

University of Nottingham
UK | CHINA | MALAYSIA

The identification of targeting peptides and tumour associated antigens for the treatment and early detection of triple negative breast cancer

Megan Wilde, MSci

Supervisors: Dr Cinzia Allegrucci, Professor Kevin Gough and Dr James Dixon

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

March 2023

Abstract

Breast cancer (BCa) is the most common type of non-cutaneous cancer globally. Despite advances in screening methods and treatment regimens, the malignancy leads to the most female cancer-related deaths each year. An aggressive subtype of breast cancer which often presents at a high histological grade is triple negative breast cancer (TNBC). TNBC is characterised by the absence of estrogen and progesterone receptor expression, as well as a lack of HER2 over-expression. TNBC currently has the worse prognosis among the molecular subtypes. However, early diagnosis and the development of targeted treatment may improve patient outcome.

Peptide-drug conjugates are an emerging class of targeted therapies that are increasing the efficacy of anticancer agents through tumour-specific delivery. This work focused on identifying small peptides which target TNBC cells for the subsequent development of a peptide-based drug delivery system. To identify small peptides which specifically target TNBC, next generation phage display was employed for the high throughput screening of a diverse, small peptide library against non-tumourigenic and TNBC cell lines. Next generation sequencing analysis led to the identification of 44 potential TNBC targeting peptide candidates. *In vitro* cellular uptake studies with the peptide candidates resulted in the prioritisation of two TNBC targeting small peptides for *in vivo* biodistribution studies and the subsequent development of TNBC targeted therapeutics.

Recent studies have highlighted the diagnostic potential of serum autoantibodies, which are produced in early carcinogenesis in response to tumour associated antigens. In this study, the presence of cancer-specific autoantibodies in TNBC patient serum samples was confirmed by flow cytometry and potential TNBC autoantibody antigens were identified using a serological proteome analysis approach. Following on from the identification of TNBC neoantigens, western blot analyses with overexpression lysate confirmed the presence of a novel TNBC autoantibody in patient sera. This highlighted the value of autoantibodies in early diagnostic screening, but also their potential use for targeted therapies when combined with novel delivery systems.

Overall, the work in this study has led to the discovery of TNBC associated antigens, their corresponding autoantibodies, and targeting small peptides for the development

of early diagnostic tools and peptide-based targeted treatments with the potential to improve the prognosis and subsequent outcome for women with TNBC.

Acknowledgements

Firstly, I would like to say a huge thank you to my supervisor, Dr Cinzia Allegrucci, for her unwavering support and guidance throughout this project. I feel incredibly lucky to have had such a wonderful supervisor and mentor who has supported my personal development as well as my development as a research scientist, I will greatly miss our catch ups. I would also like to thank my additional supervisors, Professor Kevin Gough and Dr James Dixon, for all their valuable contributions and guidance throughout this project. I am very grateful for all the support I have received from my supervisory team.

A big thank you is also due to the past and present members of Cinzia's lab group who I have had the pleasure of working alongside: Jessica Smith, Dr Ahmad Zyoud, Mercedes Vázquez Cantú, Nabeelah Almalki, Riba Thomas and Ahmed Alnaeem. Particularly, I would like to say a special thank you to Jessica Smith for welcoming me into the lab group back in 2018 and for supporting me throughout this journey, she is a fantastic mentor and a wonderful friend.

Additionally, I would like to take this opportunity to thank Dr Maria Tsoumpeli, Dr Jon Owen and Keith Bishop for all their advice and support with the phage work conducted in this project. I would also like to thank all members of labs C05, B65 and A43 in the School of Veterinary Medicine and Science for making each day in the lab enjoyable. I feel very fortunate to have studied in the vet school and to have been a part of the BBSRC DTP programme. I would like to say a huge thank you to Dr Ruby Chrisp, Dr Corinne Woodcock, Anna Harris and Caleb Marsh for being great friends and supporting me so significantly over the past four years.

I would also like to thank my wonderful family and Ricky. They have provided me with so much emotional support throughout my PhD studies, which I will be forever grateful for. I could not have completed this project without them.

Finally, I would like to say a very special thank you to my dad, David Wilde. Although he is no longer here to celebrate this moment, he always believed that I would get this far. I hope that I have made you proud Dad.

Table of Contents

Abstract…… 1

Acknowledgements 3

Abbreviations 8

List of Tables 10

List of Figures 11

Chapter 1 : Introduction 14

1.1 Breast Cancer 14

1.2 Breast Cancer Classification 15

1.3 Methods of Diagnosis 19

1.4 Treatment and Resistance 21

1.5 Therapy Resistance and Breast Cancer Stem Cells 25

1.6 Triple Negative Breast Cancer 28

1.7 Targeted therapies for TNBC 31

1.8 TNBC Unmet Clinical Needs 35

1.9 Project aims and objectives 36

Chapter 2 : Materials and Methods 38

2.1 Materials 38

2.2 Cell Culture 38

2.3 Next generation phage display 39

2.3.1 TG1 *E.coli* 39

2.3.2 Ex-phage stock preparation 39

2.3.3 Kanamycin or ampicillin 2YT titration 40

2.3.4 Library propagation 40

2.3.5 Sanger sequencing of 9-mer library clones 41

2.3.6 Mammalian cell line viability 41

2.3.7 Biopanning 42

2.3.8 NGS sample preparation 42

2.3.9 NGS analysis 44

2.4 Recovery of NGPD identified peptides 45

2.4.1 Inverse PCRs for candidate TNBC targeting peptides 45

2.4.2 Inverse PCRs for control peptides 46

2.4.3 Phosphorylation and ligation 47
2.4.4 Preparation of electrocompetent *E.coli* cells 47
2.4.5 Transformation into electrocompetent TG1 *E.coli* 47
2.4.6 Sanger sequencing of recombinant plasmid 48
2.4.7 Recombinant protein expression and purification 48
2.5 *In vitro* validation of peptides 50
2.5.1 Fluorescence microscopy 50
2.5.2 Flow cytometry 50
2.5.3 Magnetic nanoparticle (MNP) delivery 51
2.5.4 MTT assay 51
2.5.5 Prussian blue staining 51
2.5.6 Inductively coupled plasma mass spectrometry (ICP-MS) 52
2.6 Autoantibody binding 53
2.6.1 Patient IgG antibody purification 53
2.6.2 Flow cytometry 53
2.6.3 Whole cell lysate protein extraction 54
2.6.4 Nuclear, cytoplasmic and membrane protein extraction 54
2.6.5 Western blotting 54
2.6.6 LC-MS/MS data analysis 55
2.6.7 Silver staining 56
2.6.8 LC-MS/MS sample preparation 56
2.6.9 Unique normalised spectral abundance factor (uNSAF) and t-tests 56
2.6.10 DESeq2… 57
2.6.11 bc-GenExMiner 57
2.6.12 Kaplan-Meier plotter 58
2.6.13 TPX2 and KHDRBS3 western blot 58
Chapter 3 : Identification of potential TNBC targeting small peptides by next generation phage display 59
3.1 Introduction 59

3.2 Results 63
3.2.1 Library propagation and biopanning 63
3.2.2 Next generation sequencing (NGS) sample preparation 65
3.2.3 Next generation sequencing analysis 70
3.3 Discussion 75
Chapter 4 : *In Vitro* Validation of TNBC targeting peptides 81
4.1 Introduction 81
4.2 Results 85
4.2.1 Recovery of peptide encoding sequences into a monovalent red fluorescent protein expressing vector 85
4.2.2 Cellular uptake of TNBC peptides by fluorescent microscopy 92
4.2.3 Cellular uptake of TNBC peptides by flow cytometry 99
4.2.4 Qualitative analysis of TNBC peptide mediated delivery of MNPs 107
4.2.5 Quantitative analysis of TNBC peptide mediated delivery of MNPs by ICP-MS 116
4.3 Discussion 118
Chapter 5 : Identification of TNBC associated antigens and their corresponding autoantibodies 125
5.1 Introduction 125
5.2 Results 129
5.2.1 TNBC patient IgG classification 129
5.2.2 TNBC IgG pool binding to MDA-MB-231 cells 130
5.2.3 Control and patient IgG binding to MCF10A and MDA-MB-231 cells 132
5.2.4 TNBC patient IgG binding to $CD44^{+}/CD24^{-}$ MDA-MB-231 cells 135
5.2.5 Control and patient IgG binding to MCF10A and MDA-MB-231 whole cell lysate 137
5.2.6 Control and patient IgG binding to MCF10A and MDA-MB-231 nuclear, cytoplasmic and membrane proteins 140
5.2.7 LC-MS/MS sample preparation 144
5.2.8 LC-MS/MS analysis 145
5.2.9 Expression of candidate TNBC antigens in TNBC patient datasets 148

5.2.10 Prognostic values of TPX2, WRNIP1, KHDRBS3, CALU, DNAJB6, KRT19 and HLA-A in TNBC…. 153

5.2.11 Validation of control and patient IgG binding to TPX2 and KHDRBS3 overexpression lysate 155

5.3 Discussion 158

Chapter 6 : Discussion 169

Bibliography 173

Appendices 211

Appendix 1. Permission to reuse copyright material 211

Appendix 2. Materials and methods 213

Appendix 2.1 NGS barcoded primers 213

Appendix 2.2 NGS Scripts 217

Appendix 3. Identification of potential TNBC targeting small peptides by NGPD…….. 220

Appendix 3.1 Phagemid map 220

Appendix 3.2 Sanger sequencing of 9-mer library clones 221

Appendix 3.3 PCR amplification of library clone inserts 223

Appendix 3.4 Cell viability with biopanning reagents 224

Appendix 3.5 NGS barcode reads 225

Appendix 3.6 TNBC targeting peptides in naïve bacteriophage library 227

Appendix 3.7 Seq2Logo results 229

Appendix 4. In vitro validation of potential TNBC targeting peptides 231

Appendix 4.1 pGEX-6P1-mRFP expression vector 231

Appendix 4.2 pGEX-6P1-mRFP primers for cloning 232

Appendix 4.3 Inverse PCR gels for peptides 17-44 235

Appendix 4.4 Sanger sequencing of peptide-mRFP plasmids 239

Appendix 4.5 CPP prediction for TNBC 9-mer peptides 249

Appendix 5. Identification of TNBC associated antigens 252

Appendix 5.1 DESeq2 Script 252

Appendix 6. BBSRC DTP Professional Internship for PhD Students (PIPS) reflective statement 253

Abbreviations

Aabs Autoantibodies

ADC Antibody-Drug Conjugate

AHSG Alpha-2-HS-Glycoprotein

APCs Antigen Presenting Cells

ATCC American Type Culture Collection

BCa Breast Cancer

BCSC Breast Cancer Stem Cell

BL1 Basal-Like Subtype 1

BL2 Basal-Like Subtype 2

CNN3 Calponin-3

CALU Calumenin

CBE Clinical Breast Examination

CFU Colony Forming Unit

CI Confidence Interval

CPP Cell Penetrating Peptide

CSC Cancer Stem Cell

DCIS Ductal Carcinoma *In Situ*

DNAJB6 DnaJ Heat Shock Protein Family Member B6

DTT Dithiothreitol

DUSP4 Dual Specificity Phosphatase 4

EGFR Epidermal Growth Factor Receptor

EMA European Medicines Agency

EMT Epithelial to Mesenchymal Transition

ENO2 Enolase 2

ER Estrogen Receptor

FDA Food And Drugs Administration

HER2 Human Epidermal Growth Factor 2

HLA-A Major Histocompatibility Complex, Class I, A

HR Hazard Ratio

HRR Homologous Recombination Repair

IgG Immunoglobulin G

IM Immunomodulatory

KHDRBS3 KH RNA Binding Domain Containing, Signal Transduction-Associated Protein 3

KRT19 Keratin 19

LAR Luminal Androgen Receptor

LCIS Lobular Carcinoma *In Situ*

M Mesenchymal

MMAE Monomethyl Auristatin E

MSL Mesenchymal Stem-Like

NGPD Next Generaton Phage Display

NICE National Institute For Health Care Excellence

OS Overall Survival

PBS Phosphate Buffered Saline

pCR Pathological Complete Response

PDC Peptide-Drug Conjugate

PELO Pelota mRNA Surveillance and Ribosome Rescue Factor

PFA Paraformaldehyde

PPVs Personalised Peptide Vaccines

PR Progesterone Receptor

PTD Protein Transduction Domain

RCT Random Control Trial

SATB2 SATB2 Homeobox 2

SCRN1 Secernin-1

SEREX Serological Screening of cDNA Expression Libraries

SERPA Serological Proteome Analysis

SERPINB9 Serine Family B Member 9

siRNA Small Interfering RNA

SLC29A1 Solute Carrier Family 29 Member 1

TAAs Tumour Associated Antigens

TP53 Tumour Protein P53

TPX2 TPX2 Microtubule Nucleation Factor

TNBC Triple Negative Breast Cancer

VEGF Vascular Endothelial Growth Factor Receptor

VIM Vimentin

WRNIP1 WRN Helicase Interacting Protein 1

List of Tables

Table 1.1 One year survival rates based on breast cancer stage at diagnosis in England 15
Table 1.2 Common forms of breast carcinoma 16
Table 1.3 The Molecular Subtypes of Breast Cancer 18
Table 2.1 Cell line information 38
Table 2.2 Phage dilution series to establish number of phage in stock solutions 40
Table 2.3 Thermocycler conditions for first round PCRs 43
Table 2.4 Thermocycler conditions for second round PCRs 43
Table 2.5 Thermocycler conditions for rescuing 9-mer peptides 46
Table 2.6 Thermocycler conditions p21 positive control 47
Table 4.1 Summary of mRFP fluorescence observed by fluorescence microscopy 98
Table 4.2 Cellular uptake of fluorescently labelled peptides in TNBC cell lines compared to control cell lines 107
Table 4.3 Amino acid sequences of synthetic peptides used in MNP complexation 108
Table 5.1 TNBC patient IgG pools based on clinical characteristics 129
Supplementary Table 1 Ion Torrent sequencing primers 213
Supplementary Table 2 The number of NGS reads for each barcode 225
Supplementary Table 3 The number of NGS reads for each TNBC peptide in the original bacteriophage library sequencing data. 227
Supplementary Table 4 Inverse PCR primers designed for the control peptides 232
Supplementary Table 5 Inverse PCR primers designed for recovery of TNBC targeting peptides 232
Supplementary Table 6 BChemRF-CPPred results for 9-mer peptides 249

List of Figures

Figure 3.1 Production of a peptide displaying bacteriophage using the phagemid system 61
Figure 3.2 Biopanning of a random peptide bacteriophage library 62
Figure 3.3 Biopanning against TNBC cell lines 64
Figure 3.4 Amplification of 9-mer peptide inserts 66
Figure 3.5 Attachment of barcode sequences for Ion Torrent sequencing 68
Figure 3.6 Pooled DNA resolved on a 3% MetaPhor agarose gel 69
Figure 3.7 Z>2 TNBC cell-penetrating peptides 71
Figure 3.8 Frequency of each Z>2 peptide across all TNBC panning experiments 72
Figure 3.9 Z>3 TNBC cell-penetrating peptides 73
Figure 3.10 Frequency of each Z>3 peptide across all TNBC biopanning experiments 74
Figure 4.1 Insertion of 9-mer peptide encoding sequences into the mRFP expression vector 87
Figure 4.2 Translated Sanger sequencing results for recombinant plasmids 88
Figure 4.3 Insertion of the control peptide sequences into the mRFP expression vector 89
Figure 4.4 Insertion of the p21 sequence into the mRFP expression vector 90
Figure 4.5 Purification of peptide-mRFP using Glutathione Sepharose™ 4B resin 91
Figure 4.6 Binding of each mRFP labelled peptide to BJ, MCF10A, HMEC-TERT, BT-549, MDA-MB-231 and CAL-51 cells 96
Figure 4.7 Cellular uptake of control and TNBC targeting peptides by BT-549 cells. 100
Figure 4.8 Cellular uptake of control and TNBC peptides in BT-549 cells compared to HMEC-TERT and MCF10A cells 102
Figure 4.9 Cellular uptake of control and TNBC peptides in CAL-51 cells compared to HMEC-TERT and MCF10A cells 104
Figure 4.10 Cellular uptake of control and TNBC peptides in MDA-MB-231 cells compared to HMEC-TERT and MCF10A cells 106
Figure 4.11 Control and TNBC cell viability with MNPs and targeting peptides 109
Figure 4.12 TNBC peptide-mediated delivery of MNPs into HMEC-TERT cells 111
Figure 4.13 TNBC peptide-mediated delivery of MNPs into MCF10A cells 112
Figure 4.14 TNBC peptide-mediated delivery of MNPs into BT-549 cells 113
Figure 4.15 TNBC peptide-mediated delivery of MNPs into CAL-51 cells 114

Figure 4.16 TNBC peptide-mediated delivery of MNPs into MDA-MB-231 cells 115
Figure 4.17 ICP-MS results for TNBC peptide-mediated delivery of MNPs into TNBC cell lines 117
Figure 5.1 Dot plots of control IgG and TNBC patient IgG pool binding to MDA-MB-231 cells 131
Figure 5.2 Control and TNBC patient IgG pool binding to MDA-MB-231 cells 132
Figure 5.3 Dot plots of control IgG and TNBC IgG binding to MCF10A and MDA-MB-231 cells 133
Figure 5.4 Control and TNBC patient IgG binding to MCF10A and MDA-MB-231 cells 134
Figure 5.5 IF images of control and patient IgG binding to MCF10A and MDA-MB-231 cells 135
Figure 5.6 Patient IgG binding to $CD44^{+}/CD24^{-}$ MDA-MB-231 cells 136
Figure 5.7 Control and patient IgG binding to MCF10A and MDA-MB-231 cell lysate 138
Figure 5.8 Control and patient IgG binding to control and TNBC cell lysate 139
Figure 5.9 Control and patient IgG binding to MCF10A and MDA-MB-231 nuclear and cytoplasmic protein extracts 141
Figure 5.10 Control and patient IgG binding to MCF10A and MDA-MB-231 cytoplasmic and membrane protein 142
Figure 5.11 Control and patient IgG binding to MCF10A and MDA-MB-231 nuclear, cytoplasmic and membrane protein 143
Figure 5.12 Silver stained MCF10A and MDA-MB-231 proteins 144
Figure 5.13 Sample preparation for LC-MS/MS 145
Figure 5.14 Overexpressed proteins in MDA-MB-231 cells identified by NSAF 147
Figure 5.15 Overexpressed proteins in MDA-MB-231 cells identified by DESeq2 149
Figure 5.16 mRNA expression levels for NSAF and DESeq2 identified proteins in healthy, tumour adjacent, non-TNBC and TNBC tissue 153
Figure 5.17 Kaplan-Meier survival plots for potential TNBC autoantibody antigens 154
Figure 5.18 Control and Patient IgG binding to HEK293T cell lysate 156
Figure 5.19 Control and patient IgG binding to TPX2 overexpression HEK293T cell lysate 156
Figure 5.20 Control and patient IgG binding to proteins KHDRBS3 overexpression HEK293T cell lysate 157

Supplementary Figure 1 pHR2-8R phagemid map 220
Supplementary Figure 2 Sanger sequencing of bacteriophage library clones 222
Supplementary Figure 3 PCR amplified nonapeptide inserts from library clones..... 223
Supplementary Figure 4 Cell viability of non-tumourigenic and TNBC cell lines after bacteriophage stripping buffer washes. 224
Supplementary Figure 5 Kullback-Leibler Seq2Logos for TNBC targeting 9-mer peptides showing the relative frequency of each amino acid residue at each position. 230
Supplementary Figure 6 pGEX-6P1-mRFP expression vector map..... 231
Supplementary Figure 7 Insertion of the 9-mer peptide encoding sequences from the Z>3 analysis into the mRFP expression vector. 238

Chapter 1: Introduction

1.1 Breast Cancer

There was an estimated 2.3 million new cases of breast cancer (BCa) diagnosed in 2020, with BCa being the most common form of non-cutaneous cancer globally (Sung et al., 2021). The disease has been estimated to affect 1 in 8 women in their lifetime, with ageing being associated with an increased risk of the disease (Siegel et al., 2022). In 2019, BCa incidence rates were 88% higher in developed countries (Australia, New Zealand and European and North American countries) compared to developing countries (Central American, African and South Asian countries). Despite this, BCa mortality rates were 17% higher in developing countries such as Melanesian, Western African and Caribbean countries (Sung et al., 2021). Incidence rates of BCa in the United States were also higher in white women compared to Black, Asian, American Indian and Hispanic women, however BCa mortality rates were 41% higher in Black women than in white women (Siegel et al., 2022).

BCa consists of a compilation of distinct malignancies that arise in the mammary glands. Rates of survival have increased with improvements in early detection and systematic treatment, however, treatment is not curative in all individuals (Blanks et al., 2000, Cossetti et al., 2015, Siegel et al., 2022). Despite these advances, BCa accounted for 15.5% of global female cancer-related deaths in 2020 and the most female cancer-related deaths in women between 20 and 59 years old (Sung et al., 2021, Siegel et al., 2022). The metastasis of BCa is the leading cause of mortality, where the tumour disseminates and moves to secondary sites in the body such as the lungs, brain and bone (Weigelt et al., 2005).

The disease is highly heterogeneous, displaying both intertumoural and intratumoural differences (Aleskandarany et al., 2018). BCa is therefore classified into distinct subtypes whose molecular profiles and pathological characteristics vary. A patient's prognosis is dependent on such variations, as the histopathological and molecular classifications of the malignancy are used to stratify treatment (Malhotra et al., 2010). Prognosis is also dependent on the stage of a breast tumour at diagnosis, where later stage BCas relate to a poorer patient survival rate (Table 1.1). In early-stage BCa, the

Nottingham Grading System (NGS) has been recommended by clinical professional bodies for histopathological classification of the tumour (Elston and Ellis, 1991). Following the NGS, tumour histological grade has been combined with tumour size and axillary lymph node status to form the Nottingham Prognostic Index (NPI) (Galea et al., 1992). Higher grade tumours are less differentiated and correlate with an increased lymph node status, making tumour grading an important indicator of metastasis and possible recurrence (Rakha et al., 2010, Ping et al., 2016). Equally, it is largely believed that tumour size strongly correlates with metastatic risk (Sopik and Narod, 2018). It is important to note that NPI is being superseded molecular classification.

Table 1.1 One year survival rates based on breast cancer stage at diagnosis in England

Stage at diagnosis	One year net-survival
Stage I	100%
Stage II	98.9%
Stage III	95.3%
Stage VI	65.8%

Women (aged 15-99) diagnosed in 2015 and followed up in 2016. National Cancer Registration and Analysis Service with Public Health England (Bannister and Broggio, 2016).

1.2 Breast Cancer Classification

The pathology of a cancerous lesion can also be determined and utilised to stratify treatment options. Carcinomas, which arise from epithelial tissue, are the most common form of BCa, with sarcomas representing around 1% of primary tumours. The histopathological classification system defines breast carcinomas as either *in situ*, invasive or metastatic carcinomas, where invasive and metastatic carcinomas often relate to a poorer outcome (Table 1.2) (Feng et al., 2018). Carcinomas *in situ* are confined to the ductolobular system of the breast and are often grade 0 (Foote and Stewart, 1941, Lee et al., 2012). The most common form of non-invasive BCa is ductal carcinoma *in situ* (DCIS), representing 20-25% of all BCa cases and 80% of *in situ* cases (Kuerer et al., 2009). DCIS form from neoplastic luminal cells that are confined to the mammary ducts. If untreated, the ductal lesions can acquire an invasive phenotype

which allows cells to penetrate the basement membrane and infiltrate the stroma (van Seijen et al., 2019). This was observed in a population-based cohort study of women identified from the NHS Screening Programme to have DCIS. The follow up of 35,024 women over five or more years found that DCIS treated patients were at an increased risk of invasive BCa compared to the general population, but intensive treatment or larger surgical margins appeared to lower the risk (Mannu et al., 2020).

Table 1.2 Common forms of breast carcinoma

Type of Carcinoma	*In situ* or invasive	Incidence	Characteristics
Ductal Carcinoma *in situ* (DCIS)	*In situ*	20-25%	Confined to the mammary ducts. No stromal invasion.
Lobular Carcinoma *in situ* (LCIS)	*In situ*	0.5-1%	Confined to the breast lobules. No stromal invasion.
Infiltrating ductal carcinoma	Invasive	80%	Neoplastic cells invade the surrounding stroma resulting in lymphovascular invasion. Includes tubular, medullary, mucinous, papillary and cribriform carcinomas.
Infiltrating lobular carcinoma	Invasive	10-15%	Neoplastic cells invade the surrounding stroma resulting in lymphovascular invasion. More common in older women.
Inflammatory breast cancer	Invasive	1-5%	Cancerous cells block lymph vessels in the breast. Symptoms include redness and swelling of the breast. Affects young women.

Information adapted from Feng et al., 2018 and Wen and Brogi, 2018

In a similar manner, lobular carcinomas *in situ* (LCIS), which are less common, can also progress into their invasive counterpart (Wen and Brogi, 2018). Invasive carcinomas are most commonly invasive ductal (IDC) or invasive lobular (ILC), with ILCs having a lower response rate to endocrine treatment and chemotherapy (Sikora et al., 2013,

Barroso-Sousa and Metzger-Filho, 2016). Due to the infiltrating nature of invasive carcinomas, cancerous cells are likely to form lymph node metastases, leading to systemic disease progression and possible mortality.

In combination with histopathology, the gene expression profile of a tumour can be utilised as a prognostic factor tool. The molecular classification is dependent on whether a particular lesion expresses the estrogen receptor (ER) or the progesterone receptor (PR), and whether overexpression of the human epidermal growth factor 2 (HER2) is observed. In addition, the nuclear non-histone protein Ki-67 has been identified as a proliferation marker which correlates with a poorer patient prognosis in luminal cancers (Soliman & Yussif, 2016). Collectively, there are five distinct molecular subtypes of BCa (Table 1.3). Hormone receptor positive BCas account for 70% of patients and their gene expression is similar to that of the luminal epithelial component of normal breast (Perou et al., 2000). Due to differences in their relative expression of hormone receptors and proliferation- related genes, luminal cancers can be further sub classified into Luminal A and Luminal B tumours (Abubakar et al., 2019). Overall, luminal cancers have the best prognosis. The luminal B subtype is more aggressive than luminal A, where tumours are often a higher histological grade, are more likely to relapse, and possess a higher proliferative potential (Eroles et al., 2012). However, not all ER^+ tumours are luminal, as a recent 80-gene signature study led to the reclassification of 13% of IHC identified $ER^+/HER2^-$ luminal tumours to basal tumours, whose pathological complete response (pCR) rate was more similar to that of ER^- basal BCas than their luminal counterparts (Krijgsman et al., 2012, Bertucci et al., 2020). Human epidermal growth factor receptor 2 (HER2) overexpressing cancers are less common, representing 15% of patients (Figueroa-Magalhães et al., 2014). The presence of malignant cells that express abnormally high levels of HER2 correlates to poor survival (Slamon et al., 1987). Despite this, the incorporation of HER2-targeted therapy into treatment regimens has led to a 50% reduction in recurrence and a 30% increase in disease-free survival rates (Romond et al., 2005, Perez et al., 2014). In contrast, basal-like BCa is not characterised by the presence of hormone or growth factor receptors, but by the expression of gene clusters similar to that of the basal-myoepithelial layer of the breast, such as cytokeratins 5/6 and laminin (Perou et al., 2000). Basal-like tumours represent 15% of clinical cases but are highly heterogeneous and invasive, leading to the worst prognosis across the molecular subtypes (Milioli et

al., 2017, Fallahpour et al., 2017). The majority of basal-like tumours are triple negative, lacking ER, PR and HER2 overexpression. Therefore, unlike luminal and HER2-enriched tumours, basal-like BCa is poorly characterised and lacks targeted treatment (Dai et al., 2015). The least common molecular subtype is normal-like BCa, characterised by gene expression profiles similar to that of adipose and non-myoepithelial cells (Sørlie et al., 2001). This rare subtype has a prognosis between that of luminal and basal-like BCa (Yersal and Barutca, 2014).

Table 1.3 The Molecular Subtypes of Breast Cancer

Molecular Subtype	Molecular Signature	Prognosis	Specific Treatment	5-year OS rate
Luminal A	ER^+,PR^+,$HER2^-$,Ki-67 low	Good	Hormone therapy	93%
Luminal B	ER^+,PR^+,$HER2^{-/+}$, Ki-67 high	Intermediate	Hormone therapy	88%
HER2 overexpression	ER^-,PR^-,HER2+	Poor	Anti-HER2 therapy	84%
Basal	ER^-,PR^-,HER2-, basal $marker^+$	Poor	No targeted therapy	83%
Normal-like	ER^-,PR^-,HER2-, Ki-67 low	Intermediate	No targeted therapy	N/A

Adapted from Eroles et al., 2012, Dai et al.,2015 and Zuo et al., 2017

In addition to the five intrinsic subtypes originally identified by Perou *et al.*, claudin-low BCa is now being explored as an additional molecular subtype of BCa. The tumours have been identified using murine and human models and express low levels of occludin, E-cadherin, and claudins 3,4 and 7, as well as high expression of mesenchymal markers (Herschkowitz et al., 2007, Sabatier et al., 2014). They can also be distinguished by their low levels of genomic instability, mutational burden, and proliferation (Fougner et al., 2019). This subtype are ductal invasive carcinomas which adopt a triple negative like phenotype and relate to a poor prognosis, with studies showing resistance to chemotherapy (Prat et al., 2010). The variation between subtypes highlights the importance of treating BCa as a compilation of distinct diseases which require differential treatment methods.

1.3 Methods of Diagnosis

For all molecular subtypes, early-stage detection greatly improves a patient's prognosis. BCa may be detected by a mammogram or from clinical breast examination (CBE), although CBE can only detect palpable tumours and relies on self-referral. Major advances in screening programmes and primary tumour diagnosis have aided in the early detection of the disease (Milosevic et al., 2018). The early detection of the disease in this manner has contributed to a global reduction in BCa mortality, measured by random control trials (RCTs) (Paci, 2012, Marmot et al., 2013, Tabár et al., 1985, Nelson et al., 2009). It is important to note that screening alone does not account for this reduction, as there have been vast improvements in treatment in conjunction with screening (Berry et al., 2005). Furthermore, early detection of asymptomatic secondary tumours has also been shown to greatly increase survival rate (Houssami et al., 2009).

In the UK, women are invited to screening through the NHS breast screening programme within 3 years of turning 50 years old. The initial age group for screening was 50-64 years, however, this has now been extended to women who are 50-70 years old, with cancer detection rates being higher in older women (Bennett et al., 2009, Moss et al., 2001). The screening programme is estimated to have led to a 20% reduction in BCa mortality, due to early detection of the disease (Marmot et al., 2013). Additionally, the programme led to a 24% increase in BCa detection between 2000-2005 (Bennett et al., 2009). However, evidence is emerging that supports the reduction of the NHS screening programme age, with data suggesting that starting screening at 40 years old results in reduced BCa mortality (Moss et al., 2015, Duffy et al., 2020). The identification of primary tumours is achieved using a mammogram, where an X-ray of the breast tissue is taken. This is usually successful and results in asymptomatic detection, but in some cases, a mammogram is not sufficient. For example, high breast density is most frequent in younger women and affects mammogram quality, resulting in obscure results or misdiagnosis (Buist et al., 2004, Checka et al., 2012). A study by Mandelson *et al.,* observed a decrease in mammographic sensitivity from 80% in women whose breasts were predominantly fat to 30% in women who had very dense breast tissue. Furthermore, higher breast density was also associated with an increased risk of interval cancer (Mandelson et al.,

2000). The radiation risk is also higher in premenopausal women, particularly women who are less than 20 years of age (Ronckers et al., 2005). Young women are therefore not included in population-based screening programmes, but screening for women with a high genetic risk of BCa may be beneficial (Oeffinger et al., 2015). For instance, germline mutations in *BRCA1* or *BRCA2* have been reported in 8-12% of early-onset BCa cases, and women carrying *BRCA1/2* mutations have a 45-80% lifetime risk of BCa (Sanjosé et al., 2003, Loman et al., 2001, Rosenberg et al., 2016, King et al., 2003, Antoniou et al., 2003). BRCA1 positive patients are also more likely to be diagnosed pre-menopause, therefore screening younger women with a genetic susceptibility to BCa would outweigh the risks in this population (Paul and Paul, 2014, Veronesi et al., 2005).

Even though population-based screening has been linked with reductions in BCa mortality, the method has been globally scrutinised (Marmot et al., 2013). Over-diagnosis refers to the treatment of tumours identified by screening and histologically confirmed, that may not have progressed to overt malignancies. The invasive nature and psychological stress of unnecessary treatment leads to over-diagnosis being the main adverse effect of mammograms (Puliti et al., 2012). Over-diagnosis is not only unique to BCa screening, as lung and prostate cancer may also be over diagnosed by X-ray or PSA detection respectively (Welch and Black, 2010). It is difficult to determine whether BCa is over-diagnosed, and estimations vary from 1% to 10% of cancer diagnoses (Duffy et al., 2005, Puliti et al., 2012, Zackrisson et al., 2006). These figures mainly relate to non-progressive ductal carcinomas *in situ*, although only 4% of DCIS have been shown to be non-progressive (Yen et al., 2003). It is therefore widely believed that the benefits of early detection through screening greatly outweighs the chance of over-diagnosis (Marmot et al., 2013).

Another undesired factor of screening is the possibility of false-positive results, leading to unnecessary further imaging, clinical examination or, in the minority of cases, a biopsy. These results can cause psychological stress for a patient, which has been shown to last up to three years and deter patients from attending screening in the future (Bond et al., 2013). McCann *et al*. found that women who had received a false-positive result at their first screen presented with interval cancer or larger tumours at following screens (McCann et al., 2002). Other studies contradict these findings,

suggesting that an individual who has received a false-positive result previously is not less likely to attend further mammograms in the future (Johns et al., 2010). False-negative results also occur, however, they are much less frequent than false-positive scans (Nelson et al., 2016). They are influenced by high breast density and by histopathological classification, as lobular carcinomas are more difficult to identify due to their similarities with normal breast tissue (Seely and Alhassan, 2018).

Although mammography is the most common method of BCa detection, the disease can also be detected via other supplementary imaging methods, such as ultrasonography and magnetic resonance imaging (MRI). If a mammogram yields positive or inconclusive results, ultrasonography is utilised by clinicians to assess the lesion and guide biopsy (Appleton et al., 2014). For dense breast tissue that is classified as 3 or 4 by the breast imaging and reporting system (BI-RADS), ultrasound scans may also be employed as mammograms are less sensitive in these cases (Okello et al., 2014). This technique is non-invasive, does not use ionising radiation, and has been shown to have a greater sensitivity for detection in younger women (Corsetti et al., 2008). MRI scans are more sensitive than mammography, leading to earlier cancer detection, especially in high risk younger women with dense breast tissue (Ahern et al., 2014). However, the cost of an MRI scan is high, and the increased sensitivity of contrast-enhanced MRI may lead to low specificity and an increased number of over-diagnosed cases (Leach et al., 2005, Mann et al., 2019).

1.4 Treatment and Resistance

Once a patient is diagnosed with BCa, a treatment regimen is decided based on the pathological characteristics and the molecular subtype of the disease. Conventional treatment of BCa includes surgery, chemotherapy and radiotherapy. Depending on the size and histological grade of a primary tumour, an individual may be advised to have breast conserving surgery or, less frequently, a full mastectomy. The current standard treatment for early-stage BCa is breast conserving surgery followed by radiotherapy (Lovrics et al., 2009). Studies have shown that patients who receive breast conserving surgery with postoperative radiotherapy have similar rates of survival with less psychosocial impact on the individual compared to a full mastectomy (Fisher et al., 2002, Litière et al., 2012). In addition, the systemic treatment of early invasive BCa by

adjuvant chemotherapy such as CMF (cyclophosphamide, methotrexate and 5-fluorouracil) or anthracyclines reduces mortality and metastatic progression (Rossi et al., 2015). Neoadjuvant chemotherapy may also be administered in order to reduce tumour size prior to surgery, however administering chemotherapy in this manner leads to an increased risk of local recurrence compared to in an adjuvant setting (Asselain et al., 2018). In contrast, for locally advanced BCas, neoadjuvant chemotherapy can render inoperable tumours operable by breast conserving therapy (Swain et al., 1987, Klein et al., 2019). Node positive BCa neoadjuvant chemotherapy includes the use of anthracyclines and possibly taxanes (Gogas and Fountzilas, 2003, Henderson et al., 2003). Patients who are diagnosed with metastatic BCa are treated in a similar manner, with anthracyclines as a first-line treatment followed by taxanes, however metastatic BCa is rarely curable and patients eventually relapse (Crown et al., 2002, Claessens et al., 2020).

As aforementioned, prognostic markers of the disease can be utilised for molecular targeted treatment. The Her-2/neu oncogene is overexpressed in 30% of BCas, particularly in the malignant population (Slamon et al., 1987). Early *in vitro* and *in vivo* studies targeted the extracellular domain of the HER2 tyrosine kinase which subsequently inhibited cell proliferation, leading to the humanisation of a HER2 targeting monoclonal murine antibody to form trastuzumab (Carter et al., 1992, Pietras et al., 1998). A phase three trial by Slamon *et al.* found that targeting HER2 amplification with trastuzumab, along with first-line chemotherapy, increased treatment response thus patient survival in metastatic BCa (Slamon et al., 2001). When trastuzumab is administered with pertuzumab and docetaxel, a review of RCTs found that HER2-directed therapies worked synergistically and had an undeniable impact on the overall survival of patients with metastatic BCa (Mendes et al., 2015). Due to the clinical efficacy of administering trastuzumab with chemotherapy, trastuzumab has also been considered in antibody-drug conjugate therapeutics. Furthermore, HER2 is overexpressed in breast tumours and expressed at low levels in normal tissue, making HER2 monoclonal antibodies a good candidate for target treatment (Press et al., 1990). Another emerging target treatment for HER2 overexpressing BCas is lapatinib. The dual tyrosine kinase inhibitor, which interrupts both HER2 and ER signalling, progressed through phase III clinical trials and has been shown to increase progression-free survival (PFS) of $ER^+/HER2^+$ metastatic BCa patients

when administered in combination with capecitabine (Seebacher et al., 2019, Yang et al., 2020).

In ER^+ BCas, selective estrogen receptor modulators, termed SERMS, act as antagonists of the ER in mammary cells and lead to a high response rate in these patients (Johnston and Dowsett, 2003). The agonistic or antagonistic action of SERMS is dependent on the differential expression of the ER and corresponding ligand-receptor interactions, the interaction of SERMS with various co-activators and co-repressors, and the tissue specificity of the non-steroidal compounds (Martinkovich et al., 2014, Riggs and Hartmann, 2003). For instance, Shang and Brown demonstrated that the SERM tamoxifen induces the recruitment of corepressors to ER target genes in mammary cells and the recruitment of coactivators in endometrial cells (Shang and Brown, 2002). In hormone receptor positive BCas, tamoxifen is an anti-estrogen therapy which increases patient survival and lowers recurrence rate (Early Breast Cancer Trialists' Collaborative, 1998, Ward, 1973, Kooreman et al., 2018, Osborne, 1998). Adverse effects of tamoxifen include thrombosis and increased risk of endometrial cancers (Cuzick et al., 2003). In postmenopausal women, tamoxifen may be used alone or with aromatase inhibitors, which reduces recurrence risk compared to 5 years of tamoxifen alone (Burstein et al., 2010). Aromatase inhibitors result in estrogen deprivation to reduce ER activity, but also adversely lead to bone loss (osteoporosis) during treatment (Perez and Weilbaecher, 2006). In addition, the selective ER downregulator/degrader (SERD) fulvestrant was also found to increase the PFS of women with ER^+ and PR^+ locally advanced or metastatic BCa, with the median PFS being 16.6 months in the fulvestrant group vs 13.8 months in the anastozole group (Robertson et al., 2016). From this, fulvestrant became a Food and Drug Administration (FDA) approved treatment of ER^+ advanced BCas and displays clinical efficacy when administered in combination with signalling inhibitors targeting endocrine resistance pathways (Rocca et al., 2018).

There are also other emerging target therapies for BCa. Checkpoint inhibitors such as cyclin D-CDK4/CDK6 inhibitors block cell cycle progression and the subsequent proliferation of malignant cells (Toogood et al., 2005). In hormone receptor positive BCa, the inhibitors palbociclib and ribociclib significantly improved PFS when administered in conjunction with endocrine therapy (Shah and Cristofanilli, 2017,

Hortobagyi et al., 2016). Another class of inhibitors used to treat BCa are the PI3K/Akt/mTOR inhibitors. The phosphoinoside-3-kinase (PI3K)/Akt/mTOR pathway is often activated in hormone receptor positive BCa and regulates cell survival, proliferation and metabolism (Dey et al., 2017). The upregulation of this pathway is associated with endocrine therapy resistance, with PI3KCA being significantly mutated in luminal and ER^+ BCa (Miller et al., 2010, Ellis and Perou, 2013). PI3K, Akt and mTOR inhibitors have each displayed promising clinical efficacy when used in combination with trastuzumab in $HER2^+$ advanced BCa (Tolaney et al., 2015, Hudis et al., 2013, Seiler et al., 2015) . Poly (ADP-ribose) polymerase (PARP) inhibitors have also shown clinical response, particularly in cancers harbouring *BRCA1/2* mutations such as TNBC. The PARP inhibitor Olaparib prevents single strand break repairs leading to subsequent apoptosis of tumour cells. When used as a monotherapy in a phase III trial, PFS of $HER2^-$ patients with *BRCA1* mutations was extended by 2.8 months and risk of disease progression decreased by 42% when compared to standard therapy (Robson et al., 2017). Other potent PARP inhibitors are currently progressing through trials with promising clinical efficacy.

Even though prognostic markers have enabled the molecular subgrouping of BCa and thus the development of targeted therapies, many patients still experience tumour relapse after remission. Tumour relapse is believed to occur due to intrinsic or acquired resistance. With regard to systemic therapy, the use of a single chemotherapy agent often elicits a response, however acquired drug resistance and tumour progression is inevitable. In metastatic BCa patients, drug resistance leads to failure of treatment in over 90% of patients (Longley and Johnston, 2005). Anthracyclines are mainly believed to develop resistance through increased drug efflux and alteration of topoisomerase II expression, whereas resistance to taxanes may occur through increases in multi-drug resistance protein 1 (MDR1) expression and the altered expression of β-tubulin isotopes (Gonzalez-Angulo et al., 2007, Saeki et al., 2005). Patients can also acquire resistance to antibody-based and endocrine therapy. Despite the increase in overall survival of $HER2^+$ BCa patients who receive trastuzumab and chemotherapy, metastatic patients acquire resistance to trastuzumab within a year (Nahta et al., 2006). Trastuzumab resistance is believed to involve the activation of the PI3K/Akt/mTOR pathway and loss of the tumour suppressor PTEN, disruption of antibody-receptor binding via overexpression of mucin-4 (MUC4), and upregulation of

other signalling pathways like insulin-signalling (Nagy et al., 2005, Lu et al., 2001, Nagata et al., 2004). In contrast, endocrine therapy resistance is believed to occur through loss of intact ERα expression with increased expression of truncated isoforms of the receptor, activation of PI3K pathway activity, and increased AP1 activity as a result of tamoxifen-induced oxidative stress (Schiff et al., 2000, Alistair and Mitch, 2004, Musgrove and Sutherland, 2009). To overcome acquired treatment resistance to available therapies, combination therapy has been employed in clinical practice.

1.5 Therapy Resistance and Breast Cancer Stem Cells

The intrinsic chemoresistance, radioresistance and therapeutic resistance of BCa has also been linked to intratumoural heterogeneity and the existence of a subpopulation of undifferentiated cells termed cancer stem cells (CSCs) (Peitzsch et al., 2017). In normal tissue, somatic stem cells possess the ability to self-renew and to differentiate into multiple lineages to maintain tissue structure and function (Goodell et al., 2015). The existence of a similar population of cells in cancer was first alluded to in acute myeloid leukaemia, where only $CD34^{+}/CD38^{-}$ leukaemia cells could initiate tumourigenesis in NOD/SCID mice (Lapidot et al., 1994). Furthermore, these cells could differentiate into leukemic blasts, inferring that a cell hierarchy was present (Bonnet and Dick, 1997). This work led to the identification of tumour initiating cells, termed cancer stems cells (CSCs), in a range of solid tumours including BCa (Al-Hajj et al., 2003). Conversely, there has been conflicting evidence on the existence of CSCs. For instance, Quintana *et al*., found that 28% of single melanoma cells obtained from patients were able to form tumours in NOD/SCID IL2Rγ^{null} mice and suggested that differences in tumourigenicity in melanoma cells is due to phenotypic plasticity and not the presence of a cell hierarchy (Quintana et al., 2008, Quintana et al., 2010). However, there is a considerable amount of research which does correlate with the existence of a CSC subpopulation (Batlle and Clevers, 2017).

In healthy tissue, the heterogeneous collection of different cell types that collectively form tissues and organs arise from stem cells in a hierarchal manner (Nguyen et al., 2012). The CSC theory describes a CSC hierarchy similar to that of normal stem cells, where CSCs may initiate and sustain the tumour by asymmetric division to self-renew and maintain the CSC population, while producing progenitor cells which differentiate

and contribute to the tumour mass (Saeg and Anbalagan, 2018). This is controversial, as is it has also been hypothesised that neoplastic stem cells may divide symmetrically to expand the stem cell population and promote aneuploidy (Morrison and Kimble, 2006). The origin of CSCs also remains unclear. Proposed models suggest that breast cancer stem cells (BCSCs) are formed when multipotent mammary stem cells, progenitor cells or differentiated cells acquire a series of genetic mutations which lead to the conversion or de-differentiation of mammary cells to malignant cells with "stem-like" properties (Sin and Lim, 2017).

BCSCs have been suggested to be drivers of tumour progression and conventional therapy resistance (Zhang et al., 2020). This subpopulation of cells become enriched after treatment and can form mammospheres in culture, demonstrating that BCSCs may initiate tumour relapse (Li et al., 2008). As CSCs are mainly quiescent, they are protected from anti-proliferative chemotherapy (Chen et al., 2021, Chen et al., 2012). CSCs have also been shown to over express ABC transporters to increase substrate efflux, thus protecting breast cancer stem cells (BCSCs) from chemotherapeutic drugs (Begicevic and Falasca, 2017). Additionally, the inhibition of ABC transporters or reversal of CSC multi-drug resistance has been shown to sensitise chemoresistant cells to therapy (Wu et al., 2008). Radiotherapy also enriches for CSCs and promotes invasiveness through epithelial-mesenchymal transition (EMT) and increased migration (Ko et al., 2018). CSCs possess a high DNA repair capacity thorough activation of ATR-Chk1 and ATM-Chk2, allowing CSCs to repair radiation-induced lesions (Yin and Glass, 2011).

