acc:HGNC:18210]
MSRB3	ENSG00000174100	novel transcript
MT-CO1	ENSG00000198804	mitochondrially encoded cytochrome c oxidase I [Source: HGNC Symbol;Acc:HGNC:7419]
MTRNR2L10	ENSG00000256048	novel transcript
MUC5AC	ENSG00000215182	mucin 5AC, oligomeric mucus/gel-forming [Source: HGNC Symbol;Acc:HGNC:7515]
NANOGP2	ENSG00000228670	Nanog homeobox pseudogene 2 [Source: HGNC Symbol;Acc:HGNC:23100]
NDUFA4L2	ENSG00000185633	NDUFA4, mitochondrial complex associated like 2 [Source: HGNC Symbol;Acc:HGNC:29836]
NIFK-AS1	ENSG00000236859	NIFK antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:27385]
NOP56P3	ENSG00000257956	NOP56 ribonucleoprotein pseudogene 3 [Source: HGNC Symbol;Acc:HGNC:49801]
NRM	ENSG00000137404	nurim [Source: HGNC Symbol;Acc:HGNC:8003]
NSRP1P1	ENSG00000235614	novel transcript
NUTF2P2	ENSG00000258300	nuclear transport factor 2 pseudogene 2 [Source: HGNC Symbol;Acc:HGNC:19934]
OR10AD1	ENSG00000172640	olfactory receptor family 10 subfamily AD member 1 [Source: HGNC Symbol;Acc:HGNC:14819]
OR13G1	ENSG00000197437	olfactory receptor family 13 subfamily G member 1 [Source: HGNC Symbol;Acc:HGNC:14999]
OR4C46	ENSG00000185928	novel transcript
OTUB2	ENSG00000089723	OTU deubiquitinase, ubiquitin aldehyde binding 2 [Source: HGNC Symbol;Acc:HGNC:20351]
OTUD4P1	ENSG00000118976	OTUD4 pseudogene 1 [Source: HGNC Symbol;Acc:HGNC:33912]
PABPC1L	ENSG00000101104	poly(A) binding protein cytoplasmic 1 like [Source: HGNC Symbol;Acc:HGNC:15797]
PAXBP1-AS1	ENSG00000238197	PAXBP1 antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:39603]

PCDHGC5	ENSG00000240764	protocadherin gamma subfamily C, 5 [Source: HGNC Symbol;Acc:HGNC:8718]
PCOLCE-AS1	ENSG00000224729	PCOLCE antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:40430]
PHTF1	ENSG00000116793	putative homeodomain transcription factor 1 [Source: HGNC Symbol;Acc:HGNC:8939]
PIK3CD-AS1	ENSG00000179840	PIK3CD antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:32346]
PP7080	ENSG00000188242	uncharacterized LOC25845 [Source: NCBI gene;Acc:25845]
PPL	ENSG00000118898	periplakin [Source: HGNC Symbol;Acc:HGNC:9273]
PRPS1	ENSG00000147224	phosphoribosyl pyrophosphate synthetase 1 [Source: HGNC Symbol;Acc:HGNC:9462]
PRR13P1	ENSG00000232824	proline rich 13 pseudogene 1 [Source: HGNC Symbol;Acc:HGNC:50614]
PRSS8	ENSG00000052344	serine protease 8 [Source: HGNC Symbol;Acc:HGNC:9491]
PSMB8-AS1	ENSG00000204261	PSMB8 antisense RNA 1 (head to head) [Source: HGNC Symbol;Acc:HGNC:39758]
RASA4CP	ENSG00000228903	RAS p21 protein activator 4C, pseudogene [Source: HGNC Symbol;Acc:HGNC:44185]
RELL1	ENSG00000181826	RELT like 1 [Source: HGNC Symbol;Acc:HGNC:27379]
RLIMP1	ENSG00000229456	ring finger protein, LIM domain interacting pseudogene 1 [Source: HGNC Symbol;Acc:HGNC:39682]
RNF217	ENSG00000146373	ring finger protein 217 [Source: HGNC Symbol;Acc:HGNC:21487]
RNU6V	ENSG00000206832	RNA, U6 small nuclear variant sequence with SNRPE pseudogene sequence [Source: HGNC Symbol;Acc:HGNC:10230]
RP1-310O13.7	ENSG00000226239	novel transcript
RP11-101O6.2	ENSG00000234937	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7 (Mov34 homolog) (PSMD7) pseudogene
RP11-1379J22.2	ENSG00000244951	novel transcript
RP11-153N17.1	ENSG00000233961	novel transcript
RP11-15B17.1	ENSG00000245322	novel transcript
RP11-178L8.1	ENSG00000243705	ribosomal protein L39 (RPL39) pseudogene

RP11-214D15.2	ENSG00000227175	novel transcript
RP11-228B15.4	ENSG00000225032	novel transcript
RP11-23J18.1	ENSG00000258352	interferon induced transmembrane protein 3 (1-8U) (IFITM3) pseudogene
RP11-254B13.4	ENSG00000234040	ribosomal protein L10 pseudogene 12 [Source :HGNC Symbol;Acc:HGNC:52345]
RP11-281O15.7	ENSG00000253144	pseudogene similar to part of cold shock domain containing E1, RNA-binding CSDE1
RP11-293A21.2	ENSG00000248340	family with sequence similarity 64, member A (FAM64A) pseudogene
RP11-29H23.5	ENSG00000246203	novel pseudogene
RP11-307L3.4	ENSG00000233368	novel transcript
RP11-342F17.1	ENSG00000213755	ribosomal protein L29 (RPL29) pseudogene
RP11-351I21.7	ENSG00000254423	ubiquitin specific peptidase 17-like 2 (USP17L2) pseudogene
RP11-386G21.1	ENSG00000253976	novel transcript
RP11-397E7.4	ENSG00000251411	actin related protein 2/3 complex, subunit 1A, 41kDa (ARPC1A) pseudogene
RP11-409C19.2	ENSG00000253223	PRP3 pre-mRNA processing factor 3 homolog (S. cerevisiae) (PRPF3) pseudogene
RP11-460I13.2	ENSG00000227050	novel transcript
RP11-473M20.5	ENSG00000205890	novel transcript
RP11-473N11.2	ENSG00000256238	SPT16 homolog, facilitates chromatin remodeling subunit pseudogene 1 [Source: HGNC Symbol;Acc:HGNC:31388]
RP11-530C5.1	ENSG00000258048	novel transcript
RP11-603J24.7	ENSG00000237493	RAB13 member RAS oncogene family pseudogene
RP11-6B6.3	ENSG00000236942	GABA(A) receptor-associated protein (GABARAP) pseudogene

RP11-705C15.3	ENSG00000257028	novel transcript
RP11-863P13.4	ENSG00000205037	novel transcript
RP11-867G23.12	ENSG00000254756	novel transcript
RP11-996F15.2	ENSG00000257176	novel transcript
RP13-93L13.1	ENSG00000225461	novel transcript
RP3-521E19.2	ENSG00000257494	novel transcript
RP5-1063M23.1	ENSG00000250770	tetraspanin 11 (TSPAN11) pseudogene
RP5-955M13.4	ENSG00000232358	novel transcript
RPL10AP1	ENSG00000244691	ribosomal protein L10a pseudogene 1 [Source: HGNC Symbol;Acc:HGNC:19813]
RPL5P5	ENSG00000213051	ribosomal protein L5 pseudogene 5 [Source: HGNC Symbol;Acc:HGNC:35564]
RPL7P57	ENSG00000224401	ribosomal protein L7 pseudogene 57 [Source: HGNC Symbol;Acc:HGNC:35901]
RPS10P18	ENSG00000229455	ribosomal protein S10 pseudogene 18 [Source: HGNC Symbol;Acc:HGNC:36239]
RPS15A	ENSG00000134419	ribosomal protein S15a [Source: HGNC Symbol;Acc:HGNC:10389]
RPS15AP12	ENSG00000232134	ribosomal protein S15a pseudogene 12 [Source: HGNC Symbol;Acc:HGNC:36759]
RPS3AP47	ENSG00000205873	novel transcript
SEL1L	ENSG00000071537	SEL1L, ERAD E3 ligase adaptor subunit [Source: HGNC Symbol;Acc:HGNC:10717]
SEMA6A	ENSG00000092421	semaphorin 6A [Source: HGNC Symbol;Acc:HGNC:10738]
SETD9	ENSG00000155542	SET domain containing 9 [Source: HGNC Symbol;Acc:HGNC:28508]
SF3A3P2	ENSG00000254449	splicing factor 3a, subunit 3 pseudogene 2 [Source: HGNC Symbol;Acc:HGNC:23277]