BCSCs can be identified by cell surface marker expression. This was demonstrated by Al-Hajj *et al.,* who found that $CD44^+/CD24^-/lineage^-$ cells extracted from solid primary breast xenografts initiated tumorigenesis in immunocompromised mice (Al-Hajj et al., 2003). BCSCs also exhibit high cell plasticity across the molecular subtypes, with $CD44^+/CD24^-$ BCSC populations expressing enhanced mesenchymal markers (Sheridan et al., 2006). For example, $CD44^+/CD24^-$ cells possess a higher stem cell ability when identified in combination with EpCAM as a reference marker (Ghebeh et al., 2013). Furthermore, this sub-population of cells transits between either an epithelial-like proliferative state or a mesenchymal-like state which is less proliferative and highly invasive. The transition of cancer stem cells to a mesenchymal phenotype preserves

tumour ‘stemness’ and facilitates cell migration. This suggests that the microenvironment initiates EMT in BCSCs to enable metastasis and the invasion of cancer stem cells into secondary organs (Thiery et al., 2009, Liu et al., 2014). It has also been suggested that CSCs are able to readily undergo EMT as they possess epigenetic and metabolic plasticity (Walsh et al., 2019). For instance, Fiorello *et al.,* identified a hyper proliferative, energy-driven subset of BCSCs that displayed an enhanced capability to undertake anchorage-independent growth and metastasise (Fiorillo et al., 2019).

To prevent residual disease and subsequent tumour relapse, CSC targeting is now being explored. The CSC tumour microenvironment is primed by CSC mediated activation of multiple signalling pathways, leading to the release of cytokines which prime the metastatic niche for CSC invasion (López de Andrés et al., 2020). To regulate cell differentiation and self-renewal, CSCs use signalling pathways such as Sonic hedgehog, Wnt, Notch and JAK/STAT3 (Cochrane et al., 2015, Rivas et al., 2018, Wang et al., 2018, Venkatesh et al., 2018). For instance, the inhibition of Wnt signalling via frazzled receptors was found to decrease the tumourigenicity of multiple malignant xenografts (Gurney et al., 2012). Other research also includes a study by McGowen *et al.,* which found that the inhibition of Notch1 in MDA-MB-231 cells decreased the CD44+/CD24- subpopulation (McGowan et al., 2011). Interestingly, TNBC stem cells were also depleted when TNBC derived cell lines were exposed to metformlin, a type 2 diabetes drug which was shown to suppress TNBC stemness through the PKA-GSK3β-KLF5 pathway (Shi et al., 2017). With regard to targeting CSC signalling pathways, clinical trials targeting cancer stemness have been conducted, including the targeting of a hedgehog signalling pathway protein Smoothened (Smo) with Sonidegib in advanced triple negative breast cancer patients (Ruiz-Borrego et al., 2019).

Targeting epigenetic modifications in CSCs has also been speculated, where combination therapy of epigenetic drugs with chemotherapy may re-sensitise chemoresistant CSC populations to treatment. For instance, the delivery of a DNA hypermethylation inhibitor encapsulated in a nanoparticle also increased the sensitivity of BCSCs in MDA-MB-231 cells and breast xenografts to doxorubicin, decreasing the BCSC population and inhibiting tumour growth (Li et al., 2015). Furthermore, a study by Schech *et al.,* the histone deacetylating inhibitor, entinostat,

reduced the percentage of tumour initiating cells in triple negative breast cancer cells (Schech et al., 2015). Different therapeutic agents which target CSCs and the CSC microenvironment are progressing through clinical trials for cancer treatment, including the use of drugs, monoclonal antibodies and chimeric antigen receptor T-cell (CAR-T) therapies, thus highlighting the clinical interest in CSC-directed therapies (Yang et al., 2020).

1.6 Triple Negative Breast Cancer

A particular BCa subtype that has been associated with high tumour heterogeneity, BCSC enrichment and treatment resistance is triple negative breast cancer (Park et al., 2019, Bianchini et al., 2022). Around 15-20% of all BCa patients have TNBC. The disease is prevalent in younger women, particularly in African-American and Hispanic women (Feng et al., 2018, Aysola et al., 2013). Socioeconomic and biological factors are believed to account for racial disparities in TNBC. For instance, Jenkins *et al.,* demonstrated that *DARC/ACKR1* mutations which are specific to Sub-Saharan African ancestry are associated with pro-inflammatory chemokines related to tumour progression (Jenkins et al., 2019). TNBC has a higher relapse and lower survival rate compared to hormone receptor BCas, which is mainly due to a lower age of onset, lymph node involvement at diagnosis, higher tumour heterogeneity and greater metastatic potential (Dent et al., 2007, Haffty et al., 2006). Approximately 80% of TNBCs are basal-like according to gene expression profiling (Dawood, 2010, Yersal and Barutca, 2014). Even though the majority of TNBC cases are defined as basal-like through the expression of basal/myoepithelial cell markers such as cytokeratin 5, 14 and 17 and vimentin, some tumours are defined as more normal breast-like from their gene expression profiles (Dawood, 2010, Fan et al., 2006, Foulkes et al., 2010). However, Prat *et al.,* found that none of the intrinsic subtypes of TNBC led to differences in pCR rates or survival after neoadjuvant chemotherapy (Prat et al., 2014).

In a study by Dent *et al*., triple negative tumours were found in patients through clinical examination rather than imaging. This may have been due to either TNBC tumours expanding rapidly within screening intervals, or the density of breast tissue in women with TNBC effecting detectability (Dent et al., 2007). Furthermore, a study looking into interval cancers found by the population-based Norwegian Breast Cancer

Screening Programme between 1996 and 2001 discovered that triple negative, basal-like tumours were more likely to present in the interval between mammograms than the other molecular subtypes (Collett et al., 2005). This stresses the need for alternative diagnostic tools to compliment screening programmes in the detection of TNBC.

Frequent mutations observed in TNBC are related to the tumour suppressor genes *BRCA1* and *BRCA2*. The increased risk of BCa among women who harbour BRCA mutations was found to be 60% in women who had no first-degree relative and 64% in women with a first-degree relative with BCa (Metcalfe et al., 2018). With regard to TNBC, around 30% of TNBC patients carry either *BRCA1* or *BRCA2* mutations, although this proportion varies with ethnicity and age (Greenup et al., 2013, Engel et al., 2018). The prevalence of *BRCA1* mutations may relate to the development of TNBC specifically, as women with BCas that are *BRCA1* mutated are more likely to have higher grade tumours which are triple negative. Woman carrying *BRCA1* mutations are also more likely to develop TNBC than *BRCA2* mutation carriers (Chen et al., 2018). Furthermore, the majority of TNBC patients who harbour *BRCA1* mutations also have *TP53* mutations (Greenblatt et al., 2001, Holstege et al., 2009). In addition, other gene variants such as *BARD1*, *PALB2* and *RAD51D* mutations are observed in TNBC (Shimelis et al., 2018).

TNBC cannot be targeted by hormonal therapies or monoclonal antibodies that are used to treat luminal or HER2 overexpressing cancers, respectively (Eroles et al., 2012). Therefore, the main systemic treatment method for TNBC is chemotherapy (Zagami and Carey, 2022). The first-line of therapy for advanced TNBC is neoadjuvant chemotherapy, which has an objective response rate of 20-35% and therefore a poor outcome in the majority of patients (Khosravi-Shahi et al., 2019). Even though neoadjuvant chemotherapy elicits the highest response rate in TNBC compared to other BCa subtypes, TNBC patients have a higher likelihood of recurrence than non-TNBC patients within the first 5 years of diagnosis (Haffty et al., 2006). Within the TNBC subtype, patients who harbour hereditary *BRCA1* mutations have the highest response rates to neoadjuvant chemotherapy (Silver et al., 2010). Similarly, TNBC responds well to adjuvant chemotherapy post-surgery, but patients have a higher risk of relapse within 3 to 5 years after initial chemotherapy (Park et al., 2018). Overall,

chemotherapy has varying efficacy for TNBC patients, highlighting tumour heterogeneity within the molecular subtype (Masuda et al., 2013).

The differences in clinical response rate among TNBC patients has led to the identification of distinct TNBC subtypes. Using gene cluster analysis, Lehnmann *et al.* originally identified 6 intrinsic TNBC subtypes, including basal-like subtypes 1 and 2 (BL1 and BL2), immunomodulatory (IM), mesenchymal stem-like (MSL), mesenchymal (M) and luminal androgen receptor (LAR) subtypes (Lehmann et al., 2011). Furthermore, the therapeutic response of each TNBC subtype was investigated in preclinical models. Subsequent studies refined the classification to 4 subtypes (BL1, BL2, M and LAR) as the gene expression profiles of IM and MSL subtypes were influenced by infiltrating lymphocytes and stromal cells in tumour sections (Lehmann et al., 2016). Burstein *et al*. confirmed this, further identifying the basal-like subtypes as basal-like immunosuppressed (BLIS) or basal-like immune-activated (BLIA), with BLIS tumours having the worst prognosis (Burstein et al., 2015). The BL1 subtype is enriched for cell cycle and DNA damage response genes, whilst the BL2 subtype is characterised by high expression of myoepithelial markers and growth factor signalling components and resulted in lower pCR rates than in patients with LAR or BL2 TNBC (Lehmann et al., 2016). The mesenchymal subtype (M) is associated with high expression of genes involved in cell motility and ECM interaction, and share features with chemoresistant, metaplastic BCas. The most differential subtype, LAR, has gene expression patterns similar to luminal tumours and, although ER^-, the subtype is driven by hormone-related pathways such as androgen signalling (Lehmann et al., 2011).

As predicted, the TNBC subtypes respond differently to chemotherapy and targeted treatment. The least proliferative subtype, LAR, has been associated with a better prognosis than other TNBC subtypes as LAR tumours are highly differentiated, non-basal like apocrine carcinomas (Masuda et al., 2013, Yu et al., 2013, Meattini et al., 2018). Conversely, other studies have suggested that LAR tumours are the most resistant to neoadjuvant chemotherapy and are associated with poor prognosis (Santonja et al., 2018, Bareche et al., 2018). As LAR tumours are androgen receptor (AR) positive, clinical trials have targeted these tumours with anti-androgen therapy with some clinical efficacy (Gucalp et al., 2013, Bonnefoi et al., 2016, Lehmann et al., 2020). In contrast, the BL1 subtype are characterised by high tumour mutational

burden and are particularly sensitive to platinum based alkylating agents (Santonja et al., 2018). In addition, BL1 tumours exhibited the highest pCR compared to BL-2 and LAR tumours when treated with neoadjuvant anthracycline and cyclophosphamide followed by taxane-based therapy (Lehmann et al., 2016). Similar to BL1, the M-subtype is associated with high tumour mutational burden, which has been linked with increased sensitivity to immune checkpoint inhibitors (Barroso-Sousa et al., 2020, O'Meara and Tolaney, 2021). However, the M-subtype has low PD-L1 expression, indicating that PD-L1 directed immunotherapy would not be successful when used in combination with chemotherapy in patients this intrinsic subtype (Lehmann et al., 2021). The varied pCRs observed in TNBC patients to chemotherapy and current treatment regimens emphasises the need for targeted therapy which may be used for the treatment of all TNBC subtypes (Masuda et al., 2013).

1.7 Targeted therapies for TNBC

Molecular and functional studies in TNBC have uncovered pathway alterations in the disease, therefore highlighting key signalling pathways and proteins that may be promising therapeutics targets (Bianchini et al., 2022). An example of a molecular pathway which is subject to genetic alteration in TNBC is the homologous recombination repair pathway (HRR), with genetic mutations or epigenetic silencing of HRR proteins BRCA1/2, PALB2 and RAD51C being associated with TNBC (Polak et al., 2017, Liao et al., 2021). As aforementioned, PARP inhibitors are progressing through clinical trials and demonstrating clinical efficacy in the targeted treatment of TNBC. For instance, OlympiAD and EMBRACA phase III trials led to the FDA and the European Medicine Agency (EMA) approval of the PARP inhibitor olaparib and talazoparib for the treatment of $HER2^-$ metastatic breast patients harbouring *BRCA1/2* mutations, respectively (Robson et al., 2017, Litton et al., 2018). However, follow-up data for both trials revealed that neither olaparib or talazoparib improved OS compared to chemotherapy (Robson et al., 2019, Litton et al., 2020). As the PI3K/AKT signalling pathway is often activated in TNBC through PIK3CA, AKT1 or MTOR activation and/or loss of PTEN and INPP4B, the PI3K pathway has also been identified as a molecular target for TNBC therapies (Bianchini et al., 2022). Examples of inhibitors targeting this pathway in clinical trials for the treatment of TNBC include the PI3Kα inhibitor alpelisib and AKT inhibitors ipatasertib and capivasertib (Kim et al., 2017, Schmid et al., 2019,

Oliveira et al., 2019, Sharma et al., 2021). Combination therapy aimed at different TNBC molecular targets is also being explored. For instance, a phase Ib/II clinical trial found that targeting AR^+ metastatic TNBC with enzalutamide in combination with the PI3K inhibitor taselisib led to an increased clinical benefit rate being observed (Lehmann et al., 2020).

Previously, breast tumours were considered as 'cold' and immunologically quiescent due to relatively low levels of T cell infiltration (Gatti-Mays et al., 2019). However, evidence supporting the role of tumour-infiltrating lymphocytes (TILs) as a predictive and prognostic marker of a higher pCR rate to neoadjuvant chemotherapy in early TNBC is emerging (Loi et al., 2019, Criscitiello et al., 2016, Denkert et al., 2018). A study by Dieci *et al.,* found that the ten year overall survival rate of TNBC patients with a high number of stromal and intratumoural TILS was 89% compared to 68% in low TILs tumours (Dieci et al., 2015). In addition, Althobiti *et al.* found a higher percentage of TILs in TNBC samples compared to other BCa subtypes*,* with high numbers of average stromal TILs and $CD3^+$, $CD8^+$ and $CD20^+$ cells being associated with longer survival in TNBC patients (Althobiti et al., 2018). More specifically, a high number of $CD8^+$ TILS and a high $CD8^+/FOXP_3$ ratio in residual TNBC tumours after neoadjuvant chemotherapy has been shown to increase recurrence-free survival (RFS) and BCa specific survival (Miyashita et al., 2015).A similar relationship between $CD8^+/FOXP_3$ ratio and RFS has also been highlighted in residual cervical tumours (Liang et al., 2018). Individuals with primary invasive TNBC tumours possessing a high number of $CD8^+$ T cells were also found to have a better overall survival rate (Abdou et al., 2020).

As well as immune cell infiltration, the expression of the immune cell evasion molecule programme death ligand-1, or PD-L1, has been linked to TNBC patient prognosis. The presence of checkpoint molecules such as PD-L1 has resulted in the development of immune checkpoint inhibitors, a now promising class of immunotherapeutics for TNBC. PD-L1 expression is common among high grade, hormone receptor negative BCas and expression is observed in around 20% of TNBC tumours (Mittendorf et al., 2014, Dill et al., 2017). Based on results from the phase III IMpassion130 clinical trial, the anti-PD-L1 agent atezolizumab in combination with nanoparticle albumin bound paclitaxel has recently been approved by the National Institute for Health Care Excellence (NICE) and the FDA for use in $PD\text{-}L1^+$ advanced TNBC (Schmid et al., 2018,

NICE 2020, Reddy et al., 2020). Trial results demonstrated how atezolizumab with nab-paclitaxel increased the PFS of women in the intention-to-treat (7.2 vs 5.5 months) and PD-L1 positive (7.5 vs 5 months) subgroups compared to nab-paclitaxel only. With regard to early TNBC, the PD-L1 inhibitor pembrolizumab is progressing through clinical trials with promising phase 1b results, demonstrating high 12 month overall survival rates (80-100%) and manageable toxicity (Schmid et al., 2020). However, current TNBC immunotherapies are still only effective in a subset of patients.

An emerging class of immunotherapies for the targeted treatment of TNBC are antibody-drug conjugates, which may be utilised to directly deliver cytotoxic agents to cancerous cells that express a particular antigen. ADCs exploit the specificity of monoclonal antibodies for target cell antigens to deliver cytotoxic agents which are covalently linked to the monoclonal antibody. The humanised antibody must bind to the target antigen with a high affinity (Kd <10nM) to allow for efficient uptake into the target cell (Panowski et al., 2014). Upon antigen binding, the ADC is usually internalised by receptor mediated endocytosis, leading to the degradation of a cleavable or non-cleavable linker by intracellular conditions or by proteasomal or lysosomal degradation, respectively (Sievers and Senter, 2013). The target-specific delivery of ADCs improves both the efficacy and safety of cytotoxic drugs by increasing the cytotoxic load to target cells whilst decreasing adverse effects on non-target cells through minimised systemic drug exposure (Trail et al., 2018). For example, a clinically available ADC therapy for Hodgkinson's lymphoma patients is brentuximab vedotin. The ADC achieved a high efficacy and low toxicity by targeting the delivery of monomethyl auristatin E (MMAE), a highly toxic antimitotic agent which cannot be administered systemically, directly to CD30 expressing Reed-Sternberg cells (Kaloyannidis et al., 2020, Whitt et al., 2020). Non-target tumour cells may also be subjected to the cytotoxic agent when antigen expressing tumour cells release the cytotoxin into the surrounding microenvironment, further improving the efficacy of ADCs by the bystander effect (Singh et al., 2016).

Regarding BCa, trastuzumab deruxtecan is an ADC which is approved by the EMA and FDA for the treatment of $HER2^{+}$ metastatic BCa (Duranti et al., 2021, Guo et al., 2022). The ADC consists of trastuzumab conjugated to the topoisomerase I inhibitor deruxtecan via a tetrapeptide-based linker which is selectively cleaved by cathepsins

(Modi et al., 2019, Hurvitz et al., 2023). In the phase III DESTINY-Breast02 trial, trastuzumab deruxtecan led to significantly longer median PFS in patients treated with trastuzumab deruxtecan (17.8 months) compared to patients treated with capecitabine plus trasuzumab or capecitabine plus lapatinib (6.9 months) (Hurvitz et al., 2023). In addition, the ADC led to longer PFS and OS in HER2-low metastatic BCa patients compared to chemotherapy (Modi et al., 2022). Ado-trastuzumab emtansine, or TDM-1, is an FDA approved ADC for the adjuvant treatment of $HER2^+$ early BCa. The ADC TDM-1 combines trastuzumab with the antimicrotuble emtansine via a thioester linkage. Upon HER2-mediated internalisation, emtansine is released into the cytoplasm via lysosomal degradation and inhibits tubulin polymerisation, leading to cell-cycle arrest and subsequent apoptosis (Hardy-Werbin et al., 2019). Approval was based on the KATHERINE trial, where patients who received TDM-1 demonstrated a disease-free survival of 88% compared to 77% in patients who received trastuzumab alone (Wedam et al., 2020). Furthermore, the phase III EMILIA clinical trial studied the efficacy of TDM-1 compared to capecitabine–lapatinib in $HER2^+$ BCa patients who had CNS metastases and who had previously received trastuzumab and taxane combination therapy. TDM-1 was associated with a significant increase in overall survival (26.8 vs 12.9 months), resulting in FDA approval of TDM-1 for second line treatment (Amiri-Kordestani et al., 2014, Krop et al., 2015, Krop et al., 2017).

An ADC which is progressing through clinical trials for the treatment of TNBC is sacituzamab govitican, or IMMU-32, which consists of the Trop-2 targeting mAb hSR7 coupled to SN-8, a potent topoisomerase inhibitor. The pH sensitive cleavable linker allows for drug delivery to both TROP-2 overexpressing cells and bystander cells in the tumour microenvironment (Bardia et al., 2019, Bardia et al., 2017) . The $Ca2^+$ signal transducer Trop-2 is overexpressed in many solid cancers, including 90% of TNBC (Huang et al., 2005, Bardia et al., 2017). Phase II trials resulted in a 33.3% response rate in heavily pre-treated metastatic TNBC patients (Bardia et al., 2019). Other early-stage trials are also ongoing with ADCs for TNBC treatment, such as a phase I trial with SGN-LIV1a, where an anti-LIV-1 is conjugated to monomethyl auristatin E via a protease cleavable linker. Trial results found that LIV-1 was highly expressed in 75% of the metastatic BCa tumour samples and that SGN-LIV1a displayed antitumour activity in heavily pre-treated metastatic TNBC (Nagayama et al., 2020, Modi et al., 2018).

In a similar manner to ADCs, peptide-drug conjugates (PDCs) are also designed for the targeted delivery of cytotoxins to malignant cell. In contrast to antibodies, peptides can be readily synthesised and modified, display low immunogenicity and toxicity, and can possess the ability to cross cell membranes (McGregor, 2008, Boohaker et al., 2012, Marqus et al., 2017). In regard to TNBC, TH1902 is a PDC currently progressing through phase 1 clinical trials for the targeted treatment of SORT1 expressing solid cancers including triple negative tumours (Theratechnologies and PPD, 2021). When conjugated with docetaxel, the PDC was found to have potent antimigratory and cell-death inducing effects on MDA-MB-231 cells. Further *in vivo* studies also resulted in increased tumour regression in TNBC xenograft models than with docetaxel (Demeule et al., 2021). In addition, *Charfi* et al., found that TH1902/TH1904 is able to efficiently inhibit vasculogenic mimicry in TNBC cell models (Charfi et al., 2021). This highlights the potential of peptide-drug delivery systems in targeting aggressive cancers such as TNBC.

1.8 TNBC Unmet Clinical Needs

Early detection of TNBC by mammographic screening has been shown to improve patient outcome (Chen et al., 2020). However, as previously mentioned, TNBC primarily affects premenopausal women who have not yet reached the age for the NHS screening programme and the disease often presents in screening intervals (Dent et al., 2007, Bennett et al., 2009). Furthermore, denser breast tissue in young women can lead to inconclusive results being obtained by current imaging methods, which emphasises the clinical need for complementary diagnostic tests to aid in the early detection of TNBC (Checka et al., 2012).

Blood-based liquid biopsies are emerging as diagnostic tests for the early detection of cancer (Lone et al., 2022). Recent studies have focused on the presence of various tumour associated analytes in the blood and investigated their use as diagnostic and prognostic markers of TNBC, including circulating nucleic acids, circulating tumour cells, exosomes and tumour proteins (Stevic et al., 2018, Henderson et al., 2019, Radovich et al., 2020, Dass et al., 2021). When compared to invasive biopsies, screening for blood-based tumour biomarkers is low cost and minimally invasive, and can be used to monitor disease progression and response to treatment (Radovich et

al., 2020, Dass et al., 2021). Therefore, liquid biopsies could be utilised as a BCa screening method for young women at high risk of TNBC (Farina et al., 2017).

TNBC targeted treatment also remains a major unmet clinical need, as new targeted therapeutics are needed to overcome therapy resistance and to subsequently improve patient outcome. In recent years, the development of TNBC targeted therapies such as PARP inhibitors, ADCs and immune checkpoint inhibitors have improved the treatment landscape for patients with early-stage and metastatic TNBC. However, a lack of therapeutic targets has led to chemotherapy remaining as the main treatment modality for patients with TNBC, despite the disease being associated with chemotherapy resistance and disease recurrence (Nedeljkovic et al., 2019, Zagami and Carey, 2022). The site-specific delivery of anticancer agents via targeting ligands has been shown to increase the efficacy of chemotherapeutic agents against TNBC, thus highlighting the potential of drug delivery systems in TNBC treatment (Bardia et al., 2019, Demeule et al., 2021). Therefore, the work described in this thesis aims to address these unmet clinical needs and to directly investigate two main areas of early detection and targeted treatment by using advanced screening methods and novel drug delivery technologies.

1.9 Project aims and objectives

The aim of this study was to identify tumour associated antigens and small peptides with the potential to act as targeting ligands for TNBC targeted therapeutics. TNBC currently has the worse prognosis among the molecular subtypes and has a limited number of target therapies available. Therefore, the development of a TNBC targeted therapy would address a currently unmet clinical need.

To identify a TNBC selective peptide, we aimed to conduct a non-bias high-throughput screen of small peptides against TNBC cells by next-generation phage display. The identification and subsequent characterisation of a TNBC peptide could therefore lead to a peptide-conjugated or peptide-based delivery system being developed for the treatment of the disease.

We also hypothesised that TNBC specific antigens may be identified via cancer autoantibodies raised in TNBC patients and be targeted by immunotherapy or antibody-peptide fusion therapies. The identification of TNBC-specific autoantibodies in patient blood sera could also be used for the early detection of TNBC as autoantibodies are produced in early-stage disease. This method of diagnosis may therefore identify TNBC at an earlier asymptomatic stage compared to a mammogram. A diagnostic blood test also would be a low cost and relatively non-invasive test for women who are at high risk of TNBC and young women with high breast density.

The aims of this project are summarised as follows:

1. To identify peptides that target TNBC using a phage display peptide library
2. To characterise the panel of phage-derived TNBC targeting peptides so that they may be utilised in the design of peptide-based therapy.
3. To establish whether autoantibodies extracted from TNBC patient blood sera bind to TNBC cell lines
4. To identify the TNBC antigens which are leading to an autoantibody response

Chapter 2: Materials and Methods

2.1 Materials

All cell culture reagents were purchased from Thermo Fisher Scientific and all molecular biology reagents were purchased from Sigma Aldrich unless otherwise stated.

2.2 Cell Culture

Control (MCF10A, HMEC-TERT and BJ) and TNBC (BT549, CAL-51 and MDA-MB-231) cell lines (Table 2.1) were cultured at 37^{o}C in a 5% CO_2 humidified incubator and passaged using 1X trypsin/EDTA under sterile conditions. BT549 and MDA-MB-231 cells were maintained in RPMI 1640 medium and CAL-51 and BJ cells were maintained in DMEM medium. RPMI and DMEM media were supplemented with 10% (v/v) foetal calf serum, 1% (v/v) penicillin-streptomycin, 1% (v/v) L-glutamine, 1% (v/v) MEM non-essential amino acids and 1% (v/v) sodium pyruvate. MCF10A cells were maintained in DMEM-F12 medium supplemented with 5% (v/v) horse serum, 1% (v/v) penicillin-streptomycin, insulin (10 µg ml^{-1}), hydrocortisone (0.5 mg ml^{-1}), epidermal growth factor (20 ng ml^{-1}) and cholera toxin (100 ng ml^{-1}). HMEC-TERT cells were maintained in serum-free HMEC media and trypsin inhibitor was used for cell passage.

Table 2.1 Cell line information

Cell line	Source	Pathology	Site of origin	Key mutations
BJ[1]	ATCC	fibroblast	Normal neonatal foreskin	NA
HMEC-TERT[2]	HMEC purchased from Lonza and immortalised with hTERT	mammary epithelial cells immortalised with hTERT	Normal breast	NA
MCF10A[3]	ATCC	Fibrocystic	Normal breast	NA

BT-549[4]	NCI-60 Cell Line Facility, University of Nottingham	Infiltrating ductal carcinoma	Primary tumour	*p53* mutation and *PTEN* homo deletion
CAL-51[5]	NCI-60 Cell Line Facility, University of Nottingham	Adenocarcinoma	Pleural effusion	No *PTEN* expression
MDA-MB-231[6]	NCI-60 Cell Line Facility, University of Nottingham	Adenocarcinoma	Pleural effusion	*p53* and *KRas* mutations

[3-6]Information adapted from Chavez *et al.,* (Chavez et al., 2010).

[1] (Pereira da Silva et al., 2014), [2] (Kiyono et al., 1998), [3] (Soule et al., 1990), [4] (Neve et al., 2006) [5] (Gioanni et al., 1990), [6] (Cailleau et al., 1974).

2.3 Next generation phage display

2.3.1 TG1 *E.coli*

Library propagations and titrations were carried out in *Escherichia coli* TG1 supE thi-1 Δ(lac-proAB)Δ(mcrB-hsdSM)5(rK–mK)(F′ traD36 proAB lacIqZΔM15)(60502-1,Lucigen). The bacteria were grown on minimal agar at 37°C overnight. TG1 cultures were grown from a single TG1 colony in 2YT media. The same colony was also grown in 2YT media supplemented with 150 µg mL^{-1} ampicillin to confirm that there was no ampicillin resistance prior to phage infection.

2.3.2 Ex-phage stock preparation

2YT media (500ml) was inoculated with 10ml of overnight TG1 culture and grown at 37°C shaking. Once the bacterial culture reached an OD_{600} of 0.4-0.6, TG1 cells were infected with ex-phage (kindly provided by Dr Jonathan Owen, ADAS UK) at a multiplicity of infection (MOI) of 10 for 30 minutes at 37°C static. The bacterial cells were then harvested by centrifugation at 3000 x g for 10 minutes. TG1 cells were resuspended in 2YT media (1000ml) containing 50 µg ml^{-1} kanamycin and placed at 30°C and 180 rpm overnight to allow for ex-phage expression. Bacterial cell debris was

removed by spinning the culture at 3000 x g for 20 minutes. The bacteriophages were precipitated with PEG-8000/2.5M NaCl for 2 hours and harvested by centrifugation at 8000 x g for 20 minutes at 4°C. The phage pellet was slowly resuspended in PBS and titrated using the kanamycin titration method (see 2.3.3). The 1 x 10^{13} ml^{-1} ex-phage stock was stored with 20% glycerol (v/v) at -80°C.

2.3.3 Kanamycin or ampicillin 2YT titration

2YT media and 2YT media supplemented with 150 µg ml^{-1} ampicillin (negative control) was inoculated with the same single TG1 colony. Once the bacteria had entered log phase (OD_{600} 0.4-0.6), 180µl of culture was infected with 20µl of bacteriophage dilutions A-E (Table 2.2) at 37°C for 30 minutes. The infected cultures (100µl) were plated on 2YT agar plates containing 50 µg ml^{-1} kanamycin or 150 µg ml^{-1} ampicillin and 1% glucose (D-Glucose Anhydrous, Thermo Fisher Scientific) and incubated at 37°C overnight. A TG1 only negative control culture (100µl) was also plated. The number of kanamycin or ampicillin resistant colonies were used to calculate the phage colony forming unit (CFU).

Table 2.2 Phage dilution series to establish number of phage in stock solutions

Plate	PBS (µl)	Phage (µl)	Multiplication factor CFU ml^{-1}
A	9990	10	10^5
B	990	10 of dilution A	10^7
C	990	10 of dilution B	10^9
D	990	10 of dilution C	10^{11}
E	990	10 of dilution D	10^{13}

2.3.4 Library propagation

A M13 filamentous bacteriophage library constructed by Dr Robert Workman (School of Veterinary Medicine and Science, University of Nottingham) was used for all biopanning experiments. The library is based on a modified pSD3 phagemid vector which is coupled to 9 arginine residues at the N-terminus of the pIII coat protein (see

Appendix 3.1, Supplementary Figure 1)(Li et al., 1999). This was achieved using a BspQI inverse PCR cloning strategy, where 9-mer peptide diversity was introduced via NNK codons on a single primer (Tsoumpeli et al., 2022).

The naïve phagemid library was propagated in 2YT media supplemented with 150 ug ml^{-1} ampicillin and 1% (w/v) glucose (D-Glucose Anhydrous, Thermo Fisher Scientific) at 37°C shaking. For super infection, 1 x 10^{13} ex-phage were added to the bacterial culture (MOI of 10) at an OD_{600} of 0.4-0.6 and incubated at 37°C for 30 minutes. The infected TG1 cells were centrifuged at 3000 x g for 10 minutes and resuspended in 2YT media (500ml) containing 150 µg ml^{-1} ampicillin and 50 µg ml^{-1} kanamycin to allow for phage expression at 30°C shaking overnight. The phage culture was then centrifuged at 3000 x g for 20 minutes to remove TG1 cells and debris. The phage extract was precipitated with 20% (w/v) PEG-8000/2.5M NaCl for 2 hours and harvested by centrifugation at 8000 x g for 20 minutes. The phage pellet was then resuspended in PBS and the final phage concentration was calculated by ampicillin 2YT titration (see 2.3.3). The final library concentration was determined and stored at -80°C.

2.3.5 Sanger sequencing of 9-mer library clones

To confirm that each phagemid contained the nonapeptide insert, infected bacteria were plated on 2YT agar supplemented with 1% (w/v) glucose (D-Glucose Anhydrous, Thermo Fisher Scientific) and 150 µg ml^{-1} ampicillin at 37°C overnight. Single colonies were then grown in 2YT media supplemented with 150 µg ml^{-1} ampicillin at 37°C overnight. Plasmid DNA was purified using QIAprep® Spin Miniprep Kit (Qiagen) and each clone characterised by Sanger sequencing with the commercially available M13R primer (see Appendix 3.2). A 20% glycerol stock (v/v) of each overnight culture was also stored at -80°C.

2.3.6 Mammalian cell line viability

Each cell line was seeded in triplicate at a density of 2 x 10^5 cells. MCF10A and BT-549 cells were incubated for five minutes in media only (control), 2% (v/v) glycerol only, 2% (v/v) glycerol followed by stripping buffer (PBS, 150mM NaCl, 100mM glycine, pH 2.5), or 1M Tris-HCl. The viability of all cell lines was tested via incubation with

propidium iodide (50 μg ml^{-1}) after cells were washed three times with the stripping buffer for five minutes. The level of cell permeability was observed using the EVOS Cell Imaging Systems (Thermo Fisher Scientific, see Appendix 3.4).

2.3.7 Biopanning

Approximately 1 x 10^{12} phage were biopanned against 2 x 10^5 cells per well (5 replicates per cell line). Phage were incubated with each subtractive non-malignant cell line in iterative steps (BJ, MCF10A and HMEC-TERT) for 30 minutes at 37°C prior to being transferred to one of the target cell lines (BT-549, CAL-51 or MDA-MB-231) for 1 hour at 37°C. The cell media was aspirated and target cells were washed three times with PBS for five minutes and then three times with stripping buffer for five minutes to remove surface bound phage (PBS,150mM NaCl,100mM glycine, pH 2.5). Residual stripping buffer was then neutralised with 1M Tris-HCl, pH 7.4. Cells were lysed with 100mM triethyl amine, neutralised and TG1 cells (grown to an OD_{600} of 0.4-0.6) were infected with the phage containing cell lysate for 1 hour at 37°C static. Infected bacteria were used to both determine phage titre and for stock preparation. To determine output phage titre, the ampicillin titration method described in section 2.3.3 was used and the number of ampicillin resistant colonies were used to calculate the phage cfu (Table 2.2). For stock preparation, infected TG1 cells were harvested at 3000 x g for 3 minutes, resuspended in 2YT media containing 1% (w/v) glucose (D-Glucose Anhydrous, Thermo Fisher Scientific) and 150 μg ml^{-1} ampicillin, and grown at 37°C shaking overnight. The infected bacterial culture was aliquoted and stored with 20% glycerol (v/v) at -80°C. The output phage was propagated and PEG precipitated to produce sufficient phage for the subsequent round of biopanning (see section 2.3.4).

2.3.8 NGS sample preparation

For next generation sequencing, plasmid DNA was extracted from bacterial glycerol stocks using a QIAprep® Spin Miniprep Kit (Qiagen) and quantified. The correct insert was then PCR amplified using library specific forward (5' GTAATCCTTGTGGTATCGCT GGCGGCCCAACCGGCC 3') and reverse primers (5' CTAGAACATTTCACTTACCGCATGA ACTCCTTGATG 3') that flanked the gene insert and incorporated linker sequences for Ion Torrent barcode attachment (linker regions are underlined). PCR amplification was

achieved in a 50µl reaction mix containing 1 X Q5 reaction buffer, 1 X Q5 high GC enhancer, 1 U Q5 Hot Start High-Fidelity DNA polymerase (NEB), 10mM dNTPs, 10ng plasmid DNA and 10µM forward and reverse primers. Thermocycler conditions are shown in Table 2.3. An aliquot of each PCR product was analysed on a 3% agarose gel at 100V for 45 minutes to confirm amplification of the correct 175 bp insert. The remaining product was cleaned using a NucleoSpin Gel and PCR clean-up kit (Machery-Nagel).

Table 2.3 Thermocycler conditions for first round PCRs

Step	Temperature	Time	
1	95°C	3 mins	
2	95°C	30 secs	30 cycles
3	60°C	30 secs	
4	68°C	30 secs	
5	72°C	5 mins	

Each purified PCR product was amplified further using a P1 sequencing primer 5' CCTCTCTATGGGCAGTCGGTGATCTAGAACATTTCACTTAC 3' and a unique nucleotide barcoded primer designed for Ion Torrent sequencing (see Appendix 2.1). The thermocycler conditions are shown in Table 2.4. An aliquot of each barcoded PCR product was analysed on a 3% agarose gel at 100V for 45 minutes to confirm that barcodes had attached. The remaining product was quantified by Qubit (Invitrogen).

Table 2.4 Thermocycler conditions for second round PCRs

Step	Temperature	Time	
1	95°C	3 mins	
2	95°C	30 secs	12 cycles
3	63°C	30 secs	
4	68°C	30 secs	
5	72°C	5 mins	

All samples were pooled to equalise NGS reads and resolved on a 3% Metaphor agarose gel at 100V for 5h. The band was excised (~250 bps) and purified using a

NucleoSpin Gel and PCR clean-up kit (Machery-Nagel). DNA was further purified using Agencourt®AMPure®XP (Beckman Coultier) and quantified by Qubit (Invitrogen). The final amplicon pool (50µl of 100-200ng of DNA) was sequenced by Ion Torrent S5 NGS on a 540 chip (DNA Sequencing Facility, University of Pennsylvania).

2.3.9 NGS analysis

All scripts and adaptations used for NGS analysis are listed in Appendix 2.2 and were kindly provided by ADAS UK, School of Veterinary Medicine and Science, University of Nottingham. Perl scripts were applied to convert FASTQ files into FASTA files and Ion Torrent barcodes were identified using the "FASTQ/A Barcode splitter" from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). The barcoded DNA sequences were translated in all three reading frames and the 9-mer peptide inserts were isolated using the flanking motifs PAMAMSRP and PRRRR. Any amber stop codons (TAG) were replaced with glutamine (q). Peptide gene reads identified from control cell samples were pooled.

Peptide sequences that were obtained from target cell lysate were compared to the sequences obtained from the control cell pool using a two proportional Z test (see Appendix 2.2). The Z test accounted for the absolute frequency (number of reads) and ratio (frequency in the target sample vs the control pool) of a given peptide, as well as the total number of peptide reads in the target sample and control pool (Zhang et al., 2011, Naqid et al., 2016).

$$\mathrm{z} = \frac{\mathrm{p1} - \mathrm{p2}}{\sqrt{\frac{p1(1-p1)}{n1} + \frac{p2(1-p2)}{n2}}}$$

where:

p1= the number of reads obtained for a specific peptide in the target sample

p2= the number of reads obtained for a specific peptide in the control pool

n1= first input file total peptide read count (target sample)

n2= second input file total peptide read count (control pool)

This calculation led to a ranked Z score for each peptide representing the relative enrichment of the peptide specifically against the target cell line. For the first analysis, enriched peptides were selected that had a Z score above 2.0 observed in three out of four (CAL-51) or five (BT-549 and MDA-MB-231) replicates. For the second analysis, the threshold was altered to select peptides that had a Z score above 3.0 in two replicates.

The frequency of an enriched peptides identified by the Z score analysis was then analysed as a normalised percentage of the total insert reads per sample or pool (countpep script, see Appendix 2.2). Unpaired non-parametric t-tests were performed in GraphPad Prism 9.4.1.

2.4 Recovery of NGPD identified peptides

2.4.1 Inverse PCRs for candidate TNBC targeting peptides

Forward and reverse primers were designed to rescue the 44 TNBC enriched peptides identified from the next generation sequence analysis into a pGEX-6P-1 expression vector (GE Healthcare) kindly provided by Dr James Dixon (see Appendix 4.1, Supplementary Figure 6). The primers were also designed to incorporate the MSRP and 9R flanking regions present in the PSD3 phagemid vector. The 8R-mRFP sequence was cloned into the pGEX-6P-1 vector by Dr James Dixon's group (University of Nottingham).

Plasmid DNA was extracted from bacterial glycerol stocks using a QIAprep® Spin Miniprep Kit (Qiagen) and quantified. PCR amplification for each peptide was achieved in a 50µl reaction mix containing 1 X Q5 reaction buffer, 1 X Q5 high GC enhancer, 1 U Q5 DNA polymerase (NEB), 10mM dNTPs, 1ng plasmid DNA and 10µM forward and reverse primers. The thermocycler conditions are shown in Table 2.5. Each PCR product was separated on a 1% agarose gel at 100V for 45 minutes to confirm amplification of the correct insert. The 5.7kb band was gel extracted using a NucleoSpin Gel and PCR clean-up kit (Machery-Nagel) and quantified. To first assess the ligation efficiency of the PCR products, the inverse PCR was conducted with two sets of primers for peptides 1 and 2, where one reaction contained a 5' phosphorylated

Table 2.5 Thermocycler conditions for rescuing 9-mer peptides

Step	Temperature	Time	
1	95°C	3 mins	
2	94°C	30 secs	20 cycles
3	52°C	60 secs	
4	72°C	330 secs	
5	94°C	30 secs	10 cycles
6	55°C	60 secs	
7	72°C	330 secs	
8	94°C	30 secs	10 cycles
9	58°C	60 secs	
10	72°C	330 secs	
11	72°C	5 mins	

primer and the other contained no phosphorylated primers (Figure 4.1A). The non-phosphorylated PCR product was then 5' phosphorylated prior to ligation. The subsequent ligation and transformation of the ligated vector into BL21 *E.coli* led to the efficient recovery of ampicillin resistant bacterial colonies with both the phosphorylated and non-phosphorylated primers, therefore subsequent reactions used non-phosphorylated primers and the PCR products were phosphorylated prior to ligation (Figure 4.1B-E).

2.4.2 Inverse PCRs for control peptides

Forward and reverse primers were designed to incorporate sequences for the mRFP only, 9R-mRFP and p21-9R-mRFP controls into the pGEX-6P-1 vector (GE Healthcare) containing 8R-mRFP (see Appendix 4.1, Supplementary Figure 6). PCR amplification for each control peptide was achieved in a 50µl reaction mix containing 1X Q5 reaction buffer, 1X Q5 high GC enhancer, 1 U Q5 DNA polymerase (NEB), 10mM dNTPs, 1ng plasmid DNA and 10µM forward and reverse primers. The same thermocycler conditions from Table 2.5 were used for the mRFP and 9R-mRFP controls. Thermocycler conditions used for p21-9R-mRFP are shown in Table 2.6. Each PCR product was separated on a 1% agarose gel at 100V for 45 minutes to confirm

amplification of the correct insert. The 5.7kb band was gel extracted using a NucleoSpin Gel and PCR clean-up kit (Machery-Nagel) and quantified.

Table 2.6 Thermocycler conditions p21 positive control

Step	Temperature	Time	
1	95°C	3 mins	
2	94°C	30 secs	30 cycles
3	63°C	60 secs	
4	72°C	330 secs	
5	72°C	5 mins	

2.4.3 Phosphorylation and ligation

The PCR products for the 9-mer and control peptides were 5' phosphorylated in a 50µl reaction mix containing 1X T4 PNK reaction buffer, 1mM ATP, 1 U PNK (NEB) and 200ng PCR product. Each reaction mix was incubated at 37°C for 30 minutes before incubation at 65°C for 20 minutes. The phosphorylated PCR products were then ligated in a 20µl reaction mix containing 50ng of phosphorylated PCR product, 1 X T4 DNA ligase buffer and 400 U T4 DNA ligase (NEB). The ligation reaction mixes were incubated at 18°C for 48 hours before incubation at 65°C for 10 minutes.

2.4.4 Preparation of electrocompetent *E.coli* cells

A single TG1 *E.coli* colony was used to inoculate 2YT media (5ml) and was incubated overnight at 37°C shaking. The overnight bacterial culture was used to inoculate 500ml of fresh 2YT media (1:200) and was grown at 37°C shaking until the culture reached an OD_{600} 0.4-0.6. The culture was placed on ice for 15 minutes and centrifuged at 8000 x g for 10 minutes at 4°C. The supernatant was removed and the bacterial pellet was resuspended in 25ml of ice cold 10% glycerol. This process was repeated four times, decreasing the volume of 10% glycerol each time to a final resuspension volume of 5ml. The electrocompetent cells were aliquoted and stored at -80°C until required.

2.4.5 Transformation into electrocompetent TG1 *E.coli*

The 20µl ligation reaction mixes were dialysed using a 0.025 µm pore size MF-MilliporeTM Membrane Filter (Sigma Aldrich) for 30 minutes at room temperature. After dialysis, 2µl of each ligation mix was added to 50µl of electrocompetent TG1 *E.coli* cells (see section 2.4.4). The recombinant plasmids were then transformed into electrocompetent TG1 *E.coli* by electroporation (1.8kV, 25µF, 200Ω) in a 1 mm cuvette. The outgrowth of transformants was carried out in 250µl of SOC media for 40 minutes at 37°C shaking. Each transformation reaction was then plated on 2YT agar plates supplemented with 150 µg ml^{-1} ampicillin along with a pUC19 positive control (Thermo Fisher Scientific) and a TG1 *E.coli* only negative control. The plates were incubated at 37°C overnight.

2.4.6 Sanger sequencing of recombinant plasmid

To confirm that the 9-mer peptide or control peptide encoding sequences had been successfully cloned into the modified pGEX-6P1 vector (GE Healthcare), a single transformant colony was used to inoculate 5ml of 2YT media supplemented with 150 µg ml^{-1} ampicillin and grown at 37°C overnight. Plasmid DNA was purified using QIAprep® Spin Miniprep Kit (Qiagen) and insertion of the correct peptide encoding sequence was confirmed by Sanger sequencing with the following sequencing primer 5' CCCATCCTGACTTCATGTTGT 3'. A 20% (v/v) glycerol stock of each overnight culture was also stored at -80°C.

2.4.7 Recombinant protein expression and purification

Recombinant vector (10ng) was transformed into competent BL21(DE3)pLysS (L119A, Promega) at 42°C for 50 seconds. The outgrowth of transformants was carried out in 450µl of SOC media for 1 hour at 37°C shaking. Each transformation reaction was then plated onto nutrient agar plates supplemented with 34µg ml^{-1} chloramphenicol and 50µg ml^{-1}ampicillin and incubated at 37°C overnight. A single transformant colony was then used to inoculate 10ml of fresh nutrient broth containing 34µg ml^{-1} chloramphenicol and 50µg ml^{-1} ampicillin and was grown overnight at 37°C and 220rpm. A 20% glycerol stock (v/v) of each overnight transformant culture was stored at -80°C.