SLC16A13	ENSG00000174327	solute carrier family 16 member 13 [Source: HGNC Symbol;Acc:HGNC:31037]
SLC16A6P1	ENSG00000232457	SLC16A6 pseudogene 1 [Source: HGNC Symbol;Acc:HGNC:48932]
SLC16A8	ENSG00000100156	solute carrier family 16 member 8 [Source: HGNC Symbol;Acc:HGNC:16270]
SLC25A39P1	ENSG00000226148	solute carrier family 25 member 39 pseudogene 1 [Source: HGNC Symbol;Acc:HGNC:43859]
SLC9A7	ENSG00000065923	solute carrier family 9 member A7 [Source: HGNC Symbol;Acc:HGNC:17123]
SNORA65	ENSG00000201302	small nucleolar RNA, H/ACA box 65 [Source: HGNC Symbol;Acc:HGNC:10222]
SNORD45	ENSG00000200706	novel transcript
SNORD99	ENSG00000221539	small nucleolar RNA, C/D box 99 [Source: HGNC Symbol;Acc:HGNC:32762]
SPDYC	ENSG00000204710	speedy/RINGO cell cycle regulator family member C [Source: HGNC Symbol;Acc:HGNC:32681]
SPNS2	ENSG00000183018	sphingolipid transporter 2 [Source: HGNC Symbol;Acc:HGNC:26992]
SRP72P2	ENSG00000188451	signal recognition particle 72 pseudogene 2 [Source: HGNC Symbol;Acc:HGNC:31096]
STARD4-AS1	ENSG00000246859	STARD4 antisense RNA 1 [Source: HGNC Symbol;Acc:HGNC:44117]
SYCP3	ENSG00000139351	synaptonemal complex protein 3 [Source: HGNC Symbol;Acc:HGNC:18130]
TADA1	ENSG00000152382	transcriptional adaptor 1 [Source: HGNC Symbol;Acc:HGNC:30631]
TCTN1	ENSG00000204852	tectonic family member 1 [Source: HGNC Symbol;Acc:HGNC:26113]
TERF2	ENSG00000132604	telomeric repeat binding factor 2 [Source: HGNC Symbol;Acc:HGNC:11729]
TGFB3	ENSG00000119699	transforming growth factor beta 3 [Source: HGNC Symbol;Acc:HGNC:11769]
TIMM22	ENSG00000177370	translocase of inner mitochondrial membrane 22 [Source: HGNC Symbol;Acc:HGNC:17317]
TOP1	ENSG00000198900	DNA topoisomerase I [Source: HGNC Symbol;Acc:HGNC:11986]

TRIM75P	ENSG00000250374	tripartite motif containing 75, pseudogene [Source: HGNC Symbol;Acc:HGNC:32686]
UBE2L2	ENSG00000131982	ubiquitin conjugating enzyme E2 L2 (pseudogene) [Source: HGNC Symbol;Acc:HGNC:12487]
WDR55	ENSG00000120314	WD repeat domain 55 [Source: HGNC Symbol;Acc:HGNC:25971]
WDR88	ENSG00000166359	WD repeat domain 88 [Source: HGNC Symbol;Acc:HGNC:26999]
ZACN	ENSG00000186919	zinc activated ion channel [Source: HGNC Symbol;Acc:HGNC:29504]
ZBED2	ENSG00000177494	zinc finger BED-type containing 2 [Source: HGNC Symbol;Acc:HGNC:20710]
ZBTB2	ENSG00000181472	zinc finger and BTB domain containing 2 [Source: HGNC Symbol;Acc:HGNC:20868]
ZNF181	ENSG00000197841	zinc finger protein 181 [Source: HGNC Symbol;Acc:HGNC:12971]
ZNF213	ENSG00000085644	zinc finger protein 213 [Source: HGNC Symbol;Acc:HGNC:13005]
ZNF354C	ENSG00000177932	zinc finger protein 354C [Source: HGNC Symbol;Acc:HGNC:16736]
ZNF75BP	ENSG00000258212	zinc finger protein 75B, pseudogene [Source: HGNC Symbol;Acc:HGNC:13147]
ZSCAN16	ENSG00000196812	zinc finger and SCAN domain containing 16 [Source: HGNC Symbol;Acc:HGNC:20813]

Table 2: Genes associated with Breast Cancer Specific Survival (BCSS) identified by Artificial Neural Network analysis (Transcriptomic Cohort, n=112)

Gene ID	initial_alias	description
ABCA4	ENSG00000198691	ATP binding cassette subfamily A member 4 [Source:HGNC Symbol;Acc:HGNC:34]
AC003075.4	ENSG00000237773	novel transcript
AC004951.6	ENSG00000228434	novel transcript
AC020931.1	ENSG00000257110	novel transcript
AC072052.7	ENSG00000231360	novel transcript
AC073834.3	ENSG00000237655	novel transcript, antisense to TTC30A
AC079305.10	ENSG00000222043	novel transcript
AC080125.1	ENSG00000225406	pseudogene similar to part of E74-like factor 2 (ets domain transcription factor) (ELF2)
AC084809.2	ENSG00000226377	novel transcript
ACBD6	ENSG00000230124	acyl-CoA binding domain containing 6 [Source:HGNC Symbol;Acc:HGNC:23339]
ACSM4	ENSG00000215009	acyl-CoA synthetase medium chain family member 4 [Source:HGNC Symbol;Acc:HGNC:32016]
ADAMTS5	ENSG00000154736	ADAM metallopeptidase with thrombospondin type 1 motif 5 [Source:HGNC Symbol;Acc:HGNC:221]
ADORA2B	ENSG00000170425	adenosine A2b receptor [Source:HGNC Symbol;Acc:HGNC:264]
AKAP8L	ENSG00000011243	A-kinase anchoring protein 8 like [Source:HGNC Symbol;Acc:HGNC:29857]
ALG1L5P	ENSG00000226943	asparagine-linked glycosylation 1-like 5, pseudogene [Source:HGNC Symbol;Acc:HGNC:44374]
ANKRD34B	ENSG00000189127	ankyrin repeat domain 34B [Source:HGNC Symbol;Acc:HGNC:33736]
ANO5	ENSG00000171714	anoctamin 5 [Source:HGNC Symbol;Acc:HGNC:27337]
ASXL3	ENSG00000141431	ASXL transcriptional regulator 3 [Source:HGNC Symbol;Acc:HGNC:29357]
ATP1B2	ENSG00000129244	ATPase Na+/K+ transporting subunit beta 2 [Source:HGNC Symbol;Acc:HGNC:805]

B3GNT5	ENSG00000176597	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5 [Source:HGNC Symbol;Acc:HGNC:15684]
BICC1	ENSG00000122870	BicC family RNA binding protein 1 [Source:HGNC Symbol;Acc:HGNC:19351]
BLVRB	ENSG00000090013	biliverdin reductase B [Source:HGNC Symbol;Acc:HGNC:1063]
BMS1	ENSG00000165733	BMS1, ribosome biogenesis factor [Source:HGNC Symbol;Acc:HGNC:23505]
C11orf24	ENSG00000171067	chromosome 11 open reading frame 24 [Source:HGNC Symbol;Acc:HGNC:1174]
C19orf12	ENSG00000131943	chromosome 19 open reading frame 12 [Source:HGNC Symbol;Acc:HGNC:25443]
C1orf147	ENSG00000162888	chromosome 1 open reading frame 147 [Source:HGNC Symbol;Acc:HGNC:32061]
C1QBP	ENSG00000108561	complement C1q binding protein [Source:HGNC Symbol;Acc:HGNC:1243]
C2orf70	ENSG00000173557	chromosome 2 open reading frame 70 [Source:HGNC Symbol;Acc:HGNC:27938]
CCDC54	ENSG00000138483	coiled-coil domain containing 54 [Source:HGNC Symbol;Acc:HGNC:30703]
CEP350	ENSG00000135837	centrosomal protein 350 [Source:HGNC Symbol;Acc:HGNC:24238]
CGNL1	ENSG00000128849	cingulin like 1 [Source:HGNC Symbol;Acc:HGNC:25931]
CMB9-22P13.1	ENSG00000173727	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed (FAU) pseudogene
CMYA5	ENSG00000164309	cardiomyopathy associated 5 [Source:HGNC Symbol;Acc:HGNC:14305]
CNTN4-AS1	ENSG00000237990	CNTN4 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:39985]
CNTNAP2	ENSG00000174469	contactin associated protein like 2 [Source:HGNC Symbol;Acc:HGNC:13830]
CSDC2	ENSG00000172346	cold shock domain containing C2 [Source:HGNC Symbol;Acc:HGNC:30359]
CTAGE12P	ENSG00000215441	CTAGE family member 12, pseudogene [Source:HGNC Symbol;Acc:HGNC:37297]
CTC-498J12.3	ENSG00000248664	novel transcript
CTD-2026K11.6	ENSG00000203394	novel transcript
CTD-2228K2.2	ENSG00000214278	mitochondrial translational initiation factor 3 (MTIF3) pseudogene