Overnight cultures (5ml) were used to inoculate fresh nutrient broth (125ml) and grown at 37^{o}C and 220rpm. Once the bacterial cultures had reached OD_{600} 0.4-0.6, mRFP expression was induced with 1mM isopropyl-β-D-1-thioglactopyranoside (IPTG) at 20^{o}C, 180rpm for 24 hours. The cultures were then centrifuged at 2000 x g for five minutes and the supernatant removed. Bacterial pellets were resuspended in 1 X STE buffer (8ml) (10mM Tris-HCl, 1mM EDTA and 100mM NaCl) containing 1mM dithiothreitol (DTT) and freeze thawed on dry ice.

The bacterial pellets were sonicated on ice five times for 30 seconds at 5 amplitudes. The insoluble protein fractions were then pelleted at 2000 x g for five minutes at 4^{o}C and discarded. The soluble protein fractions were then incubated overnight at 4^{o}C rotating with 600µl of Glutathione Sepharose™ 4B resin (Cytiva) washed in 1 X STE buffer and 1mM DTT. The resin was centrifuged at 3,600 x g for five minutes at 4^{o}C and washed with 5ml of PBS for 1 hour. The beads were washed three more times with PBS and left in the final wash overnight at 4^{o}C. The Glutathione Sepharose™ 4B resin was centrifuged at 3600 x g for five minutes and resuspended in 588µl of cleavage buffer (50mM Tris-HCL, 150mM NaCl, 1mM EDTA and 1mM DTT) and 40 units ml^{-1} of PreScission protease (Cytiva). GST-tag cleavage was achieved overnight at 4^{o}C. The Glutathione Sepharose™ 4B resin was then pelleted at 3600 x g for five minutes and the supernatant containing recombinant protein fractions were collected and protein was quantified by Bradford assay (Sigma Aldrich) using a FLUOstar OPTIMA plate reader. The mRFP recombinant proteins were stored at -80^{o}C until further use.

GST tag cleavage and recombinant protein purification was then confirmed for three recombinant proteins by SDS-PAGE. To confirm recombinant protein purification, 10µl of Glutathione Sepharose 4B resin, 1µl of PreScission protease, 10µl of cleaved Glutathione Sepharose 4B resin for peptides 2, 8 and 9 and 500ng of purified protein for peptides 2, 8 and 9 were denatured with 1X Lamelli buffer and 5% (v/v) β-mercaptoethanol at 95^{o}C for 10 minutes. All samples were loaded onto a 10% polyacrylamide gel (Mini-PROTEAN® TGX™, BioRad) a long with a broad range molecular weight marker (NEB #P7712). Proteins were separated in 1 X Tris-glycine-SDS running buffer at 150V for 45 minutes. The polyacrylamide gel was then stained with 0.5% (w/v) Coomassie in a 40% (v/v) methanol and 10% (v/v) acetic acid solution overnight at RT. Excess stain was removed by three 30 minute washes with a 40%

methanol (w/w) and 10% acetic acid (w/w) solution and two washes with milli-Q water before imaging using a BioRad ChemiDoc Imaging System.

2.5 *In vitro* validation of peptides

2.5.1 Fluorescence microscopy

BJ, HMEC-TERT, MCF10A, CAL-51, BT-549 and MDA-MB-231 cells were seeded at a density of 5,000 cells per well in a 96 well plate and incubated with each mRFP recombinant protein (20 μg ml^{-1}) for 24 hours at 37°C in a 5% CO2 humidified incubator. HMEC-TERT cells were seeded on plates which had been coated with 0.1% (w/v) gelatin in PBS. Hoescht (1:200) was added to each well for 15 minutes and cells were washed in PBS and fixed with 4% PFA for 10 minutes. Cells were then washed with PBS prior to imaging. Microscope images were obtained using a Zeiss AxioObserver Z1 inverted widefield microscope with a 40x objective. DAPI and mRFP staining was visualised at 465nm and 561nm, respectively. The images were processed by applying background subtractions (100μM radius) and contrast alterations in Zeiss Zen 3.0 software. The total number of nuclei in each image were counted using ImageJ software and the number of cells with mRFP specific staining were calculated as a percentage of the total number of cells within a given image (Schneider et al., 2012).

2.5.2 Flow cytometry

5 x 10^4 HMEC-TERT, MCF10A, CAL-51, BT-549 and MDA-MB-231 cells were seeded and incubated with each mRFP recombinant protein (20 μg ml^{-1}) for 24 hours at 37°C in a 5% CO2 humidified incubator. Cells were then washed with PBS and harvested with 1X trypsin/EDTA. Cell suspensions were centrifuged at 400 x g for four minutes and resuspended in 4% PFA for 10 minutes at RT. After fixation, the cells were centrifuged at 400 x g for four minutes and PFA was removed. Excess PFA was then washed with PBS and cells were pelleted at 400 x g for four minutes. Cell pellets were then resuspended in a final volume of 100 μl of PBS and flow cytometric analysis was carried out using a ID7000™ spectral cell analyser at the University of Nottingham Flow Cytometry Facility. Forward and side scatter data was acquired at 488nm (blue laser) and gated for single cells. mRFP fluorescence data acquired at 561nm with 20,000 single cell events recorded per sample. The flow cytometry data was analysed using

Kaluza Analysis 2.1 software and Mann-Whitney tests were performed in GraphPad Prism 9.4.1.

2.5.3 Magnetic nanoparticle (MNP) delivery

The magnetic nanoparticles used for this study were 250nm Fe_3O_4 Nanomag®-D MNPs (Micromod,09-02-252). MNPs (25 µg ml^{-1}) were complexed with 0.5nM of synthetically produced 9R, p21-8R, TLYRMVTVTS-9R, LMQDMALRL-9R and CSTTYYTVK-9R (Synpeptide). The peptides were complexed to MNPs (20 nmol mg^{-1} MNP) in 50µl of Opti-MEM (Thermo Fisher Scientific) for 15 minutes at RT. Peptide-MNPs were then delivered to HMEC-TERT, MCF10A, CAL-51, BT-549 and MDA-MB-231 cells in 500µl of supplemented RPMI 1640 media (see section 2.1) for 1 hour at 37°C 5% CO_2. The peptide-MNP containing media was then removed and fresh RPMI 1640 media was added to cells and incubated at 37°C 5% CO_2 for 23 hours.

2.5.4 MTT assay

5 x 10^4 HMEC-TERT, MCF10A, CAL-51, BT-549 and MDA-MB-231 cells were seeded and were incubated with 50 µg ml^{-1} MNPs or 1nM of synthetically produced 9R, p21-8R, TLYRMVTVTS-9R, LMQDMALRL-9R or CSTTYYTVK-9R (Synpeptide) for 24 hours at 37°C in a 5% CO_2 humidified incubator. MTT ((3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazoliumbromide) was then added to each well (0.5 mg ml^{-1}) and incubated for 3 hours at 37°C in a 5% CO2 humidified incubator. The MTT containing media was removed and cells were washed with PBS. The insoluble formazan crystals were then dissolved in isopropanol and 0.04M HCl and absorbances at 570nm and 650nm (background) were measured using a Varioskan LUX multimode microplate reader. Cell viability for each cell line was statistically assessed by ordinary one-way ANOVA performed in GraphPad Prism 9.4.1.

2.5.5 Prussian blue staining

6 x 10^4 HMEC-TERT, MCF10A, CAL-51, BT-549 and MDA-MB-231 cells were seeded and grown overnight at 37°C in a 5% CO_2 humidified incubator. MNPs were then delivered to cells as described in section 2.4.3. Following MNP delivery, cell culture media was

removed from wells and cells were washed with PBS. Cells were then fixed with 4% PFA for 20 minutes at RT. After fixation, the cells were washed with milli-Q water and incubated in 2.5% (w/v) potassium ferrocyanide (Sigma-Aldrich) in 2.5% (v/v) hydrochloric acid for 30 minutes at RT. Prussion blue stained cells were then washed with milli-Q water to reduce background staining and imaged using a Leica DMi8 inverted microscope. All images were obtained using a 10x objective and scale bars were added in Leica Application Suite X (LAS X).

2.5.6 Inductively coupled plasma mass spectrometry (ICP-MS)

6×10^5 HMEC-TERT, MCF10A, CAL-51, BT-549 and MDA-MB-231 cells were seeded and incubated overnight at 37°C in a 5% CO_2 humidified incubator. MNPs were then delivered to cells as described in section 2.5.3. Following MNP delivery, cells were washed with PBS and harvested with 1X trypsin/EDTA. The trypsin was then neutralised with supplemented RPMI 1640 media (see section 2.1) and cells were pelleted at 400 x g for five minutes. The media was removed and cells were lysed to release the Fe content from MNPs in 12M hydrochloric acid and 6M nitric acid for 2 hours at RT. Samples were then diluted in milli-Q water to a final volume of 10ml and a final acid concentration of < 2% (v/v) for ICP-MS.

Multi-elemental analysis of samples was undertaken by ICP-MS (Thermo Fisher Scientific iCAP-Q; Thermo Fisher Scientific, Bremen, Germany). Samples were introduced (flow rate 1.2 mL min^{-1}) from an autosampler (Cetac ASX-520) incorporating an ASXpress rapid uptake module through a perfluoroalkoxy Microflow PFA-ST nebulizer (Thermo Fisher Scientific, Bremen, Germany). Internal standards were used to correct for instrumental drift and were introduced to the sample stream on a separate line (flow rate 1.2 mL min^{-1}) via the ASXpress unit. Internal standards included combinations of Sc (10 µg L^{-1}), Ge (10 µg L^{-1}), Rh (5 µg L^{-1}), Re (5 µg L^{-1}) and Ir (5 µg L^{-1}). Calibration standards included a multi-element solution with Ag, Al, As, Ba, Be, Cd, Ca, Co, Cr, Cs, Cu, Fe, K, Li, Mg, Mn, Mo, Na, Ni, P, Pb, Rb, S, Se, Sr, Ti, Tl, U, V and Zn (0, 20, 40, 100 µg L^{-1}) (Claritas-PPT grade CLMS-2 from SPEX Certiprep Inc., Metuchen, NJ, USA), a multi-element calibration solution (PlasmaCAL, SCP Science, France) with Ca, Mg, Na and K (0-30 mg L^{-1}) and a mixed phosphorus, boron and sulphur standard made in-house (ICP-MS Facility, University of Nottingham). All

internal and calibration standards were diluted in 2% Primar grade HNO_3 (Fisher Scientific, UK). Sample processing was undertaken using Qtegra™ software (Thermo Fisher Scientific) utilising external cross-calibration between pulse-counting and analogue detector modes where required. A collision cell charged with He gas with kinetic energy discrimination (KED) to remove polyatomic interferences was employed to measure Fe and results were reported in ug L^{-1}.

2.6 Autoantibody binding

2.6.1 Patient IgG antibody purification

Human serum extracted from 40 triple negative breast cancer patients obtained from The Nottingham Health Science Biobank (Ethical approval 2124-70914) was pooled and immunoglobulins G were purified using Nab protein G spin columns (Thermo Fisher Scientific). The elution buffer was exchanged via Zeba spin desalting columns (Thermo Fisher Scientific) to the storage buffer (PBS, 150mM NaCl) and the IgG concentration was measured by nanodrop. Purified IgGs were stored with 0.15% (w/v) BSA at -80^{o}C until further use.

2.6.2 Flow cytometry

Cells were harvested from culture flasks and counted. 300,000 cells were incubated at 4^{o}C for 1h with human IgG isotype control (0.2 µg µl^{-1}, Thermo Fisher Scientific, 02-7102) or TNBC patient IgG (0.2 µg µl^{-1}, see 2.6.1). Cells were washed with RPMI 1640 medium and pelleted at 300 x g for four minutes. Cells were then fixed in ice cold acetone for 10 minutes at 4^{o}C. After two further washes with media, cells were incubated with anti-human IgG-APC (1:20, Biolegend) for 30 minutes at 4^{o}C. Excess antibody was washed with medium and after centrifugation, cells were resuspended in PBS for flow cytometry. Flow cytometric analysis was carried out using a ID7000™ spectral cell analyser at the University of Nottingham Flow Cytometry Facility. Forward and side scatter data was acquired at 488nm (blue laser) and gated for single cells. APC fluorescence data was acquired at 637nm (red laser) with 20,000 single cell events recorded per sample. The flow cytometry data was analysed using Kaluza Analysis 2.1

software and Kruskal-Wallis tests (Figure 5.2) or unpaired t-tests (Figure 5.4) analyses were performed in GraphPad Prism 9.4.1.

2.6.3 Whole cell lysate protein extraction

Control and TNBC cells were harvested using 1X trypsin/EDTA and washed with PBS. Cells were pelleted at 300 x g for four minutes and stored at -80^{o}C until further use. All cells were lysed using 1X RIPA buffer (Cell Signalling) and incubated on ice for 40 minutes. To remove cell debris, samples were centrifuged at 10000 x g for 10 minutes at 4^{o}C and the protein supernatant was collected. Protein extracted from each cell line was then quantified by Qubit (Invitrogen) and stored at -80^{o}C.

2.6.4 Nuclear, cytoplasmic and membrane protein extraction

MCF10A and MDA-MB-231 cells were harvested using 1X trypsin/EDTA and washed with PBS. Cells were pelleted at 300 x g for four minutes and stored at -80^{o}C until further use. Nuclear and cytoplasmic fractionation of MCF10A and MDA-MB-231 lysate was achieved using a NE-PER Nuclear and Cytoplasmic Extraction kit (Thermo Fisher Scientific) according to manufacturer's instructions. Cytoplasmic fraction of MCF10A and MDA-MB-231 lysate was also achieved using a MEM-PER Plus Membrane Protein Extraction kit (Thermo Fisher Scientific) according to manufacturers instructions. Final nuclear, cytoplasmic and membrane fractions were quantified by Qubit and Halt Protease Inhibitor Cocktail (100X) (Thermo Fisher Scientific) was added to each fraction (10 µl ml^{-1}). Protein aliquots were then sorted at -80^{o}C.

2.6.5 Western blotting

Protein extracts (20 µg) were denatured with 1 X Lamelli buffer and 5% (v/v) β-mercaptoethanol at 95^{o}C for 10 minutes. The protein extracts were loaded onto 12% polyacrylamide gels (Mini-PROTEAN TGX, BioRad) along with a broad range molecular weight marker (NEB, #P7712). Proteins were separated in 1 X Tris-glycine-SDS running buffer at 150V for 30 to 45 minutes. The separated proteins were transferred to a PVDF membrane (Trans-blot turbo transfer pack, BioRad) using a BioRad Trans-Blot® Turbo™ transfer system (Standard SD, 1.0A, 25V for 30 minutes). Membranes were blocked with 5% (w/v) skimmed milk for 1 hour at RT and incubated with human IgG

isotype control (0.02 µg µl^{-1}, Thermo Fisher Scientific) or TNBC patient IgG (0.02 µg µl^{-1}) in 5% (w/v) milk at 4°C overnight. The membranes were then washed three times with 0.1% (v/v) TBS-Tween 20 before incubation with goat anti-human (H+L) HRP (1:5000, Thermo Fisher Scientific) for 40 minutes at RT. After three more washes with 0.1% (v/v) TBS-Tween 20, membranes were incubated in Amersham ECL Prime (GE Healthcare) for five minutes to enable the detection of specific IgG binding by chemiluminescence using a BioRad ChemiDoc Imaging System.

Membranes were stripped with 0.2M glycine pH 2.8 for 30 minutes at 50°C and 0.2M glycine pH 7.4 for 30 minutes at RT. The membranes were washed three times with 0.1% TBS-Tween 20 and probed with either mouse anti-β actin (0.01 µg ml^{-1}, R&D, P37215), mouse anti-GAPDH (0.05 µg ml^{-1}, Biolegend, 649201), mouse anti-HDAC1 (0.125 µg ml^{-1}, Biolegend, 815103) or mouse anti-Na^+/K^+ ATPase a-1 (0.05 µg ml^{-1}, Sigma-Aldrich 05-369) in 5% (w/v) milk overnight at 4°C. Membranes were then washed three times with 0.1% TBS-Tween 20 and probed with anti-mouse HRP (1:10,000, GE Healthcare) for 1 hour at RT.

2.6.6 LC-MS/MS data analysis

All LC-MS/MS samples were analysed using Mascot (Matrix Science, London, UK; version 2.7.0) by the Cambridge Centre for Proteomics. Mascot was set up to search the CCP_UniProt_homo_sapiens_proteome_20180409 peptide sequence database and a cRAP_20190401 database (Cambridge Centre for Proteomics) to remove contaminant proteins such as common proteins that are introduced in the sample preparation process or through contact with the sample. Mascot was searched with a fragment ion mass tolerance of 0.100 Da and a parent ion tolerance of 20 PPM. Carbamidomethyl of cysteine was specified as a fixed modification and deamidated of asparagine and glutamine and oxidation of methionine were specified in Mascot as variable modifications.

Scaffold (version Scaffold_5.0.1, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications (Searle, 2010). Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Peptide Prophet algorithm (Keller et al., 2002) with Scaffold delta-

mass correction. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

2.6.7 Silver staining

10 µg of nuclear, cytoplasmic and membrane protein from MCF10A and MDA-MB-231 protein fractions (see 2.5.4) were denatured with 1 X Lamelli buffer and 5% (v/v) β-mercaptoethanol at 95°C for 10 minutes. The protein extracts were loaded onto 12% polyacrylamide gels (Mini-PROTEAN TGX, BioRad) along with a broad range molecular weight marker (New England Biolabs, #P7712). Proteins were separated in 1 X Tris-glycine-SDS running buffer at 150V for 30 to 45 minutes. Proteins were then detected using a ProteoSilver Silver Stain kit (Sigma-Aldrich) according to manufacturer's instructions. Stained SDS-PAGE gels were then imaged using a BioRad ChemiDoc Imaging System.

2.6.8 LC-MS/MS sample preparation

10 µg of nuclear and cytoplasmic protein from MCF10A and MDA-MB-231 protein fractions (see 2.6.4) were separated by SDS-PAGE and silver stained as described in section 2.6.7. Polyacrylamide gel slices were then excised at the correct molecular weight and submitted for liquid chromatography tandem mass spectrometry (LC-MS/MS) at the Cambridge Centre for Proteomics (n=3).

2.6.9 Unique normalised spectral abundance factor (uNSAF) and t-tests

Proteins identified were normalised between LC-MS/MS samples by calculating the unique normalised spectral abundance factor (uNSAF) for each protein in Scaffold. The software calculated uNSAF values from the spectral abundance factor (SAF) for each protein based on the following calculation:

$$SAF = \frac{uSPc}{uL + sL}$$

Strategy 2-a (Zhang et al., 2010) for uNSAF calculation where:
uSPC = number of unique spectral counts mapping to a protein
uL = protein lengths mapping to unique peptides (amino acid length)
sL = protein lengths mapping to shared peptides (amino acid length)

SAF values were then normalised to NSAFs using a normalisation process within the Scaffold software. Zero data across all replicates was removed before statistical analysis. Differential expression was detected by applying multiple unpaired t-tests with Benjamini, Krieger and Yekutieli 1% FDR (False Discovery Rate) correction (Benjamini et al., 2006) in GraphPad Prism 9.4.1. Volcano plots were also produced in GraphPad Prism 9.4.1.

2.6.10 DESeq2

Exclusive spectral counts were also tested for differential expression by DESeq2. DESeq2 was implemented in R using the DESeq2 bioconductor package (Anders and Huber, 2010, Love et al., 2014). The DESeq2 script used is shown in Appendix 5.1. Exclusive spectrum count data was analysed by negative binomial distribution model fitting and Wald statistics. Statistical significance in log2 fold changes between a given protein in MCF10A and MDA-MB-231 samples was assessed by Benjamini-Hochberg adjusted p-values (padj) (Benjamini and Hochberg, 1995). Volcano plots were produced in GraphPad Prism 9.4.1.

2.6.11 bc-GenExMiner

Gene expression levels for proteins identified from uNSAF and t-tests and DESeq2 analyses were analysed by bc-GenExMiner v4.8 (Jézéquel et al., 2021) (accessed 20th November 2022). Standardised mRNA expression levels were compared in healthy (n=92, GTEx), tumour-adjacent (n=89, TCGA), non-TNBC (n=578, TCGA) and TNBC (n=87, TCGA) transcriptomic data from the Genotype-Tissue Expression (GTEx) project (Lonsdale et al., 2013) and The Cancer Genome Atlas (TCGA) Network. Gene

expression comparisons were conducted between clinical groups by Welch's t-tests in bc-GenExMiner v4.8. Multiple pairwise comparisons between each clinical group were performed by Dunnett-Tukey-Kramer's test for two-by-two comparisons if Welch p-values were significant ($p < 0.0001$ for all genes).

2.6.12 Kaplan-Meier plotter

The correlation between the expression of TPX2, WRNIP1, KHDRBS3, CALU, DNAJB6, KRT19 and HLA-A and overall survival in ER negative, PGR negative and HER2 negative patients (n=126) was assessed using the online Kaplan-Meier plotter tool (Lánczky and Győrffy, 2021). The analysis was conducted using BCa mRNA RNA-seq data and Kaplan-Meier plots were generated using the best performing threshold (auto select best cutoff).

2.6.13 TPX2 and KHDRBS3 western blot

20µg of TPX2 (NM_012112) human over-expression lysate (OriGene, LC415965) and 20µg of KHDRBS3 (NM_0006558) human over-expression HEK293T lysate (OriGene, LC41673) was resuspended in PBS and denatured with 1 X Lamelli buffer and 5% (v/v) β-mercaptoethanol at 95^{o}C for 10 minutes. 10µg of lysate was then loaded per well onto a 12% polyacrylamide gel (Mini-PROTEAN TGX, BioRad) along with a broad range molecular weight marker (New England Biolabs, #P7712). Proteins were separated in 1 X Tris-glycine-SDS running buffer at 150V for 30 to 45 minutes. The separated proteins were then transferred to a PVDF membrane using a BioRad Trans-Blot® Turbo™ transfer system (Standard SD, 1.0A, 25V for 30 minutes). Membranes were blocked with 5% (w/v) skimmed milk for 1 hour at RT and incubated with human IgG isotype control (0.02 µg µl^{-1}, Thermo Fisher Scientific) or TNBC patient IgG (0.02 µg µl^{-1}) in 5% (w/v) milk at 4^{o}C overnight. The membranes were then washed three times with 0.1% (v/v) TBS-Tween before incubation with goat anti-human (H+L) HRP (1:5000, Thermo Fisher Scientific) for 40 minutes at RT. After three more washes with 0.1% (v/v) TBS-Tween 20, membranes were incubated in Amersham ECL Prime (GE Healthcare) for five minutes to enable the detection of specific IgG binding by chemiluminescence using a BioRad ChemiDoc Imaging System.

Chapter 3: Identification of potential TNBC targeting small peptides by next generation phage display

3.1 Introduction

A high-throughput screening method, which has promising potential in the discovery of tumour targeting peptides for cancer therapeutics, is Next Generation Phage Display (NGPD). NGPD is a powerful technology that isolates and identifies target specific peptides or antibodies from an unbiased combinatorial phage-peptide library of up to 10^{10} clones (Scott and Smith, 1990, Dias-Neto et al., 2009).

NGPD has successfully been applied to several cancer therapeutic studies. For instance, phage display screening for novel autoantigens for seminoma detected the highly conserved Kelch-like protein 11 (KLH11) and found that KLHL11 autoantibodies reflect paraneoplastic encephalitis in testicular cancer patients (Mandel-Brehm et al., 2019). Additionally, *Sun et al.* identified the novel peptide LS-7 from a phage display peptide library, which specifically binds to the colon CSC marker $CD133^{+}$. LS-7 binding reduced migration of both mouse colon cancer and BCa cells in *in vitro* assays, however LS-7 based therapeutics have not been clinically developed(Sun et al., 2012). Wada *et al.,* also applied phage display technology to identify a homing peptide, termed LN1, which targeted LNCaP xenografts. The peptide was also found to selectively deliver a pro-apoptotic peptide to inhibit tumour growth in *in vivo* prostate cancer cells (Wada et al., 2019).

Phage display technology was first described by George P. Smith who demonstrated that foreign DNA could be inserted into the filamentous phage gene III to create a fusion protein which is displayed on the surface of a bacteriophage without hindering the infectivity of the virion (Smith, 1985). His work led to the construction of epitope libraries and the subsequent affinity purification of target specific phages, as well as the expansion of the technology to display recombinant antibody fragments (Parmley and Smith, 1988, Devlin et al., 1990, McCafferty et al., 1990).

The lysogenic family of filamentous Ff bacteriophages are the most extensively applied in phage display systems, particularly M13 bacteriophages. M13 bacteriophages

consist of a 6.4 kb ssDNA genome encapsulated by the minor (pIII, pVI, pVII and PIX) and major (pVIII) coat proteins, with ligands being displayed on the pVIII or pIII coat protein of M13 filamentous phage (Ledsgaard et al., 2018, van Wezenbeek et al., 1980). There are up to 2700 copies of the major coat protein pVIII, therefore display of ligands on pVIII N-terminus can enhance their avidity (Bazan et al., 2012). However, the multivalent display of fusion proteins can lead to the identification of low affinity binding ligands through such high avidity selection (Falciani et al., 2005). Furthermore, only peptides up to seven amino acids in length can be displayed on the alpha-helical pVIII coat protein, as larger peptides impact phage viability (Saw and Song, 2019, Iannolo et al., 1995). Even though hybrid phage systems and coat protein modifications have been employed to overcome this, the majority of phage display libraries are therefore engineered to display peptides and antibody fragments on the pIII coat protein, as larger polypeptides or proteins (e.g antibody fragments) can be displayed and the low copy number of pIII (two to five copies per virion) enables the selection of high affinity ligands (Lowman, 1997, Fagerlund et al., 2008, Bazan et al., 2012, Hess et al., 2012).

In addition, there are also different display systems which impact fusion protein valency. In early studies, the generation of a combinatorial peptide library was achieved by the in-frame insertion of a peptide encoding gene directly into the phage genome (phage vector), which led to the foreign peptide being presented on all copies of the chosen coat protein and the identification of low affinity ligands through multivalent binding, termed the chilate-effect (Cwirla et al., 1990). Other display formats (33 and 88) reduce the copy number of the displayed protein by incorporating two pIII or pVIII genes into the phage genome, resulting in mosaic expression of both recombinant and wild-type protein on a single virion (Smith and Petrenko, 1997, Chang et al., 2020). Similarly, the display format types 3+3 and 8+8 are encoded by two copies of the pIII or pVIII gene, respectively. A key difference, however, is that the two genes are located on different replicons, with the wild-type pIII or pVIII gene incorporated in the helper phage genome and the recombinant gene included as part of a phagemid vector (Smith and Petrenko, 1997).

The M13 bacteriophage library used in this study expressed 9 amino acid peptides displayed on the pIII coat protein flanked by MSRP and 9R sequences. The expression

was via a phagemid vector as described in Figure 3.1. Here, the wild-type pIII coat protein supplied by the helper phage competes with the recombinant gene for incorporation into the virion, resulting in less than three copies of recombinant pIII being expressed per bacteriophage particle (usually zero or one copies). The monovalent display achieved by this type of display system subsequently minimises avidity effects and allows for relatively high affinity binding ligands to be selected (Marks et al., 1992). In addition, phagemid vectors have a higher transformation efficiency than phage vectors and therefore facilitate the construction of larger and subsequently more diverse libraries for screening (O'Connell et al., 2002, Ledsgaard et al., 2018).

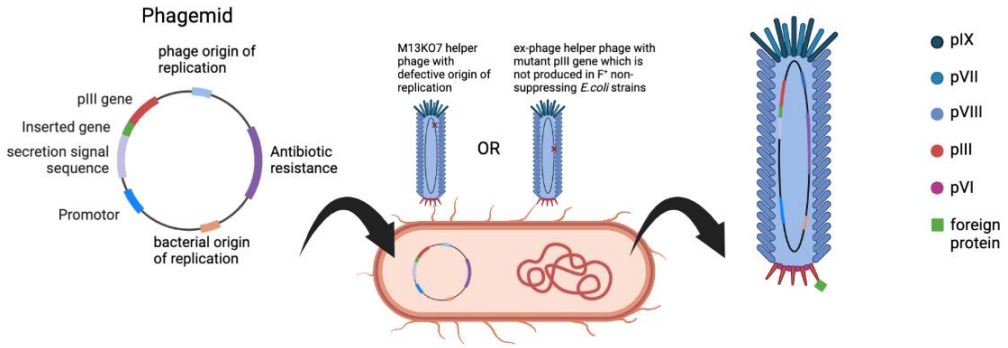


Figure 3.1 Production of a peptide displaying bacteriophage using the phagemid system

F^+ non-suppressing *E.coli* are transformed with a phagemid vector containing the fusion protein encoding gene and propagated. Helper phage (M13K07 or the mutant ex-phage) are supplied *in trans* and infect the phagemid containing antibiotic resistant *E.coli* through the F-pilus. The helper phage genome encodes for all phage proteins required for the replication and assembly of phage progeny. The phagemid derived pIII-peptide protein competes with helper phage derived wild-type pIII for incorporation into the mature virion, leading to the monovalent display of the peptide for affinity selection (3+3). The helper phage is packaging deficient and is packaged at low levels compared to the phagemid vector. Figure adapted from Jaroszewiz et al., and created with BioRender.com. (Jaroszewicz et al., 2022, Ledsgaard et al., 2018).

In order to isolate target specific peptides from a diverse library, the library undergoes affinity selection with a target molecule to enrich for target specific phage over three to five biopanning rounds (Figure 3.2). The target molecule selected may be *in situ*, such as a cell surface receptor, *in vitro,* such as purified protein, or *in vivo* (Gallo et al., 2020, Zhou et al., 2015b, Arap et al., 1998) .

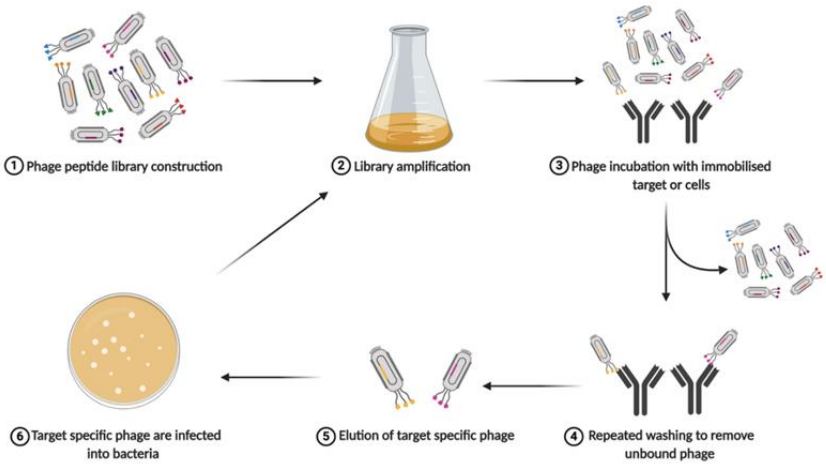


Figure 3.2 Biopanning of a random peptide bacteriophage library

1) Combinatorial phage peptide library is generated 2) and amplified. 3) The library is then incubated with an immobilised target or adherent cells to allow for binding (example given is an antibody). 4) Repeated washing removes unbound, non-target specific phage. 5) Target specific-phage are eluted or harvested and 6) propagated in bacteria to produce an amplified selective phage library for following rounds of biopanning. Iterative rounds are required to enrich for target binders as washing is not fully efficient. Figure adapted from Saw & Song, 2019 and created in BioRender.com.

In this project, a filamentous M13 bacteriophage pIII library containing 10^8 9 amino acid peptide variants was screened against control and TNBC cell lines to identify small peptides which selectively target TNBC cell and lead to the subsequent internalisation of their respective bacteriophage particle via a cell penetrating peptide (CPP). To achieve this, the phage display peptide library was enriched for TNBC over three rounds of subtractive biopanning with control cell lines, followed by the isolation of target specific phage pools from cell lysate and the identification of each peptide by next generation sequencing (NGS). Furthermore, the screening assay aimed to identify the phage-displayed peptides that best discriminated between normal and TNBC cell lines using a Z score bioinformatic analysis. The most promising peptides identified from this study may therefore lead to the development of a peptide-drug conjugate or peptide-based therapy with the capability of delivering a high cytotoxic payload to TNBC tumours with minimal off-target side effects in healthy tissue.

3.2 Results

3.2.1 Library propagation and biopanning

To first confirm that the naïve pIII peptide library contained the correct insert to allow for nonapeptide expression, phagemids purified from infected TG1 *E.coli* clones were sequenced. The library was constructed by Dr Robert Workman (University of Nottingham, Dixon Group) using a variation of the phagemid vector pSD3, named pHR2-8R (see Appendix 3.1, Supplementary Figure 1), where expression of the phagemid led to pIII fusion protein display. In addition to the 9-mer peptide inserts, the region of display also contained sequences encoding for a monomeric red fluorescent protein (mRFP) and a polyHis tag, as well as the 9-mer peptide flanking regions MSRP and a cationic CPP sequence, PRRRRRRRRR (Mitchell et al., 2000, Futaki et al., 2001). Sanger sequencing confirmed that all clones contained the peptide gene correctly inserted between the MSRP and polyarginine PRRRRRRRRR flanking regions. However, 25% of clones (4/16) contained multiple pelB-MSRP-peptide-PRRRR sequences found in different reading frames upstream to the correct MSRP-peptide-PRRRRRRRRR encoding region (see Appendix 3.2, Supplementary Figure 2).

To test whether the biopanning procedure would impact the viability of cells used in this study, each cell line was incubated with all reagents from the mammalian cell panning protocol and cell viability was assessed via a propidium iodide permeability test. The final stripping buffer wash, which removes unbound phage, appeared to increase cell permeability to propidium iodide, however, the cells remained intact (see Appendix 3.4, Supplementary Figure 4). As the stripping buffer is essential for the removal of background phage, this step was not removed from the biopanning protocol even though the buffer washes increased cell permeability. However, the protocol was adapted so that cells were lysed in the plate to minimise the loss of target specific phage by further manipulation of already permeabilised cells.

To select for peptides which were specific to TNBC cell lines, three biopanning experiments were designed where the naïve peptide-phage library was separately enriched against three TNBC cell lines, BT-549, CAL-51 or MDA-MB-231 cells over three rounds of affinity selection (Figure 3.3).

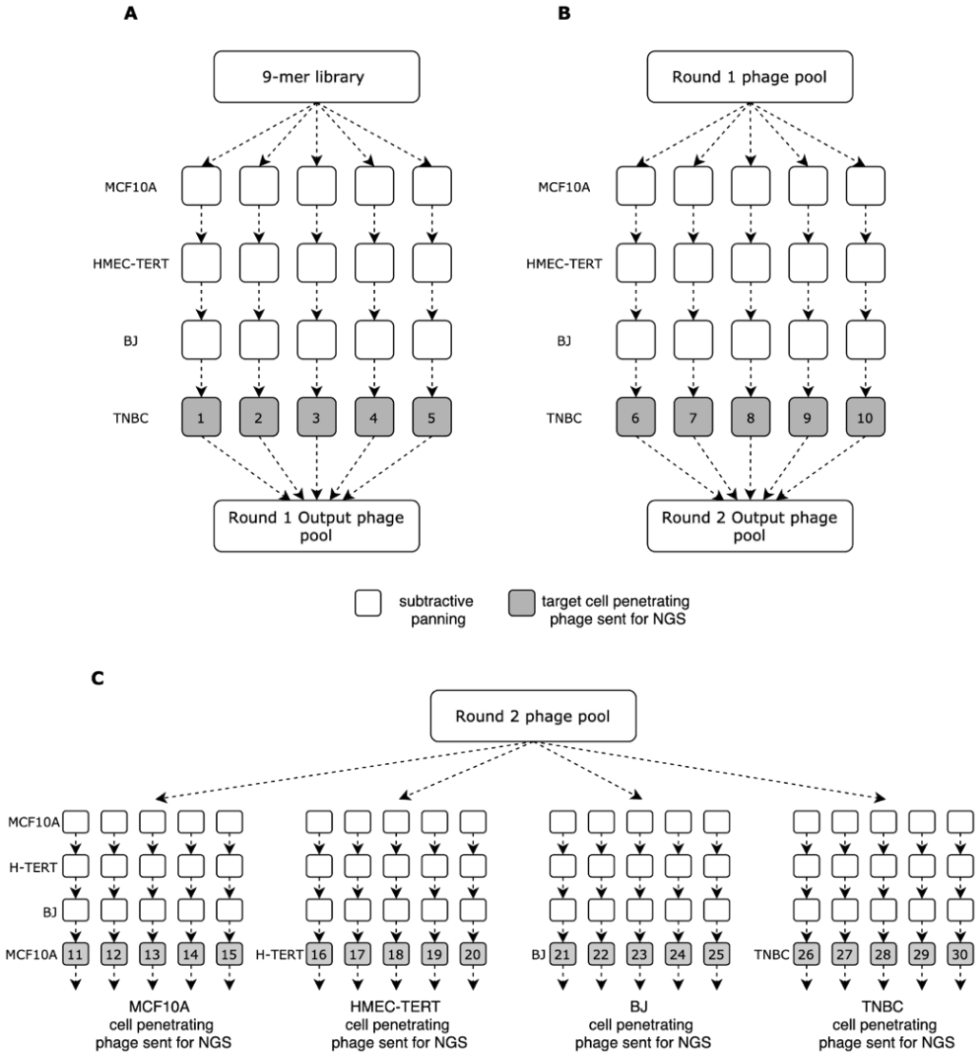


Figure 3.3 Biopanning against TNBC cell lines

(A) The naïve library was firstly panned against the three control cell lines MCF10A, HMEC-TERT and BJ (white boxes) before being transferred to either BT-549, CAL-51 or MDA-MB-231 cells (grey boxes). (B) Target-specific phage were harvested by TNBC cell lysis and propagated to produce a TNBC selective sub-library for the second round of panning. (C) In the final round of panning, round two output phage underwent subtractive panning against the three control cell lines and the unbound phage were transferred to either MCF10A, HMEC-TERT, BJ cells or TNBC cells (shown in grey). Cells were lysed and the control cell-penetrating phage (wells 11-25) and target cell penetrating phage (wells 26-30) were analysed by next generation sequencing. This method was repeated for each TNBC cell line.

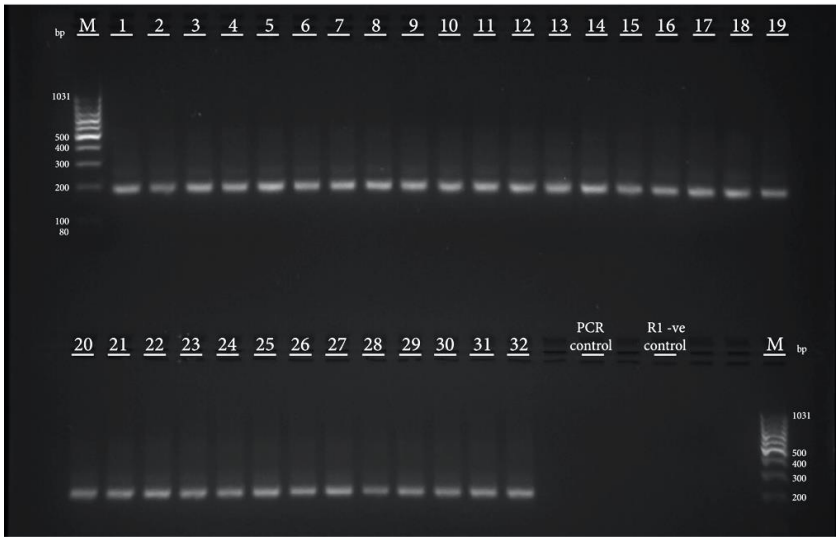
Prior to incubation with either BT-549, CAL-51 or MDA-MB-231 cells, the input phage library for each round of biopanning was screened against control cell lines for the removal of non-target specific phage, referred to as subtractive depletion. The three control cell lines used were the immortalised non-tumourigenic epithelial cell lines MCF10A and HMEC-TERT and a normal fibroblast cell line BJ (see Table 2.1). After each

round of biopanning, cell-penetrating phage were harvested by target cell lysis. The resulting output phage titre harvested from each target cell line well was calculated to be between 10^5-10^6 CFU by the ampicillin titration method (see 2.3.3). The target specific bacteriophages were then propagated in bacteria to produce an amplified selective peptide-phage library for the subsequent biopanning round. The high-throughput screen resulted in the production of 90 individual sub-libraries (30 sub-libraries for BT-549, CAL-51 and MDA-MB-231 experiments) for subsequent Ion Torrent sequencing.

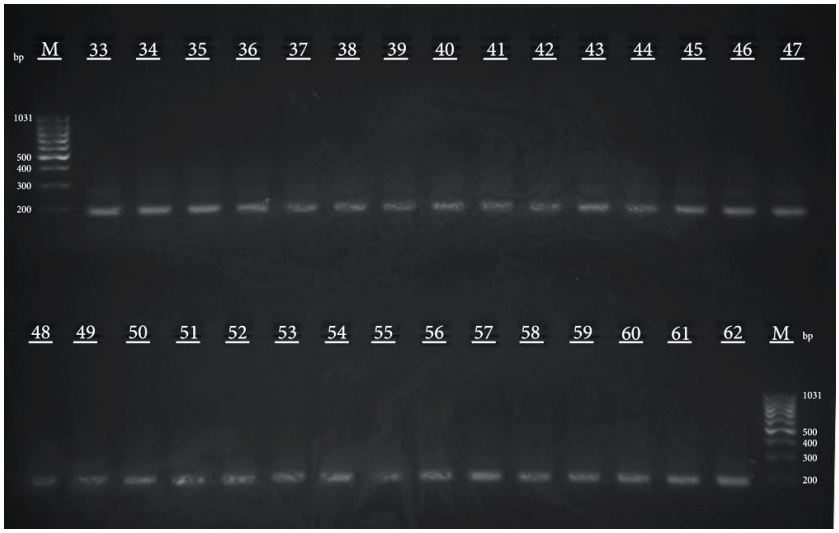
3.2.2 Next generation sequencing (NGS) sample preparation

To prepare each selective sub-library for NGS, plasmid DNA from all 90 biopanning samples was purified and the peptide gene inserts were PCR amplified using library specific primers (Figure 3.4A-C). The separation of the PCR products by gel electrophoresis confirmed that the correct 176 base pair insert was successfully amplified across all 90 samples. The negative controls included for the PCR confirmed that there was no phage contamination of the PCR master mix components or primers (PCR control, Figure 3.4A), and of the reagents used in the plasmid purification kit (R1 negative control, Figure 3.4A). Plasmid DNA purified from two separate infected TG1 stocks of the original library was also successfully amplified in the same manner (Appendix 3.3, Supplementary Figure 3).

A



B



C

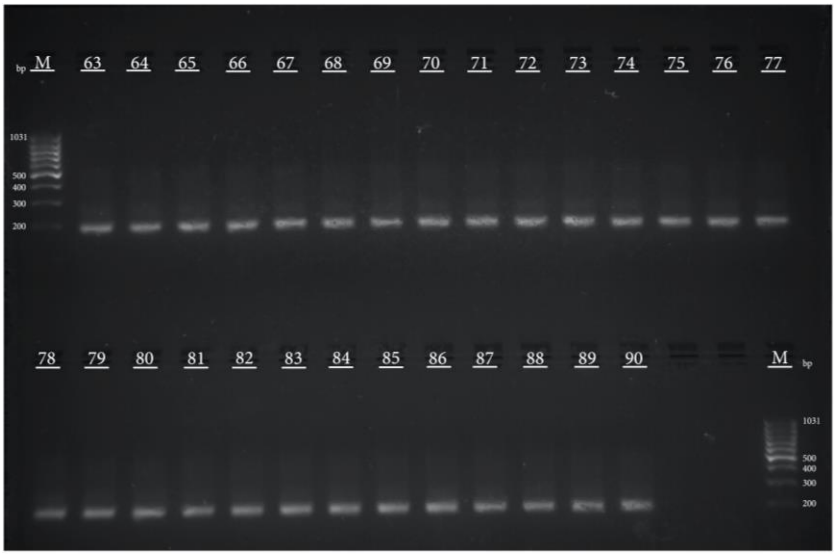
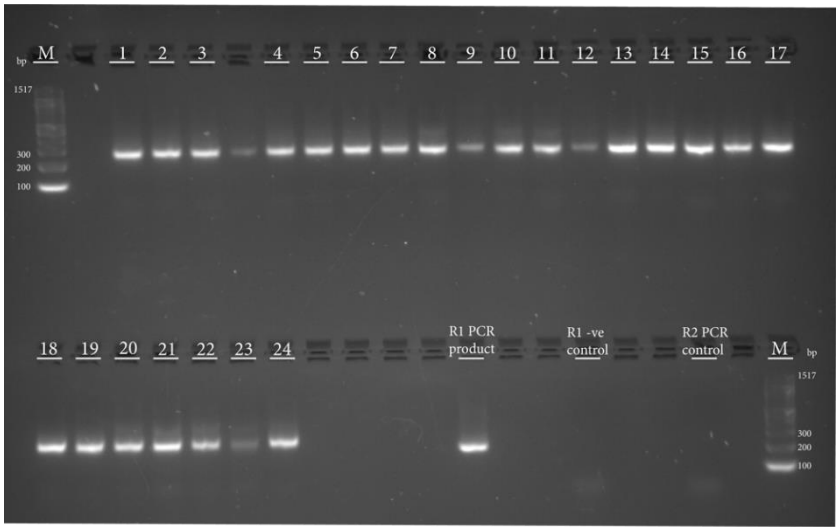


Figure 3.4 Amplification of 9-mer peptide inserts

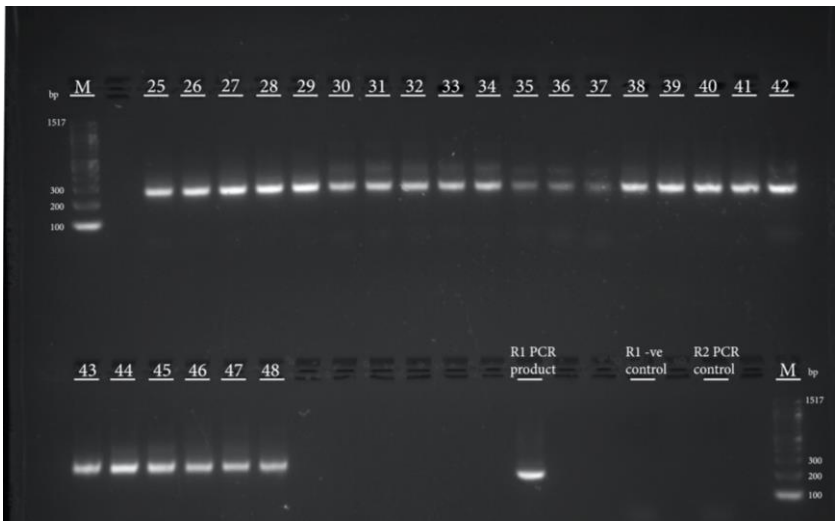
The nonapeptide gene inserts and flanking regions were amplified using library-specific primers. Amplification of the 176 base pair insert was confirmed across all 90 panning samples on a 1% agarose gel at 100V. The negative PCR and round 1 controls are also shown (A). MassRuler™ DNA ladder was used as the molecular marker.