CTD-2290C23.2	ENSG00000241739	ribosomal protein L21 (RPL21) pseudogene
CTNS	ENSG00000040531	cystinosin, lysosomal cystine transporter [Source:HGNC Symbol;Acc:HGNC:2518]
CUBN	ENSG00000107611	cubilin [Source:HGNC Symbol;Acc:HGNC:2548]
CYP1B1-AS1	ENSG00000232973	CYP1B1 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:28543]
DCTN1-AS1	ENSG00000237737	DCTN1 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:44151]
DDA1	ENSG00000130311	DET1 and DDB1 associated 1 [Source:HGNC Symbol;Acc:HGNC:28360]
DNAH7	ENSG00000118997	dynein axonemal heavy chain 7 [Source:HGNC Symbol;Acc:HGNC:18661]
DNAJC17	ENSG00000104129	DnaJ heat shock protein family (Hsp40) member C17 [Source:HGNC Symbol;Acc:HGNC:25556]
DOCK10	ENSG00000135905	dedicator of cytokinesis 10 [Source:HGNC Symbol;Acc:HGNC:23479]
DST	ENSG00000151914	dystonin [Source:HGNC Symbol;Acc:HGNC:1090]
EEF1D	ENSG00000104529	eukaryotic translation elongation factor 1 delta [Source:HGNC Symbol;Acc:HGNC:3211]
EGFR	ENSG00000146648	epidermal growth factor receptor [Source:HGNC Symbol;Acc:HGNC:3236]
ERV3-1	ENSG00000213462	endogenous retrovirus group 3 member 1, envelope [Source:HGNC Symbol;Acc:HGNC:3454]
ETV3	ENSG00000117036	ETS variant 3 [Source:HGNC Symbol;Acc:HGNC:3492]
EXOSC5	ENSG00000077348	exosome component 5 [Source:HGNC Symbol;Acc:HGNC:24662]
F8	ENSG00000185010	coagulation factor VIII [Source:HGNC Symbol;Acc:HGNC:3546]
FAM151A	ENSG00000162391	family with sequence similarity 151 member A [Source:HGNC Symbol;Acc:HGNC:25032]
FAM47C	ENSG00000198173	family with sequence similarity 47 member C [Source:HGNC Symbol;Acc:HGNC:25301]
FAM86KP	ENSG00000163612	family with sequence similarity 86 member K, pseudogene [Source:HGNC Symbol;Acc:HGNC:44098]
FAM86MP	ENSG00000186234	family with sequence similarity 86 member M, pseudogene [Source:HGNC Symbol;Acc:HGNC:44100]
FANCF	ENSG00000183161	FA complementation group F [Source:HGNC Symbol;Acc:HGNC:3587]
FANCG	ENSG00000221834	novel transcript
FCRLB	ENSG00000162746	Fc receptor like B [Source:HGNC Symbol;Acc:HGNC:26431]

FIGN	ENSG00000182263	fidgetin, microtubule severing factor [Source:HGNC Symbol;Acc:HGNC:13285]
FMR1-IT1	ENSG00000236338	novel transcript
FOXD1-AS1	ENSG00000248003	novel transcript
FOXD3	ENSG00000187140	forkhead box D3 [Source:HGNC Symbol;Acc:HGNC:3804]
FZD5	ENSG00000163251	frizzled class receptor 5 [Source:HGNC Symbol;Acc:HGNC:4043]
GIT1	ENSG00000108270	novel transcript
GPR183	ENSG00000169508	G protein-coupled receptor 183 [Source:HGNC Symbol;Acc:HGNC:3128]
GPR33	ENSG00000214943	G protein-coupled receptor 33 (gene/pseudogene) [Source:HGNC Symbol;Acc:HGNC:4489]
GPRIN3	ENSG00000185477	GPRIN family member 3 [Source:HGNC Symbol;Acc:HGNC:27733]
GS1-251I9.3	ENSG00000253358	novel transcript
GTF3C6	ENSG00000155130	general transcription factor IIIC subunit 6 [Source:HGNC Symbol;Acc:HGNC:20872]
GUCY2F	ENSG00000101890	guanylate cyclase 2F, retinal [Source:HGNC Symbol;Acc:HGNC:4691]
HAMP	ENSG00000105697	hepcidin antimicrobial peptide [Source:HGNC Symbol;Acc:HGNC:15598]
HARBI1	ENSG00000180423	harbinger transposase derived 1 [Source:HGNC Symbol;Acc:HGNC:26522]
HDDC2	ENSG00000111906	HD domain containing 2 [Source:HGNC Symbol;Acc:HGNC:21078]
HEPACAM	ENSG00000165478	hepatic and glial cell adhesion molecule [Source:HGNC Symbol;Acc:HGNC:26361]
HEXDC	ENSG00000169660	hexosaminidase D [Source:HGNC Symbol;Acc:HGNC:26307]
IGFBP5	ENSG00000115461	insulin like growth factor binding protein 5 [Source:HGNC Symbol;Acc:HGNC:5474]
IGKV1OR10-1	ENSG00000237592	immunoglobulin kappa variable 1/OR10-1 (pseudogene) [Source:HGNC Symbol;Acc:HGNC:44978]
IKZF5	ENSG00000095574	IKAROS family zinc finger 5 [Source:HGNC Symbol;Acc:HGNC:14283]
ITGA9	ENSG00000144668	integrin subunit alpha 9 [Source:HGNC Symbol;Acc:HGNC:6145]
KIAA0355	ENSG00000166398	KIAA0355 [Source:HGNC Symbol;Acc:HGNC:29016]
KIAA1462	ENSG00000165757	junctional cadherin 5 associated [Source:HGNC Symbol;Acc:HGNC:29283]
KLHL41	ENSG00000239474	kelch like family member 41 [Source:HGNC Symbol;Acc:HGNC:16905]
KRTAP3-3	ENSG00000212899	keratin associated protein 3-3 [Source:HGNC Symbol;Acc:HGNC:18890]

LGALS4	ENSG00000171747	galectin 4 [Source:HGNC Symbol;Acc:HGNC:6565]
LINC01074	ENSG00000227612	novel transcript
LIPE	ENSG00000079435	lipase E, hormone sensitive type [Source:HGNC Symbol;Acc:HGNC:6621]
LRRC3C	ENSG00000204913	leucine rich repeat containing 3C [Source:HGNC Symbol;Acc:HGNC:40034]
MAP2K6	ENSG00000108984	mitogen-activated protein kinase kinase 6 [Source:HGNC Symbol;Acc:HGNC:6846]
MAP3K4	ENSG00000085511	mitogen-activated protein kinase kinase kinase 4 [Source:HGNC Symbol;Acc:HGNC:6856]
MED14	ENSG00000180182	mediator complex subunit 14 [Source:HGNC Symbol;Acc:HGNC:2370]
MED22	ENSG00000148297	mediator complex subunit 22 [Source:HGNC Symbol;Acc:HGNC:11477]
MEX3A	ENSG00000254726	mex-3 RNA binding family member A [Source:HGNC Symbol;Acc:HGNC:33482]
MIR199A1	ENSG00000207752	microRNA 199a-1 [Source:HGNC Symbol;Acc:HGNC:31571]
MMP25	ENSG00000008516	matrix metallopeptidase 25 [Source:HGNC Symbol;Acc:HGNC:14246]
MPND	ENSG00000008382	MPN domain containing [Source:HGNC Symbol;Acc:HGNC:25934]
MRGBP	ENSG00000101189	MRG domain binding protein [Source:HGNC Symbol;Acc:HGNC:15866]
NAPA	ENSG00000105402	NSF attachment protein alpha [Source:HGNC Symbol;Acc:HGNC:7641]
NDUFA4L2	ENSG00000185633	NDUFA4, mitochondrial complex associated like 2 [Source:HGNC Symbol;Acc:HGNC:29836]
NIPSNAP1	ENSG00000184117	nipsnap homolog 1 [Source:HGNC Symbol;Acc:HGNC:7827]
NLRP5	ENSG00000171487	NLR family pyrin domain containing 5 [Source:HGNC Symbol;Acc:HGNC:21269]
NPAP1	ENSG00000185823	nuclear pore associated protein 1 [Source:HGNC Symbol;Acc:HGNC:1190]
ODF2-AS1	ENSG00000225951	ODF2 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:49461]
OR10K1	ENSG00000173285	olfactory receptor family 10 subfamily K member 1 [Source:HGNC Symbol;Acc:HGNC:14693]
OR1I1	ENSG00000094661	olfactory receptor family 1 subfamily I member 1 [Source:HGNC Symbol;Acc:HGNC:8207]
OR1S1	ENSG00000172774	novel transcript
OR2A1	ENSG00000221970	olfactory receptor family 2 subfamily A member 1 [Source:HGNC Symbol;Acc:HGNC:8229]

OR7A17	ENSG00000185385	olfactory receptor family 7 subfamily A member 17 [Source:HGNC Symbol;Acc:HGNC:8363]
OR8B4	ENSG00000198657	novel transcript
PAXBP1-AS1	ENSG00000238197	PAXBP1 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:39603]
PAXIP1-AS2	ENSG00000214106	PAXIP1 antisense RNA 2 [Source:HGNC Symbol;Acc:HGNC:48958]
PCDHB1	ENSG00000171815	protocadherin beta 1 [Source:HGNC Symbol;Acc:HGNC:8680]
PDIA4	ENSG00000155660	protein disulfide isomerase family A member 4 [Source:HGNC Symbol;Acc:HGNC:30167]
PHOSPHO1	ENSG00000173868	phosphoethanolamine/phosphocholine phosphatase [Source:HGNC Symbol;Acc:HGNC:16815]
POM121L3P	ENSG00000167390	POM121 transmembrane nucleoporin like 3, pseudogene [Source:HGNC Symbol;Acc:HGNC:16440]
PPIAP31	ENSG00000217094	peptidylprolyl isomerase A pseudogene 31 [Source:HGNC Symbol;Acc:HGNC:44962]
PPL	ENSG00000118898	periplakin [Source:HGNC Symbol;Acc:HGNC:9273]
QRSL1P3	ENSG00000257957	QRSL1 pseudogene 3 [Source:HGNC Symbol;Acc:HGNC:43669]
RLBP1	ENSG00000140522	retinaldehyde binding protein 1 [Source:HGNC Symbol;Acc:HGNC:10024]
RNF5	ENSG00000204308	ring finger protein 5 [Source:HGNC Symbol;Acc:HGNC:10068]
ROPN1B	ENSG00000114547	rhophilin associated tail protein 1B [Source:HGNC Symbol;Acc:HGNC:31927]
RP1-274L7.1	ENSG00000229702	novel transcript
RP11-252O18.3	ENSG00000213155	zinc finger, CCHC domain containing 10 (ZCCHC10) pseudogene
RP11-262H14.11	ENSG00000219693	fibroblast growth factor 7 pseudogene 8 [Source:HGNC Symbol;Acc:HGNC:34516]
RP11-286N22.8	ENSG00000256591	novel transcript
RP11-298I3.1	ENSG00000257285	novel transcript
RP11-29H23.5	ENSG00000246203	novel transcript
RP11-409C19.2	ENSG00000253223	PRP3 pre-mRNA processing factor 3 homolog (S. cerevisiae) (PRPF3) pseudogene

RP11-429J17.5	ENSG00000254549	novel transcript
RP11-466G12.2	ENSG00000249758	novel transcript
RP11-542G1.3	ENSG00000251384	novel transcript
RP11-544M22.3	ENSG00000232879	glutaredoxin 5 (GLRX5) pseudogene
RP11-592N21.1	ENSG00000212673	novel transcript
RP11-619A14.2	ENSG00000254933	novel transcript
RP11-798K23.3	ENSG00000251545	novel pseudogene
RP11-967K21.2	ENSG00000255953	novel transcript
RP3-508I15.9	ENSG00000228274	novel transcript, antisense to CBY1
RP4-622L5.7	ENSG00000224066	novel transcript
RPL15P18	ENSG00000228501	ribosomal protein L15 pseudogene 18 [Source:HGNC Symbol;Acc:HGNC:36515]
RPL21P134	ENSG00000233254	ribosomal protein L21 pseudogene 134 [Source:HGNC Symbol;Acc:HGNC:36006]
RPL27A	ENSG00000166441	ribosomal protein L27a [Source:HGNC Symbol;Acc:HGNC:10329]
RPL32P34	ENSG00000239524	ribosomal protein L32 pseudogene 34 [Source:HGNC Symbol;Acc:HGNC:35903]
RPL3P7	ENSG00000225093	ribosomal protein L3 pseudogene 7 [Source:HGNC Symbol;Acc:HGNC:36797]
RPS10P18	ENSG00000229455	ribosomal protein S10 pseudogene 18 [Source:HGNC Symbol;Acc:HGNC:36239]
RPS3AP47	ENSG00000205873	novel transcript
RPS8	ENSG00000142937	ribosomal protein S8 [Source:HGNC Symbol;Acc:HGNC:10441]

SCN10A	ENSG00000185313	sodium voltage-gated channel alpha subunit 10 [Source:HGNC Symbol;Acc:HGNC:10582]
SDPR	ENSG00000168497	caveolae associated protein 2 [Source:HGNC Symbol;Acc:HGNC:10690]
SIPA1L3	ENSG00000105738	signal induced proliferation associated 1 like 3 [Source:HGNC Symbol;Acc:HGNC:23801]
SKP1P1	ENSG00000231234	S-phase kinase associated protein 1 pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:33696]
SLC27A4	ENSG00000167114	solute carrier family 27 member 4 [Source:HGNC Symbol;Acc:HGNC:10998]
SLC39A4	ENSG00000147804	solute carrier family 39 member 4 [Source:HGNC Symbol;Acc:HGNC:17129]
SLC7A9	ENSG00000021488	solute carrier family 7 member 9 [Source:HGNC Symbol;Acc:HGNC:11067]
SMIM5	ENSG00000204323	small integral membrane protein 5 [Source:HGNC Symbol;Acc:HGNC:40030]
SNORD99	ENSG00000221539	small nucleolar RNA, C/D box 99 [Source:HGNC Symbol;Acc:HGNC:32762]
SNRPGP15	ENSG00000224543	small nuclear ribonucleoprotein polypeptide G pseudogene 15 [Source:HGNC Symbol;Acc:HGNC:49371]
SNU13	ENSG00000100138	small nuclear ribonucleoprotein 13 [Source:HGNC Symbol;Acc:HGNC:7819]
SORBS2	ENSG00000154556	sorbin and SH3 domain containing 2 [Source:HGNC Symbol;Acc:HGNC:24098]
SPAG7	ENSG00000091640	sperm associated antigen 7 [Source:HGNC Symbol;Acc:HGNC:11216]
SPATA31A6	ENSG00000185775	SPATA31 subfamily A member 6 [Source:HGNC Symbol;Acc:HGNC:32006]
SPDYC	ENSG00000204710	speedy/RINGO cell cycle regulator family member C [Source:HGNC Symbol;Acc:HGNC:32681]
SPRED1	ENSG00000166068	sprouty related EVH1 domain containing 1 [Source:HGNC Symbol;Acc:HGNC:20249]
SRP72P2	ENSG00000188451	signal recognition particle 72 pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:31096]
ST13P19	ENSG00000228110	ST13, Hsp70 interacting protein pseudogene 19 [Source:HGNC Symbol;Acc:HGNC:38862]
ST6GAL1	ENSG00000073849	ST6 beta-galactoside alpha-2,6-sialyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:10860]
STAT1	ENSG00000115415	signal transducer and activator of transcription 1 [Source:HGNC Symbol;Acc:HGNC:11362]