The purified PCR products were then further amplified using a forward primer containing a unique barcode sequence and A-key sequence for Ion Torrent sequencing, and a reverse P1 primer containing a P1 adapter sequence also designed for Ion Torrent Sequencing (Figure 3.5A-D). The unique Ion Torrent barcode sequences used to label DNA from each of the 90 panning samples are shown in Appendix 2.1. The PCR products were resolved on 1% agarose gels and the

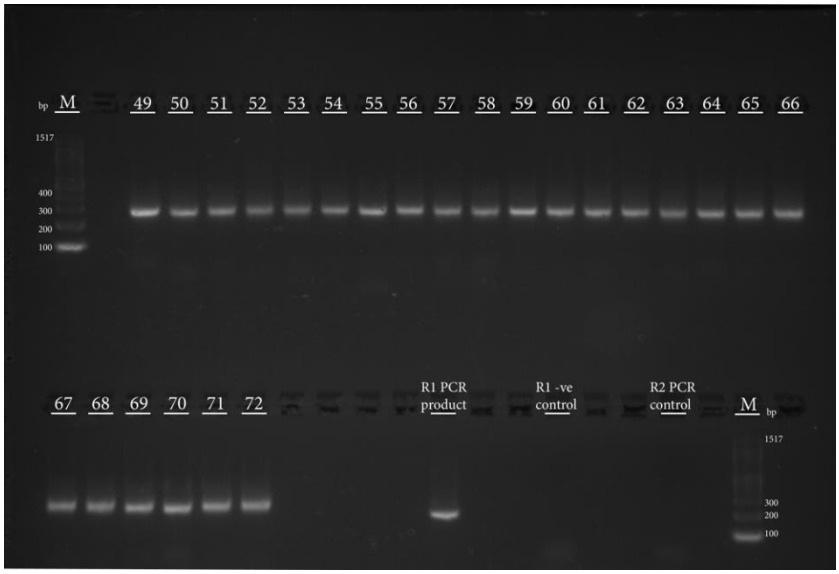
A


bp M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
1517
300
200
100
18 19 20 21 22 23 24
R1 PCR product
R1 -ve control
R2 PCR control
M bp
1517
300
200
100

B


bp M 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
1517
300
200
100
43 44 45 46 47 48
R1 PCR product
R1 -ve control
R2 PCR control
M bp
1517
300
200
100

C


bp M 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66
1517
400
300
200
100
67 68 69 70 71 72
R1 PCR product
R1 -ve control
R2 PCR control
M bp
1517
300
200
100

D

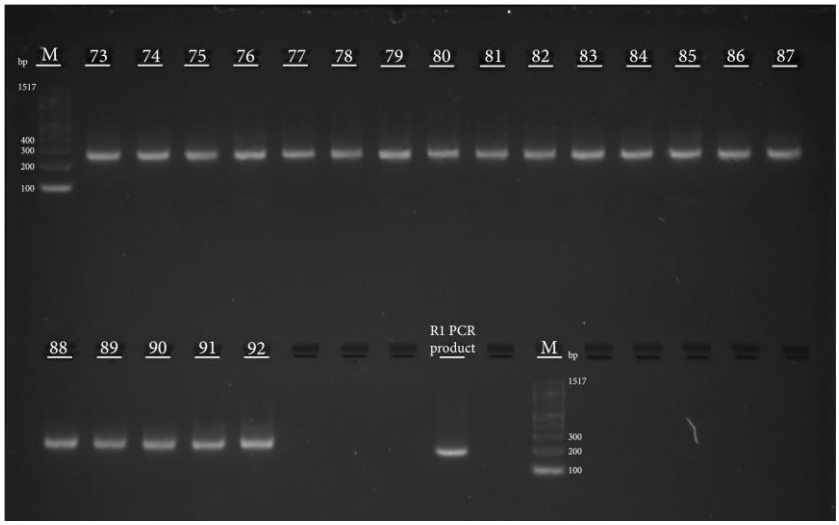


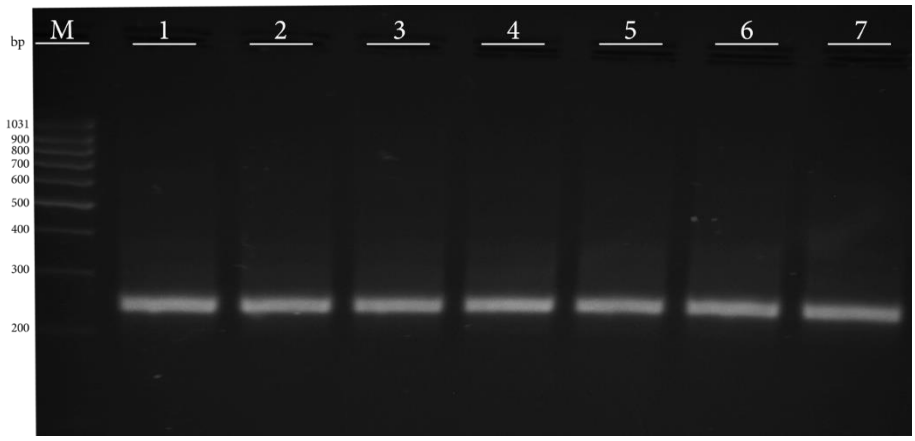
Figure 3.5 Attachment of barcode sequences for Ion Torrent sequencing

A unique barcoded sequence was attached to each first round PCR product and separated by electrophoresis on a 1% agarose gel. Amplified inserts from the original library were also barcoded (D, Lanes 91 and 92). The first-round negative controls and PCR controls were also included in each PCR run (A-C). A first-round PCR product (176 bp) was resolved on each 1% agarose gel (A-D) to confirm that the second-round PCR products (275 bp) were longer. A 100bp ladder was used as the molecular marker.

attachment of the barcoded linkers was confirmed as the size of each insert increased by 99 base pairs. A round one negative control and a no template PCR control were included in each PCR run to confirm that there was no background phage contamination of the negative control after round one and no contamination of the PCR components, respectively.

The barcoded peptide encoding inserts were pooled equally and further resolved to ensure that the correct insert had been amplified (Figure 3.6). The DNA pool did resolve into two bands after 5 hours (Figure 3.6B), however the base pair difference between the two bands was small and therefore both bands were gel extracted and purified. The pooled amplicon was then purified using paramagnetic beads and sequenced by NGS on a S5 Ion Torrent platform.

A



B

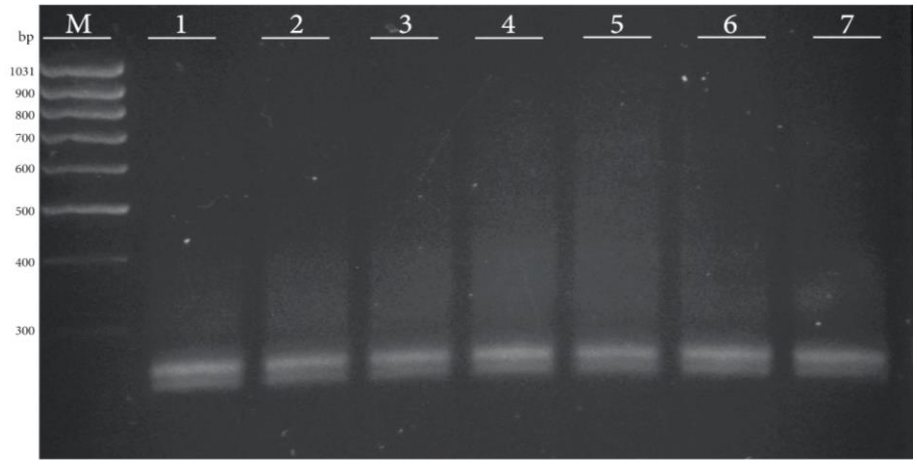


Figure 3.6 Pooled DNA resolved on a 3% MetaPhor agarose gel

The amplicon pool was resolved at 100V on a 3% MetaPhor agarose gel for (A) three hours or (B) five hours to confirm that a single product had been amplified. The pooled DNA was resolved using 7 wells (Lanes 1-7). MassRuler™ DNA ladder was used as the molecular marker.

3.2.3 Next generation sequencing analysis

Demultiplexing of the NGS data revealed that 85.6% of the total number of sequences contained an identifiable barcode. When translated, 35.3% of these sequences contained the correct PAMAMSRP and PRRRR flanking motifs and a minimum of a one amino acid insert. On average, ~250,000-350,000 barcoded sequences were identified for each target sample and ~3.5 million sequences were identified in each control pool (15 samples). Deep sequencing of the original library (see Appendix 3.2, Supplementary Figure 4) revealed that there was no overrepresentation of any peptide in the 9-mer library, confirming that the naïve library was unbiased.

To compare the frequency of nonapeptides isolated from either BT-549, CAL-51 or MDA-MB-231 target samples that are likely to be specific for the TNBC cell line, a two proportional Z test was applied (see methods 2.3.9). From the BT-549 and MDA-MB-231 biopanning experiments, five target sample replicates were sequenced, whereas four target sample replicates were sequenced from the CAL-51 panning. The Z score ranking system considered the total frequency of each peptide and the frequency of the peptide in the target and control replicates. TNBC binding peptides were isolated using a comparison script which was firstly set to select for peptides whose Z score was above the arbitrary value of 2.0 across three target cell replicates (Z>2). These peptides, apart from LESLSSMSP and SPVSSPPIT, did not appear more than twice in the original library sequences (Appendix 3.6, Supplementary Table 3).

From this analysis, a total of 13 peptides were identified. This total included four BT-549 targeting peptides, three CAL-51 targeting peptides, and six MDA-MB-231 targeting peptides (Figure 3.7). The variance in Z scores across all biological replicates revealed that the Z scores of the BT-549 peptide TLYRMVVTS, the CAL-51 peptide MGSVRDTKF and the MDA-MB-231 peptide LSLPqHAPQ varied the least across replicates (Figure 3.7B). The frequency of each peptide was then analysed across the target samples and the control pool as a percentage of the total sequences obtained for the relevant barcode or pool. As alluded to by the Z score analysis, the average percentages calculated across all biological replicates for each TNBC cell line revealed that there was a significant increase in the frequency of each peptide found in the

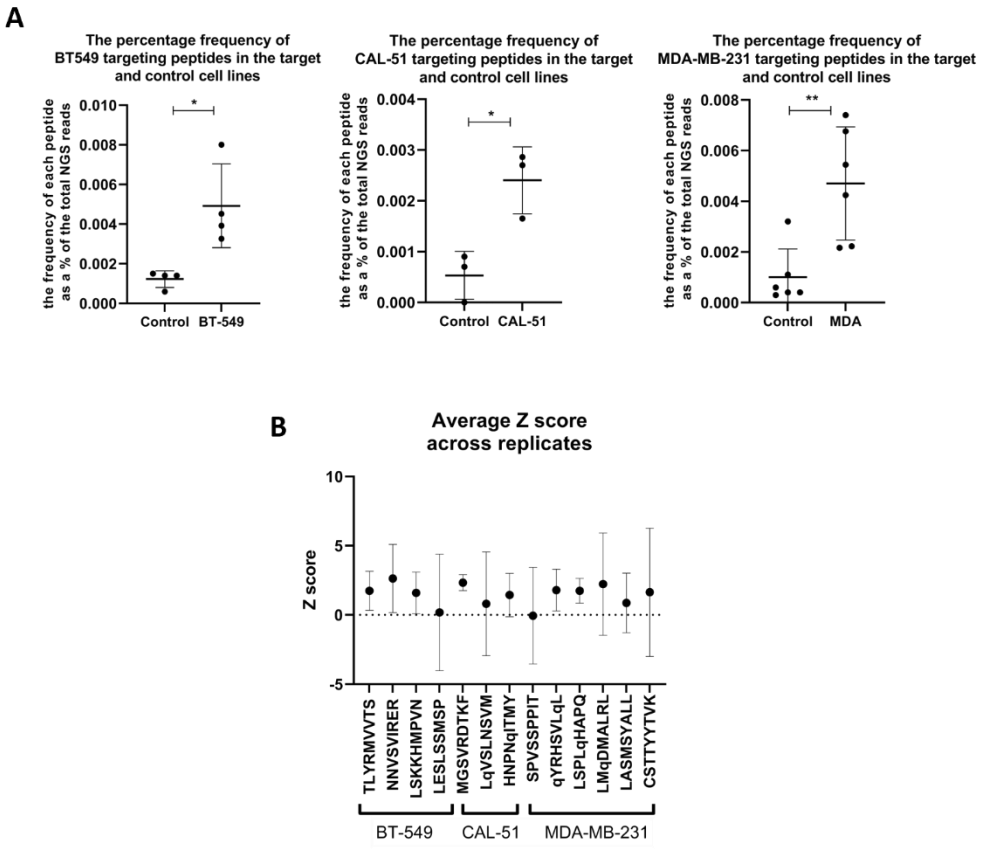


Figure 3.7 Z>2 TNBC cell-penetrating peptides

(A) The average percentage frequency of each peptide from the Z>2 analysis. Statistical significance was calculated by unpaired t-test. * $p = 0.0139$ and ** $p = 0.0046$. (B) The variance in Z score values across the biological replicates shown as the mean ± SD.

target samples compared to across the 15 control samples (Figure 3.7A, BT-549 0.0012 vs 0.0049; CAL-51 0.00053 vs 0.0024; MDA-MB-231 0.001 vs 0.0047). The normalised frequencies of the individual nonapeptides isolated from TNBC cells are shown in Figure 3.8. The peptides with the highest normalised frequency values were the BT-549 peptide NNSVIRER (0.0065%) and the MDA-MB-231 peptides qYRHSVLqL (0.0036%), LMqDMALRL (0.0071%) and CSTTYYTVK (0.0057%). The CAL-51 identified peptides MGSVRDTKF, LqVSLNSVM and HNPNqITMY had a lower frequency in CAL-51 samples than the peptides identified in BT-549 or MDA-MB-231 samples.

Of the 13 peptides identified, eight appeared to show cross-reactivity with other TNBC cell lines (Figure 3.8). The peptides that appear to show the most cross-

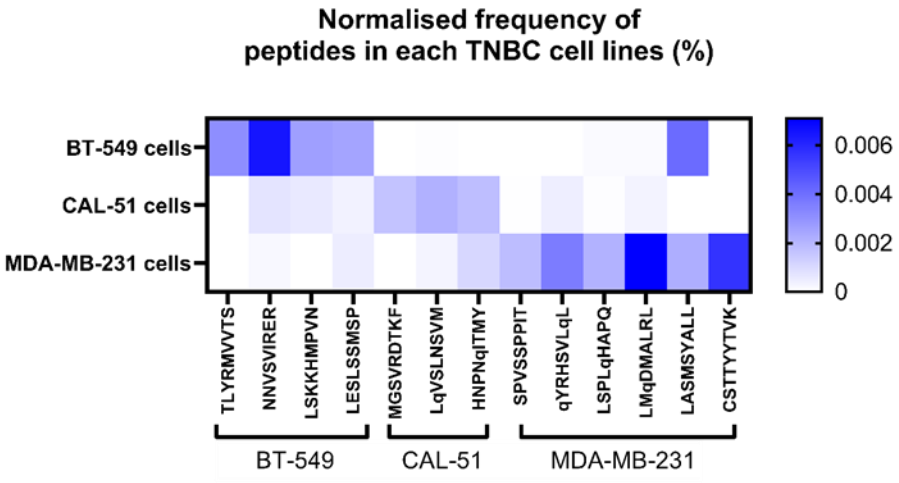


Figure 3.8 Frequency of each Z>2 peptide across all TNBC panning experiments

Heatmap displaying the normalised frequency of peptides identified to have a Z score greater than 2 across three replicates in BT-549, CAL-51 and MDA-MB-231 cells.

reactivity are the CAL-51 peptide HNPNqITMY, which is also seen in the MDA-MB-231 barcodes, and the MDA-MB-231 peptide LASMSYALL, whose normalised frequency is higher in BT-549 cells. In contrast, TLYRMVVTS, MGSVRDTKF, SPVSSPPIT, LSPLqHAPQ and CSTTYYTVK did not show cross-reactivity with other TNBC cell lines.

In order to increase the number of TNBC targeting peptides, the stringency of the analysis was altered by adjusting the threshold of the Z score criteria. Firstly, peptides with a Z score greater than 2.0 across two biological replicates were isolated. However, from this analysis, 172 BT549 targeting peptides, 189 CAL-51 targeting peptides and 198 MDA-MB-231 targeting peptides were identified. The Z score comparison script was therefore adjusted so that peptides with a Z score greater than 3.0 across two out of five biological replicates of BT-549 and MDA-MB-231 cells, or two out of four biological replicates of CAL-51 cells, were highlighted. A total of 13 BT-549 peptides, 10 CAL-51 peptides and 12 MDA-MB-231 peptides fit the Z score greater than 3 (Z>3) across two biological replicates criteria.

Of the 35 TNBC targeting peptides from the Z>3 analysis, three peptides were also present in the Z>2 analysis (NNVSVIRER, LMqDMALRL and CSTTYYTVK). In addition, the MDA-MB-231 peptide LASMSYALL from the Z>2 analysis was found in BT-549 cell replicates in the Z>3 analysis. The only peptides to be seen more than twice in the

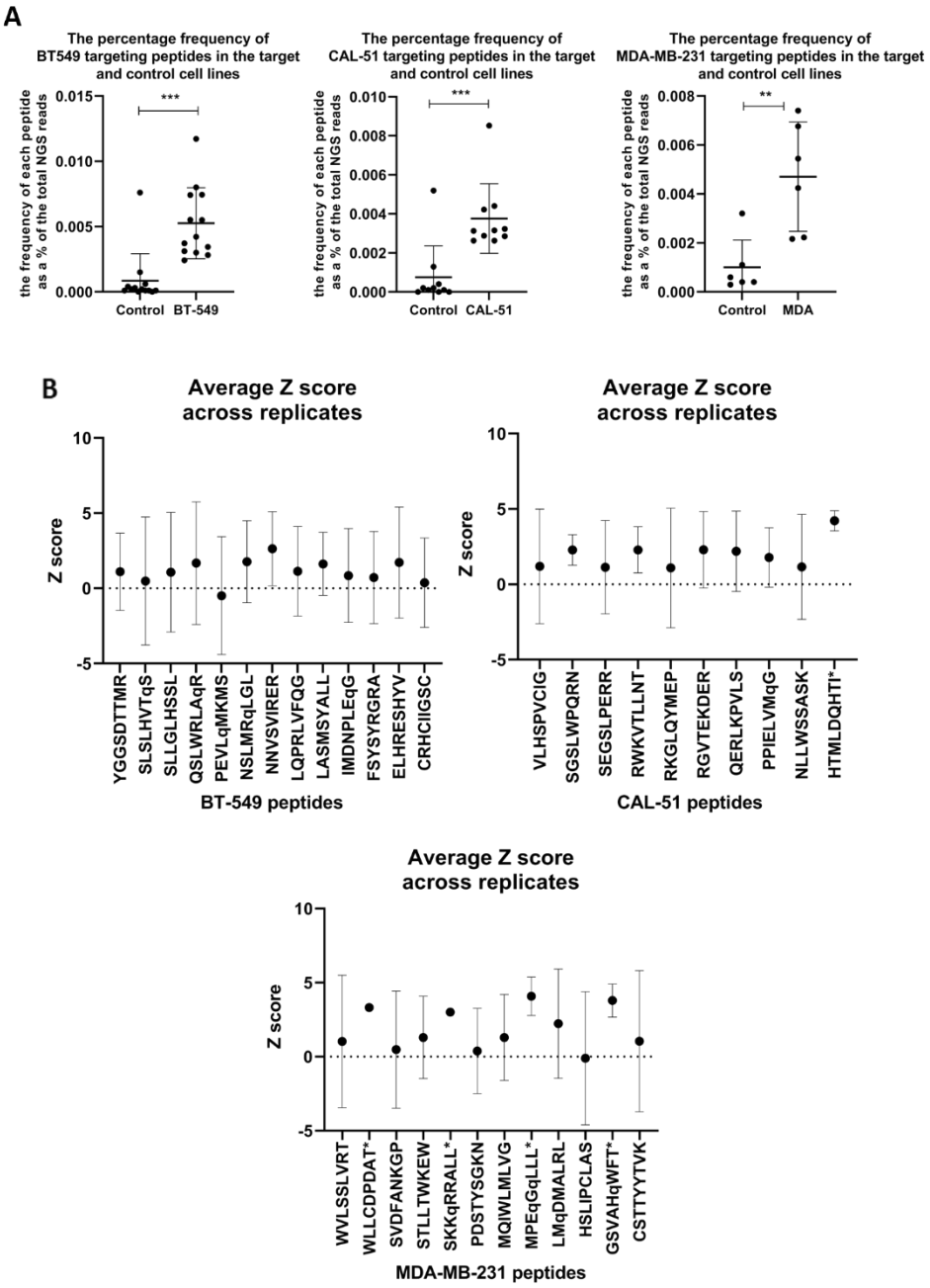


Figure 3.9 Z>3 TNBC cell-penetrating peptides

(A) The average percentage frequency of each peptide from the Z>3 analysis. Statistical significance was calculated by unpaired t-test. *** p = 0.0001 and **** p = <0.0001. (B) The variance in Z score values across the biological replicates shown as the mean ± SD. Peptides marked with an asterisk were only seen in two out of four (CAL-51) or five (BT-549 and MDA-MB-231) replicates.

original library files were the CAL-51 targeting peptides RWKVTLLNT and PPIELVMqG, which were seen 12 and 5 times, respectively (see Appendix 3.6, Supplementary Table 3).

The Z score variance for each peptide is shown in Figure 3.9B. The smallest variance in Z scores across the replicates is observed in CAL-51 targeting peptides and the largest variance in the MDA-MB-231 peptides. Some peptides were only seen in two of the TNBC replicates, therefore the mean Z score was not representative of all biological replicates. The peptides with the smallest variance in Z scores across all replicates are LASMSYALL, SGSLWPQRN, RWKVTLLNT and PPIELVMqG.

For all three TNBC cell lines, there was a significant increase in the normalised frequency of the peptides in the TNBC cell lines compared to their respective control pool (Figure 3.9A). The peptides with the highest normalised frequency, again given as a percentage of the total sequences, were the BT549 penetrating peptides QSLWRALAqR (0.0074%), NNVSVIRER (0.0065%) and ELHRESHYV (0.0073%), as well as the MDA-MB-231 targeting peptides WVLSSLVRT (0.0057%), LMqDMALRL (0.0071%) and CSTTYYTVK (0.0057%). The CAL-51 targeting peptide with the highest normalised frequency was QERLKPVLS (0.0044%). Four BT549 peptides were also observed in another TNBC cell line, with NNVSVIRER also being observed in CAL-51 replicates, and LASMSYALL and CRHCIIGSC in MDA-MB-231 replicates. In addition, the CAL-51 peptide RWKVTLLNT was also found in BT-549 samples (Figure 3.10).

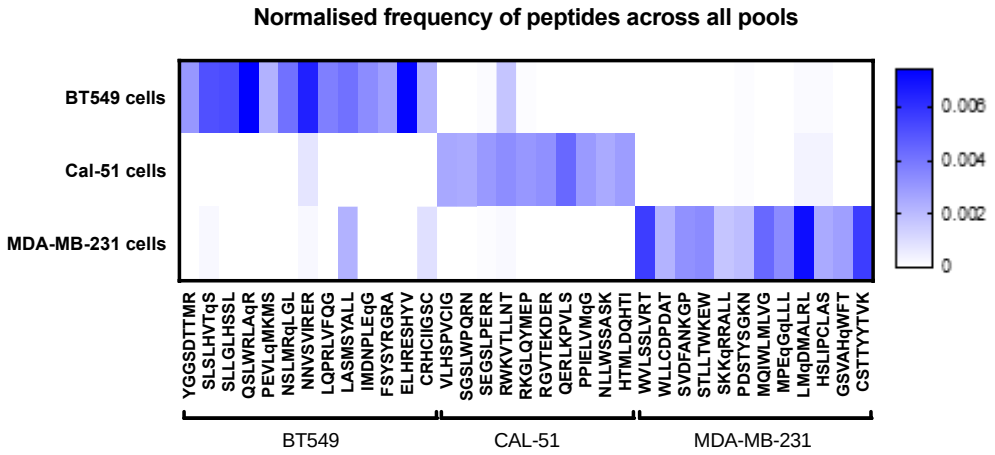


Figure 3.10 Frequency of each Z>3 peptide across all TNBC biopanning experiments

Heatmap displaying the normalised frequency of peptides identified to have a Z score greater than 3 across two replicates in BT-549, CAL-51 and MDA-MB-231 cells.

3.3 Discussion

Deep sequence analysis of the naïve and amplified selective libraries by Ion Torrent sequencing allowed for the characterisation of the diversity of the phage library and the identification of TNBC specific ligands which may not have been identified from small-scale screening (’t Hoen et al., 2012, Matochko et al., 2012, Yang et al., 2017). The use of next generation sequencing to identify the phage isolated TNBC targeting 9-mer peptide genes led to over 68 million reads being obtained. Of these reads, 85.6% were successfully labelled with a sample specific barcode (see Appendix 3.5). The multiplexing of the amplicon using barcode identifier primers therefore led to most sequences being assigned to a sample library. This is consistent with other studies where the percentage of barcoded reads ranged from 69 to >95% (Galan et al., 2010, Bybee et al., 2011, Shokralla et al., 2014). For example, Guo et al., achieved pentanucleotide labelling of 86.8% of 894 NGS samples (Guo et al., 2019). However, there were no sequences for barcode 52 which is believed to be through human error where the PCR amplified product was not included in the final DNA pool. In addition, although the concentration of each amplicon was measured and equalised for NGS, there were still differences in the number of reads obtained for each barcode (see Appendix 3.5, Supplementary Table 2). In part, the variation can be explained by differences in PCR amplification even though the amplicons were equalised, as lower reads for barcodes 35, 36 and 37 appeared to be consistent with the fainter bands seen in Figure 3.4C for these samples. The quantification method may have also been less accurate for lower DNA concentrations. Unequal amplicon concentrations may not be the only influence for differences in the barcodes reads, as a study by Binladen *et al.,* found that sequencing PCR products of equal concentration produced unequal NGS reads of each product (Binladen et al., 2007).

When translated in all three reading frames, only 35% of sequences contained the correct flanking motifs with at least one amino acid insert (see Appendix 3.5). A proportion of sequences may not have contained the correct PAMAMSRP-peptide-PRRRRRRRR when translated as not all the reads obtained from NGS will have been long enough to encapsulate the peptide gene. Additionally, the Sanger sequencing data of 16 library clones revealed that, for 4 out of 16 clones, there were multiple binding sites for the forward primer in different reading frames upstream to the

correct peptide encoding sequence (see Appendix 3.2, Supplementary Figure 2). All the Sanger sequenced clones contained a single site for reverse primer binding. Multiple PAMAMSRP sequences in some clones were incorporated during library construction and as a result of this, the annealing of the forward PCR primer in the incorrect reading frame may have resulted in the final number of sequences with the correct insert being somewhat less than the total number of NGS reads. In addition, a proportion of PCR products may have been slightly longer than expected, which is supported by the presence of two DNA bands shown on the MetaPhor agarose gel when run for 5 hours (Figure 3.6). However, as the base pair difference between the two bands was small, it was experimentally difficult to separate the two bands. Therefore, if one band had not been sequenced, a proportion of the correct peptide encoding sequences would have been lost. Although the percentage of peptide encoding sequences is quite low, there were still approximately 20.6 million nonapeptide sequences, which were sufficient to investigate TNBC-peptide binders.

The naïve random peptide library should be of a high diversity to increase the possibility of identifying target binding ligands but also be unbiased towards particular phage displayed peptides (Krumpe and Mori, 2007). In order to investigate whether the naïve library was indeed unbiased, peptide encoding genes were sequenced from two TG1 *E.coli* stocks of the original library (samples 91 and 92). A total of 532,972 peptide sequences were obtained from the library stocks which were used to study library bias (see Appendix 3.5, Supplementary Table 2). Bias can occur through the presence of parasitic phage in the peptide library, which have a particular propagation advantage and would lead to the over-representation of some peptides irrespective of target selection (Matochko et al., 2014). Sequencing revealed that no TNBC peptide identified in this study appeared to be over represented in the initial 9-mer library, therefore confirming that the library was unbiased, and no fusion peptide led to the growth advantage of a particular bacteriophage (see Appendix 3.6, Supplementary Table 3).

Biopanning against control cell lines for non-target cell phage depletion and target phage selection against the TNBC cell lines BT-549, CAL-51 and MDA-MB-231 successfully isolated an amplified target specific phage pool for each TNBC cell line. Phage peptide libraries can, however, be selectively enriched against *in situ*, *in vitro* or

in vivo targets. Biopanning against immobilised *in vitro* targets, such as purified receptors, has been shown to reduce background interference from non-specific binders and leads to the identification of peptides with a high affinity for the target molecule. However, the homogenous environment and artificial presentation of a target molecule may result in the identification of high affinity binders which may not have been isolated through *in situ* or *in vivo* studies. In contrast, screening against *in situ* and *in vivo* targets retains the structure, function and expression levels of the peptide target (Wu et al., 2016). For example, an *in situ* phage display study conducted by Gallo *et al.,* where integrin-α11/β1 was overexpressed on the cell surface, led to the identification of a diverse panel of anti-integrin-α11/β1 antibodies which targeted multiple epitopes that were able to block collagen binding (Gallo et al., 2020).

In this study, the 9-mer peptide library was screened *in situ* against whole cells as cell specific peptides can be identified without knowing the specific cell surface target and whole cell screening in this manner may lead to the identification of a novel cancer targeting ligand (McGuire et al., 2009, Jäger et al., 2007). For example, the ovarian cancer specific peptide SWQIGGN was isolated from whole cell phage screening and the tumour targeting RLQLKL cell-penetrating peptide was identified when panning against tissue extracts (Whitney et al., 2010, Zhou et al., 2015a). In addition, a phage peptide library also led to the discovery of a 9 amino acid peptide, named RPMrel, which was found to specifically bind intact colon cancer cells. Furthermore, the peptide coupled to a mitochondrial toxin induced cell death in a carcinoma cell line but not in a well-differentiated cell line (Kelly and Jones, 2003). Target specific phage identified in this project were peptides which targeted unknown TNBC antigens and facilitated the internalisation of their corresponding phage particle into TNBC cells. Thus, the peptides identified by this peptide-phage screening study may alone exhibit anti-cancer effects or be utilised in the development of a therapeutic to deliver a cytotoxic agent directly into TNBC cells.

A total of 44 peptides were determined as TNBC targeting through the bioinformatic pipeline analysis. As the bacteriophage biopanning setup resulted in the collection of phage extracted from control and target cell replicates, a Z score analysis script was able to detect statistically significant differences in the frequency of a given peptide in target cell replicates compared to control cell replicates and rank peptides accordingly.

As mentioned, the Z score threshold was a value given to rank peptides based on their enrichment across target cell replicates. Therefore, the peptides identified from this analysis need to be verified experimentally to confirm that enriched peptides are indeed promising candidates for TNBC targeting and to test the true TNBC cell penetrating ability of each peptide. Many NGPD studies have utilised Z score thresholds in a similar manner (Yuan et al., 2018, Vekris et al., 2019). Naqid *et al.,* found that applying a Z score cut off of 2.0 to rank potential target specific phage-peptides isolated a number of peptides which, when synthesised, were shown to be discriminatory for their corresponding target by ELISA, thus indicating that the bioinformatic strategy applied to our study is effective in isolating target-specific peptides (Naqid et al., 2016). All the nonapeptides identified were also not observed in the PepBank database, suggesting that the novel peptides were not enriched due to being parasitic phage or due to the peptides binding to biopanning components such as plastic (Shtatland et al., 2007). There was also no overall consensus between peptides from Seq2Logo analyses, indicating that the peptides did not share a common TNBC binding domain (Thomsen and Nielsen, 2012) (see Appendix 3.7, Supplementary Figure 5).

Firstly, the peptides which had a Z score greater than 2.0 in three or more TNBC replicates (Z>2) when compared to the control cell lines were deemed to be enriched and indicative of TNBC cell specific penetration (Figure 3.7A). When considering all biological replicates, CAL-51 (n=4), BT-549 (n=5) and MDA-MB-231 (n=5), the Z score demonstrated variation (Figure 3.7B). The Z scores may have differed due to general variations that can occur across replicates, such as the number of cells seeded per well, the efficiency of bacteriophage pool transfer across wells and the success of the final isolation and amplification of target-specific bacteriophage through target cell lysis and TG1 *E.coli* infection. However, the stringency of the threshold resulted in peptides possessing a Z score higher than 2.0 in most replicates being selected and the normalised frequency of peptides across all samples supports this, with the average frequency of peptides in the target samples being significantly higher than the average frequency of a given peptide in the control pool (Figure 3.7A). Therefore, it is assumed that these peptides are indeed enriched for the TNBC target despite some variation across replicates.

The normalised frequency of each Z>2 peptide was also assessed to compare the TNBC binding ability of peptides and whether a peptide identified to be BT549, CAL-51 or MDA-MB-231 targeting was also present in other TNBC cell line samples (Figure 3.8). The peptides with the highest normalised frequencies, NNVSVIRER, LMqDMALRL and CSTTYYTVK may penetrate their respective TNBC cell line more readily and therefore may be more promising candidates for PDC development. In addition, peptides that have been internalised by more than one TNBC cell line, such as LASMSYALL, may target different TNBC subtypes or cell subpopulations and therefore have a higher efficacy in cancer therapy than peptides which only target one of the cell lines.

To expand the list of TNBC peptide binders and to ensure the stringency of the analysis was not leading to the omission of potential candidates, the Z score threshold was altered to an arbitrary value of 3.0 across two or more replicates (Z>3). Even though the number of replicates was decreased, which may decrease the reliability and the reproducibility of the peptide binding data, the Z score cut off was increased so that only the most frequently observed peptides were selected. Therefore, possible TNBC binders which may not be present in all replicates due to handling errors or insufficient cell lysis were still considered. Surprisingly, the variation in Z score for each peptide across all the replicates in the Z>3 analysis (Figure 3.9B) was similar to the peptides from the Z>2 analysis (Figure 3.7B), where more replicates were considered in the Z score ranking. The small variation, such as for the CAL-51 peptides SGSLWPQRN and RWKVTLLNT, shows that even though the peptides did not meet the Z score threshold of 2.0 in three biological replicates criteria, the phage-peptides were still present and still targeted TNBC cells in these samples. The average normalised frequency for each peptide across all replicates is shown in Figure 3.9A, where again the average obtained from the BT549, CAL-51 or MDA-MB-231 peptide frequencies is significantly higher than the average in either of their respective control pool. This data infers that the peptides isolated from the bioinformatics analysis are discriminatory for TNBC cell lines, meaning the Z>3 peptides are also promising candidates for TNBC targeting therapy.

Of the 35 peptides which had a Z score greater than three across two replicates, four of the peptides were also seen in the Z>2 analysis (NNVSVIRER, LASMSYALL, LMqDMALRL and CSTTYYTVK), suggesting that these peptides readily bind to either

BT-549, CAL-51 or MDA-MB-231 cells as they are still identified when the stringency of the Z score statistical analysis was altered. Again, the peptide sequences with the highest normalised frequencies, QSLWRLAqR, NNVSVIRER, WVLSSLVRT and LMqDMALRL, may be internalised more readily upon binding to a specific TNBC surface marker compared to other nonapeptides. The four peptides which facilitated phage internalisation in more than one cell line may, as previously mentioned, be the better candidates for TNBC target treatment. Interestingly, the Z>3 analysis clearly highlights that the peptide LASMSYALL targets both BT549 and MDA-MB-231 cells and not CAL-51 cells (Figure 3.10). This cell line-specific targeting may occur through the expression of a particular cell surface antigen in BT-549 and MDA-MB-231 cells but not in CAL-51 cells.

In summary, the implementation of NGPD in this study led to the identification of a small number of peptide candidates being selected from a diverse bacteriophage library in a high-throughput screen. The Z score analysis inferred that the peptides identified discriminate TNBC from non-malignant mammary cells and therefore may be better candidates for the development of a TNBC targeted drug delivery system. Furthermore, the NGPD biopanning and bioinformatic strategy applied have previously been shown to successfully identify highly specific peptide ligands (Naqid et al., 2016). However, target-unspecific ligands can still be identified from NGPD screening methods, therefore potential target-specific ligands need to be validated by other experimental approaches (Vodnik et al., 2011). Furthermore, the binding of a pIII displayed peptide as part of the phage protein scaffold may differ from the binding of the peptide when synthesised or recombinantly expressed (’t Hoen et al., 2012). Therefore, to confirm these findings and further characterise the 9-mer peptide candidates, further *in vitro* validation assays were performed and presented in the following chapter.

Chapter 4: *In Vitro* Validation of TNBC targeting peptides

4.1 Introduction

The first aim of this study was to identify TNBC discriminatory small peptides to be utilised in peptide-based delivery systems for TNBC targeted treatment. In relation to cancer therapeutics, there are three classes of peptides which have been studied for drug delivery: fusogenic peptides, cell penetrating peptides (CPPs) and tumour targeting peptides. Fusogenic peptides are derived from the N-terminal domain of the HA2 subunit of the influenza viral protein, hemagglutinin. The amphipathic, pH-sensitive peptides are able to mediate cellular endocytic uptake and lead to the efficient release of drugs or therapeutic agents by destabilising the endosomal membrane once internalised (Plank et al., 1994, Plank et al., 1998, Samec et al., 2022a). This class of peptides has mainly been applied for the delivery of siRNA-mediated therapies to overcome the retention and subsequent lysosomal degradation of siRNA in target cells (Cantini et al., 2013). For example, Samec *et al.,* efficiently delivered siRNA targeting casein kinase II which resulted in a 94% knockdown of CSNK2A1 mRNA compared to non-targeted siRNA in ovarian cancer cells (Samec et al., 2022a). Additionally, Cantini *et al.,* found that delivering siRNA targeting CIP2A using a fusogenic peptide sequence led to 18-fold higher amounts of siRNA being delivered into oral cancer cells than siRNA alone which significantly decreased CIP2A mRNA and protein expression levels (Cantini et al., 2013). However, clinically available siRNA and mRNA delivery involves the use of lipid nanoparticles, such as lipid nanoparticle-encapsulated mRNA vaccines (Adams et al., 2018, Baden et al., 2020, Schoenmaker et al., 2021).

CPPs have also been applied to cancer therapy to increase the transduction efficiency of cell-impermeable therapeutic cargos across the plasma membrane. The 10 to 30 amino acid peptides consist mainly of basic residues and can translocate drugs, nucleic acids, proteins and nanoparticles across the membrane by energy-dependant endocytosis (Drin et al., 2003, Dissanayake et al., 2017). One of the first discovered CPPs was the trans-activating transcriptional activator TAT sequence, GRKKRRQRRRPPQ**,** from human immunodeficiency virus-1 (HIV-1) (Frankel and Pabo, 1988). Zhang *et al.,* found that conjugating palmitoylated TAT to doxorubicin could

increase the accumulation and subsequent retention of the anticancer drug in drug sensitive and resistant cervical cancer cell lines (Zhang et al., 2014). Leading on from the discovery of the protein transduction domain (PTD) TAT, further research found that arginine plays an important role in CPP translocation, with octa-arginine and nona-arginine sequences leading to efficient internalisation (Mitchell et al., 2000, Futaki et al., 2001). Zhang *et al.,* showed that oct-arginine resulted in a high transfection efficiency of HDM2-siRNA in lung tumour cells, thus leading reduced cell proliferation through HDM2 gene knockdown (Zhang et al., 2006). 8R has also been applied to functionalise drug delivery systems such as to improve the intracellular delivery of doxorubicin-loaded PEGylated liposomes, resulting in increased cancer cell cytotoxicity *in vitro,* and to deliver taxol into drug-resistant ovarian cancer cell lines (Dubikovskaya et al., 2008, Biswas et al., 2013).

As mentioned, CPPs play a vital role in translocating therapeutic agents across cell membranes, but they lack specificity. Tumour targeting peptides, however, lead to the specific homing of a therapeutic to cancerous tissue by targeting cells that overexpress a particular cancer-specific antigen, such as integrins, or G protein or tyrosine kinase coupled receptors (Worm et al., 2020). The interaction of a targeting peptide with its corresponding receptor can then lead to the translocation of an anticancer drug or therapy by clathrin-mediated endocytosis, where the therapeutic agent is released into the cytosol by lysosomal degradation. For example, targeting the vascular endothelial growth factor receptor (VEGF) with a targeting peptide led to VEGF-receptor mediated endocytosis and the subsequent release of MMAE into target cells, leading to inhibition of VEGF activity (Michigami et al., 2021).

Several peptides have been identified to target receptors which are overexpressed in TNBC cells. For instance, p160, NGR and RGD target have been shown to target keratin 1, CD13 and integrin, respectively (Sugahara et al., 2010, Soudy et al., 2012, Soudy et al., 2017). The epidermal growth factor receptor (EGFR) is also of interest in TNBC therapeutics, as the receptor is overexpressed in a large number of patients and has been correlated with higher grade tumours and a poor clinical outcome (Park et al., 2014, Changavi et al., 2015). Hossein-Nejad Ariani *et al.,* found that four analogues of the dodecapeptide YHWYGYTPQNVI had a high uptake in BCa cell lines which expressed high levels of EGFR, making the EGFR peptides promising candidates in the

development of EGFR targeting PDCs (Hossein-Nejad-Ariani et al., 2019). Furthermore, when YHWYGYTPQNVI was reversibly cross linked to polymersomal doxorubicin (GE11-PS-DOX), *in vivo* biodistribution experiments revealed that GE11-PS-DOX achieved 2.5 fold higher accumulation of doxorubicin in tumours compared to liposomal doxorubicin in EGFR overexpressing SKOV3 ovarian tumour xenografts in mice (Zou et al., 2018). In addition to targeting cytotoxic drugs to cancer receptor expressing tissue, tumour targeting peptides can deliver other molecular cargos such as siRNAs. For instance, Gandioso *et al.,* were able to induce selective gene silencing in $HER2^+$ cell lines when covalently coupling the HER2 targeting peptide, TAT-AHNP, to siRNAs via click chemistry (Gandioso et al., 2017).

In addition, antibody-peptide fusions combining the specificity of monoclonal antibodies and the cell penetrating ability of small peptides may improve the efficacy of targeted therapy. For instance, when a cell-penetrating peptide was fused to the C-terminal of a Hepatitis B virus X protein targeting mAb, the recombinant antibody achieved efficient cell penetration which enabled the suppression of viral transcription and the promotion of an antiviral state in both *in vitro* and *in vivo* studies (Zhang et al., 2018). Regarding tumour targeting, Gaston *et al.,* fused known cell-penetrating peptides to a carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5) specific mAb and found that PEP-1, an amphipathic 21-residue peptide previously reported by Morris *et al.,* facilitated the internalisation of the IgG antibody (Morris et al., 2001, Gaston et al., 2019).

A major limitation to the use of peptides is their susceptibility to proteases and their short serum half-life. However, small peptides can be chemically modified or cyclised to improve their bioavailability (McGregor, 2008, Le Joncour and Laakkonen, 2018). An example of this is the decapeptide WXEAAYQkFL (where k is D-lysine), a cyclic analogue of the BCa cell line targeting peptide p160, which was identified when a random peptide phage library was screened against BCa cells. As well as displaying a high uptake in BCa cells, the N to C-terminus amide cyclised peptide was found to be stable to proteolysis and to preferentially accumulate in mouse orthotropic MDA-MB-231 tumours (Raghuwanshi et al., 2017). Chemical modifications of peptides include lipidation or the addition of polyethylene glycol (PEG) polymers to increase the serum stability of cell penetrating peptides (Frokjaer and Otzen, 2005, Zhang et al., 2012).

Small peptides have also been shown to selectively bind to cancer cell subpopulations. A phage display screening study by Jones *et al.,* demonstrated the selective binding of peptides to MCF-7 cells which had undergone TGFβ-induced EMT through subtractive depletion against breast epithelial cells and fibroblasts. The EMT specific phages were further screened against EMT enriched cell lines MDA-MB-231 and T47D-shNMI and primary and metastatic TMAs. This resulted in the identification of a 7 amino acid peptide LGLRGSL which showed specific binding to EMT BCa cells (Jones et al., 2018). Small peptides may also be utilised in stem cell targeted therapy, as an *in vitro* study found that MDA-MB-231 isolated stem cells could be specifically targeted by a dodecamer peptide (Liu et al., 2016).

In the previous chapter, potential TNBC targeting peptides were identified when a diverse small peptide bacteriophage library was selectively enriched for the TNBC cell lines, BT-549, CAL-51 and MDA-MB-231. However, as previously mentioned, NGPD can lead to the identification of non-target specific ligands or ligands which are unable to bind to a target outside of the bacteriophage protein scaffold, meaning further *in vitro* validation of peptide candidates is required (Vodnik et al., 2011, ’t Hoen et al., 2012). Therefore, this chapter firstly aimed to validate the phage identified TNBC peptide candidates by confirming the binding of each peptide to TNBC cells. To achieve this, the TNBC targeting peptides were expressed with the CPP sequence 9R and monovalent red fluorescent protein (mRFP) for TNBC cell specific binding and intracellular delivery studies. Peptide-mediated binding and subsequent intracellular delivery of each peptide-mRFP into control (MCF10A and HMEC-TERT) and TNBC (BT-549, CAL-51 and MDA-MB-231) cell lines was then assessed by fluorescent microscopy and flow cytometry. The most promising peptide candidates from the flow cytometric analysis were also complexed to magnetic nanoparticles (MNPs) and MNP uptake was assessed by Prussian blue staining and inductively coupled plasma mass spectrometry (ICP-MS). Overall, the *in vitro* validation and characterisation of the NGPD peptides in this chapter aimed to identify the more promising TNBC targeting peptides for their use in TNBC targeted therapeutics.