SYCE1L	ENSG00000205078	synaptonemal complex central element protein 1 like [Source:HGNC Symbol;Acc:HGNC:37236]
SYNGR1	ENSG00000100321	synaptogyrin 1 [Source:HGNC Symbol;Acc:HGNC:11498]
SZT2	ENSG00000198198	SZT2, KICSTOR complex subunit [Source:HGNC Symbol;Acc:HGNC:29040]
TBX1	ENSG00000184058	T-box 1 [Source:HGNC Symbol;Acc:HGNC:11592]
TDRKH	ENSG00000182134	tudor and KH domain containing [Source:HGNC Symbol;Acc:HGNC:11713]
TEX13B	ENSG00000170925	testis expressed 13B [Source:HGNC Symbol;Acc:HGNC:11736]
TMA16P2	ENSG00000232467	translation machinery associated 16 homolog pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:43781]
TMCO1	ENSG00000143183	transmembrane and coiled-coil domains 1 [Source:HGNC Symbol;Acc:HGNC:18188]
TMED10	ENSG00000170348	transmembrane p24 trafficking protein 10 [Source:HGNC Symbol;Acc:HGNC:16998]
TMEM78	ENSG00000177800	transmembrane protein 78 [Source:HGNC Symbol;Acc:HGNC:32307]
TMEM80	ENSG00000177042	transmembrane protein 80 [Source:HGNC Symbol;Acc:HGNC:27453]
TOR1AIP2	ENSG00000169905	torsin 1A interacting protein 2 [Source:HGNC Symbol;Acc:HGNC:24055]
TRBC2	ENSG00000211772	T cell receptor beta constant 2 [Source:HGNC Symbol;Acc:HGNC:12157]
TTPA	ENSG00000137561	alpha tocopherol transfer protein [Source:HGNC Symbol;Acc:HGNC:12404]
TUBA3E	ENSG00000152086	tubulin alpha 3e [Source:HGNC Symbol;Acc:HGNC:20765]
TUBB3P1	ENSG00000220418	tubulin beta 3 class III pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:42339]
U82670.9	ENSG00000229979	high-mobility group nucleosomal binding domain 2 (HMGN2) pseudogene
UPP2	ENSG00000007001	uridine phosphorylase 2 [Source:HGNC Symbol;Acc:HGNC:23061]
URB2	ENSG00000135763	URB2 ribosome biogenesis homolog [Source:HGNC Symbol;Acc:HGNC:28967]
USP51	ENSG00000247746	ubiquitin specific peptidase 51 [Source:HGNC Symbol;Acc:HGNC:23086]
VSIG8	ENSG00000243284	V-set and immunoglobulin domain containing 8 [Source:HGNC Symbol;Acc:HGNC:32063]
XXbac-B444P24.10	ENSG00000161132	proline dehydrogenase (oxidase) 1 (PRODH) pseudogene
YIF1B	ENSG00000167645	Yip1 interacting factor homolog B, membrane trafficking protein [Source:HGNC Symbol;Acc:HGNC:30511]

ZCCHC2	**ENSG00000141664**	**zinc finger CCHC-type containing 2 [Source:HGNC Symbol;Acc:HGNC:22916]**
ZMIZ1	**ENSG00000108175**	**zinc finger MIZ-type containing 1 [Source:HGNC Symbol;Acc:HGNC:16493]**
ZNF397	**ENSG00000186812**	**zinc finger protein 397 [Source:HGNC Symbol;Acc:HGNC:18818]**
ZNF446	**ENSG00000083838**	**zinc finger protein 446 [Source:HGNC Symbol;Acc:HGNC:21036]**
ZNF609	**ENSG00000180357**	**zinc finger protein 609 [Source:HGNC Symbol;Acc:HGNC:29003]**
ZNF804A	**ENSG00000170396**	**zinc finger protein 804A [Source:HGNC Symbol;Acc:HGNC:21711]**
ZSWIM1	**ENSG00000168614**	**novel transcript**

7.3 Chapter 4: Comprehensive Transcriptomic Profiling Reveals PPL as a Novel, Independent Predictor of Outcomes in TNBC

Mansour A Alsaleem (MAA), Chitra Joseph (CJ), Yousif A Kariri (YK) ,Pavan L Narasimha (PN), Michael S Toss (MS), Jennie Jeyapalan (JNJ), Daisy B Haigh (DBH), Corinne L Woodcock (CLW), Anna E Harris (AEH), Jennifer Lothion-Roy (JLR), Mohammed Aleskandarany (MA), Padmashree C G Rida (PR), Ritu Aneja (RA) , Graham Ball (GB) , Ian O Ellis (IE), Andrew Green (AR) , Nigel P Mongan (NPM) and Emad Rakha (ER). "Mechanistic and clinical evidence supports a key role for Periplakin (PPL) as a novel independent predictor of outcome in Triple Negative Breast Cancer". post-viva publication

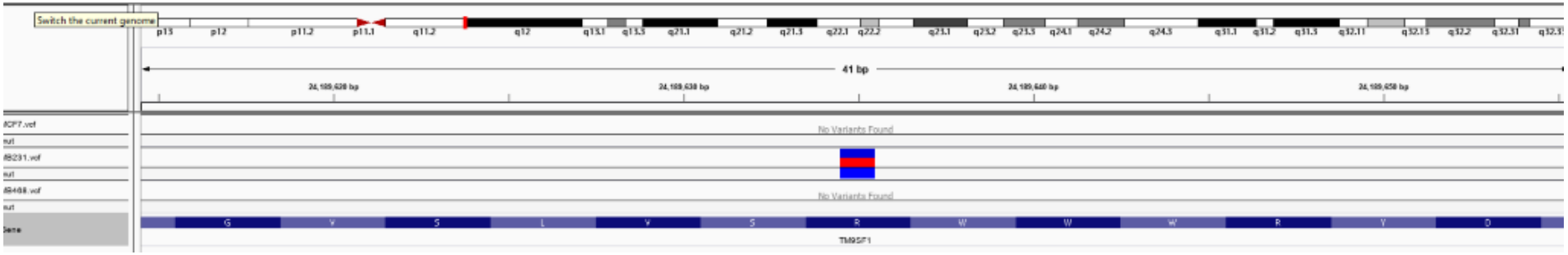
Table 1: MB-MDA-231 unique variants

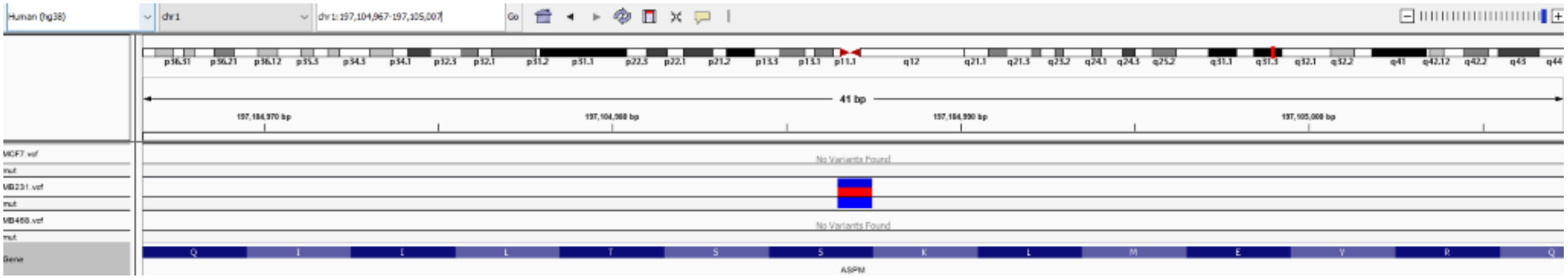
Gene	Transcript	Census Tier 1	AA Mutation	CDS Mutation	Somatic status	Zygosity	Validated	Type	Position	Found
HAUS3	ENST00000243706.8	No	p.D487H	c.1459G>C	Confirmed Somatic	Heterozygous	Unverified	Substitution - Missense	4:2236347-2236347	4:2236347-2236347
IDUA	ENST00000247933.8	No	p.F352L	c.1054T>C	Confirmed Somatic	Heterozygous	Unverified	Substitution - Missense	4:1002350-1002350	4:1002350-1002350
EIF2B4	ENST00000493344.6	No	p.C458S	c.1373G>C	Confirmed Somatic	Heterozygous	Unverified	Substitution - Missense	2:27364780-27364780	2:27364780-27364780
LONP1	ENST00000360614.7	No	p.R542Q	c.1625G>A	Confirmed Somatic	Heterozygous	Unverified	Substitution - Missense	19:5699087-5699087	19:5699087-5699087
GAS8	ENST00000268699.8	No	p.E423G	c.1268A>G	Confirmed Somatic	Homozygous	Unverified	Substitution - Missense	16:90042516-90042516	16:90042516-90042516
GTF3C1	ENST00000356183.8	No	p.R1988H	c.5963G>A	Confirmed Somatic	Homozygous	Unverified	Substitution - Missense	16:27462448-27462448	16:27462448-27462448
AL136295.1	ENST00000530611.1	No	p.R743Q	c.2228G>A	Confirmed Somatic	Heterozygous	Unverified	Substitution - Missense	14:24189635-24189635	14:24189635-24189635
ASPM	ENST00000367409.8	No	p.S1422P	c.4264T>C	Confirmed Somatic	Heterozygous	Unverified	Substitution - Missense	1:197104987-197104987	1:197104987-197104987
ATP8B2	ENST00000368489.3	No	p.P495L	c.1484C>T	Confirmed Somatic	Heterozygous	Unverified	Substitution - Missense	1:154342893-154342893	1:154342893-154342893
LRIG2	ENST00000361127.5	No	p.H628R	c.1883A>G	Confirmed Somatic	Heterozygous	Unverified	Substitution - Missense	1:113112563-113112563	1:113112563-113112563

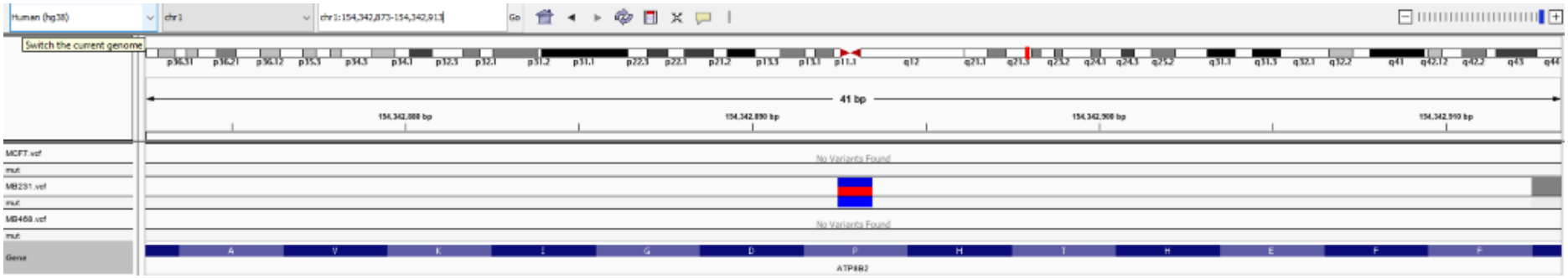
Table 2 : MB-MDA-231 unique variants

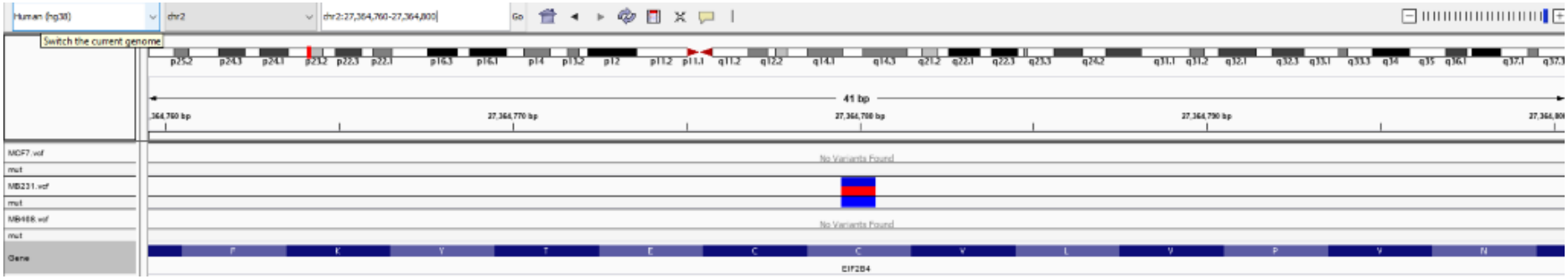
Gene	Transcript	Census Tier 1	AA Mutation	CDS Mutation	Somatic status	Zygosity	Validated	Type	Position	Found
CAMK4	ENST00000282356.8	No	p.R338C	c.1012C>T	Confirmed Somatic	Homozygous	Unverified	Substitution - Missense	5:111484056-111484056	5:111484056-111484056