4.2 Results

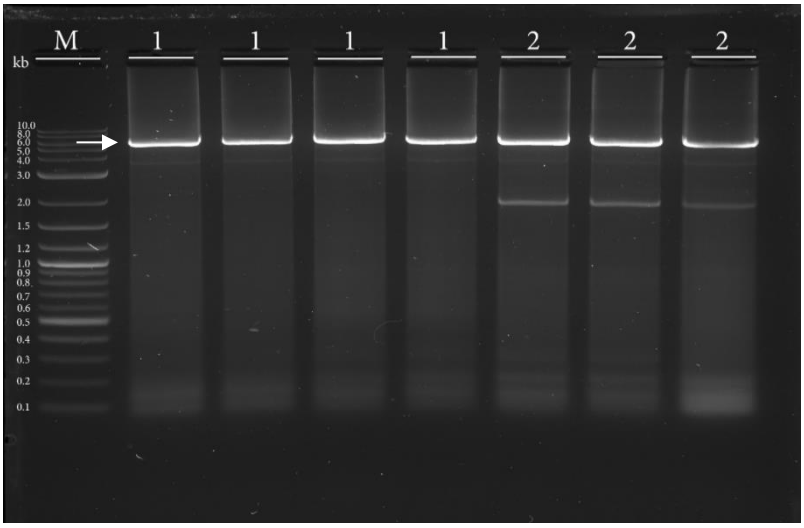
4.2.1 Recovery of peptide encoding sequences into a monovalent red fluorescent protein expressing vector

To confirm whether the peptide candidates identified from the NGPD screening and bioinformatic analysis target TNBC cells, the 44 candidate 9-mer peptides were recovered by the insertion of the individual peptide encoding sequences into a pGEX-6P1-8R-mRFP expression vector by inverse PCR (see Appendix 4.1, Supplementary Figure 6). Cloning the peptide encoding sequences directly upstream to the monomeric red fluorescent protein (mRFP) gene therefore led to peptide-mRFP recombinant protein expression for fluorescent microscopy and flow cytometric analysis. The peptides that were identified from the first Z score analysis (Z score greater than 2.0 across three or more replicates) were referred to as peptides 1 to 13, whereas peptides identified from the less stringent Z score analysis (Z score greater than 3.0 across two or more replicates) were referred to as peptides 14 to 44.

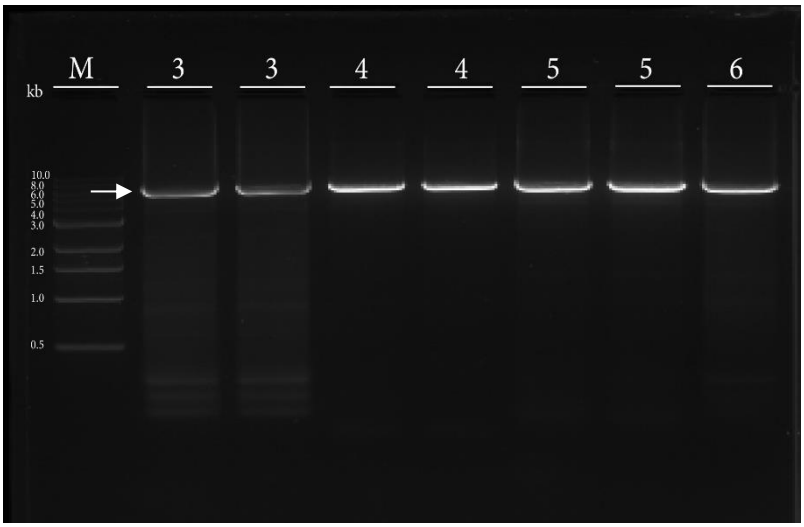
Individual primers were designed to incorporate the TNBC specific 9-mer peptide sequences and the phagemid vector flanking regions, MSRP and PRRRRRRRRR, into the pGEX-6P1-mRFP by inverse PCR (see Appendix 4.2, Supplementary Figure 6). The nona-arginine CPP was also included to enhance the endocytic uptake of the final TNBC peptide-mRFP recombinant proteins as the identified TNBC targeting peptides may not be able to cross the plasma membrane (Dixon et al., 2016, Dissanayake et al., 2017). The separation of the resulting PCR products by gel electrophoresis confirmed that the correct size sequence had been inserted into the expression vector for both the Z score greater than 2.0 peptides (Figure 4.1), and the Z score greater than 3.0 peptides (see Appendix 4.3).

Ampicillin resistant bacterial clones for each peptide were then miniprepped for sequencing. Sanger sequencing analysis of each recombinant plasmid revealed that all the peptide encoding sequences were successfully cloned into the expression vector (Figure 4.2).

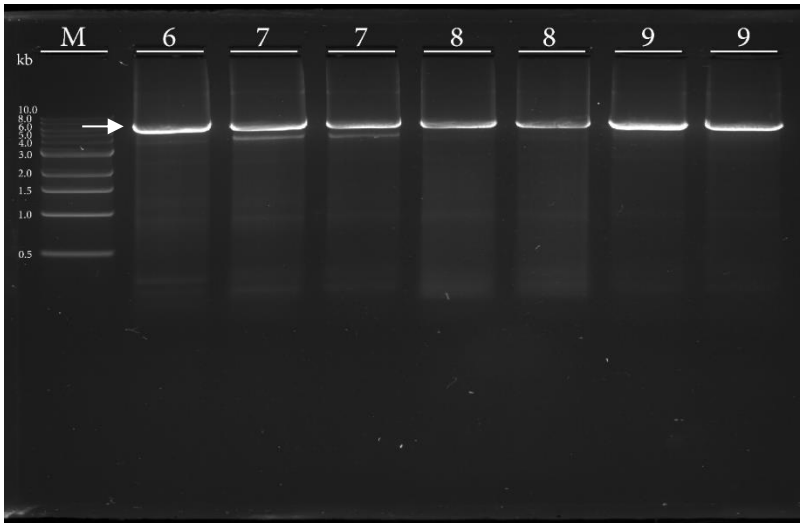
A


M
1
1
1
1
2
2
2
kb
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

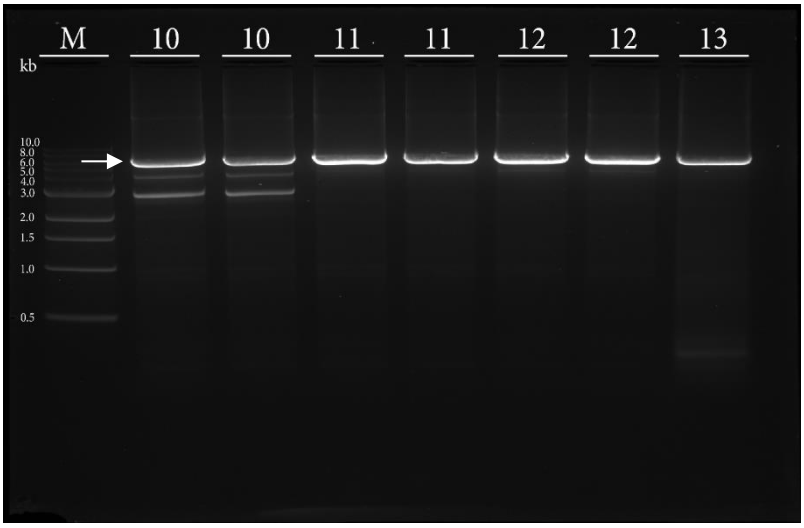
B


M
3
3
4
4
5
5
6
kb
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5

C


M
6
7
7
8
8
9
9
kb
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5

D



E

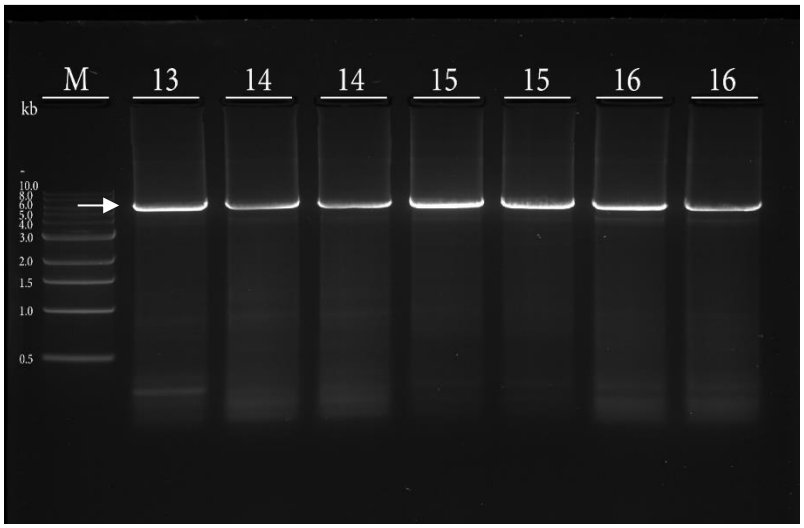


Figure 4.1 Insertion of 9-mer peptide encoding sequences into the mRFP expression vector

The DNA encoding sequences for the 13 peptides identified from the Z>2 criteria were cloned into the pGEX-6P1-mRFP expression vector by inverse PCR. The resulting PCR products are observed at around 5.8kb on a 1% agarose gel (A-E). A 1kb ladder was used as the molecular marker.

PEPTIDE 1 LEVLFQGPMSRP**TLYRMVVTS**PRRRRRRRRRTSTMASSE

PEPTIDE 2 LEVLFQGPMSRP**NNVSVIRER**PRRRRRRRRRTSTMASSE

PEPTIDE 3 LEVLFQGPMSRP**LSKKHMPVN**PRRRRRRRRRTSTMASSE

PEPTIDE 4 LEVLFQGPMSRP**LESLSSMSP**PRRRRRRRRRTSTMASSE

PEPTIDE 5 LEVLFQGPMSRP**SPVSSPPIT**PRRRRRRRRRTSTMASSE

PEPTIDE 6 LEVLFQGPMSRP**QYRHSVLQL**PRRRRRRRRRTSTMASSE

PEPTIDE 7 LEVLFQGPMSRP**LSPLQHAPQ**PRRRRRRRRRTSTMASSE

PEPTIDE 8 LEVLFQGPMSRP**LMQDMALRL**PRRRRRRRRRTSTMASSE

PEPTIDE 9 LEVLFQGPMSRP**LASMSYALL**PRRRRRRRRRTSTMASSE

PEPTIDE 10 LEVLFQGPMSRP**CSTTYYTVK**PRRRRRRRRRTSTMASSE

PEPTIDE 11 LEVLFQGPMSRP**MGSVRDTKF**PRRRRRRRRRTSTMASSE

PEPTIDE 12 LEVLFQGPMSRP**LQVSLNSVM**PRRRRRRRRRTSTMASSE

PEPTIDE 13 LEVLFQGPMSRP**HNPNQITMY**PRRRRRRRRRTSTMASSE

PEPTIDE 14 LEVLFQGPMSRP**YGGSDTTMR**PRRRRRRRRRTSTMASSE

PEPTIDE 15 LEVLFQGPMSRP**SLSLHVTQS**PRRRRRRRRRTSTMASSE

PEPTIDE 16 LEVLFQGPMSRP**SLLGLHSSL**PRRRRRRRRRTSTMASSE

PEPTIDE 17 LEVLFQGPMSRP**QSLWRLAQR**PRRRRRRRRRTSTMASSE

PEPTIDE 18 LEVLFQGPMSRP**PEVLQMKMS**PRRRRRRRRRTSTMASSE

PEPTIDE 19 LEVLFQGPMSRP**NSLMRQLGL**PRRRRRRRRRTSTMASSE

PEPTIDE 20 LEVLFQGPMSRP**LQPRLVFQG**PRRRRRRRRRTSTMASSE

PEPTIDE 21 LEVLFQGPMSRP**IMDNPLEQG**PRRRRRRRRRTSTMASSE

PEPTIDE 22 LEVLFQGPMSRP**FSYSYRGRA**PRRRRRRRRRTSTMASSE

PEPTIDE 23 LEVLFQGPMSRP**ELHRESHYV**PRRRRRRRRRTSTMASSE

PEPTIDE 24 LEVLFQGPMSRP**CRHCIIGSC**PRRRRRRRRRTSTMASSE

PEPTIDE 25 LEVLFQGPMSRP**VLHSPVCIG**PRRRRRRRRRTSTMASSE

PEPTIDE 26 LEVLFQGPMSRP**SGSLWPQRN**PRRRRRRRRRTSTMASSE

PEPTIDE 27 LEVLFQGPMSRP**SEGSLPERR**PRRRRRRRRRTSTMASSE

PEPTIDE 28 LEVLFQGPMSRP**RWKVTLLNT**PRRRRRRRRRTSTMASSE

PEPTIDE 29 LEVLFQGPMSRP**RKGLQYMEP**PRRRRRRRRRTSTMASSE

PEPTIDE 30 LEVLFQGPMSRP**RGVTEKDER**PRRRRRRRRRTSTMASSE

PEPTIDE 31 LEVLFQGPMSRP**QERLKPVLS**PRRRRRRRRRTSTMASSE

PEPTIDE 32 LEVLFQGPMSRP**PPIELVMQG**PRRRRRRRRRTSTMASSE

PEPTIDE 33 LEVLFQGPMSRP**NLLWSSASK**PRRRRRRRRRTSTMASSE

PEPTIDE 34 LEVLFQGPMSRP**HTMLDQHTI**PRRRRRRRRRTSTMASSE

PEPTIDE 35 LEVLFQGPMSRP**WVLSSLVRT**PRRRRRRRRRTSTMASSE

PEPTIDE 36 LEVLFQGPMSRP**WLLCDPDAT**PRRRRRRRRRTSTMASSE

PEPTIDE 37 LEVLFQGPMSRP**SVDFANKGP**PRRRRRRRRRTSTMASSE

PEPTIDE 38 LEVLFQGPMSRP**STLLTWKEW**PRRRRRRRRRTSTMASSE

PEPTIDE 39 LEVLFQGPMSRP**SKKQRRALL**PRRRRRRRRRTSTMASSE

PEPTIDE 40 LEVLFQGPMSRP**PDSTYSGKN**PRRRRRRRRRTSTMASSE

PEPTIDE 41 LEVLFQGPMSRP**MQIWLMLVG**PRRRRRRRRRTSTMASSE

PEPTIDE 42 LEVLFQGPMSRP**MPEQGQLLL**PRRRRRRRRRTSTMASSE

PEPTIDE 43 LEVLFQGPMSRP**HSLIPCLAS**PRRRRRRRRRTSTMASSE

PEPTIDE 44 LEVLFQGPMSRP**GSVAHQWFT**PRRRRRRRRRTSTMASSE

Figure 4.2 Translated Sanger sequencing results for recombinant plasmids

Each peptide encoding sequence is shown in bold. The sequence was inserted between the HRV 3C site (shown in purple) and the mRFP1 (shown in red) original vector sequences. Sequencing data was analysed using SnapGene viewer (see Appendix 4.4).

In order to determine the specificity of each peptide for TNBC cells by fluorescent microscopy and flow cytometric analysis, the binding ability of three control mRFP recombinant proteins was also assessed. The two negative control proteins included in this study were mRFP only and mRFP coupled to the 9R protein transduction domain, named 9R-mRFP. The positive control protein selected included a lysine rich heparin binding domain peptide, p21. The positively charged 21-residue peptide, KRKKKGKGLGKKRDPCLRKYK, has been shown to greatly increase cell-penetrating peptide mediated endocytosis by Dixon *et al.,* (Dixon et al., 2016). The p21 peptide was therefore expressed as the recombinant protein p21-9R-mRFP.

In a similar manner to the 9-mer peptides, inverse PCR primers were designed to clone the control peptide sequences into the pGEX-6P1-mRFP vector (see Appendix 4.2). As the MRSP and 9RRRRRRRRR flanking regions were inserted into the pGEX-6P1-mRFP vector as part of the recovery of the 9-mer peptide clones, the expression construct for peptide 1 was used as a template to obtain the 9R, 9R-mRFP and p21-9R-mRFP recombinant vectors. The amplification of the correct inverse PCR product for mRFP and 9R-mRFP was confirmed by gel electrophoresis (Figure 4.3) and the successful cloning of each negative control peptide was confirmed by Sanger sequencing (see Appendix 4.4).

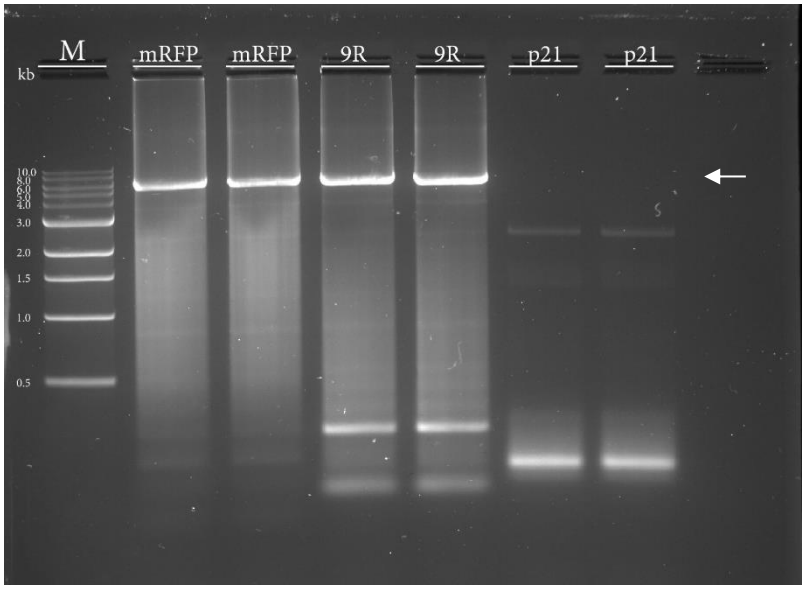


Figure 4.3 Insertion of the control peptide sequences into the mRFP expression vector

The DNA encoding sequences for mRFP, 9R-mRFP and p21-9R-mRFP were cloned into the pGEX-6P1-mRFP expression vector by inverse PCR. The correct size PCR product was observed at around 5.8kb (indicated by a white arrow). A 1kb ladder was used as the molecular marker.

However, the correct inverse PCR product was not observed for p21-9R-mRFP. A new set of primers were therefore designed to incorporate this sequence into the original pGEX-6P1-mRFP vector. The inverse PCR with the original vector designed primers also did not lead to the amplification of a 5.8kb product. As the p21 primers were long, the PCR conditions were altered to increase the number of annealing cycles, which led to the amplification of the correct p21 PCR product when both sets of primers were tested (Figure 4.4) and the insertion of the correct p21 encoding sequence (see Appendix 4.4).

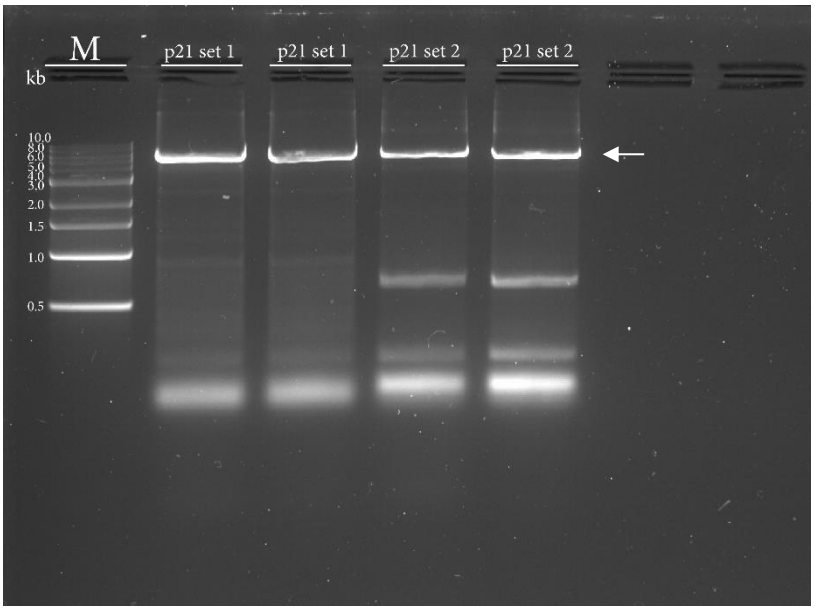


Figure 4.4 Insertion of the p21 sequence into the mRFP expression vector

The DNA encoding sequence for the 21-mer positive control peptide was cloned into the expression vector by inverse PCR. The correct size PCR product was observed at around 5.8kb (indicated by a white arrow) and gel excised. A 1kb ladder was used as the molecular marker.

Next, fluorescently labelled recombinant proteins were expressed from the individual peptide encoding constructs and purified. To ensure that IPTG induced protein expression and Glutathione Sepharose™ 4B resin purification successfully led to recombinant protein being obtained, three peptides were randomly selected (peptides 2, 8 and 9) and the purification protocol followed. Separation of 500ng of purified recombinant protein for each peptide by 1D SDS-PAGE confirmed that purified peptide-mRFP was present in the eluted fractions (Figure 4.5). However, the Coomassie stained polyacrylamide gel revealed that peptide-mRFP remained on the

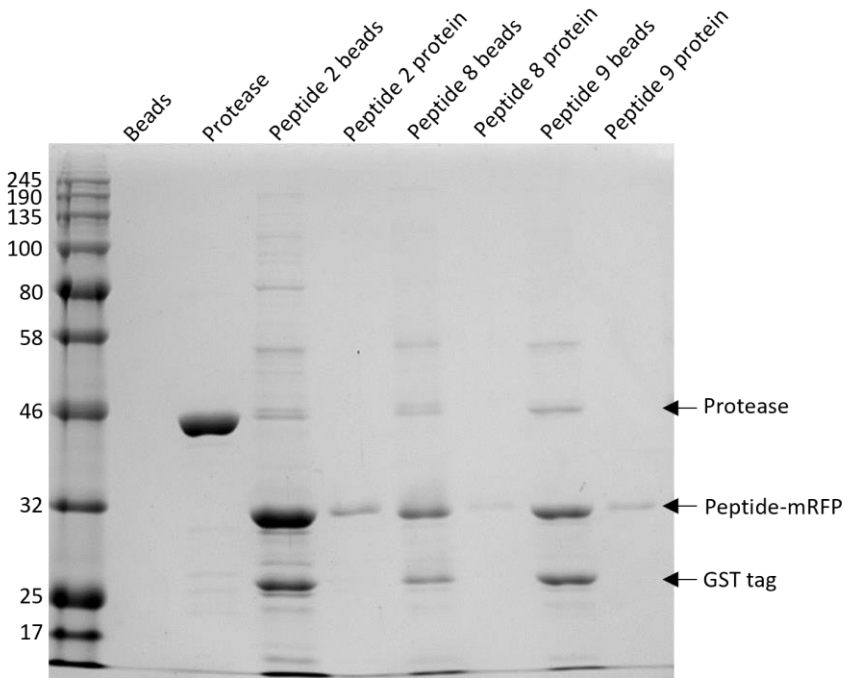


Figure 4.5 Purification of peptide-mRFP using Glutathione Sepharose™ 4B resin

10ul of unused resin, 1ul of PreScission protease, 10ul of cleaved resin and 500ng of eluted recombinant protein were loaded onto a 12% polyacrylamide gel and proteins were separated by SDS-PAGE and stained with Coomassie blue. A colour-stained broad range marker was used as the molecular weight marker.

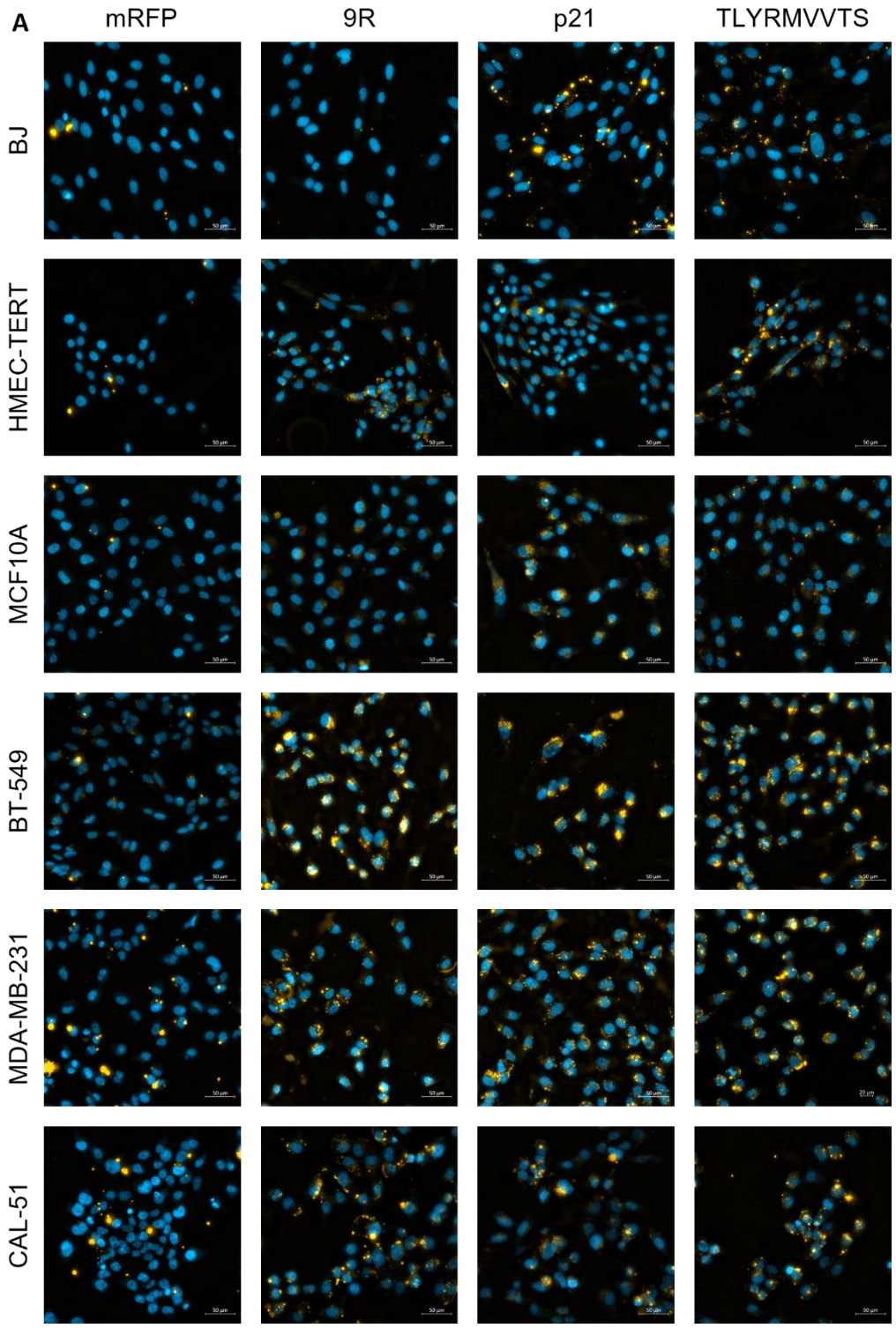
glutathione beads after cleavage with PreScission protease. Therefore, the protein expression and purification protocol was altered by increasing the number of bacteria in the TG1 expression culture and the concentration of PreScission protease was increased by decreasing the volume of cleavage buffer for subsequent elutions. This resulted in higher concentrations of the final purified proteins being obtained.

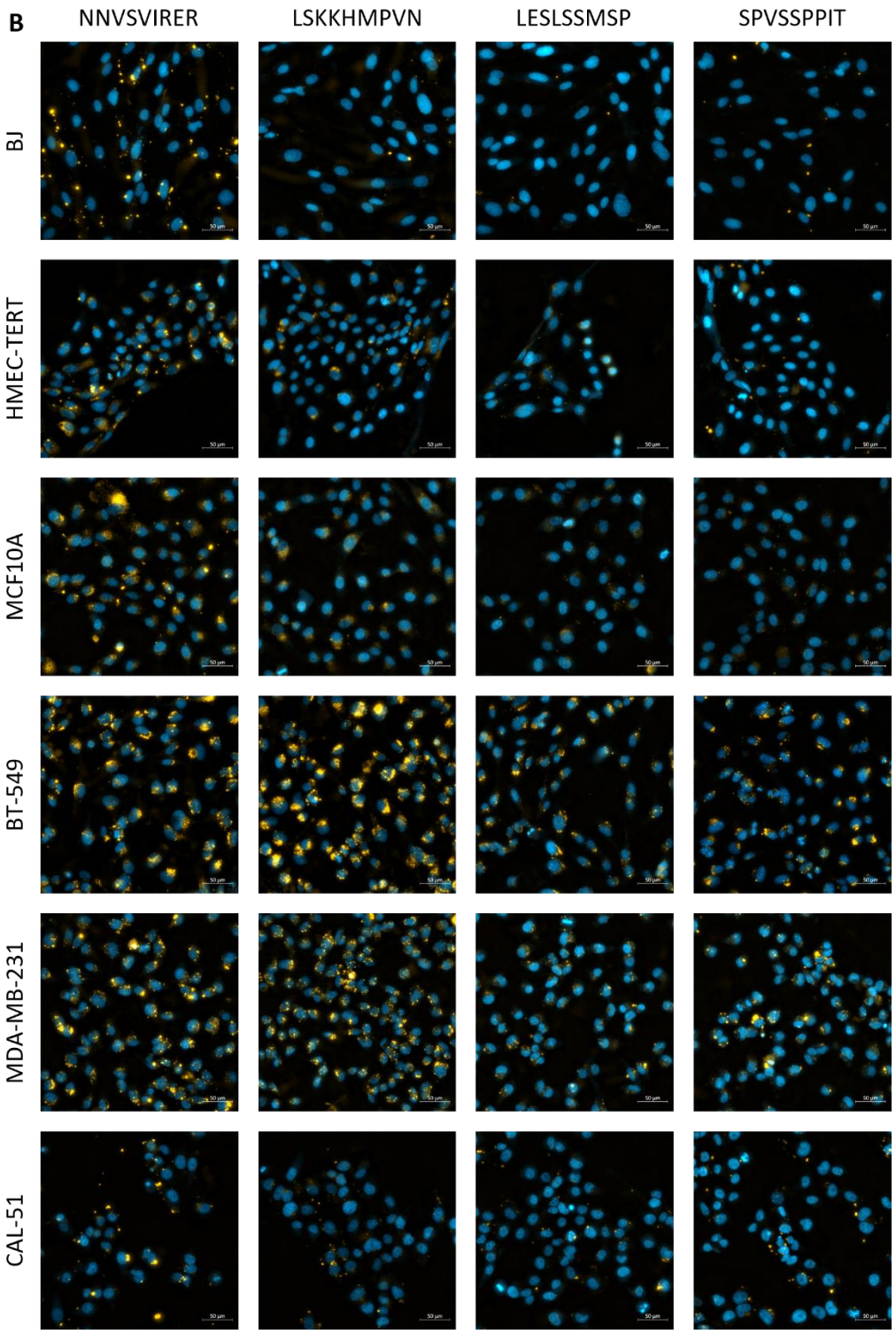
As the peptides identified from the first Z score criteria were found to be enriched in more target cell replicates than the second Z score analysis, peptides 1 to 13 were prioritised for *in vitro* validation. Therefore, the BT-549 targeting peptides TLYRMVVTS, NNVSVIRER, LSKKHMPVN, LESLSSMSP, the CAL-51 targeting peptides MGSVRDTKF, LQVSLNSVM and HNPNQITMY and the MDA-MB-231 targeting peptides, SPVSSPPIT, QYRHSVLQ, LSPLQHAPQ, LMQDMALRL, LASMSYALL, CSTTYYTVK were expressed and purified for subsequent peptide uptake analyses.

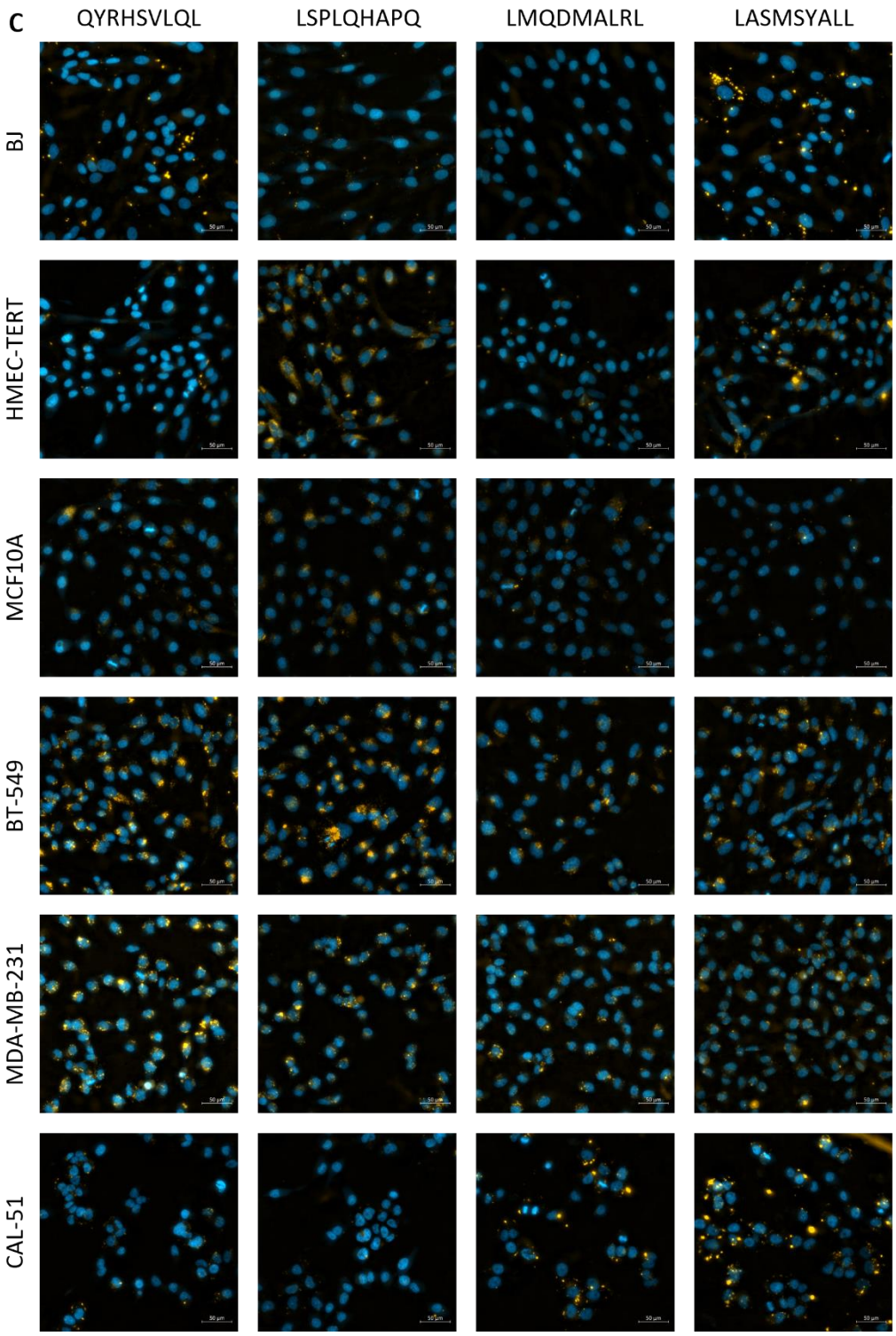
4.2.2 Cellular uptake of TNBC peptides by fluorescent microscopy

The endocytic uptake of the recombinantly expressed peptide-mRFPs was first visualised by fluorescent microscopy to assess whether the peptides interacted with the control and TNBC cell lines used in the NGPD screen. To confirm that the recombinant proteins could bind to the cells and that internalised peptide-mRFPs could be visualised by fluorescent microscopy, the positive control p21-9R-mRFP was incubated with the control cell lines BJ, MCF10A, HMEC-TERT, and the TNBC cell lines BT-549, CAL-51 and MDA-MB-231 cells. This resulted in mRFP fluorescence being observed for all cell types (Figure 4.6A, p21). In contrast, there was only a small amount of mRFP fluorescence observed when each cell type was incubated with the mRFP only negative control (Figure 4.6A, mRFP). The 9R-mRFP control resulted in an increase in mRFP fluorescence compared to the mRFP only control in MCF10A, HMEC-TERT, BT-549, CAL-51 and MDA-MB-231 wells. However, no mRFP or 9R-mRFP was observed in BJ cells (Figure 4.6A, 9R).

The BT-549 targeting peptide, TLYRMVTTS, appeared to lead to an increased uptake of mRFP in the TNBC cell lines as mRFP was observed around the nuclei of BT-549, CAL-51 and MDA-MB-231 cells. Fluorescence could also be seen in the non-tumourigenic cell lines, with fluorescence being observed in HMEC-TERT and MCF10A cells. However, the TLYRMVTTS-mRFP fluorescence appeared to be lower in the control cell lines compared to the TNBC cell lines (Figure 4.6A, TLYRMVTTS). The peptides NNVSVIRER and LSKKHMPVN appeared to target BT-549 and MDA-MB-231 cells the most compared to the other cell lines, with considerably less staining observed in CAL-51 cells. The NNVSVIRER recombinant protein did, however, appear to lead to an increase in mRFP seen in all the control cell lines compared to the fluorescence observed with the mRFP and 9R-mRFP negative controls (Figure 4.6B, NNVSIRER). In contrast, there was less mRFP observed in the control cell lines with LSKKHMPVN. The peptide also appeared to lead to the most fluorescence in BT-549 cells compared to the other Z>2 peptides (Figure 4.6B, LSKKHMPVN). There was limited mRFP observed with the LESLSSMSP-mRFP recombinant protein, with the least fluorescence observed in the control and CAL-51 cell lines. The peptide did appear to result in more fluorescence in the BT-549 and MDA-MB-231 cells, however there was less mRFP observed with LESLSSMSP when compared to the other BT-549 targeting peptides


A
mRFP
9R
p21
TLYRMVVTS
BJ
HMEC-TERT
MCF10A
BT-549
MDA-MB-231
CAL-51


B
NNVSVIRER
LSKKHMPVN
LESLSSMSP
SPVSSPPIT
BJ
HMEC-TERT
MCF10A
BT-549
MDA-MB-231
CAL-51


C
QYRHSVLQL
LSPLQHAPQ
LMQDMALRL
LASMSYALL
BJ
HMEC-TERT
MCF10A
BT-549
MDA-MB-231
CAL-51

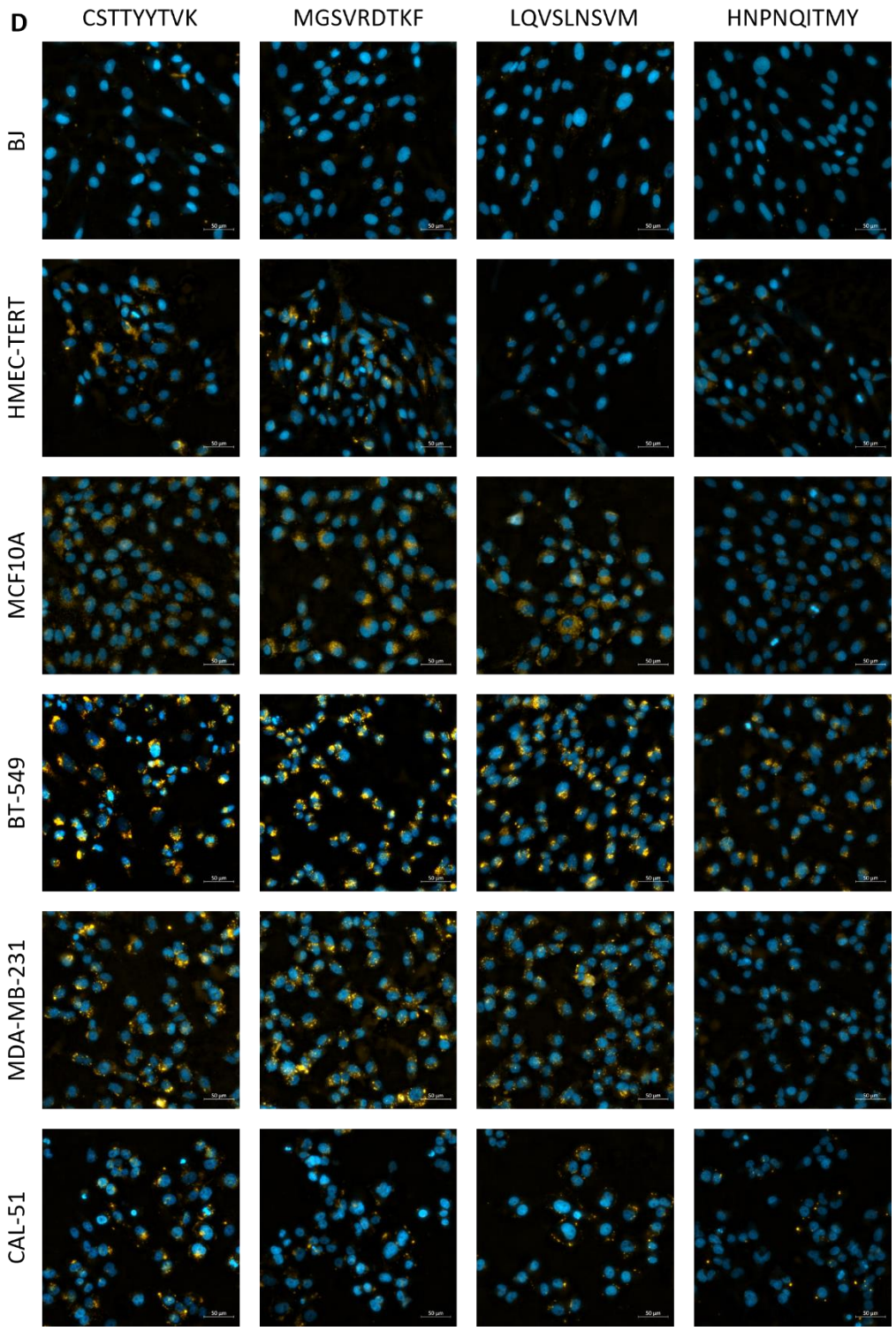


Figure 4.6 Binding of each mRFP labelled peptide to BJ, MCF10A, HMEC-TERT, BT-549, MDA-MB-231 and CAL-51 cells

Cell nuclei were stained with Hoechst 33342 (blue) to identify cell specific mRFP fluorescence (orange). Cells were fixed with 4% PFA following incubation with each peptide for 24 hours (n=3). All images were captured using a 40X objective.

(Figure 4.6B, LESLSSMSP).

SPVSSPPIT-mRFP binding is also shown in Figure 4.6. Like LESLSSMSP, there was only a small amount of mRFP observed with SPVSSPITT. The peptide did appear to lead to more mRFP observed in the BT-549 and MDA-MB-231 samples, however the fluorescence was less than that observed with the fluorescently labelled TLYRMVTTS, NNVSVIRER and LSKKHMPVN peptides (Figure 4.6B, SPVSSPITT). The MDA-MB-231 targeting peptides QYRHSVLQL and LSPLQHAPQ did not lead to any mRFP in CAL-51 cells, but mRFP was observed around the nuclei of BT-549 and MDA-MB-231 cells. Some mRFP was also seen when the control cell lines were incubated with QYRHSVLQL-mRFP or LSPLQHAPQ-mRFP, with QYRHSVLQL resulting in mRFP in BJ cells and LSPLQHAPQ resulting in fluorescence being observed in HMEC-TERT cells. Although faint, there was also some mRFP fluorescence seen in MCF10A wells with both peptides (Figure 4.6C, QYRHSVLQL and LSPLQHAPQ). In contrast, there was limited fluorescence observed in each control cell line with LMQDMALRL-mRFP. Furthermore, this peptide led to intracellular mRFP being observed in all the TNBC cell lines, albeit at low intensity (Figure 4.6C, LMQDMALRL). TNBC cellular uptake of mRFP was also present with LASMSYALL, however a considerable amount of mRFP was also observed in BJ and HMEC-TERT cells (Figure 4.6C, LASMSYALL). The final MDA-MB-231 peptide, CSTTYYTVK, appeared to lead to slightly more fluorescence observed in MCF10A cells than the other potential MDA-MB-231 targeting peptides, however intracellular fluorescence observed in the BT-549, CAL-51 and MDA-MB-231 cells appeared brighter than that observed in the control cells (Figure 4.6D, CSTTYYTVK).

The final Z score greater > 2 peptides, MGSVRDTKF, LQVSLNSVM and HNPNQITMY, are shown in Figure 4.6D. Even though the peptides were identified to be CAL-51 cell targeting from the phage study, all three of the peptides only lead to a small amount of mRFP being observed in CAL-51 cells. Regarding the control cell lines, mRFP fluorescence was seen in HMEC-TERT cells when MGSVRDTKF was the targeting peptide. Additionally, MGSVRDTKF and LQVSLNSVM led to fluorescence being observed in MCF10A cells. In contrast, there was little fluorescence observed when all the CAL-51 targeting recombinant proteins were incubated with BJ cells. A similar amount of fluorescence was observed in BT-549 cells with MGSVRDTKF-mRFP and LQVSLNSVM-mRFP, with both peptides appearing to lead to specific mRFP binding.

The peptides also led to the cellular uptake of their corresponding protein in MDA-MB-231 cells. On the other hand, there was little fluorescence observed in the non-tumourigenic and TNBC cell lines when the cells were subjected to HNPNQITMY-mRFP. However, cell specific uptake of HNPNQITMY-mRFP could still be seen in all TNBC cell lines.

The fluorescence microscopy data for all fluorescently labelled peptides is summarised in Table 4.1. The arbitrary values of +, ++ and +++ were assigned to each image based on the level of cell specific intracellular fluorescence observed in each cell line for the control and TNBC targeting peptides. Image analysis found that the peptide-mediated delivery of mRFP did vary depending on the control or TNBC peptide used, however differences in the level of mRFP uptake could not be clearly determined, particularly in BT-549 and MDA-MB-231 cells. Therefore, flow cytometry was employed to quantify peptide-mediated mRFP in the non-tumourigenic and TNBC cell lines.

Table 4.1 Summary of mRFP fluorescence observed by fluorescence microscopy

	Control cell lines			TNBC cell lines		
Peptides	BJ	HMEC-TERT	MCF10A	BT-549	CAL-51	MDA-MB-231
mRFP				+		
9R		++	++	+++	++	+++
p21		+	+++	+++	++	+++
TLYRMVVTS		+++	++	+++	++	+++
NNVSVIRER	+	++	+++	+++	+	+++
LSKKHMVN		+	++	+++		+++
LESLSSMSP		+	+	+++		+++
SPVSSPPIT				+++		+++
QYRHSVLQL			+	+++	+	+++
LSPLQHAPQ		++	++	+++		+++
LMQDMALRL				+++	++	+++
LASMSYALL	+			+++	++	+++
CSTTYYTVK		+	+++	+++	++	+++
MGSVRDTKF		+	+++	+++	+	+++
LQVSLNSVM			+++	+++	++	+++
HNPNQITMY			+	+++		+++

+ cell specific fluorescence in 25-50% of cells, ++ cell specific fluorescence in 50-75% of cells, +++ cell specific fluorescence in greater than 75% of cells

4.2.3 Cellular uptake of TNBC peptides by flow cytometry

Fluorescence microscopy confirmed that each recombinant protein was able to lead to the cell specific uptake of mRFP in one or more control or TNBC cell line. However, the differential uptake of each peptide-mRFP between the cell lines used in this study could not be determined from the qualitative data. Therefore, flow cytometry was employed so that the peptide-mediated endocytic uptake of the individual peptide-mRFPs by HMEC-TERT, MCF10A, BT-549, CAL-51 and MDA-MB-231 cells could be quantified.