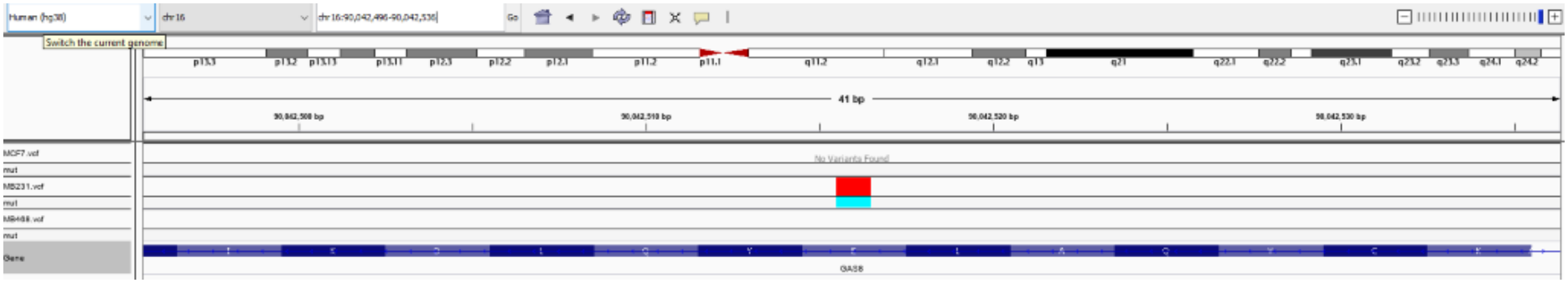
MDA-MB-231 unique variants

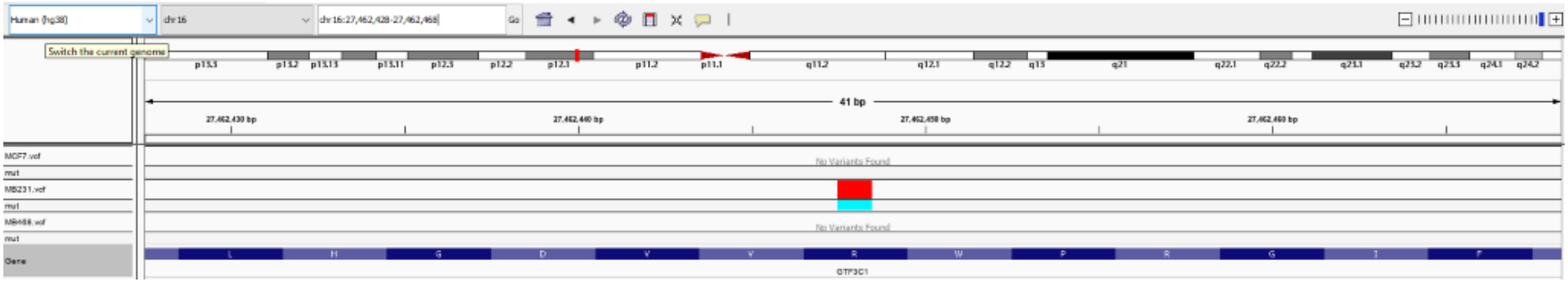


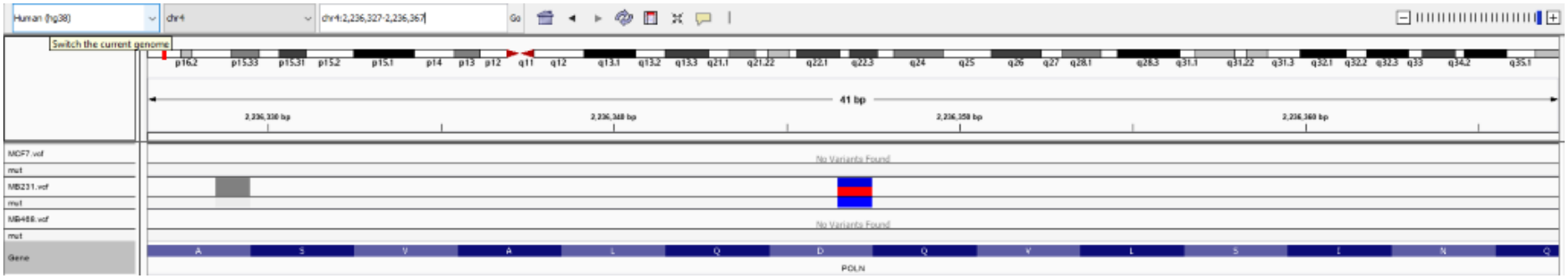


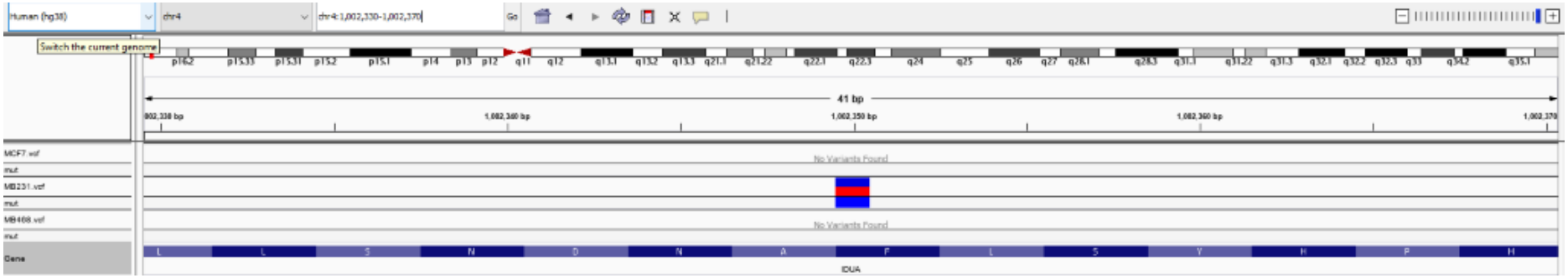

Human (hg38)
chr1
chr1:154,342,873-154,342,913
Go
Switch the current genome
41 bp
154,342,888 bp
154,342,890 bp
154,342,900 bp
154,342,910 bp
MCF7.vcf
mut
MB231.vcf
mut
MB468.vcf
mut
Gene
No Variants Found
No Variants Found
A V K I G D P H T H E F F
ATP8B2

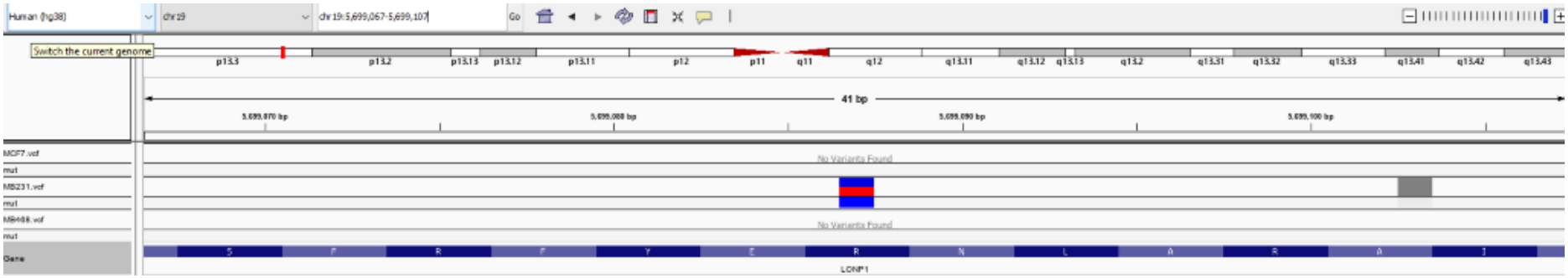

Human (hg38)
chr2
chr2:27,364,760-27,364,800
Go
Switch the current genome
41 bp
27,364,770 bp
27,364,780 bp
27,364,790 bp
MCF7.vcf
mut
MB231.vcf
mut
MB468.vcf
mut
Gene
No Variants Found
No Variants Found
F K Y T E C C V L V P V N
EIF2B4

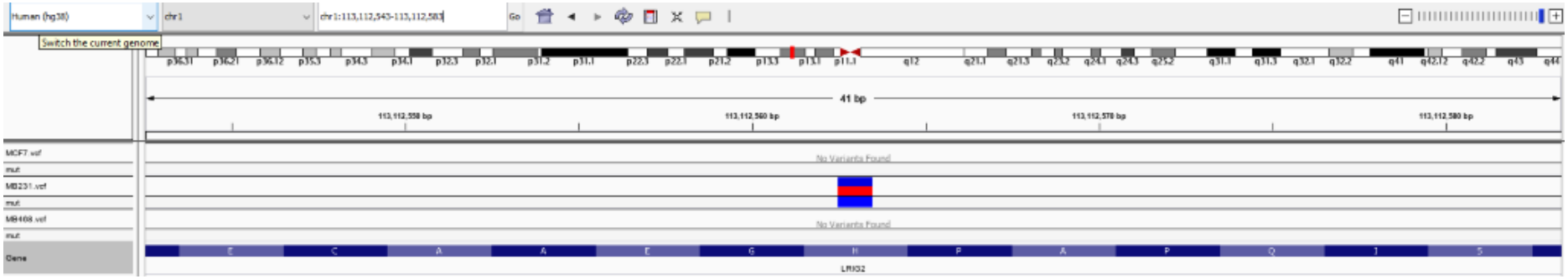

Human (hg38)
chr16
chr16:90,042,496-90,042,536
Go
Switch the current genome
41 bp
90,042,500 bp
90,042,510 bp
90,042,520 bp
90,042,530 bp
MCF7.vcf
mut
MB231.vcf
mut
MB468.vcf
mut
Gene
No Variants Found
GAS8

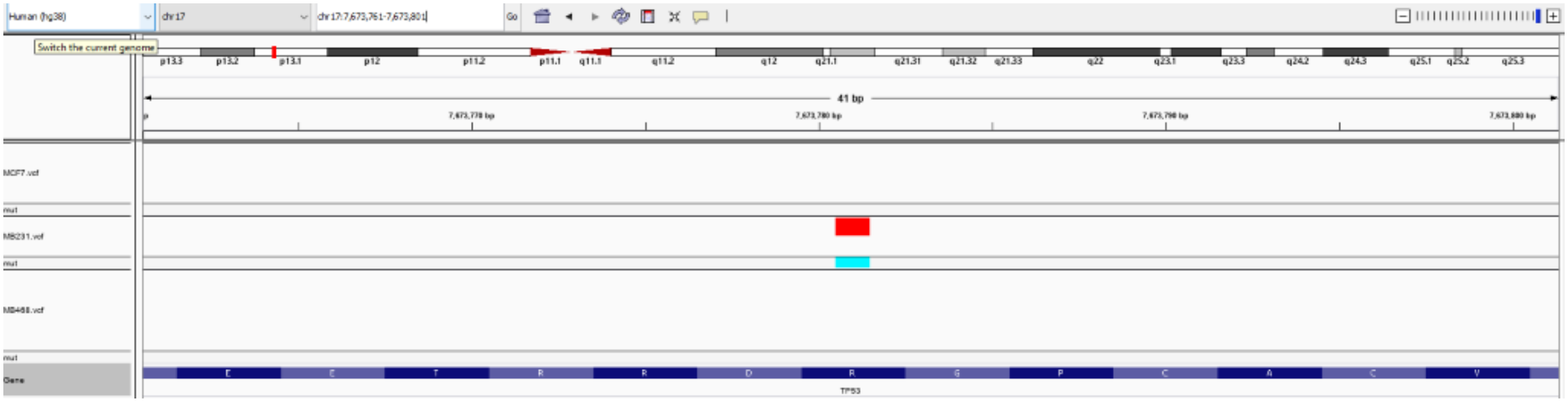

Human (hg38)
chr16
chr16:27,462,428-27,462,468
Go
Switch the current genome
41 bp
27,462,430 bp
27,462,440 bp
27,462,450 bp
27,462,460 bp
MCF7.vcf
mut
MB231.vcf
mut
MB468.vcf
mut
Gene
No Variants Found
No Variants Found
GTF3C1