To determine whether each peptide specifically targeted one or more TNBC cell line, the cellular uptake of the Z score greater than 2.0 peptides in each TNBC cell line was compared to the uptake of the same peptide in HMEC-TERT and MCF10A cells. Even though BJ cells were used as a control cell line for the phage display screening study, this cell type was not used in flow cytometric analysis as fluorescence microscopy revealed that there was limited cellular uptake of all peptide-mRFPs in BJ cells (Figure 4.6).

An example of the flow cytometry data collected for BT-549 cells is shown in Figure 4.7. Flow cytometric analysis revealed that an increase in mRFP positive binding events, defined as binding events found above the mRFP only gate, were observed with 9R-mRFP, p21-mRFP and TLYRMMVTS-mRFP, with p21 (p = < 0.0001) and TLYRMMVTS (p = 0.0038) leading significant increases in the percentage of mRFP positive cells (Figure 4.7E). To measure the uptake of each peptide candidate by non-tumourigenic and TNBC cell lines, mean mRFP fluorescence values were calculated for each peptide-mRFP to assess fluorescence intensity. Mean fluorescence values were calculated as an average of the raw fluorescence values recorded, with more than 20,000 single cells per sample. The single cell population was assessed for each sample based on forward and side scatter characteristics (Figure 4.7A, gate A). As the 9R CPP may lead to differing levels of endocytic uptake in each cell line, the mean fluorescence values that were measured for each TNBC peptide-mRFP with each cell line were normalised based on the mean fluorescence value that was recorded when the cell penetrating peptide mRFP which possessed no targeting peptide (9R-mRFP) was

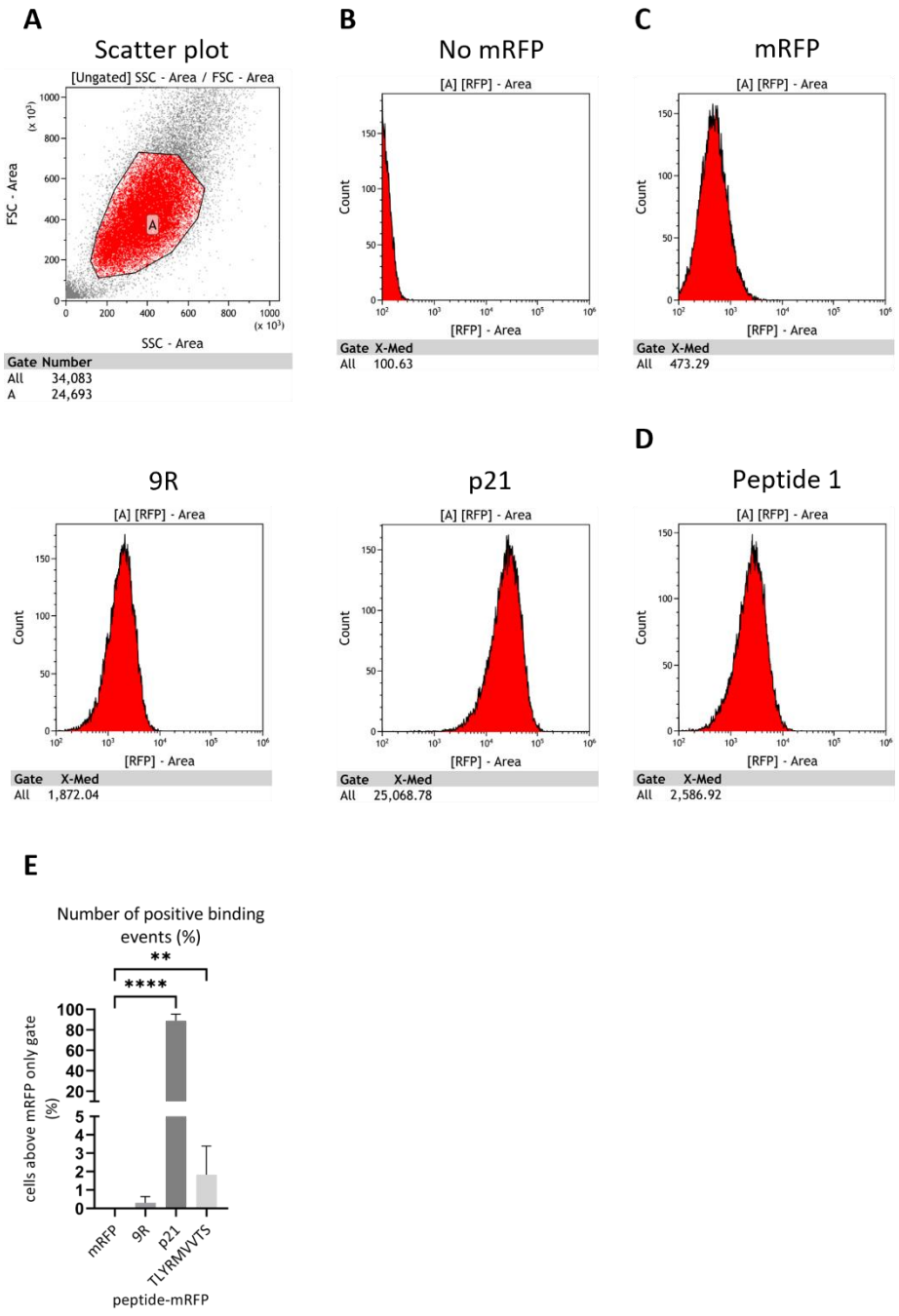


Figure 4.7 Cellular uptake of control and TNBC targeting peptides by BT-549 cells

(A) More than 20,000 single cell events were analysed (gate A) and the level of mRFP fluorescence within the single cell population was quantified by the mean fluorescence value obtained when cells were incubated with (B) no peptide, (C) the control peptides (mRFP, 9R and p21) and (D) the individual TNBC targeting peptides. Mean fluorescence values for each sample are shown (X-Med). Flow cytometric data is presented for one TNBC targeting peptide (TLYRMVVTS) and from one experimental repeat for reference. (E) Number of positive binding events found above the mRFP only gate shown as the mean ± SD (n=6–9). Statistical significance was calculated by Kruskall-Wallis test. ** p = 0.0038, **** p = < 0.0001.

incubated with the same cell line (Figure 4.7C). The endocytic uptake of mRFP only (negative control protein) and p21-9R-mRFP (positive control protein) by each control and TNBC cell line was also measured and the raw fluorescence values were normalised to 9R-mRFP (Figure 4.8, Figure 4.9 and Figure 4.10).

The uptake of p21-9R-mRFP led to high intracellular fluorescence in both the non-tumourigenic cell lines HMEC-TERT (4.79 ± 0.321) and MCF10A (32.3 ± 0.738) and in the TNBC cell lines BT-549 (10.7 ± 0.991), CAL-51 (57.7 ± 12.3) and MDA-MB-231 (40.6 ± 7.47). The mean fluorescence values obtained for p21-9R-mRFP varied the most across all the cell lines, however the values were much higher than the fluorescence values obtained from the TNBC targeting peptides. In contrast, the negative control recombinant protein, mRFP, resulted in little fluorescence in each cell line, with mRFP leading to normalised fluorescence values of 0.032 ± 0.002, 0.091 ± 0.007, 0.236 ± 0.020, 0.354± 0.026 and 0.273 ± 0.030 in HMEC-TERT, MCF10A, BT-549, CAL-51 and MDA-MB-231 cells, respectively.

The ability of each fluorescently labelled TNBC peptide to target TNBC was firstly assessed with BT-549 cells (Figure 4.8). Four peptides, TLYRMVVTS (0.472 ± 0.070 vs 1.14 ± 0.078; P = 0.0004), LSKKHMPVN (0.507 ± 0.104 vs 1.46 ± 0.231; p = 0.0008), CSTTYYTVK (1.12 ± 0.058 vs 1.36 ± 0.068; p = 0.0360) and MGSVRDTKF (1.91 ± 0.193 vs 2.48 ± 0.165; p = 0.0256), resulted in a significant increase in mRFP fluorescence observed in BT-549 cells compared to HMEC-TERT cells. There were also four peptides, SPVSSPITT (1.15 ± 0.146 vs 0.474 ± 0.025; p = 0.0004), LASMSYALL (0.767 ± 0.028 vs 0.498 ± 0.032; p = 0.0004), LQVSLNSVM (2.79 ± 0.257 vs 1.26 ± 0.125; p = 0.0004) and HNPNQITMY (0.94 ± 0.166 vs 0.422 ± 0.044; p = 0.0076), which resulted in significantly less mRFP fluorescence in BT-549 cells than HMEC-TERT cells (Figure 4.8A). In contrast, no peptides were found to target BT-549 cells more than MCF10A cells, whereas CSTTYYTVK (2.38 ± 0.172 vs 1.36 ± 0.068; p = 0.0002), LQVSLNSVM (2.88 ± 0.300 vs 1.26 ± 0.125; p = 0.0004) and HNPNQITMY (0.77 ± 0.043 vs 0.42 ± 0.044; p = 0.0007) led to a higher mean mRFP fluorescence value measured from MCF10A cells compared to BT-549 cells (Figure 4.8B).

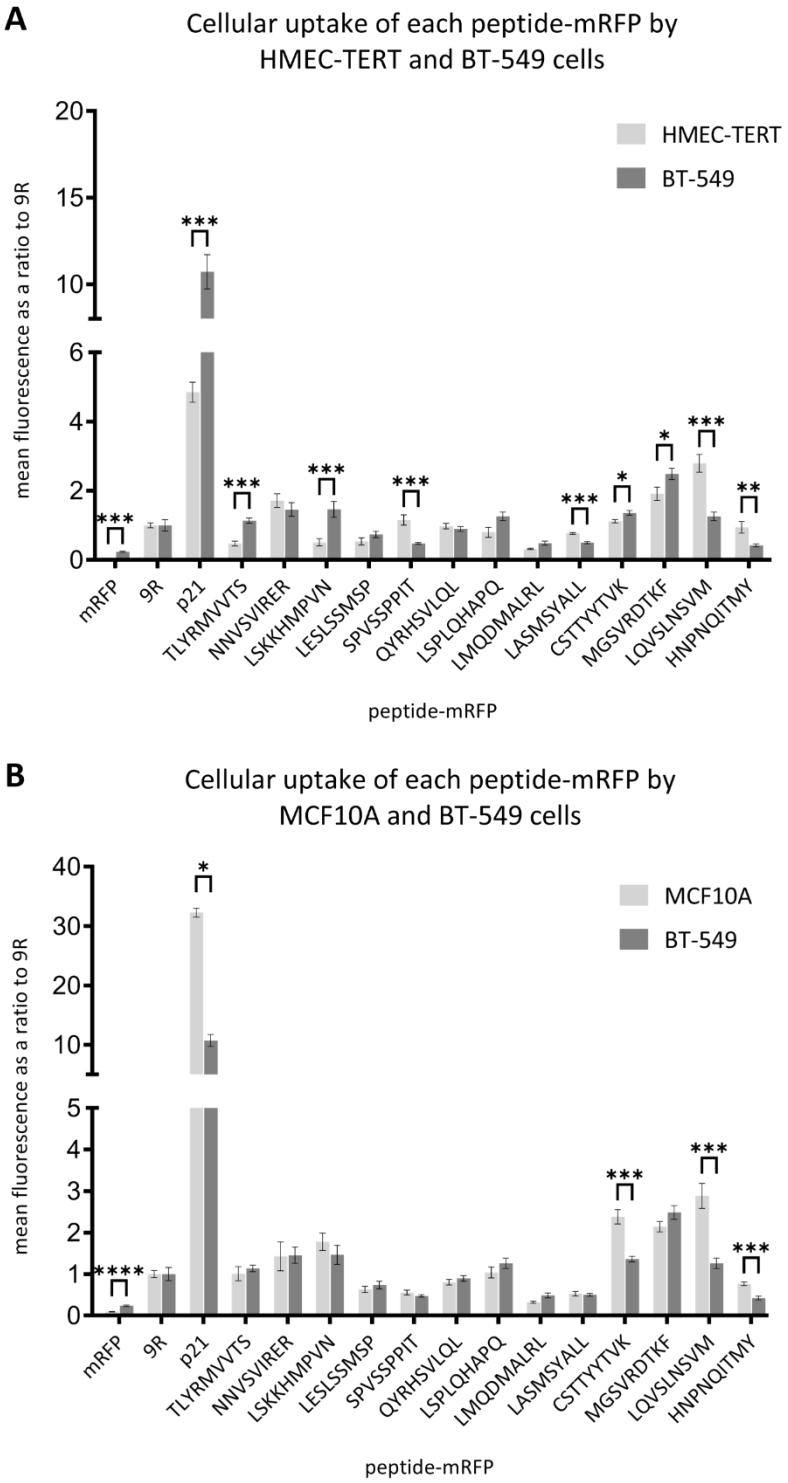


Figure 4.8 Cellular uptake of control and TNBC peptides in BT-549 cells compared to HMEC-TERT and MCF10A cells

mRFP fluorescence was measured by flow cytometry with 20,000 single cell events recorded per sample. Mean fluorescence data is shown as the mean ± SEM (n=6-9). Statistical significance between (A) HMEC-TERT and (B) MCF10A compared to BT-549 cell samples was calculated by Mann-Whitney test. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ and **** $p \leq 0.0001$.

With regard to CAL-51 targeting peptides, TLYRMVVTS (0.472 ± 0.070 vs 2.56 ± 0.426; *p* = 0.0003), LSKKHMPVN (0.507 ± 0.104 vs 1.93 ± 0.272; *p* = 0.0007), LESLSSMSP (0.538 ± 0.102 vs 0.985 ± 0.091; *p* = 0.0160), LMQDMALRL (0.316 ± 0.020 vs 1.14 ± 0.315; *p* = 0.0003) and CSTTYYTVK (1.12 ± 0.047 vs 4.67 ± 0.733; *p* = 0.0003) led to a significantly higher mean mRFP fluorescence value being observed in CAL-51 cells than in HMEC-TERT cells (Figure 4.9A). In addition, TLYRMVVTS (1.01 ± 0.171 vs 2.56 ± 0.426; *p* = 0.0004), LESLSSMSP (0.632 ± 0.072 vs 0.985 ± 0.091; *p* = 0.0220), LMQDMALRL (0.321 ± 0.026 vs 1.14 ± 0.315; *p* = 0.0003) and CSTTYYTVK (2.38 ± 0.172 vs 4.67 ± 0.733; *p* = 0.0232) also led to more mRFP fluorescence being observed in CAL-51 cells compared to MCF10A cells (Figure 4.9B). The 9-mer peptides NNVSVIRER, LSPLQHAPQ, LASMSYALL and MGSVRDTKF resulted in a higher mean fluorescence value in CAL-51 samples than in both MCF10A and HMEC-TERT samples, however the increase was not significant. SPVSSPPIT was the only peptide in this analysis which resulted in significantly less mRFP being observed in CAL-51 cells than a control cell line, with a mean fluorescence value of 1.15 ± 0.146 and 0.65 ± 0.078 (*p* = 0.0176) being observed in CAL-51 and HMEC-TERT cells, respectively.

All TNBC targeting peptides, excluding SPVSSPITT and HNPNQITMY, resulted in a higher mean fluorescence value in MDA- MB-231 cells being observed when compared to both HMEC-TERT and MCF10A cells (Figure 4.10). Statistical analysis of the flow cytometric data revealed that peptides TLYRMVVTS (0.51 ± 0.075 vs 1.72 ± 0.150; *p* = 0.001), LSKKHMPVN (0.51 ± 0.104 vs 1.98 ± 0.323 *p* = 0.001), LESLSSMSP (0.54 ± 0.102 vs 0.97 ± 0.111; *p* = 0.0200), QYRHSVLQL (0.97 ± 0.080 vs 1.35 ± 0.163; *p* = 0.0312), LSPLQHAPQ (0.80 ± 0.143 vs 1.53 ± 0.152; *p* = 0.0120) and LMQDMALRL (0.32 ± 0.024 vs 0.88 ± 0.141; *p* = 0.001) delivered more mRFP intracellularly into MDA-MB-231 cells than into HMEC-TERT cells (Figure 4.10B). NNVSVIRER, LASMSYALL, CSTTYYTVK, MGSVRDTKF and LQVSLNSVM also led to more fluorescence in MDA-MB-231 cells when compared to HMEC-TERT cells, although the increases were not statistically significant.

In addition, NNVSVIRER, LSKKHMPVN, SPVSSPITT, QYRHSVLQL, LSPLQHAPQ, LASMSYALL, CSTTYYTVK, MGSVRDTKF, LQVSLNSVM and HNPNQITMY resulted in non-significant increases in fluorescence in MDA-MB-231 cells compared to MCF10A cells. However, significant increases were observed with TLYRMVVTS (1.01 ± 0.171 vs 1.72

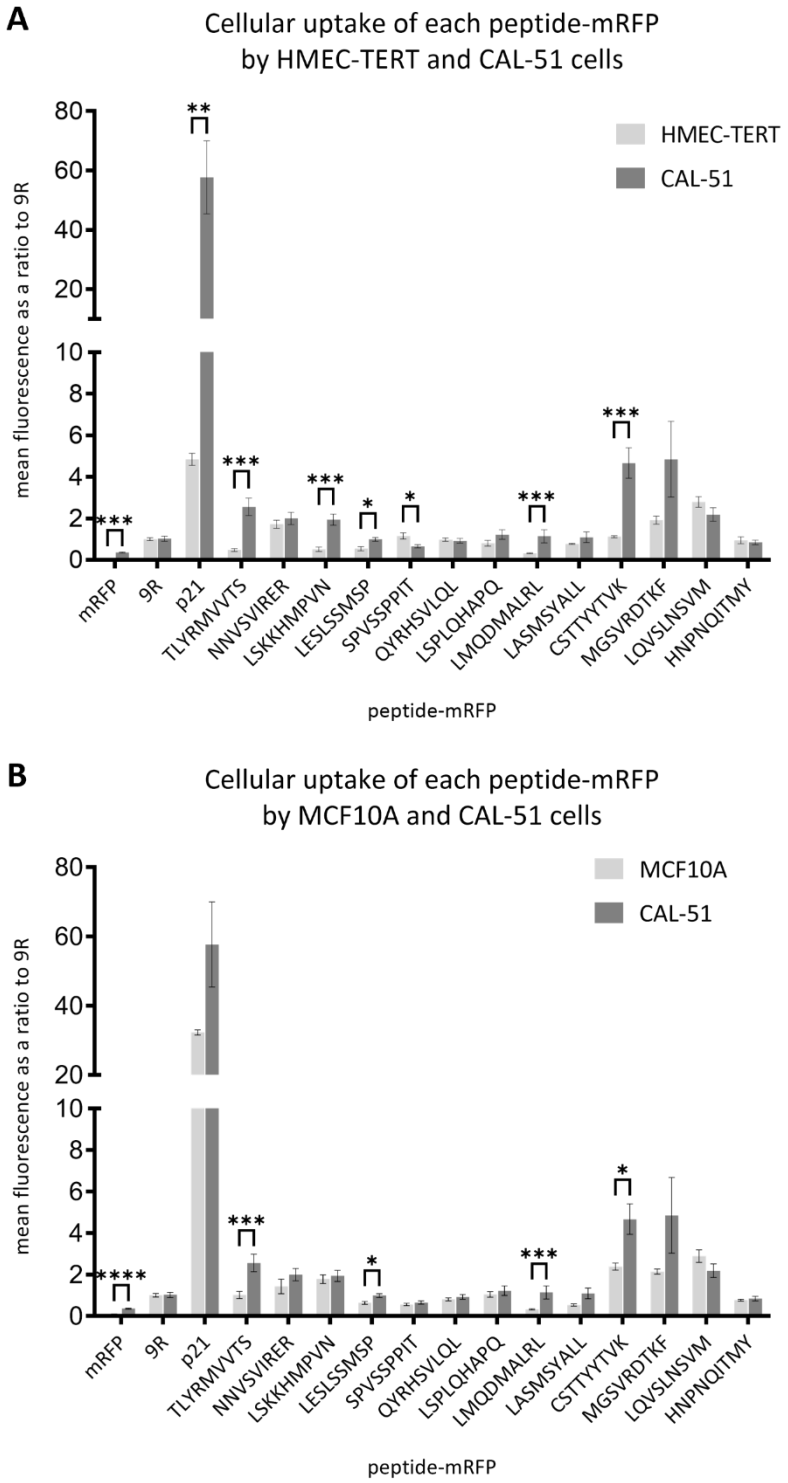


Figure 4.9 Cellular uptake of control and TNBC peptides in CAL-51 cells compared to HMEC-TERT and MCF10A cells

mRFP fluorescence was measured by flow cytometry with 20,000 single cell events recorded per sample. Mean fluorescence data is shown as the mean ± SEM (n=6-9). Statistical significance between (A) HMEC-TERT and (B) MCF10A compared to CAL-51 cell samples was calculated by Mann-Whitney test. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ and **** $p \leq 0.0001$.

± 0.150; p = 0.0315), LESLSSMSP (0.632 ± 0.072 vs 0.967 ± 0.111; p = 0.0220), QYRHSVLQL (0.804 ± 0.067 vs 1.35 ± 0.163; p = 0.0021) and LMQDMALRL (0.321 ± 0.026 vs 0.881 ± 0.141; p = 0.00004), with LMQDMALRL leading to the most significant increase in the MDA-MB-231 versus MCF10A analysis (Figure 4.10B).

The 9-mer TNBC peptides which led to a significantly higher mean mRFP fluorescence in a TNBC cell line compared to one (MCF10A or HMEC-TERT) or both (MCF10A and HMEC-TERT) of the control cell lines are highlighted in Table 4.2. From this, TLYRMVVTS, LMQDMALRL and CSTTYYTVK were selected for further TNBC targeting analyses, as each of these peptides resulted in significantly higher mRFP uptake in two or more TNBC cell lines when compared to either MCF10A cells, HMEC-TERT cells or both non-tumourigenic cell lines. Peptides LSKKHMPVN and LESSLSSMSP were also considered for further validation, however, LSKKHMPVN did not lead to any significant fluorescent increases in the MCF10A vs TNBC cell line analyses and all LESLSSMSP comparisons led to less significant p-values ($p \leq 0.05$) than observed with LMQDMALRL and CSTTYYTVK.

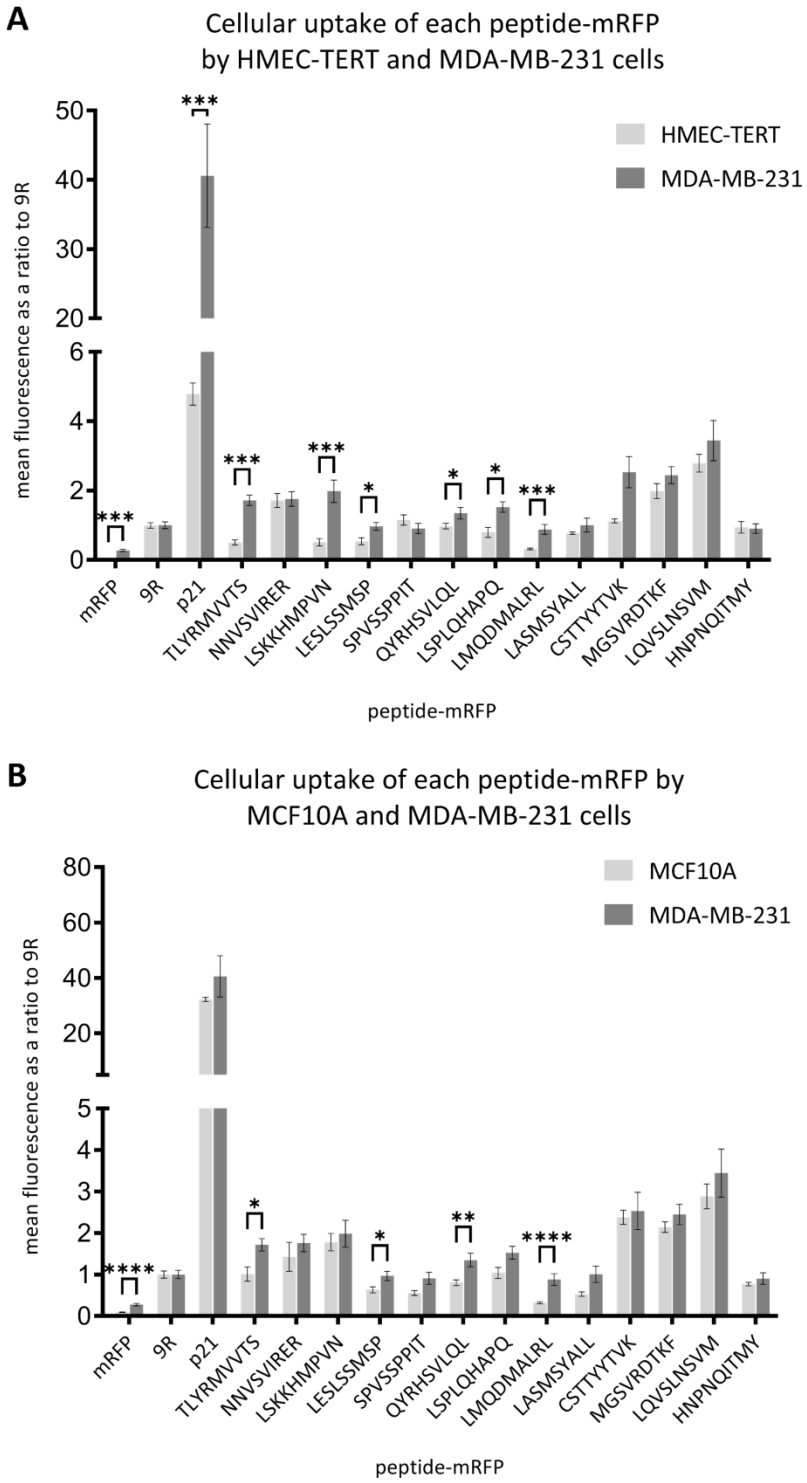


Figure 4.10 Cellular uptake of control and TNBC peptides in MDA-MB-231 cells compared to HMEC-TERT and MCF10A cells

mRFP fluorescence was measured by flow cytometry with 20,000 single cell events recorded per sample. Mean fluorescence data is shown as the mean ± SEM (n=6-9). Statistical significance between (A) HMEC-TERT and (B) MCF10A compared to MDA-MB-231 cell samples was calculated by Mann-Whitney test. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ and **** $p \leq 0.0001$.

Table 4.2 Cellular uptake of fluorescently labelled peptides in TNBC cell lines compared to control cell lines

Peptides	TNBC cell lines		
	BT-549	CAL-51	MDA-MB-231
TLYRMVVTS	✓	✓✓	✓✓
NNVSVIRER			
LSKKHMVN	✓	✓	✓
LESLSSMSP		✓✓	✓✓
SPVSSPPIT			
QYRHSVLQL			✓✓
LSPLQHAPQ			✓
LMQDMALRL		✓✓	✓✓
LASMSYALL			
CSTTYYTVK	✓	✓✓	
MGSVRDTKF	✓		
LQVSLNSVM			
HNPNQITMY			

✓ significantly increased uptake in TNBC cell line when compared to one control cell line.

✓✓ significantly increased uptake in TNBC cell line when compared to both control cell lines.

4.2.4 Qualitative analysis of TNBC peptide mediated delivery of MNPs

Following flow cytometric analysis, the specificity of the targeting peptides TLYRMVVTS, LMQDMALRL and CSTTYYTVK for TNBC was further investigated by assessing their ability to enhance cellular uptake of magnetic nanoparticles (MNPs). MNPs are superparamagnetic iron oxide nanoparticles (SPIONs) with high *in vivo* biocompatibility and biodegradability (Reddy et al., 2012). The nanoparticles have been translated to numerous biomedicine applications including cancer diagnostics and therapies such as MRI contrast enhancement, controlled and targeted drug release and magnetic fluid induced hyperthermia cancer treatment (Thiesen and Jordan, 2008, Sun et al., 2008, Estelrich et al., 2015, Avasthi et al., 2020).

Blokpoel Ferreras *et al.,* have previously demonstrated that positively charged homing peptides or CPPs can be complexed to negatively charged MNPs via electrostatic interactions. The formation of peptide-MNP nanocomplexes can then enhance the

intracellular delivery of MNPs (Blokpoel Ferreras et al., 2020). Therefore, to investigate whether TLYRMVVTS, LMQDMARL and CSTTYYTVK can lead to increases in cellular MNP uptake, the three potential TNBC targeting peptides were complexed to MNPs and delivered into MCF10A, HMEC-TERT, BT-549, CAL-51 and MDA-MB-231 cells.

As the recombinantly expressed proteins used in the initial cellular uptake analyses (see 4.2.2 and 4.2.3) were fluorescently labelled with mRFP, the TNBC peptides used for MNP complexation were purchased as synthetic peptides. In a similar manner to the recombinant proteins used for the fluorescent microscopy and flow cytometric analyses, the synthetic peptides included the MRSP and PRRRRRRRRR flanking sequences (Table 4.3). A 9R control peptide was also synthetically produced as a negative control protein and the p21-8R peptide investigated by Blokpoel Ferraras *et al.,* in previous MNP delivery studies was used as the positive control, referred to here as p21 (Blokpoel Ferreras et al., 2020).

Table 4.3 Amino acid sequences of synthetic peptides used in MNP complexation

Peptides	Amino acid sequence
9R	RRRRRRRRR
p21	KRKKKGKGLGKKRDPCLRKYKRRRRRRRR
TLYRMVVTS	MSRPTLYRMVVTSPRRRRRRRRR
LMQDMALRL	MSRPLMQDMALRLPRRRRRRRRR
CSTTYYTVK	MSRPCSTTYYTVKPRRRRRRRRR

To ensure that the MNPs and synthetically produced peptides did not cause cytotoxicity, cell viability assays were firstly performed (Figure 4.11). Blokpoel Ferraras *et al.,* delivered 50 ug ml^{-1} MNPs and p21-8R:MNPs (4 nmol mg^{-1} MNP) into NIH3T3 cells with no effect on cell viability or proliferation observed and *Markides et al.,* demonstrated that 20 nmol mg^{-1} of p21-8R:MNPs (1nM of peptide per 50 ug of MNP) could be delivered to mesenchymal stromal cells without impacting cell viability or proliferation (Markides et al., 2019). Therefore, in this study, 50 ug ml^{-1} of MNPs and 1nM of each peptide were delivered into the non-tumourigenic breast cell lines (Figure 4.11A) and the TNBC cell lines (Figure 4.11B).

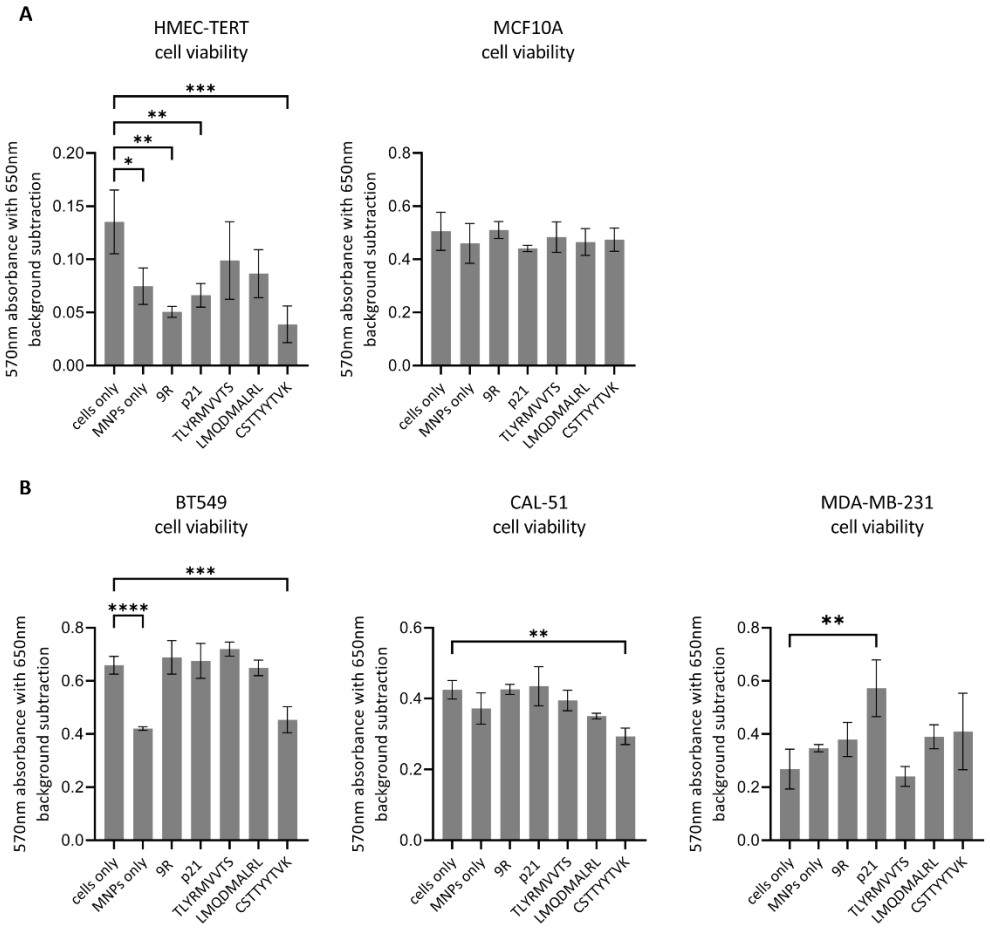


Figure 4.11 Control and TNBC cell viability with MNPs and targeting peptides

Cell viability was assessed by the conversion of MTT to formazan crystals in (A) control and (B) TNBC cells incubated with 50 µg ml^{-1} MNPs or 1nM 9R, p21, TLYRMVVTS, LMQDMALRL or CSTTYYTVK for 24 hours (mean ± SD). Statistical significance was calculated by ordinary one-way ANOVA (n=3). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ and **** $p \leq 0.0001$.

The addition of MNPs did not appear to impact cell viability in MCF10A, CAL-51 and MDA-MB-231 cells, however some reduction in cell viability was observed in HMEC-TERT and BT-549 cells. The control peptides 9R and p21 also appeared to result in reduced cell viability in HMEC-TERT cells. In contrast, peptides TLYRMVVTS and LMQDMALRL did not lead to decreases in cell viability being observed in any of the five cell lines. CSTTYYTVK did, however, lead to reduced cell viability in HMEC-TERT, BT-549 and CAL-51 cells.

MNPs were then complexed to each control and TNBC targeting peptide at 20nmol mg^{-1} (1nM peptide : 50 µg MNPs) and delivered into the non-tumourigenic (Figure 4.12

and Figure 4.13) and TNBC cells lines (Figure 4.14, Figure 4.15 and Figure 4.16). Each cell line was exposed to the peptide:MNP complexes for 1 hour and removed so that only highly specific binding events would lead to MNP endosomal uptake. Cells which were not exposed to MNPs or any peptide:MNP complex were also analysed as negative controls. Peptide:MNP delivery was analysed by Prussian blue staining, a chemical stain which can detect iron oxide magnetic nanoparticles in cells or tissue (Liu et al., 2010). As expected, no Prussian blue staining was observed when the cells were not exposed to MNPs in all cell lines (Figure 4.12 to Figure 4.16; no peptide or MNPs).

Prussian blue staining of the iron content in each non-tumourigenic cell line after peptide:MNP incubation revealed that MNP increases were observed when comparing MNPs only, 9R:MNPs and p21:MNPs, with p21:MNPs leading to the most MNP uptake in HMEC-TERT and MCF10A cells. With regard to HMEC-TERT cells, TLYRMVVTS and LMQDMALRL resulted in a similar amount of cellular MNP to cells incubated with 9R:MNPs. In contrast, CSTTYYTVK:MNPs resulted in an increase in Prussian blue staining observed when compared to the 9R:MNPs control (Figure 4.12). Similarly, complexing p21 and MNPs resulted in high iron content in MCF10A cells, with the majority of cells containing MNPs (Figure 4.13). CSTTYYTVK:MNPs also led to more Prussian blue staining compared to 9R:MNPs but the staining was substantially less than that observed with p21:MNPs. In contrast, TLYRMVVTS:MNPs and LMQDMALRL:MNPs complexes delivered less MNPs to MCF10A cells, however some intracellular iron from MNPs was seen.

Prussian blue staining also highlighted intracellular MNPs in BT-549 cells when subjected to the peptide:MNP complexes. As with the control cell lines, p21:MNPs appeared to lead to higher levels of intracellular iron than that observed with the other TNBC peptide:MNPs (Figure 4.14). When compared to the 9R:MNPs control, LMQDMALRL appeared to result in less Prussian blue staining, however the staining observed appeared to be cell specific. In contrast, TLYRMVVTS:MNPs appeared to lead to an increase in MNPs delivered into BT-549 cells than the CPP peptide only (9R:MNPs). CSTTYYTVK:MNPs resulted in the most Prussian blue staining observed out of all the TNBC targeting peptide complexes, inferring that the peptide may have led to more intracellular MNP delivery.

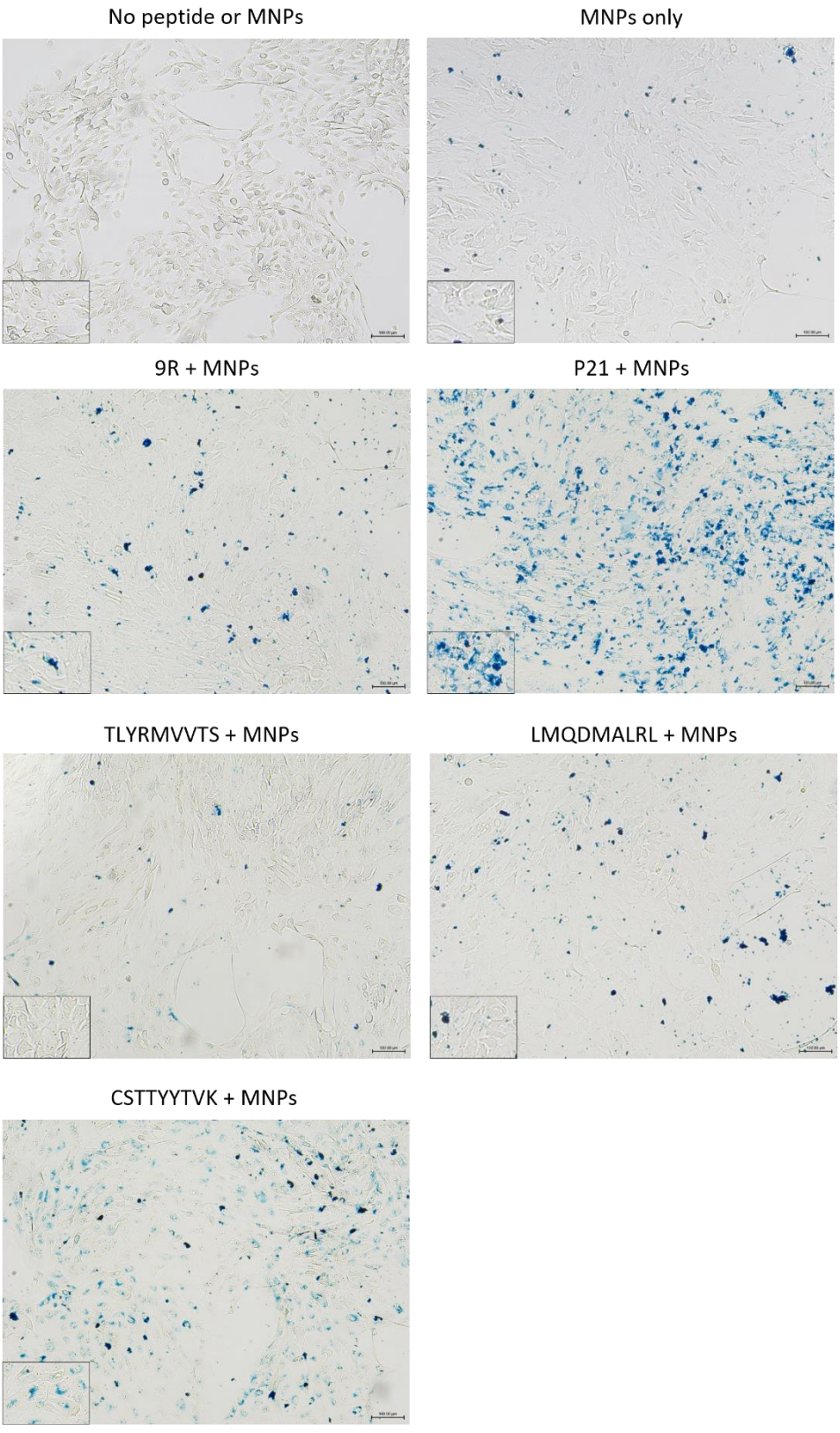


Figure 4.12 TNBC peptide-mediated delivery of MNPs into HMEC-TERT cells

Peptide:MNP complexes were delivered to cells for 1 hour to allow for initial cell peptide:MNP binding with bound MNPs internalised overnight. Intracellular MNPs were stained with Prussian blue (n=3). Images taken with 10x objective. Scale bar = 100 µM.

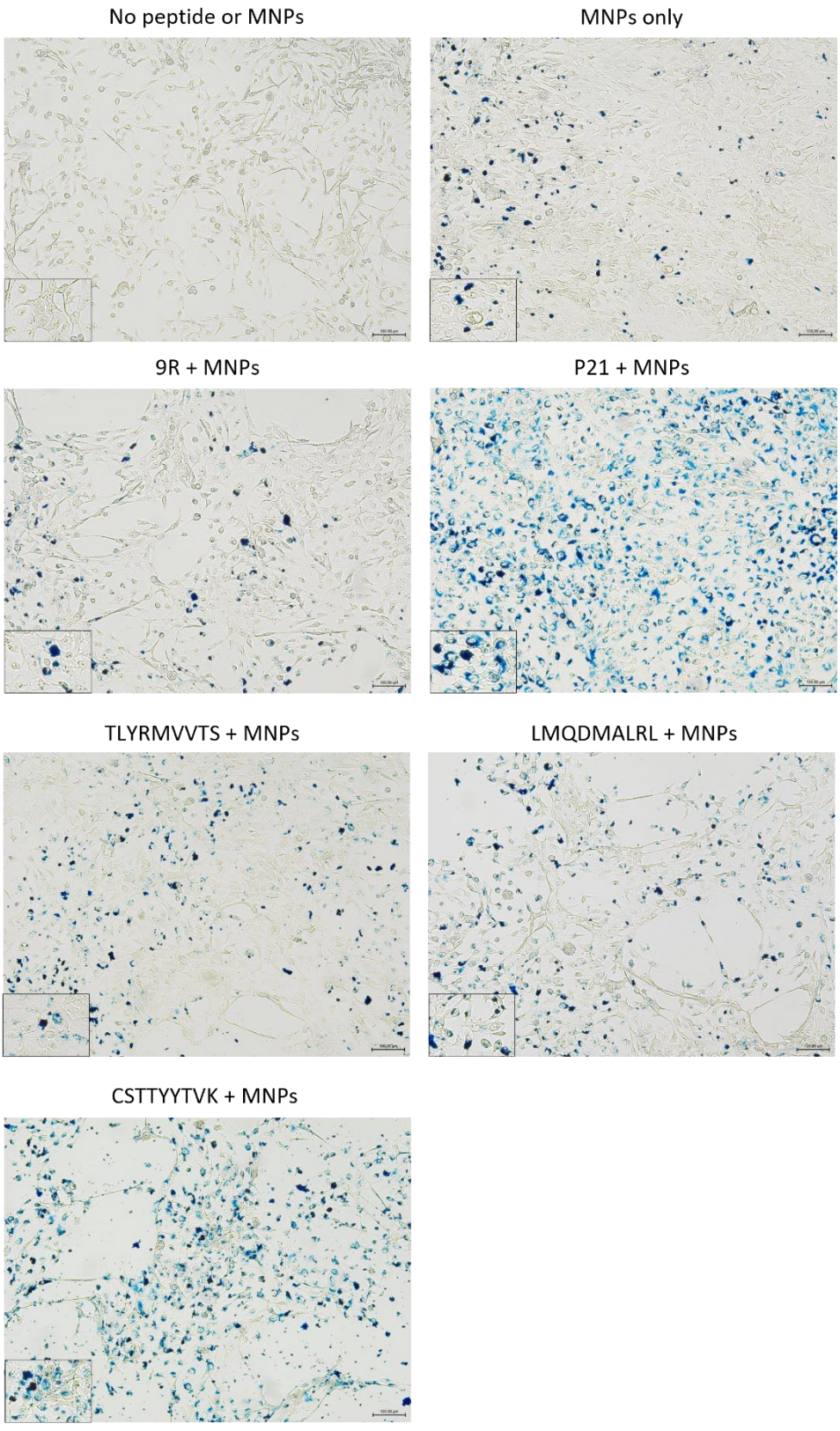


Figure 4.13 TNBC peptide-mediated delivery of MNPs into MCF10A cells

Peptide:MNP complexes were delivered to cells for 1 hour to allow for initial cell peptide:MNP binding with bound MNPs internalised overnight. Intracellular MNPs were stained with Prussian blue (n=3). Images taken with 10x objective. Scale bar = 100 µM.

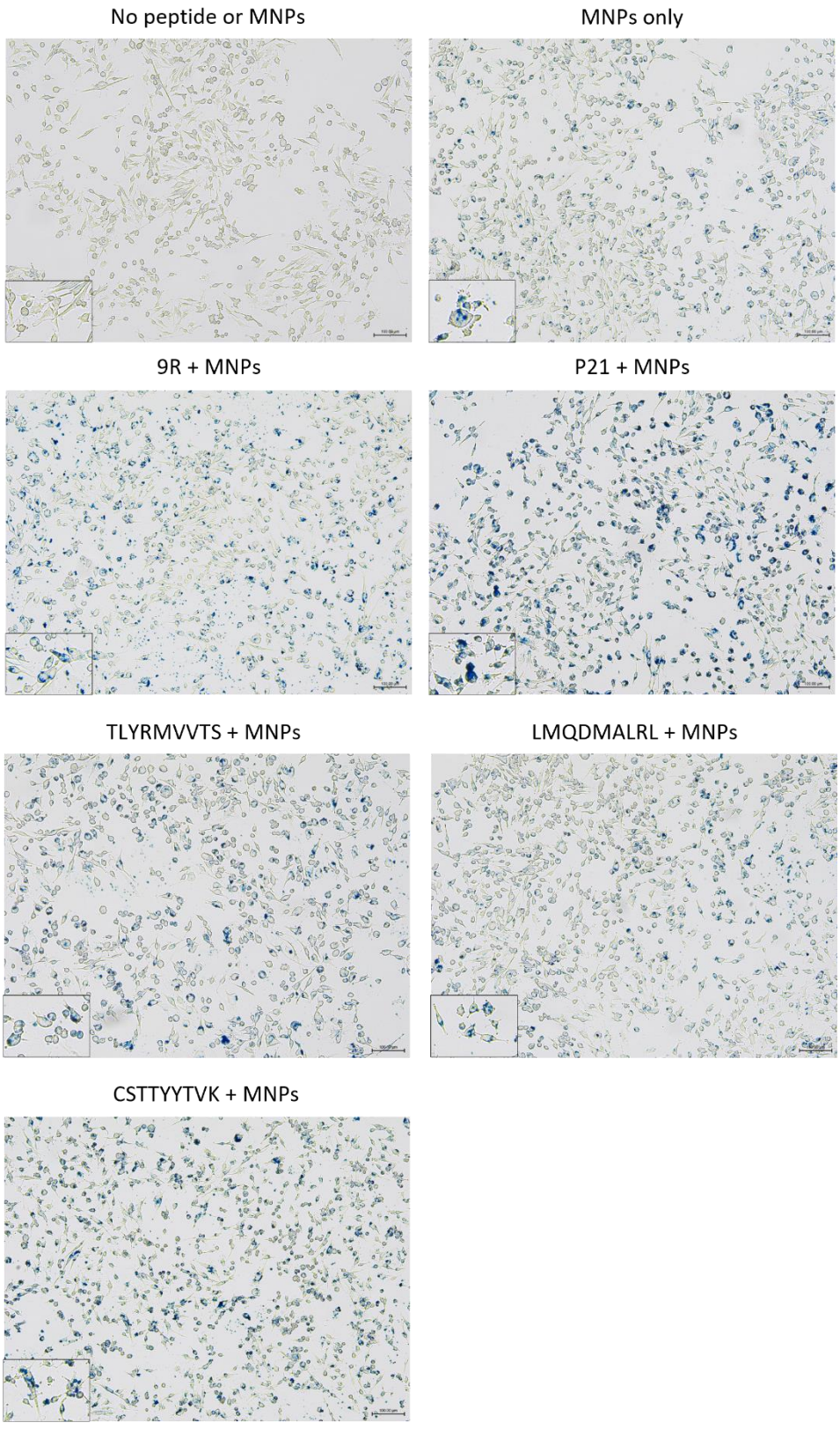


Figure 4.14 TNBC peptide-mediated delivery of MNPs into BT-549 cells

Peptide:MNP complexes were delivered to cells for 1 hour to allow for initial cell peptide:MNP binding and bound MNPs were internalised overnight. Intracellular MNPs were stained with Prussian blue (n=3). Images taken with 10x objective. Scale bar = 100 µM.

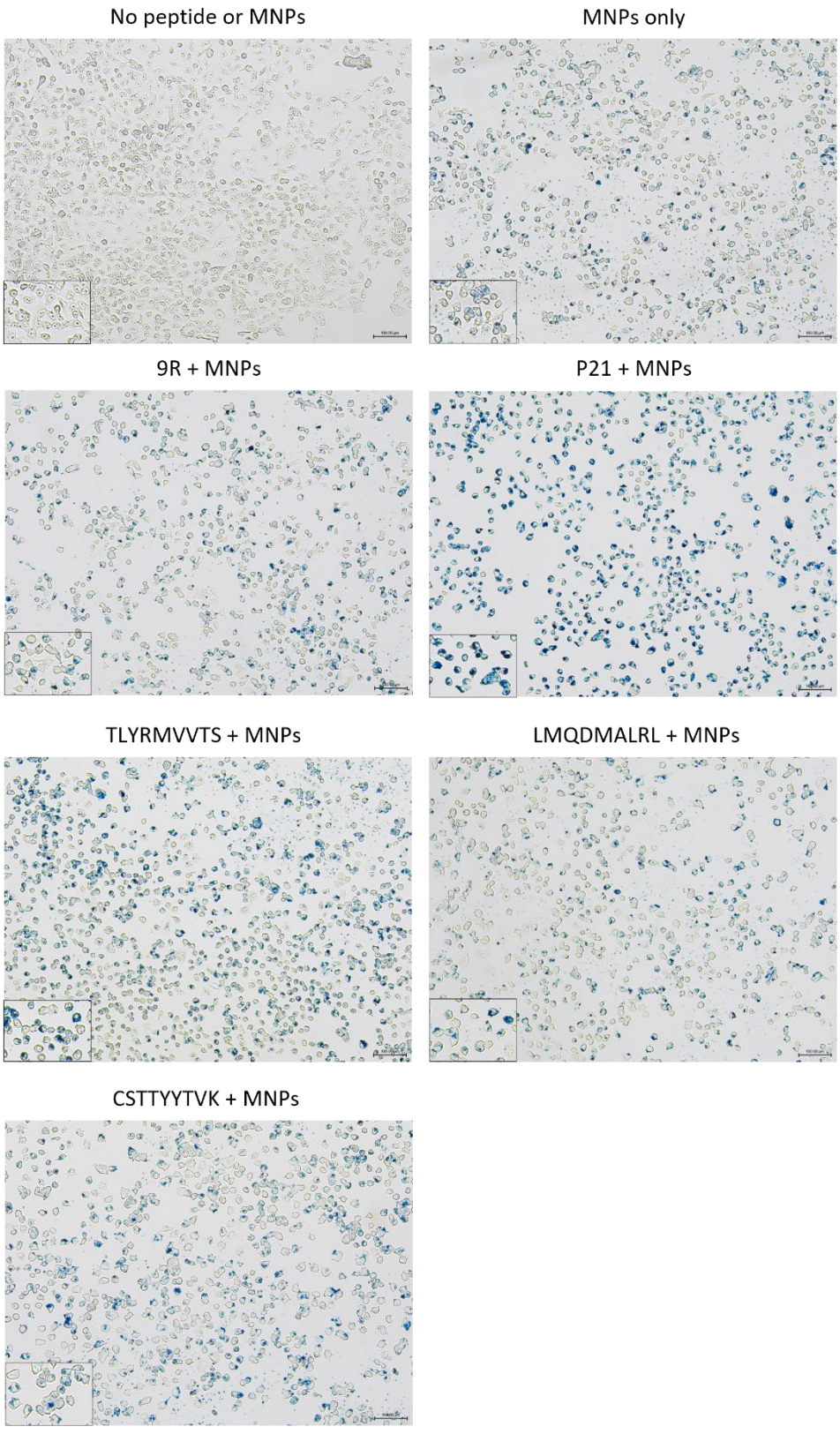


Figure 4.15 TNBC peptide-mediated delivery of MNPs into CAL-51 cells

Peptide:MNP complexes were delivered to cells for 1 hour to allow for initial cell peptide:MNP binding and bound MNPs were internalised overnight. Intracellular MNPs were stained with Prussian blue (n=3). Images taken with 10x objective. Scale bar = 100 µM.

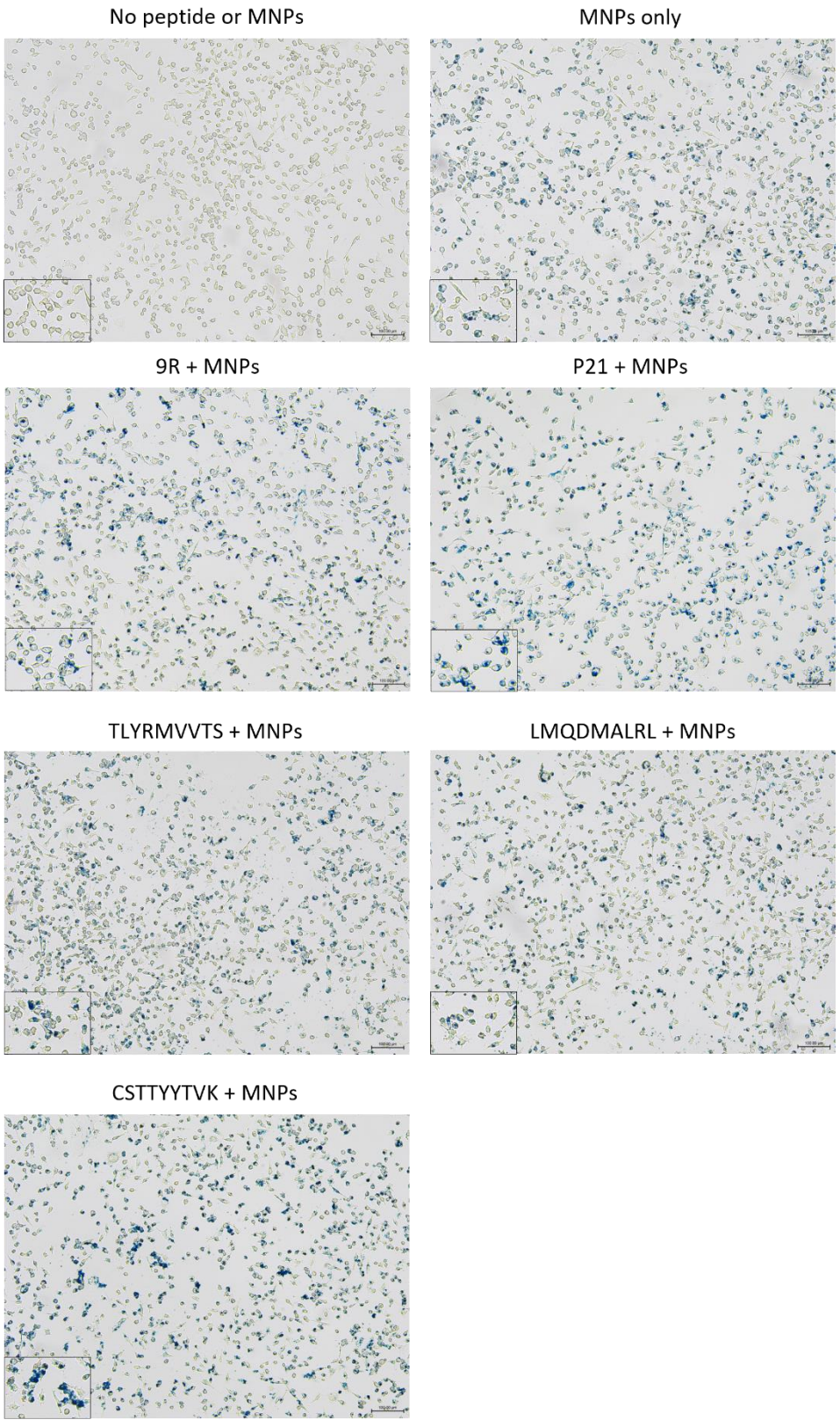


Figure 4.16 TNBC peptide-mediated delivery of MNPs into MDA-MB-231 cells

Peptide:MNP complexes were delivered to cells for 1 hour to allow for initial cell peptide:MNP binding and bound MNPs were internalised overnight. Intracellular MNPs were stained with Prussian blue (n=3). Images taken with 10x objective. Scale bar = 100 µM.

When considering CAL-51 MNP delivery, TLYRMVVTS led to higher levels of intracellular MNP derived iron than LMQDMALRL and CSTTYYTVK (Figure 4.15). Additionally, both TLYRMVVTS and CSTTYYTVK resulted in more Prussian blue staining being seen than in cells subjected to 9R:MNPs. In a similar manner to HMEC-TERT, MCF10A and BT-549 cells, the complexation of p21 to MNPs resulted in the most Prussian blue staining, with intracellular iron being observed in almost all CAL-51 cells.

Unlike the other cell lines in this study, p21:MNPs did not appear to lead to the most Prussian blue staining in MDA-MB-231 cells compared to other peptide:MNP complexes (Figure 4.16). Instead, CSTTYYTVK complexed MNPs resulted in the most cell staining observed. p21 does, however, appear to result in higher levels of cellular iron content through MNP delivery than 9R only. The qualitative analysis did not highlight differences in TLYRMVVTS or LMQDMALRL mediated MNP uptake, with both peptides appearing to lead to similar levels of cell associated iron in MDA-MB-231 cells.

4.2.5 Quantitative analysis of TNBC peptide mediated delivery of MNPs by ICP-MS

To quantify differences in the uptake of TLYRMVVTS:MNPs, LMQDMALRL:MNPs and CSTTYYTVK:MNPs seen by the TNBC cell lines with Prussian blue staining, the iron content of each cell after exposure to the peptide:MNPs was then investigated by ICP-MS . Cells were exposed to the same peptide and MNPs conditions as in section 4.2.4. After incubation with no MNPs, MNPs only or peptide:MNPs, the cells were harvested and the iron content of each sample was measured. The iron content for each cell was calculated and normalised to the Fe values measured in the 9R:MNPs samples for each cell line (Figure 4.17).

There were no significant differences observed between cells subjected to MNPs only or 9R:MNPs for any of the TNBC cell lines (Figure 4.17). As expected, the positive control peptide:MNP complex, p21:MNPs, led to significantly higher Fe values in BT-549 (Figure 4.17A, P=0.0068), CAL-51 (Figure 4.17B, P = 0.0006) and MDA-MB-231 (Figure 4.17C, P = 0.0215) cells. There was also a significant increase in Fe content in BT-549 cells subjected to CSTTYYTVK:MNPs compared to 9R:MNPs (Figure 4.17A, P = 0.003). However, no other increases in Fe content were observed with TLYRMVVTS:MNPs, LMQDMALRL:MNPs or CSTTYYTVK:MNPs.

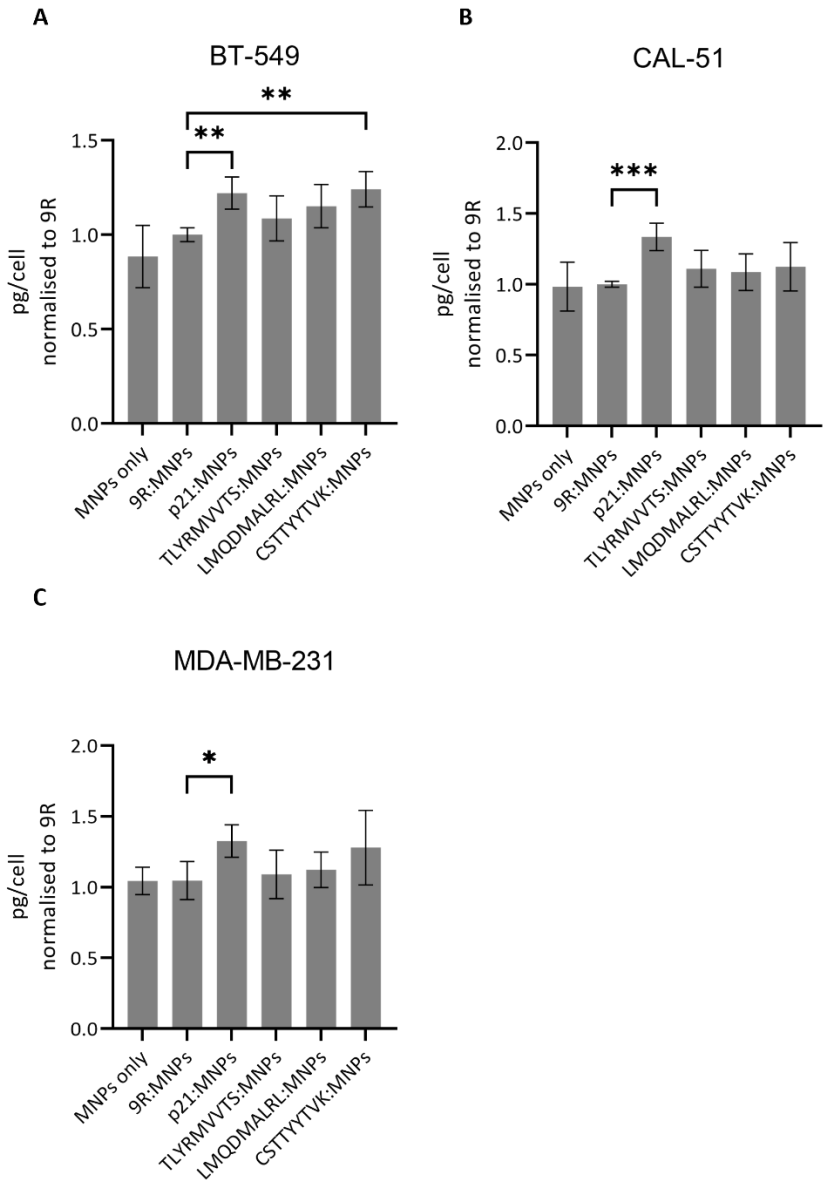


Figure 4.17 ICP-MS results for TNBC peptide-mediated delivery of MNPs into TNBC cell lines

Peptide:MNP complexes were delivered to (A) BT-549, (B) CAL-51 or (C) MDA-MB-231 cells for 1 hour to allow for initial cell peptide:MNP binding and bound MNPs were internalised overnight. Intracellular MNPs were quantified by ICP-MS and each experimental run was normalised to the uptake of 9R:MNPs for each cell line (mean ± SD). Statistical significance was calculated by ordinary one-way ANOVA (n=6). * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$.

4.3 Discussion

In this chapter, the 44 9-mer TNBC targeting peptides identified from NGPD analysis in chapter 4 were successfully sub-cloned into a mRFP expressing vector for cell delivery studies (Figure 4.2). Additionally, the CPP sequence was also introduced into the vector downstream of the 9-mer sequence, as removing the 9R CPP may have resulted in the decreased intracellular delivery of mRFP and poor TNBC cell penetration. The CPP ability of each 9-mer peptide was assessed using BChemRF-CPPred, but the online predictor tool revealed that only one of the peptides, SKKQRRALL, was predicted to be a CPP when all groups of structure-based and sequence-based descriptors were considered (see Appendix 4.5, Supplementary Table 6 (de Oliveira et al., 2021). However, removal of the 9R sequence may have still led to endocytic uptake of the TNBC targeting peptides and their corresponding cargos via TNBC-specific receptor mediated endocytosis, therefore TNBC-specific delivery without the 9R sequence warrants further investigation. Furthermore, the removal of the CPP sequence may increase TNBC cell selectivity by reducing the uptake of therapeutic cargos into non-TNBC cells.

Following the insertion of each 9-mer peptide into the mRFP expressing vector, the 13 peptides from the Z>2 analysis (Figure 3.7) were expressed and purified. The Z>2 TNBC targeting peptides were selected for *in vitro* validation studies as this Z score criteria was the most stringent threshold used, with each peptide identified in the Z>2 analysis being enriched in three target cell replicates compared to the control pool. Therefore, the Z>2 peptides were prioritised as the most promising TNBC targeting candidates for cell delivery studies. Fluorescence microscopy confirmed mRFP expression and purification and demonstrated that each peptide-mRFP led to more cell specific mRFP being observed in the non-tumourigenic (BJ, MCF10A and HMEC-TERT) and TNBC cell lines (BT-549, CAL-51 and MDA-MB-231) than with mRFP only, suggesting that the peptide-9R sequences were facilitating mRFP delivery into the cells (Figure 4.6). However, as previously mentioned, peptide-mediated differences in intracellular mRFP could not be clearly determined from the qualitative analysis.

To quantify differences in mRFP uptake and to validate the specificity of the 9-mer

peptides for TNBC cells, the ability of each TNBC targeting peptide to deliver mRFP intracellularly in the non-tumourigenic (MCF10A and HMEC-TERT) and tumourigenic (BT-549, CAL-51 and MDA-MB-231) cell lines was then assessed by flow cytometry (Figure 4.8, Figure 4.9 and Figure 4.10). Other studies have successfully utilised this technique to assess the binding of tumour targeting peptides to different cell lines (Xiao et al., 2016, Mansour et al., 2022, Allred et al., 2023). For example, Xiao *et al.,* investigated the binding of five phage displayed peptides to EGRVIII overexpressing and EGRVIII wild-type cell lines. This led to the prioritisation of a targeting peptide, VLGREEWSTSYW, which was found to be selective for the cancer-specific EGR variant (Xiao et al., 2016).

In this project, the recombinantly expressed p21-8R sequence was utilised as a positive control peptide and 9R-mRFP was used as a negative control peptide. The high mean fluorescence values observed with p21-8R are consistent with findings by Dixon *et al.,* who confirmed that fusing p21 to the CPP 8R greatly increased mRFP intracellular delivery, resulting in a high level of mRFP fluorescence being measured in a range of cell lines (NIH3T3, HUES7, IPS2, CGR-8 and HL1) by flow cytometry (Dixon et al., 2016). The high mean fluorescence values observed with p21-8R positive control and the low fluorescence values observed with the 9R-mRFP negative control therefore confirmed that the flow cytometry experiments were set up correctly (Figure 4.7).

When compared to the 9R-mRFP, eight of the phage display identified peptides led to an increase in mRFP fluorescence being observed, indicating that the addition of the TNBC targeting peptides to the 9R sequence may increase the cellular uptake of a therapeutic agent (Table 4.2). However, some TNBC targeting peptides resulted in less fluorescence being observed than with the 9R-mRFP control peptide, such as LMQDMALRL. Although less than 9R-mRFP, this peptide did lead to significantly more mRFP fluorescence in the TNBC cell line samples compared to the controls, suggesting that the TNBC peptide may have increased TNBC cell specificity.

The flow cytometry experiments led to three peptides being prioritised for further study: TLYRMVVTS, LMQDMALRL and CSTTYYTVK (Table 4.2). TLYRMVVTS, which was identified from biopanning with the 9-mer peptide library against BT-549 cells, resulted in significant increases ($p \leq 0.05$) in mRFP fluorescence being observed in

TNBC cell lines when compared to MCF10A and HMEC-TERT cells. TLYRMVVTS targeting of BT-549 correlates with the NGPD analysis for this cell line, with TLYRMVVTS resulting in the least Z score variation across the five replicate samples (Figure 3.7). However, this peptide was not identified to target CAL-51 and MDA-MB-231 from biopanning experiments (Figure 3.8), which may have been due to the loss of the peptide from the library in earlier biopanning rounds. LMQDMALRL and CSTTYYTVK were identified to be MDA-MB-231 targeting from the NGPD analysis, with both peptides displaying high normalised frequencies in MDA-MB-231 cell lysate compared to the control pool (Figure 3.8). The flow cytometric analysis confirmed that LMQDMALRL targets MDA-MB-231 cells, as the peptide led to significantly higher mRFP fluorescence being observed in MDA-MB-231 samples compared to HMEC-TERT and MCF10A cells (Figure 4.10). This peptide was also found to target CAL-51 cells more than the control cell lines by flow cytometry, which was not observed from the CAL-51 phage biopanning experiment. Again, this may have been due to the peptide being removed in early rounds of biopanning and therefore the peptide would not have been enriched in subsequent rounds of selection. In contrast, CSTTYYTVK was found to target BT-549 and CAL-51 cells. There was an increase in the delivery of mRFP into MDA-MB-231 cells compared to the control cell lines, however this difference was not significant. Additionally, CSTTYYTVK led to high levels of fluorescence in MCF10A and HMEC-TERT cells, indicating that this peptide would deliver high levels of a therapeutic cargo into healthy tissue as well as TNBC cells. Each of the peptides were found to target one or more TNBC cell line, suggesting that the peptides may be targeting a cell surface protein which may be overexpressed or aberrantly expressed in different TNBC tumours.

Even though a phagemid derived pIII 9-mer peptide library was used in this study, each phage virion may have displayed multiple copies of a single peptide. Therefore, the affinity of each peptide for the TNBC specific target may have been decreased when each peptide was expressed monomerically with mRFP (Brown, 2010). To investigate whether increasing the presentation of each peptide enhanced TNBC cell delivery, the valency of the peptides was increased by complexing each peptide with MNPs. Increasing the valency of the peptide via a nanoparticle system may also increase the avidity of the peptide for the TNBC specific target, enhancing differences observed in the flow cytometry experiments (Xiao et al., 2016). However, Allfred *at al.,* found that

the affinity of a phage display selected non-small cell lung cancer (NSCLC) targeting peptide for H1299 cells was not increased when the peptide was tetramerised (McGuire et al., 2014, Allred et al., 2023).

As previously mentioned, MNPs have biomedical applications in cancer diagnosis and therapy, such as MRI enhancement and controlled drug release (Farzin et al., 2020). Therefore, increasing the selectivity of MNPs for cancer cells by specifically targeting tumour antigens is highly relevant in the development of targeted therapeutics. Studies have demonstrated that functionalising MNPs with targeting ligands, such as tumour targeting peptides or antibodies, can result in cancer-specific delivery of MNPs (Hathaway et al., 2011, Sader et al., 2015, Zhu et al., 2019). For instance, Sader *et al.,* found that functionalising MNPs through the covalent attachment of the tumour targeting, multivalent pseudopeptide N6L resulted in N6L-MNP accumulation in MDA-MB-231 cells both *in vitro* and in tumour bearing mice, thus highlighting the potential of TNBC targeting peptides in therapeutic delivery (Sader et al., 2015).

In this study, 250nm Fe_3O_4 Nanomag®-D magnetic nanoparticles were electrostatically complexed to the most promising TNBC targeting peptides, TLYRMVVTS, LMQDMALRL and CSTTYYTVK. Nanomag®-D MNPs have been utilised in other studies with no effects on cell viability. For instance, other studies have shown that 250nm MNP incubation did not impact the viability of adipose-derived stem cells, mesenchymal stromal cells or primary dendritic cells (Yang et al., 2016, Markides et al., 2019, Marcos-Campos et al., 2011). In addition, the viability of h-TERT-BJ1 cells in collagen gel cultures was not impacted by 100nm, 200nm or 500nm MNPs (Lewis et al., 2015). However, MTT assays conducted in this study revealed that exposure to MNPs led to decreases in cell viability in HMEC-TERT and BT-549 cells, indicating that MNP cell viability effects may differ between cell lines (Figure 4.11).

Prior to complexation, TLYRMVVTS or LMQDMALRL did not impact the viability of the non-tumourigenic or TNBC cells, however incubation with CSTTYYTVK resulted in decreased cell viability being observed in HMEC-TERT, BT-549, CAL-51 and MDA-MB-231 cell lines (Figure 4.11). This indicated that CSTTYYTVK may be a cytotoxic peptide, however further toxicity experiments need to be conducted to confirm the results. The CSTTYYTVK-mediated decrease in cell viability in the TNBC cell lines indicates that

this peptide could be used as an anticancer agent by facilitating the delivery of high amounts of MNPs into TNBC cells. However, the decreased viability of HMEC-TERT cells suggests that CSTTYYTVK is not selective for malignant cells and could lead to toxicity in healthy mammary tissue. The viability of HMEC-TERT cells was also impacted by exposure to MNPs only, 9R and p21, therefore the toxicity observed in HMEC-TERTs with CSTTYYTVK may not be directly related to this peptide. Furthermore, the peptide did not impact the viability of MCF10A cells.

Following the MNP complexation method developed by Blokpoel Ferreras *et al.,* TNBC cell targeting by peptide:MNPs was firstly assessed by Prussian blue staining. Blokpoel Ferreras *et al.,* demonstrated that complexing p21-8R to MNPs via electrostatic interactions led to significant increases in intracellular MNP delivery compared to non-functionalised MNPs (Blokpoel Ferreras et al., 2020). As expected, incubation with p21:MNPs resulted in high levels of intracellular iron being observed in all cell lines compared to MNPs only, confirming that p21-8R was successfully complexed to MNPs by the method used in this study (Figure 4.12—Figure 4.16). The TNBC peptides also delivered MNPs into the cells, albeit much less than p21. This correlates with the flow cytometric analysis, where p21-mediated mRFP delivery led to much higher mean fluorescence values in all cell lines compared to the TNBC peptides. For example, p21 resulted in a 10-fold higher mean fluorescence value being measured in BT-549 cells compared to TLYRMVVTS (Figure 4.7).

The peptides TLYRMVVTS and LMQDMALRL generally did not lead to significant increases in Prussian blue staining compared to 9R:MNPs. Some differences in staining were observed, such as TLYRMVVTS:MNPs resulting in more Prussian blue staining than 9R:MNPs in CAL-51 cells (Figure 4.15). This is consistent with the cellular uptake of TLYRMVVTS-mRFP, as this peptide delivered more mRFP into CAL-51 cells than in the non-tumourigenic and other TNBC cell lines. The similar amounts of MNP delivery with TLYRMVVTS:MNPs and LMQDMALRL:MNPs compared to 9R:MNPs in each cell line relates to the flow cytometry data, where, apart from TLYRMVVTS-mRFP in CAL-51 cells, TLYRMVVTS-mRFP and LMQDMALRL-mRFP delivery did not lead to significant increases in mRFP delivery compared to the delivery of 9R-mRFP in the same cell line. However, when comparing the uptake of TNBC peptide:MNPs in MCF10A and HMEC-TERT cells to in BT-549, CAL-51 and MDA-MB-231 cells, TLYRMVVTS:MNPs and

LMQDMALRL:MNPs resulted in more cell specific MNP delivery in the TNBC cell lines than in the control cell lines, suggesting that the 9 mer peptides are TNBC specific. In a similar manner to CSTTYYTVK-mRFP, CSTTYYTVK:MNPs resulted in the highest delivery of the corresponding cargo into the TNBC cell lines in relation to the other TNBC peptide candidates. The cellular uptake of CSTTYYTVK:MNPs was also more in MCF10A and HMEC-TERT than TLYRMVVTS:MNPs and LMQDMALRL:MNPs, which is also consistent with the flow cytometry findings. Again, this suggests that CSTTYYTVK may target healthy tissue as well as TNBC cells, indicating that this peptide may not be a suitable targeting ligand for TNBC targeted therapeutics.

Quantitative analysis of MNP delivery found that TLYRMVVTS and LMQDMALRL did not lead to any increases in MNP uptake in the TNBC cell lines compared to 9R only, with CSTTYYTVK only leading to increased MNP delivery in BT-549 cells (Figure 4.17). Although significant, p21:MNPs did not lead to a large increase in intracellular MNP as expected, as Blokpoel Ferreras *et al.,* reported an up to 8-fold increase in cellular iron content with p21:MNPs compared to MNPs only (Blokpoel Ferreras et al., 2020). ICP-MS analysis found that p21:MNPs resulted in less than a 0.5-fold increase in cellular iron content compared to MNPs in the TNBC cell lines, which does not correlate with the ICP-MS results obtained by Blokpoel Ferreras *et al.,* or the Prussian blue staining results shown in Figure 4.12—Figure 4.16. These inconclusive results may have been due to non-specific, background MNPs being detected, as the MNP only experimental conditions led to much higher Fe values than expected when considering the Prussian blue staining images. Therefore, a greater number of cell washes may be required to remove any unbound MNPs and the ICP-MS method used in this study requires further optimisation before conclusive results can be obtained. Once optimised, the uptake of peptide:MNPs by MCF10A and HMEC-TERT also needs to be assessed to confirm the qualitative results obtained by Prussian blue staining.

The *in vitro* validation studies conducted in this chapter highlighted the 9-mer peptides TLYRMVVTS and LMQDMALRL as potential TNBC targeting peptides. The work presented here also demonstrated that peptides identified from NGPD can be efficiently recovered by inverse PCR and validated. Although multiple Z>2 peptides were found to target TNBC cells more readily than non-malignant cells, TLYRMVVTS, LMQDMALRL and CSTTYYTVK were found to lead to significant increases in two or

more TNBC cell lines compared to MCF10A and HMEC-TERT cells. Particularly, TLYRMVVTS resulted in higher mRFP delivery in all three TNBC cell lines compared to one or more of the control cell lines (Table 4.2). TLYRMVVTS and LMQDMALRL were also non-toxic to non-tumourigenic cells and appeared to lead to cell specific MNP delivery into TNBC cell lines (Figure 4.12—Figure 4.16). The ability of TLYRMVVTS and LMQDMALRL to facilitate the delivery of various molecular cargos specifically into TNBC cells whilst limiting off-target toxicity *in vitro* indicates that the peptides may be good targeting ligands for peptide-based targeted therapeutics. Therefore, to ensure the serum stability of the peptides and the peptide-mediated, selective delivery into TNBC tumours, the peptides require further *in vivo* validation.

Chapter 5: Identification of TNBC associated antigens and their corresponding autoantibodies

5.1 Introduction

The high-throughput screening of peptides against cancer cells *in vitro* aimed to identify a small peptide to target TNBC without the cancer-specific protein target being known. However, the discovery of tumour associated antigens (TAA), such as over-expressed self-antigens, differentiation antigens or cancer-testis antigens (CTA), may also lead to the development of a TNBC therapy which targets a TAA (Feola et al., 2020). Furthermore, TAA derived autoantibodies, which are produced in early-stage disease, are potential biomarkers of cancer and may be used in the early detection of TNBC (Kathrikolly et al., 2022).

Autoantibodies recognise self-antigens and are produced as part of the humoral immune response in cancer patients (Scanlan et al., 1998, Scanlan et al., 1999). They arise from the overexpression or aberrant expression of a particular antigen in cancerous cells compared to normal tissue (Qiu et al., 2018). Autoantibodies are now being explored as potential biomarkers of cancer, including lung, colorectal, liver and breast cancer (Chapman et al., 2008, Negm et al., 2016, Koziol et al., 2018, Zhong et al., 2008). Autoantibodies do not generally undergo proteolysis, are highly specific and are easily purified from blood serum, making cancer associated autoantibodies attractive biomarkers of early-stage cancer (Anderson et al., 2011). Furthermore, the autoantibody signal is amplified by the immune response and can be detected when their corresponding antigen can no longer be detected (Anderson et al., 2015).

In early carcinogenesis, the expression of TAAs in malignant cells is identified by the immune system via immune surveillance. Tumour cells expressing neoantigens are engulfed, lysed and TAAs are presented on the surface of antigen presenting cells (APCs) for recognition by $CD4^+$ T cells. The presentation of TAAs and interaction between APCs and T cells activates the humoral immune response, leading to T cell activation and proliferation. B cells with a high affinity for the TAA will also present the antigen for T cell recognition, with T cell binding resulting in B cell proliferation. A proportion of the TAA recognising B cells will serve as memory cells, whereas others

will differentiate into plasma cells, leading to the subsequent production and amplification of autoantibodies against the corresponding cancer antigen (Wardemann and Nussenzweig, 2007, Zaenker et al., 2016). The rapid increase in autoantibody titre can be detected in the blood serum of a patient. Autoantibodies can be detected prior to cancer diagnosis, as Liu *et al.,* detected autoantibodies against HER2/neu and p53 in prediagnostic sera collected from BCa patients more than 150 days prior to diagnosis and Chapman *et al.,* detected cancer associated autoantibodies four years prior to mammographic diagnosis in women at high risk of BCa (Chapman et al., 2007, Liu et al., 2012). In addition, Stanton *et al.,* identified a panel of autoantibodies (VPS35, ARPC2, SERBP1, KRT8 and PDIA6) from SEREX and natural protein array serological studies which were present prior to a palpable tumour in mouse models inoculated with mouse mammary tumour cells. Vaccination with the corresponding antigens elicited a immune response in the TgMMTV-neu and C3(1)Tag mouse models and thus inhibited tumour growth. Furthermore, IgG autoantibodies against VPS35, ARPC2, SERBP1, KRT8 and PDIA6 were found to be elevated in DCIS and invasive BCa patient serum compared to control serum (Stanton et al., 2021).

The detection of autoantibodies has also been found to be highly specific and sensitive (Henderson et al., 2019, Kostianets et al., 2017). For instance, a study by Negm *et al.,* identified a panel of autoantibodies raised against 12 TAAs in early-stage BCa patients. The autoantibodies were found in 57% of BCa patients with a high specificity of 85%. Moreover, a different panel of autoantibodies against TAAs were able to detect ER^+ BCa with 93% specificity (Negm et al., 2018). Research conducted by Kostianets *et al.,* also demonstrated that autoantibodies can be sensitive and specific markers of BCa. The study concluded that a panel of six antigens (RAD50, PARD3, SPP1, SAP30BP, NY-BR-62 and NY-CO-58) was able to differentiate between BCa patient and healthy donor sera with 70% sensitivity and 90% specificity (Kostianets et al., 2017).

Due to the heterogenous nature of cancer, a single biomarker is not sufficient, so often a panel of autoantibodies or TAAs are investigated. The clinically available *Early*CDT-Lung autoantibody test is most sensitive when autoantibodies to p53, NY-ESO-1, CAGE, GBU4-5, Annexin 1 and SOX2 are studied as a panel. The EarlyCDT-Lung test can detect up to 40% of lung cancers, demonstrating that autoantibodies can act as immunobiomarkers of cancer (Lam et al., 2011). Similarly, autoantibodies detected

early-stage gastric cancer from a panel of six TAAs (Hoshino et al., 2017). In regard to BCa, Bassaro *et al.,* detected significantly higher levels of a panel of autoantibodies against survivin, osteoprotegrin, interleukin-29, growth hormone and resistin in blood plasma extracted from BCa patients compared to blood plasma extracted from donors with no history of BCa (Bassaro et al., 2017). A recent study also found that increasing a panel of TAAs from four (p53, PRDX6, c-Myc, Hsp70) to five antigens (p53, PRDX6, c-Myc, Hsp70, Nm23) led to an increase in sensitivity from 28% to 34% when screened against BCa patient sera and healthy sera (Pagaza-Straffon et al., 2020). As early detection of BCa in younger patients can be challenging due to high breast density, a secondary diagnostic method is required for younger women. Lacombe *et al.,* found that a panel of eight well recognised breast autoantigens were able to detect tumour associated autoantibodies in patients under 50 years old with node-negative early-stage BCa, however large-scale trails are needed to validate these results (Lacombe et al., 2014).

The potential clinical application of the use of autoantibody detection in the early diagnosis of BCa has recently been shown through clinical trials, such as a current study which is investigating the use of autoantibodies in BCa detection (ClinicalTrials.gov/NCT03623945, 2018). In addition, the clinical trials Provista-001 and Provista-002 reported that a proteomic panel comprised of eight serum protein biomarkers and ten tumour associated autoantibodies was able to detect BCa in women under 50 years old with BI-RADS 3 or 4 with 87.5% sensitivity, 83.8% specificity and 99.1% negative predictive value (Lourenco et al., 2017). Furthermore, the assay was not impacted by breast density, implying that a serum immunobiomarker based assay would be able to diagnosis BCa in women with dense breast tissue, unlike imaging (Reese et al., 2017). The combinatorial proteomic marker assay, Videssa® Breast™, was further developed to aid in the diagnosis of women over 50 years old with indeterminate imaging results and will be clinically available in 2023 by Provista Diagnostics (Henderson et al., 2019). However, Chen *et al.,* found that Videssa® Breast™ was not robust enough to deter from tissue biopsy (Chen et al., 2019). This recent research highlights how tumour autoantibodies may be a valuable tool in the early diagnosis of difficult to detect BCas such as TNBC, yet findings may need to be confirmed by other diagnostic methods.

It has also been speculated that TAAs may be utilised for the design of personalised peptide vaccines (PPVs) and there are ongoing clinical trials investigating the efficacy of TAA-derived PPVs for cancer treatment (Sasada et al., 2012, Bezu et al., 2018). For example, a study by Ott *et al.,* reported that four out of six melanoma patients experienced no reoccurrence 25 months-post neoantigen vaccination. Additionally, the remaining two patients experienced complete tumour regression when the vaccine was administered in combination with anti-PD-1 therapy (Ott et al., 2017). PPVs have also displayed clinical efficacy in immunologically cold tumours, with a phase Ib clinical trial in glioblastoma patients leading to increased tumour infiltrating T cells and a phase III trial in castration-resistant prostate cancer resulting in survival benefits in a subset of patients, respectively (Keskin et al., 2019, Noguchi et al., 2021).

The research presented here supports the possibility of utilising autoantibodies in the early detection of BCa. The identification of their corresponding neoantigens may also be used for early detection of TNBC or for the development of biomarker targeted treatment or PPVs. The work presented in this chapter was therefore focused on detecting an autoantibody present in TNBC patient sera through identification of the corresponding TNBC neoantigen. To investigate autoantibody binding, flow cytometry and western blotting analyses were employed followed by LC-MS/MS proteomics to determine the identity of potential TNBC TAAs.

5.2 Results

5.2.1 TNBC patient IgG classification

Immunoglobulins G were purified from TNBC patient sera (n=40) sourced from the Nottingham Health Centre Biobank and the binding of the antibody fraction was investigated against the TNBC cell line MDA-MB-231. To firstly investigate whether the detection of autoantibodies is patient subgroup specific, individual patient IgGs were pooled depending on the clinical data provided with each serum sample. The global pool contained sera derived IgG from all 40 patients, with an equal volume of serum being used from each sample to form the pool. The other IgG pools were formed based on different clinical criteria (Table 5.1). Pools 1 and 2 separated the patient samples by age, where pool 1 represented individuals that were assumed to be premenopausal (≤50 years old) and pool 2 represented the women that were assumed to be postmenopausal (>50 years old). The next criteria considered was recurrence, where pool 3 contained serum from patients who experienced recurrence or disease progression and patients in pool 4 did not. Patients were also pooled based on whether they had received treatment (chemotherapy, radiotherapy or both) prior to sample collection in pool 5, or whether they had not received any prior treatment in pool 6. It is important to note that individual patient samples did not fit into just one pool, with each sample fitting the criteria of at least four different pools.

Table 5.1 TNBC patient IgG pools based on clinical characteristics

Pool	Number of patients	Clinical characteristics
Global	40	All patients
1	15	≤50 years old
2	25	>50 years old
3	13	Recurrence
4	26	No recurrence
5	25	Treatment
6	6	No Treatment

To identify TNBC IgG specific events, a commercially available human IgG isotype control was also employed as the negative control antibody in this study to

differentiate between non-specific background patient IgG binding and autoantibody binding events to TNBC antigens.

5.2.2 TNBC IgG pool binding to MDA-MB-231 cells

Firstly, to assess differences in TNBC patient antibody pool binding, TNBC MDA-MB-231 cells were incubated with either the control IgG or one of the patient IgG pools and any binding events were assessed by flow cytometry (Figure 5.1). Preliminary experiments with various conditions and fixative agents found that antibody binding was only observed in fixed/permeabilised cells and the permeabilisation of MDA-MB-231 cells with ice cold acetone resulted in an increase in the number of autoantibody binding events recorded. Therefore, all cells were fixed with ice cold acetone in the following experiments (Belinda Wang and Derek Hardie, University of Nottingham). Human IgG bound cells were detected using an APC conjugated anti-human IgG antibody. The forward and side scatter characteristics of the cell population was used to gate for the main cell population, therefore leading to cell aggregates and debris being discarded from sample analysis (Figure 5.1A). An unstained and secondary control were included in each experimental repeat to identify MDA-MB-231 cell autofluorescence and any non-specific binding events, respectively (Figure 5.1B and Figure 5.1C). Additionally, the binding of the control IgG was assessed in each experimental run to set a baseline gate that was used as a threshold to detect specific binding events (Figure 5.1D). Therefore, any binding events observed above the control gate suggested patient-derived IgG specific binding (Figure 5.1E).

Flow cytometric analysis revealed that, when compared to the control IgG, there were no significant differences in binding between the patient IgG antibody pools (Figure 5.2). However, as shown in the dot plots in Figure 5.2, binding events above the control IgG threshold were observed with each patient antibody pool. Furthermore, the percentage of the cell population with positive binding events in the global pool and pool 2 was significantly higher than the control (control vs global $p = 0.039$, control vs pool 2 $p = 0.021$). As there was no significant difference in the percentage of cells above the control gate across the patient pools, the global pool containing IgG antibodies from all patients (patient IgG) was used for all subsequent experiments.

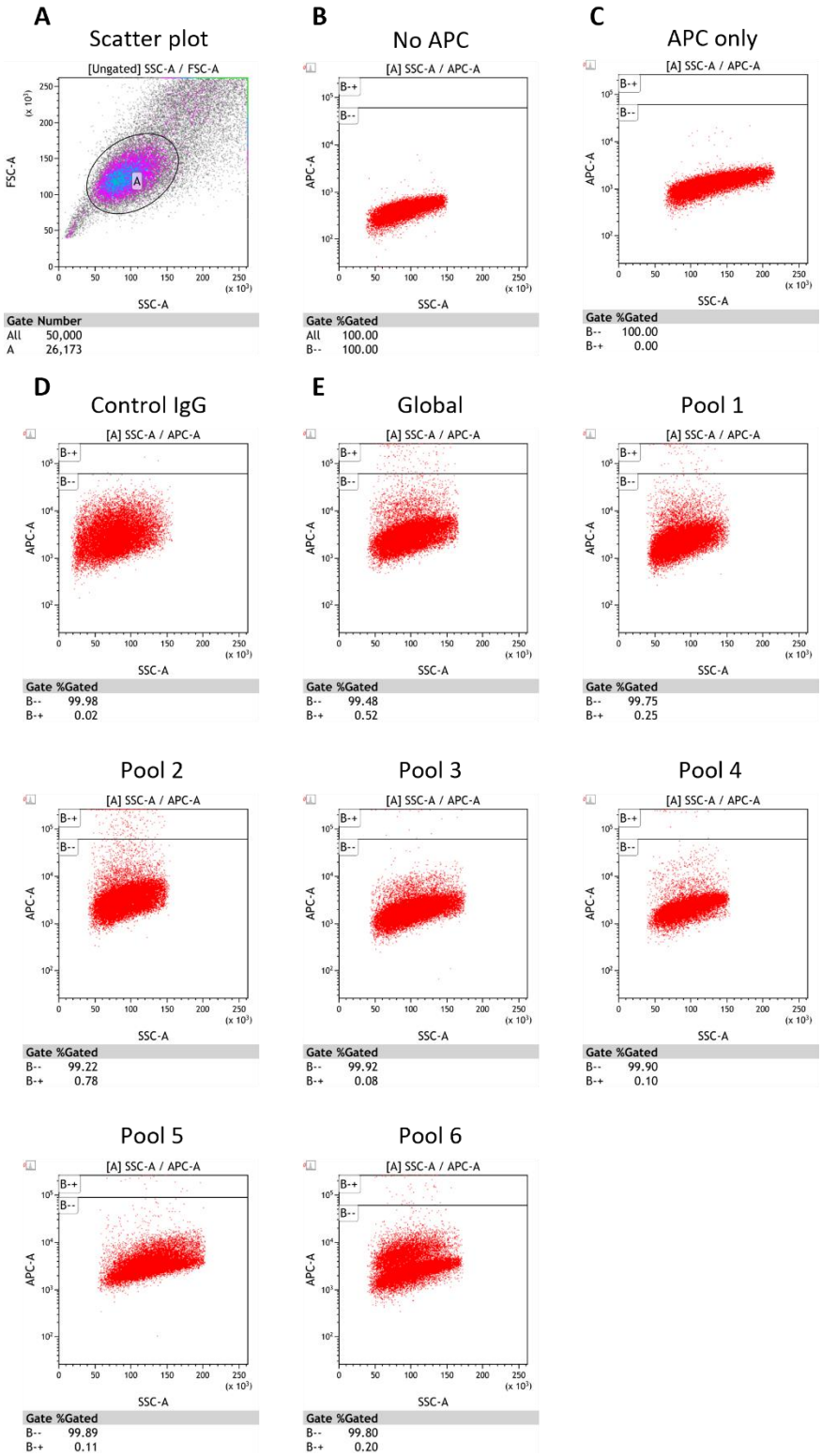


Figure 5.1 Dot plots of control IgG and TNBC patient IgG pool binding to MDA-MB-231 cells

A total of 20,000 single cell events were recorded by flow cytometry and identified from (A) forward and side scatter plots. Dot plots present (B) no anti-human IgG-APC, (C) anti-human IgG-APC only, (D) control IgG and (E) patient IgG pool data from one experimental repeat for reference. The experimental gate was set based on (D) control IgG binding data.