Human (hg38)
chr4
chr4:2,236,327-2,236,367
Go
Switch the current genome
41 bp
MCF7.vcf
mut
MB231.vcf
mut
MB468.vcf
mut
Gene
No Variants Found
No Variants Found
POLN

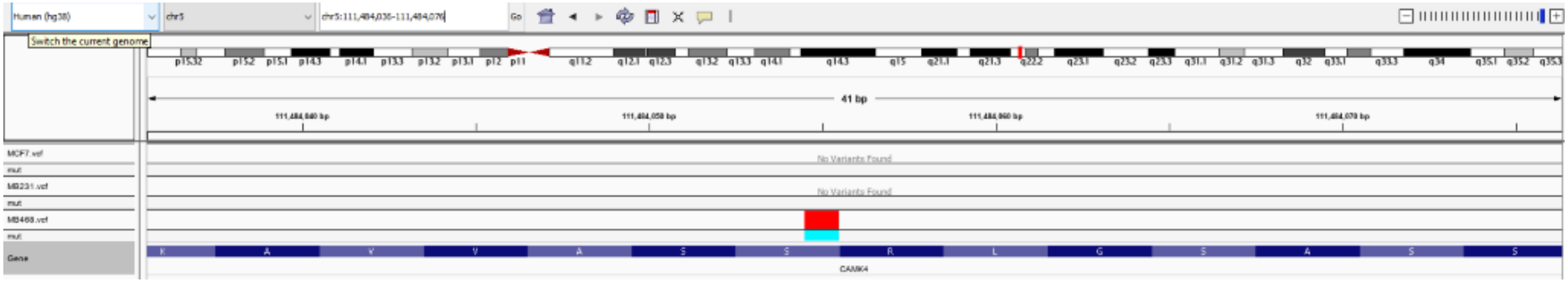

Human (hg38)
chr4
chr4:1,002,330-1,002,370
Go
Switch the current genome
41 bp
MCF7.vcf
mut
MB231.vcf
mut
MB468.vcf
mut
Gene
No Variants Found
No Variants Found
IDUA


Human (hg38)
chr19
chr19:5,699,067-5,699,107
Switch the current genome
41 bp
5,699,070 bp
5,699,080 bp
5,699,090 bp
5,699,100 bp
MCF7.vcf
mut
MB231.vcf
mut
MB468.vcf
mut
No Variants Found
No Variants Found
Gene
LDNF1


Human (hg38)
chr1
chr1:113,112,543-113,112,583
Switch the current genome
41 bp
113,112,550 bp
113,112,560 bp
113,112,570 bp
113,112,580 bp
MCF7.vcf
mut
MB231.vcf
mut
MB468.vcf
mut
No Variants Found
No Variants Found
Gene
LRIG2

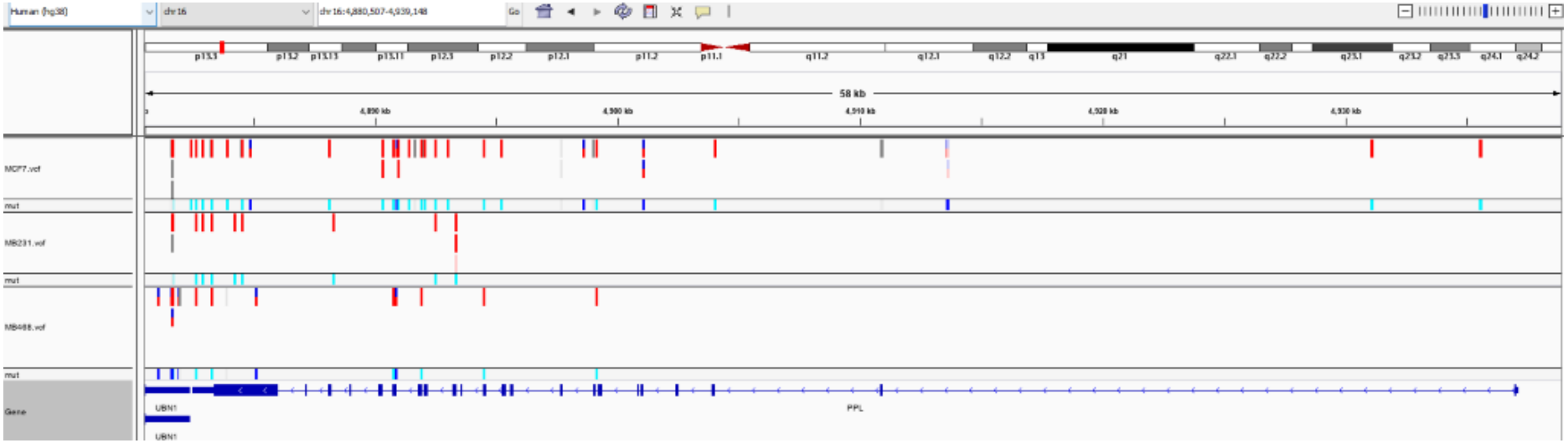

Human (hg38)
chr17
chr17:7,673,761-7,673,801
Go
Switch the current genome
41 bp
MCF7.vcf
mut
MB231.vcf
mut
MB468.vcf
mut
Gene
TP53

MDA-MB-436 unique variants



PPL

variants



7.4 **Chapter 5: The Molecular Landscape Of E-Cadherin in Ductal Carcinoma of The Breast**

Authors' list and contributions

Mansour Alsaleem (MAA) , Michael S Toss (MS), Chitra Joseph (CJ), Mohammed Aleskandarany (MA) , Sasagu Kurozumi (SK) , Ibrahim Alshankyty (IA), Angela Ogden(AO), Padmashree C G Rida (PR), Ian O Ellis (IE), Ritu Aneja (RA), Andrew R. Green (AR), Nigel P Mongan (NPM) and Emad A. Rakha (ER).

SPRINGER NATURE

The molecular mechanisms underlying reduced E-cadherin expression in invasive ductal carcinoma of the breast: high throughput analysis of large cohorts

Author: Mansour Alsaleem et al
Publication: Modern Pathology
Publisher: Springer Nature
Date: Feb 13, 2019

Copyright © 2019, Springer Nature

Author Request

If you are the author of this content (or his/her designated agent) please read the following. If you are not the author of this content, please click the Back button and select no to the question "Are you the Author of this Springer Nature content?".
Ownership of copyright in original research articles remains with the Author, and provided that, when reproducing the contribution or extracts from it or from the Supplementary Information, the Author acknowledges first and reference publication in the Journal, the Author retains the following non-exclusive rights:
To reproduce the contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).
The author and any academic institution, where they work, at the time may reproduce the contribution for the purpose of course teaching.
To reuse figures or tables created by the Author and contained in the Contribution in oral presentations and other works created by them.
To post a copy of the contribution as accepted for publication after peer review (in locked Word processing file, of a PDF version thereof) on the Author's own web site, or the Author's institutional repository, or the Author's funding body's archive, six months after publication of the printed or online edition of the Journal, provided that they also link to the contribution on the publisher's website.
Authors wishing to use the published version of their article for promotional use or on a web site must request in the normal way.

If you require further assistance please read Springer Nature's online author reuse guidelines.

For full paper portion: Authors of original research papers published by Springer Nature are encouraged to submit the author's version of the accepted, peer-reviewed manuscript to their relevant funding body's archive, for release six months after publication. In addition, authors are encouraged to archive their version of the manuscript in their institution's repositories (as well as their personal Web sites), also six months after original publication.

v1.0

BACK

CLOSE WINDOW