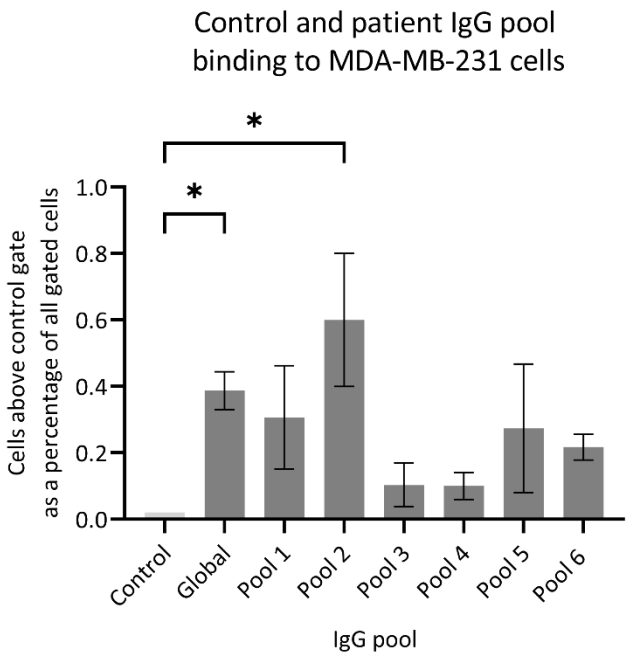


Figure 5.2 Control and TNBC patient IgG pool binding to MDA-MB-231 cells

The number of cells above the control gate were compared for each antibody pool. Binding data is shown for the control IgG or patient IgG antibody pools as the mean ± SD (n=3) with statistical significance calculated by a Kruskall-Wallis test. * $p = \leq 0.05$.

5.2.3 Control and patient IgG binding to MCF10A and MDA-MB-231 cells

Next, the binding of the control IgG and the patient IgG to both non-tumourigenic mammary MCF10A cells (Figure 5.3A) and TNBC MDA-MB-231 cells (Figure 5.3B) was investigated. Similar to the TNBC IgG pool binding analysis, forward and side scatter characteristics of the cell population were used to gate for the main cell population and an unstained and secondary control were included in each experimental repeat to identify MCF10A and MDA-MB-231 cell autofluorescence and any non-specific binding events. As differences in control IgG and patient IgG binding were being assessed, the threshold gate was determined on the binding of APC conjugated anti-human IgG without the primary antibodies, with any binding events above this gate being considered (Figure 5.3, APC only).

Differences in control and patient IgG binding to MCF10A and MDA-MB-231 cells are shown in Figure 5.4. Statistical analysis found that even though there was an increase

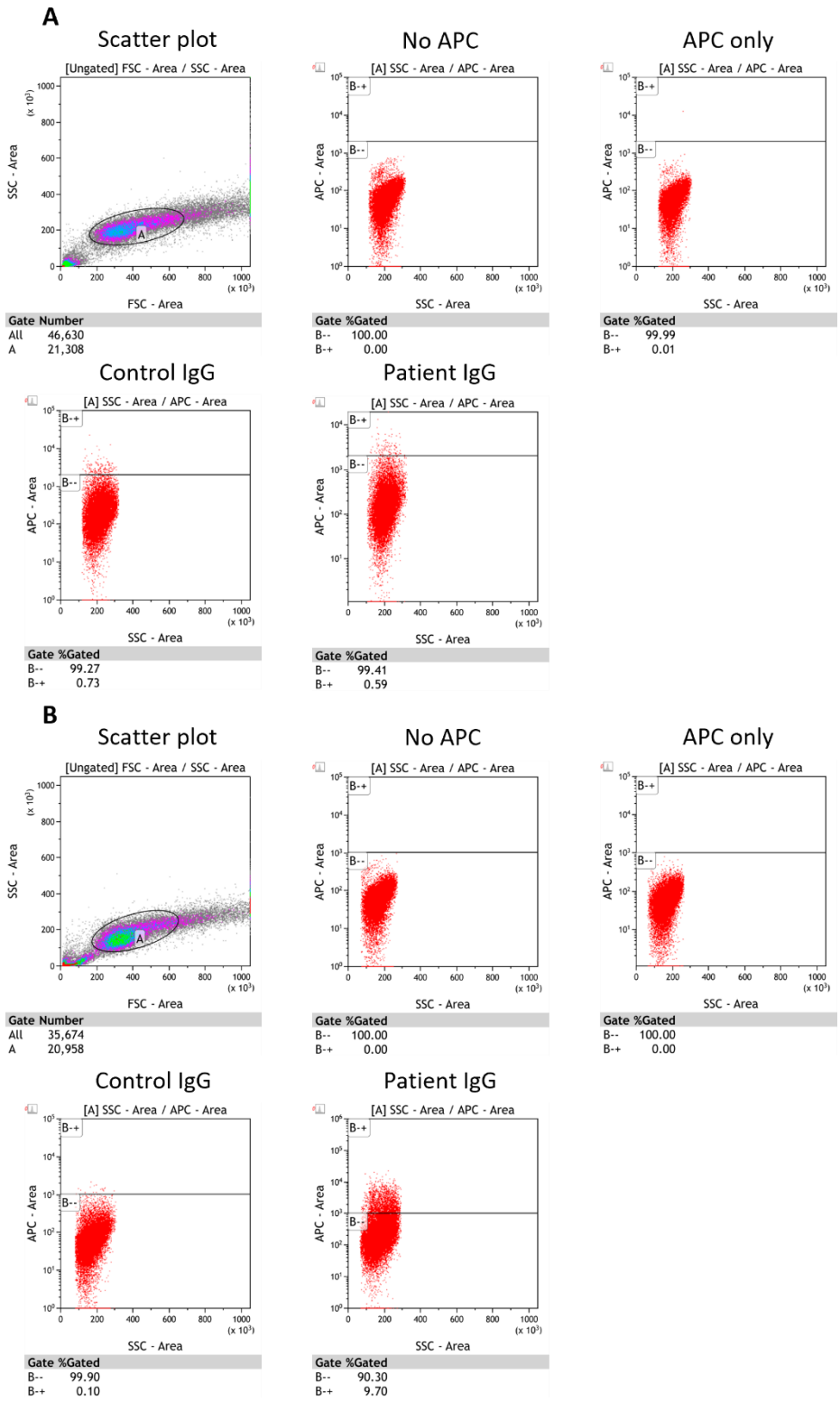


Figure 5.3 Dot plots of control IgG and TNBC IgG binding to MCF10A and MDA-MB-231 cells

A total of 20,000 single cell events were recorded for (A) MCF10A and (B) MDA-MB-231 samples by flow cytometry and identified from forward and side scatter plots. Dot plots present data from one experimental repeat for reference. Cell gates were set based on anti-human IgG-APC only samples.

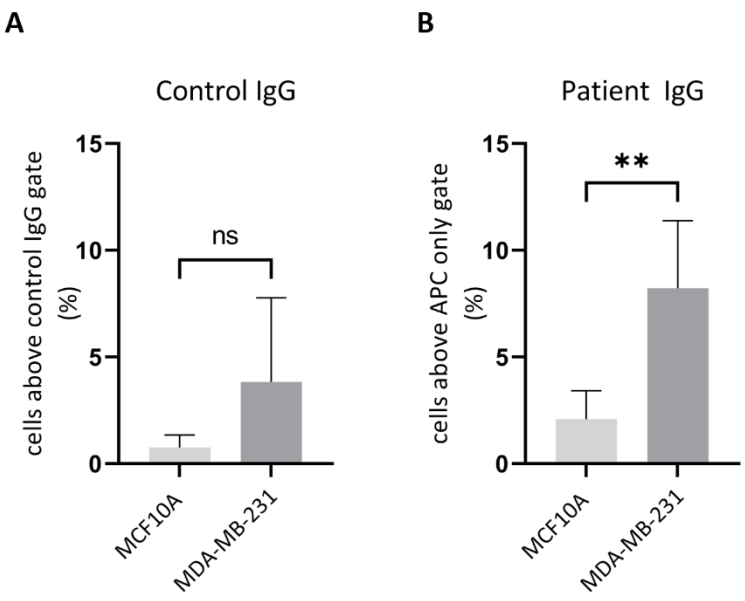


Figure 5.4 Control and TNBC patient IgG binding to MCF10A and MDA-MB-231 cells

The binding of the control and patient IgG to (A) MCF10A and (B) MDA-MB-231 cells was analysed by flow cytometry where 20,000 single cell events were recorded per sample. The number of control IgG and patient IgG binding events were assessed (mean ± SD). Statistical significance was calculated by unpaired t-test (n=6) where ** = 0.0014.

control IgG cell binding events observed to MDA-MB-231 cells when compared to MCF10A cells, the increase was not significant (Figure 5.4, 1.447 ± 0.9381). In contrast, more patient IgG binding events were observed with MDA-MB-231 cells than with MCF10A cells (Figure 5.4B, 6.510 ± 2.099; p = 0.0014).

Immunofluorescence studies confirmed that TNBC patient IgG binds MDA-MB-231 cells, as inferred by flow cytometric analyses (Figure 5.5B). The fluorescent images obtained confirmed that the anti-human IgG antibody did not lead to non-specific fluorescence being observed when cells were not bound by the control or patient IgG. The images did demonstrate, however, that both the control IgG and TNBC patient IgG bound to fixed MDA-MB-231 cells, with no visible difference in antibody binding observed with the patient IgG compared to the control IgG. Both IgGs appeared to bind to the cytoplasm, however the control IgG was also observed in cell nuclei. Similarly, control IgG binding localised to the nucleus in MCF10A cells, with spots of APC fluorescence observed (Figure 5.4A). Patient IgG also led to antibody fluorescence in this cell line, although the fluorescence did appear to be less than that observed in MDA-MB-231 cells.

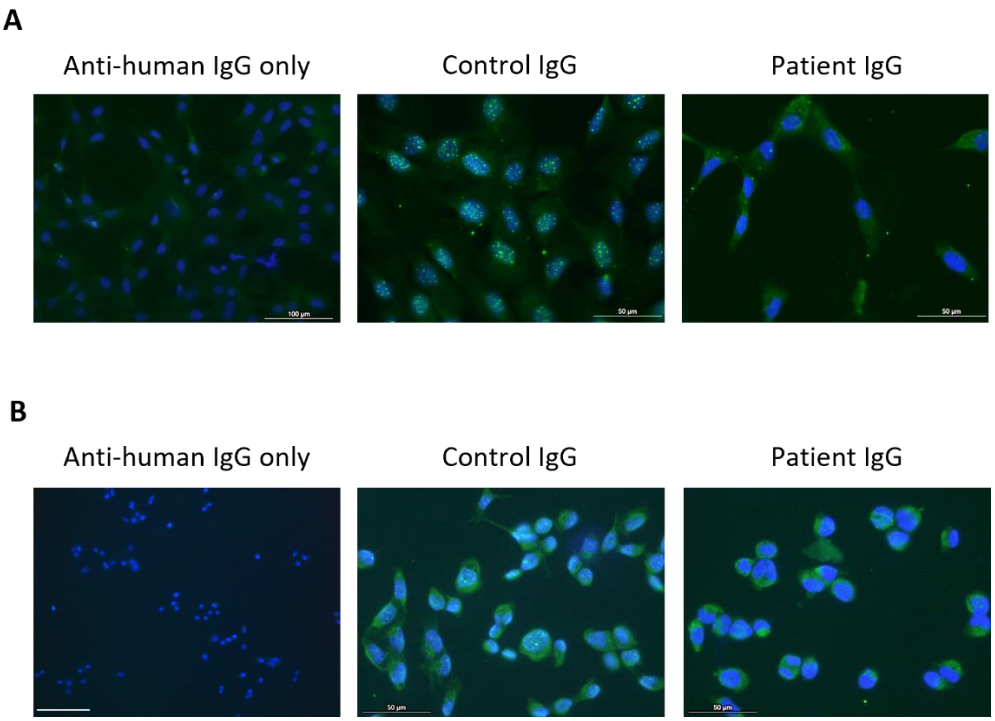


Figure 5.5 IF images of control and patient IgG binding to MCF10A and MDA-MB-231 cells
The binding of control or patient IgG to fixed and (A) MCF10A cells and (B) MDA-MB-231 was also observed via immunocytochemistry (n=3). Anti-human IgG only scale bar = 100µM and control IgG and patient IgG scale bar = 50µM.

5.2.4 TNBC patient IgG binding to $CD44^+/CD24^-$ MDA-MB-231 cells

Once patient IgG binding to the whole MDA-MB-231 cell population had been confirmed, antibody binding to the CSC population was assessed (Figure 5.6). Breast CSCs can be detected by using a combination of specific markers, as this population is defined as $CD44^+/CD24^-$ and $EpCAM^+$. In this study, the antibodies CD44-APCCy7, CD24-FITC and EpCAM-PE allowed for the detection of the $CD44^+/CD24^-/EpCAM^+$ cell population. However, due to a small percentage of MDA-MB-231 cells being $EpCAM^+$, there were only a small amount of $CD44^+/CD24^-/EpCAM^+$ cells identified per sample. Therefore, the percentage of cells bound to either the control or patient IgG was highly variable across replicates as few events were analysed. This led to $EpCAM^+$ not being used as an additional stem cell marker in the following experiments, with CSCs being defined as $CD44^+/CD24^-$ (Figure 5.6B). The $CD44^+$ and $CD24^-$ cells were determined by threshold gates which were set depending on the autofluorescence of the markers in an unstained MDA-MB-231 sample.

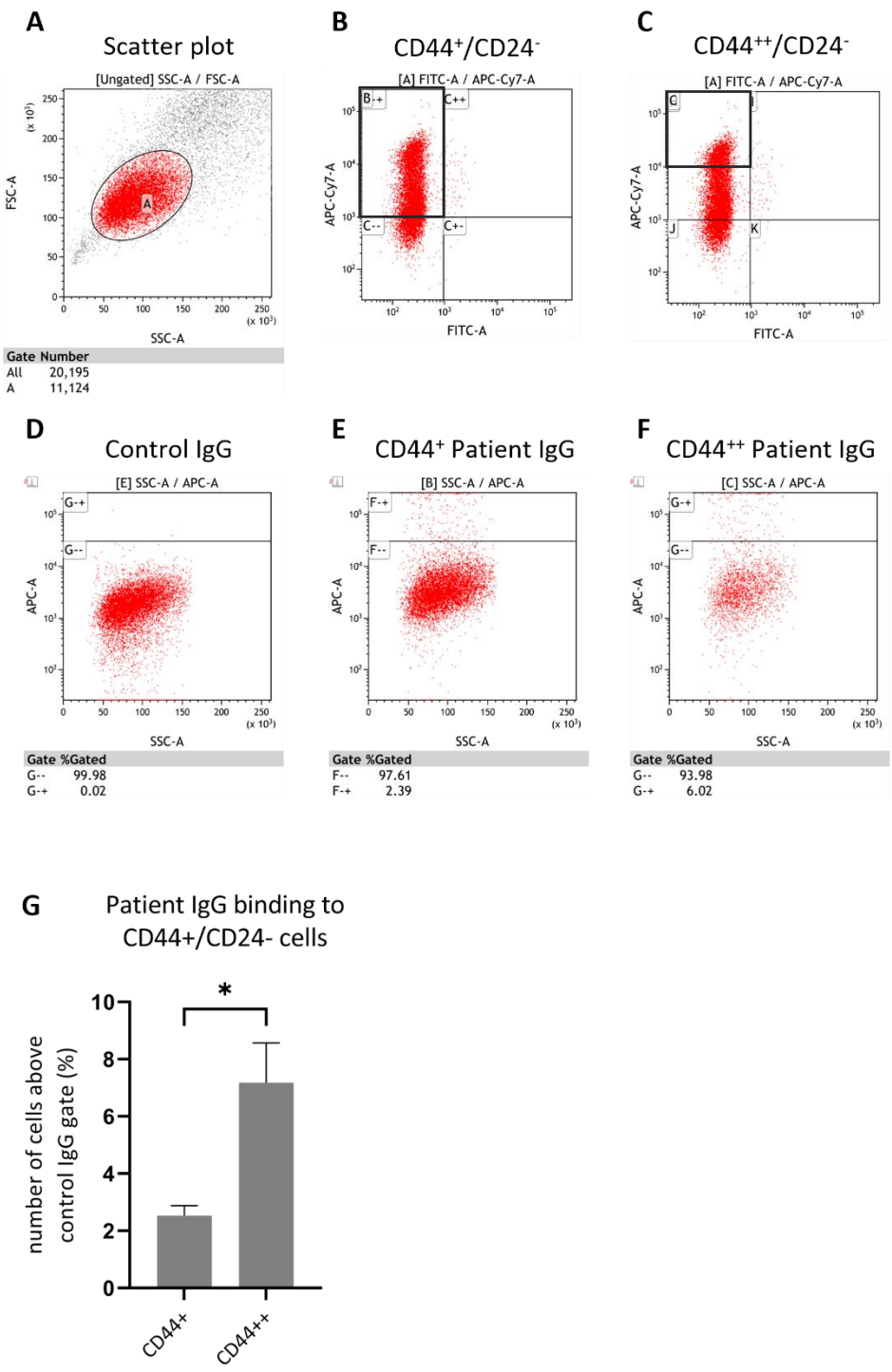


Figure 5.6 Patient IgG binding to CD44^{+}/CD24^{-} MDA-MB-231 cells

(A) The binding of the TNBC patient IgG to the MDA-MB-231 CD44^{+}/CD24^{-} population was analysed by flow cytometry where 10,000 single cell events were recorded per sample. (B) CD44^{+}/CD24^{-} and (C) CD44^{++}/CD24^{-} cell populations were identified and patient IgG binding events observed above the (D) control IgG gate in the (E) CD44^{+}/CD24^{-} (F) and CD44^{++}/CD24^{-} populations were assessed (mean ± SD). (G) Statistical significance was calculated by paired t-test (n=3) where * = 0.0377.

The very bright $CD44^{+}$ were termed $CD44^{++}$ (Figure 5.6C) and the binding of the patient IgG to both the $CD44^{+}$ and $CD44^{++}$ populations was assessed. The analysis revealed that a small percentage of both the $CD44^{+}/CD24^{-}$ (2.53% ± 0.35%) and $CD44^{++}/CD24^{-}$ (7.17% ± 1.40%) cell populations uniquely bound to patient IgG when compared to control IgG samples. Furthermore, a higher percentage of cells bound patient IgG in the $CD44^{++}$ population compared to the $CD44^{+}$ population (Figure 5.6G, $p = 0.0377$).

5.2.5 Control and patient IgG binding to MCF10A and MDA-MB-231 whole cell lysate

In order to identify neoantigens which are recognised by TNBC patient IgG, protein lysate extracted from MCF10A and MDA-MB-231 cells was separated by 1D SDS-PAGE and probed against the control and TNBC patient-derived IgG (Figure 5.7). When comparing the binding of patient IgG to MCF10A and MDA-MB-231 cells, the main difference observed was that the signal of a band at around 50kDa was more intense when bound to MDA-MB-231 samples compared to MCF10A samples. Furthermore, the same band in the control IgG blot was less intense than the patient IgG blot in MDA-MB-231 lanes. Interestingly, the 50kDa band was more intense in MCF10A lanes compared to MDA-MB-231 lanes when probed with control IgG. The patient IgG also bound to an MDA-MB-231 specific antigen between 80kDa and 100kDa which is indicated by the top arrow in Figure 5.7. The 80-100kDa band was not observed in MCF10A cell lysate probed with patient IgG, or in MCF10A or MDA-MB-231 protein lanes on the control IgG blot. In addition, a faint band at ~40kDA was present in patient IgG MDA-MB-231 lanes that was no observed in any other lanes on the control IgG or patient IgG blots (Figure 5.7, indicated by the bottom arrow).

The binding of TNBC patient-derived IgG to TNBC specific antigens was then investigated further by expanding the number of cell lines to include the non-tumourigenic cell lines BJ and HMEC-TERT (Figure 5.8A) and the TNBC cell lines BT-549 and CAL-51 (Figure 5.8B). Western blot analysis revealed that the patient IgG binding to the 50kDa band resulted in a more intense band in the MDA-MB-231 cell lysate lane compared to the other control and TNBC cell lines (Figure 5.8B).

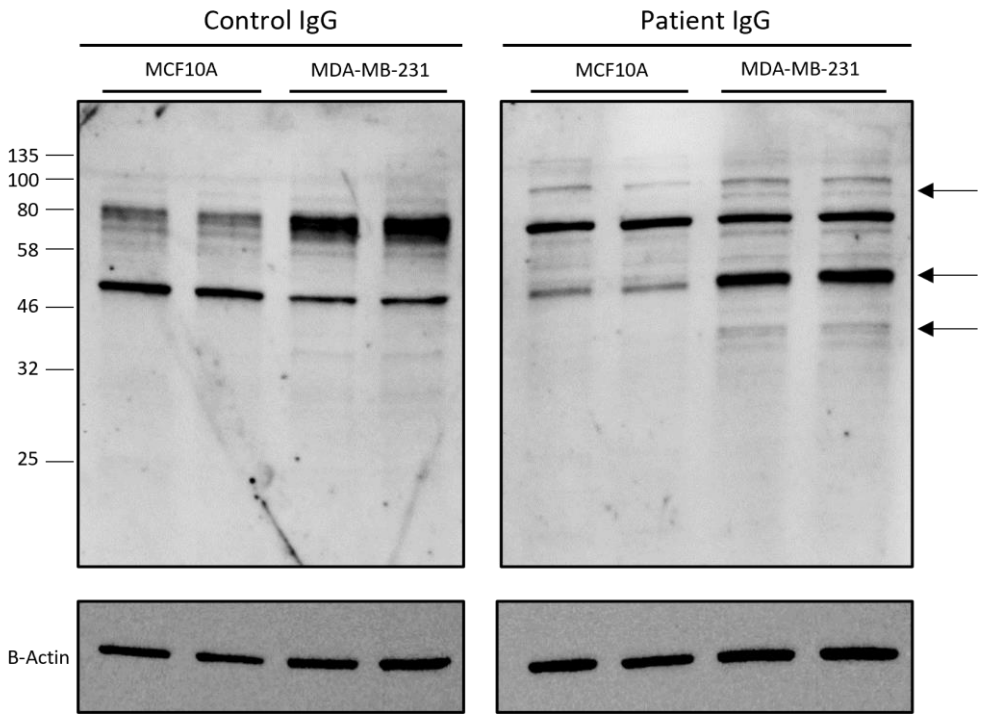


Figure 5.7 Control and patient IgG binding to MCF10A and MDA-MB-231 cell lysate
Protein extracts from both cell lines were separated by SDS-PAGE on a 12% polyacrylamide gel and probed with either control IgG or the TNBC patient derived IgG pool. B-actin was used as a protein loading control. Protein samples were loaded in duplicate.

However, this band is also observed in all cell line lanes when probed with control IgG (Figure 5.8A). The immunoreactive protein band observed at around 40kDa is also more intense when patient IgG is bound to the MDA-MB-231 lysate compared to protein extracted from the other TNBC cell lines (Figure 5.8B). As seen in Figure 5.7, the 40kDa protein does not appear to interact with the control IgG. Similarly, the immunoreactive band observed between 80-110kDa on the patient IgG blot in Figure 5.7 was also not observed when all cell lines were probed with control IgG (Figure 5.8A). The 80-100kDa band which is observed in Figure 5.7B is also present in the BT-549 lane, inferring that the antigen is also present in BT-549 cells.

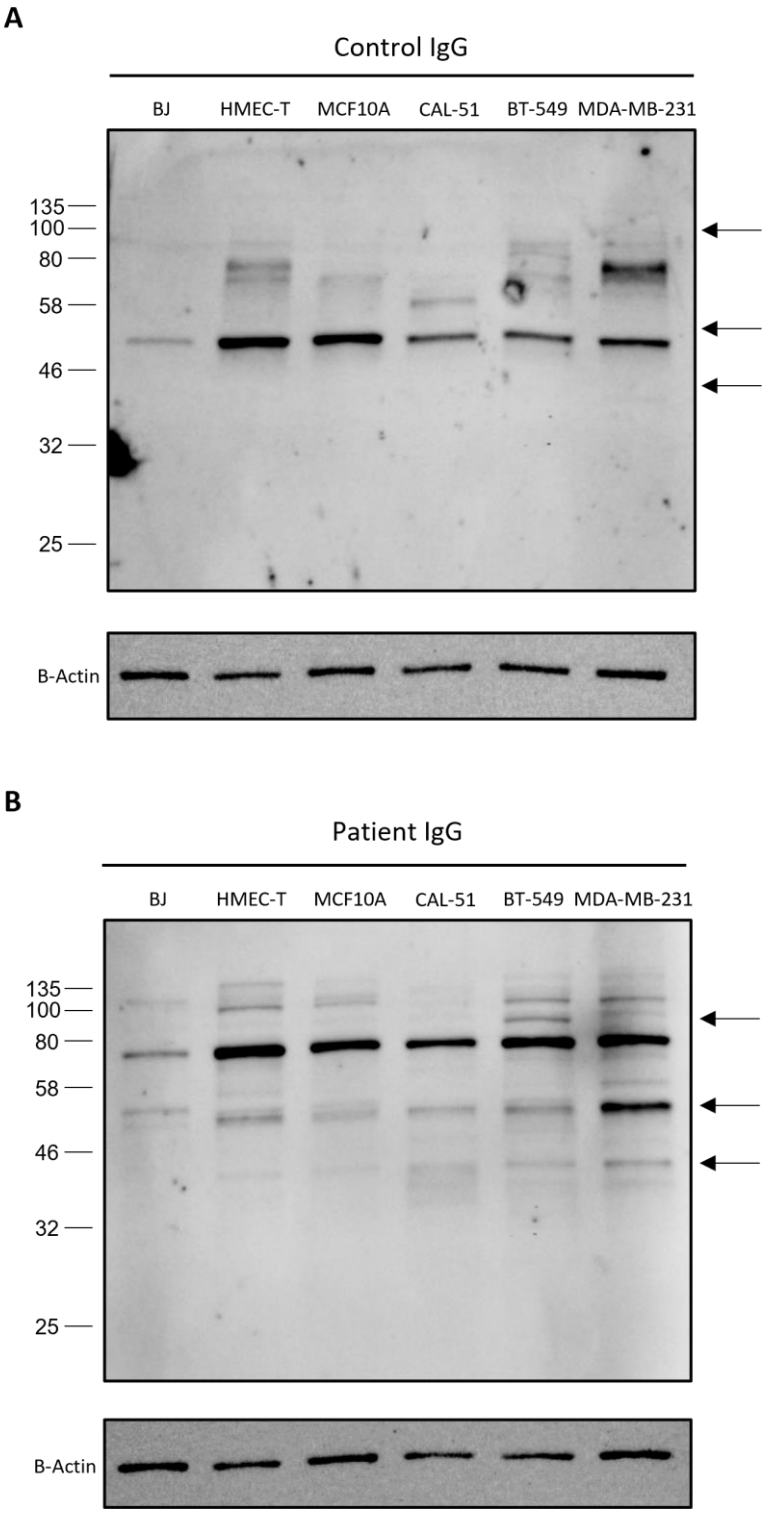


Figure 5.8 Control and patient IgG binding to control and TNBC cell lysate

Protein extracts from all cell lines was separated by SDS-PAGE on a 12% polyacrylamide gel and probed with either control IgG or the TNBC patient-derived IgG pool. Beta actin was used as a protein loading control.

5.2.6 Control and patient IgG binding to MCF10A and MDA-MB-231 nuclear, cytoplasmic and membrane proteins

Probing HMEC-TERT, MCF10A, BT-549 and CAL-51 extracted protein with control and TNBC patient IgG did not lead to any additional bands of interest being observed, therefore only protein extracted from MCF10A and MDA-MB-231 cells was used in further studies to determine if a neoantigen-derived autoantibody was present in TNBC patient IgG.

Following on from whole cell lysate western blot analyses, nuclear and cytoplasmic protein was extracted from MCF10A and MDA-MB-231 cell lysate to further characterise the TNBC patient IgG immunoreactive proteins and to determine the cellular localisation of potential TNBC neoantigens. Further separation of MCF10A and MDA-MB-231 cell lysates was also important to identify potential TNBC specific proteins by liquid chromatography with tandem mass spectrometry (LC-MS/MS).

The nuclear and cytoplasmic protein blot is shown in Figure 5.9. Clean separation of the nuclear and cytoplasmic fractions was confirmed by probing the two protein fractions with anti-histone deacetylase 1 (HDAC1) and anti-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) antibodies, as no HDAC1 or GAPDH bands were present in the cytoplasmic and nuclear fractions, respectively. The blot of NE-PER extracted protein revealed that the 80-100kDA band observed in MDA-MB-231 cells was nuclear, as the band was enriched in both MCF10A and MDA-MB-231 nuclear fractions compared to the cytoplasmic fractions (Figure 5.9, top arrow). In contrast to previous blots, the 80-110kDa band was not more intense in the MDA-MB-231 lane than the MCF10A lane when exposed to patient IgG. Nuclear and cytoplasmic fractionation of the cell lysate also identified the 40kDA band to be cytoplasmic, with a more intense band being observed in the MDA-MB-231 cytoplasmic fraction in the patient IgG blot compared to the control IgG blot (Figure 5.9, bottom arrow). In addition, only a faint band was observed at 40kDa in the MCF10A cytoplasmic protein lane of the patient IgG blot. The 50kDa band was present in both the nuclear and cytoplasmic fractions, inferring that the protein was localised in both cellular compartments (Figure 5.9, middle arrow). The 50kDa protein band was more intense in the MDA-MB-231 nuclear fraction than

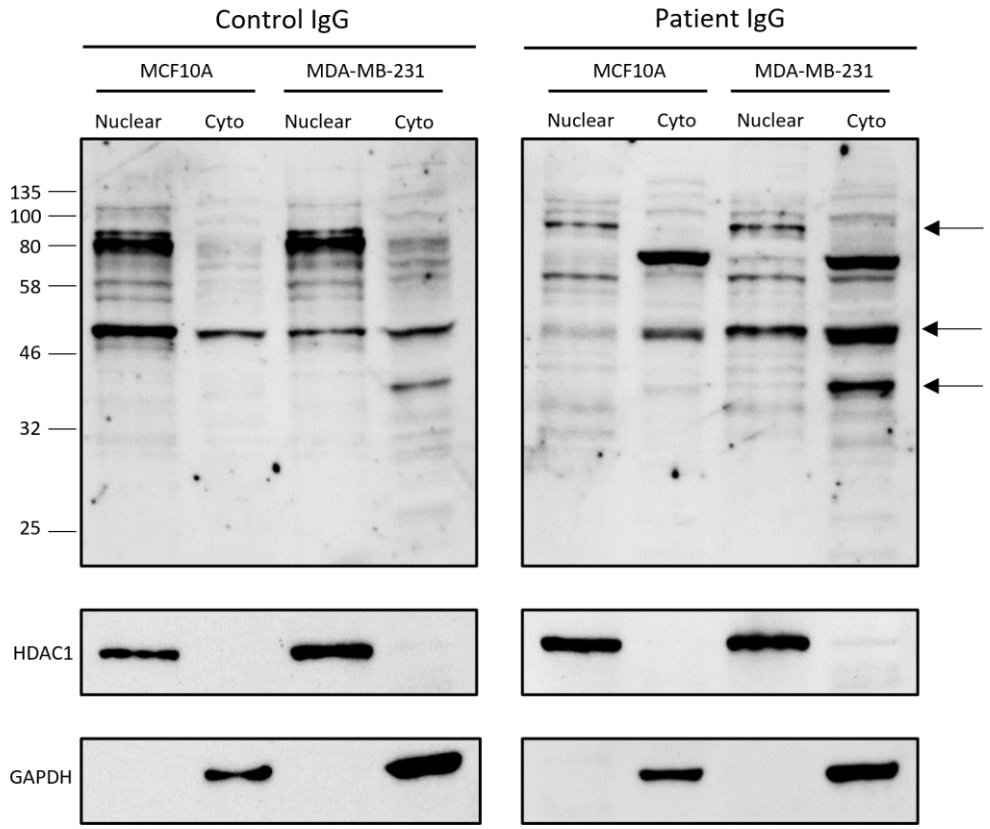


Figure 5.9 Control and patient IgG binding to MCF10A and MDA-MB-231 nuclear and cytoplasmic protein extracts

Nuclear and cytoplasmic extracts from both cell lines were separated by SDS-PAGE and probed with either control IgG or the TNBC patient derived IgG pool. HDAC1 was used as a nuclear protein control and GAPDH was used as a cytoplasmic protein control.

the MCF10A nuclear fraction when bound to patient IgG. Similarly, the band observed in the MDA-MB-231 cytoplasmic lane was also more intense than in the MCF10A cytoplasmic lane. As for the control IgG blot, the 50kDa band was also present in all protein fractions, but the most intense band was observed in the MCF10A nuclear fraction.

Cell lysate was also separated into cytoplasmic and membrane fractions and fractionation was confirmed with GAPDH and Na^+/K^+-ATPase antibodies, respectively (Figure 5.10). Fractionation of MCF10A and MDA-MB-231 cell lysate by membrane and cytoplasmic extraction found that the 40kDa band present in the cytoplasmic lanes in

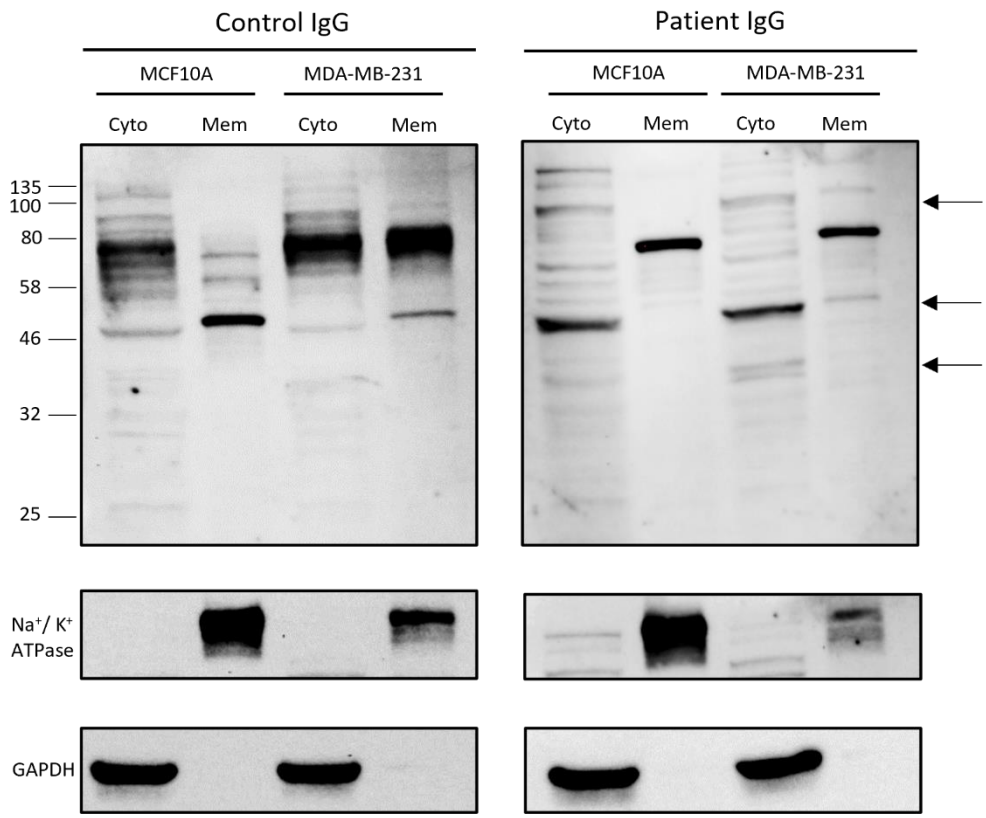


Figure 5.10 Control and patient IgG binding to MCF10A and MDA-MB-231 cytoplasmic and membrane protein

Cytoplasmic and membrane protein extracts from both cell lines were separated by SDS-PAGE and probed with either control IgG or the TNBC patient derived IgG pool. GAPDH was used as a cytoplasmic protein control and Na^+/K^+-ATPase was used as a membrane protein control.

Figure 5.9 is not observed in the membrane fractions (Figure 5.10, bottom arrow). The patient IgG reactive protein band at approximately 50kDa is also enriched in the cytoplasmic lanes, inferring that the proteins were not membrane bound (Figure 5.10, middle arrow). It is important to note that a band at around 50kDa was present in MCF10A and MDA-MB-231 lanes in the control IgG blot, however this was found in the membrane protein fractions. The 80-110kDa band of interest was observed in the cytoplasmic fraction, as this fraction also contained nuclear proteins (Figure 5.10, top arrow). The patient IgG did interact with an approximately 70kDa membrane protein, however the intensity of the band observed is similar in both MCF10A and MDA-MB-231 membrane protein lanes.

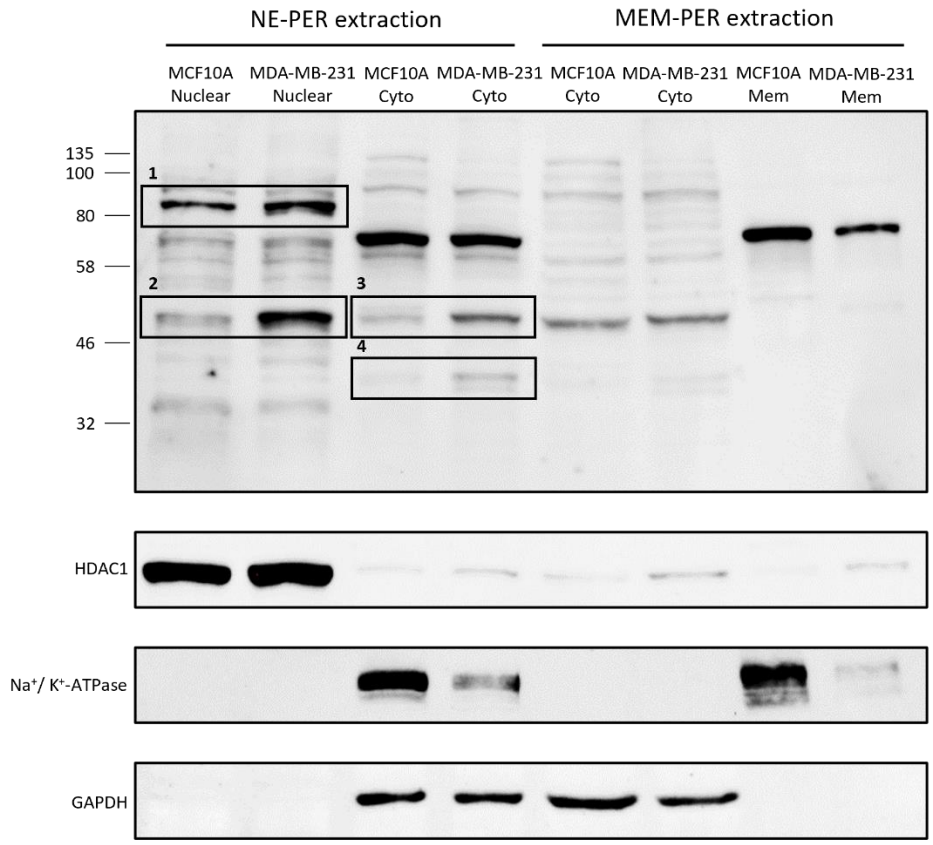


Figure 5.11 Control and patient IgG binding to MCF10A and MDA-MB-231 nuclear, cytoplasmic and membrane protein

Nuclear, cytoplasmic and membrane protein extracts from both cell lines were separated by SDS-PAGE and probed with either control IgG or the TNBC patient derived IgG pool. HDAC1 was used as a nuclear protein control, GAPDH was used as a cytoplasmic protein control and Na^+/K^+-ATPase was used as a membrane protein control.

Western blot analyses suggested that TNBC antigens whose molecular weights were approximately 40kDa, 50kDa and 80-100kDa are recognised by a TNBC specific antibody. To determine the identity of the potential TNBC antigens, the gel areas of interest were extracted for proteomic analysis by LC-MS/MS. Areas of interest were confirmed in nuclear, cytoplasmic and membrane fractions separated on a 1D SDS-PAGE gel probed with patient IgG (Figure 5.11). This blot confirmed that the areas of interest for mass spectrometry are the nuclear protein bands observed at 80-100kDa (Figure 5.11, box 1), the 50kDA bands observed in the nuclear (Figure 5.11, box 2) and cytoplasmic fractions (Figure 5.11, box 3), and the band present at around 42kDa in the MDA-MB-231 cytoplasmic fraction (Figure 5.11, box 4). The patient IgG blot also verified that all fractions were well separated, with HDAC1, GAPDH and Na^+/K^+-ATPase

only being detected in nuclear, cytoplasmic and membrane fractions, respectively. In addition, there were no additional bands observed in MEM-PER extracted lanes, confirming that no TNBC specific membrane neoantigens elicited an antibody response.

5.2.7 LC-MS/MS sample preparation

Prior to extracting gel slices for mass spectrometry, a 12% polyacrylamide gel containing the nuclear, cytoplasmic and membrane fractions was separated by electrophoresis and silver stained (Figure 5.12). Silver staining detected all separated proteins and confirmed that the proteins in each fraction resolved well. Gel slices were then excised as shown in Figure 5.13 and submitted for LC-MS/MS analysis (Cambridge Centre for Proteomics).

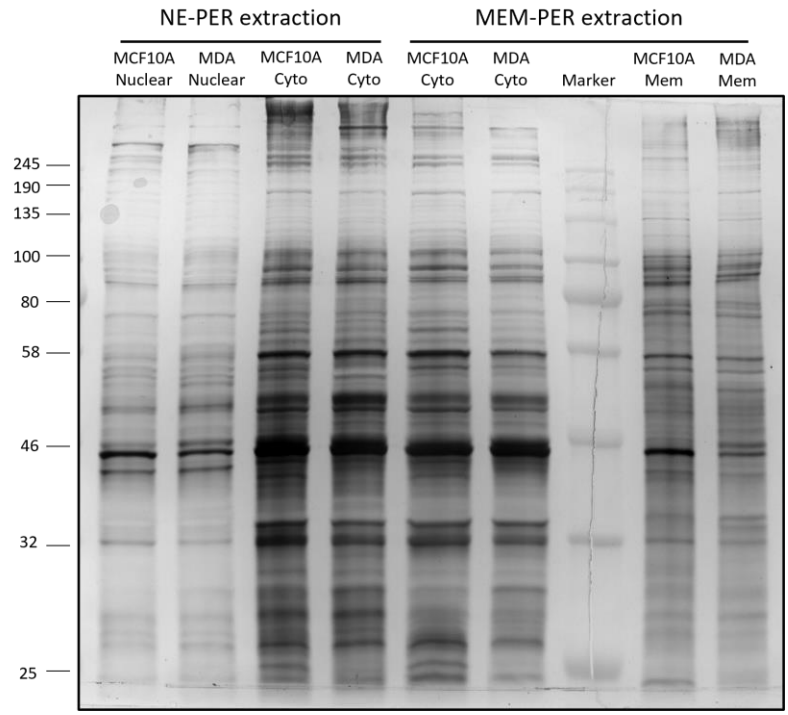


Figure 5.12 Silver stained MCF10A and MDA-MB-231 proteins

NE-PER and MEM-PER extracted fractions were loaded onto a 12% polyacrylamide gel (10µg per lane) and proteins were separated by SDS-PAGE. Proteins were then detected by silver staining. MDA = MDA-MB-231.

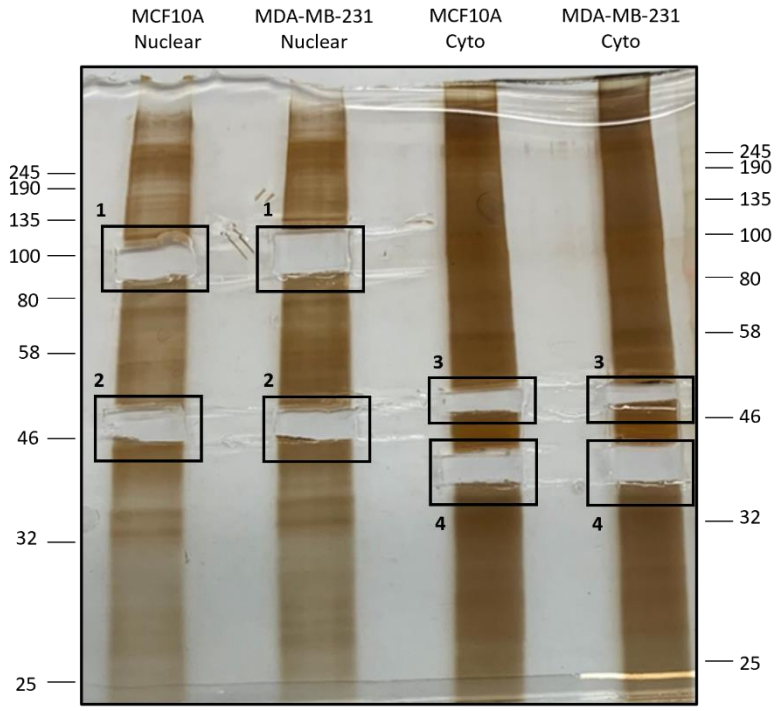


Figure 5.13 Sample preparation for LC-MS/MS

Polyacrylamide gel slices containing nuclear or cytoplasmic MCF10A and MDA-MB-231 proteins were excised based on molecular weight and submitted for LC-MS/MS. Gel slices for MCF10A and MDA-MB-231 protein were excised to encapsulate the nuclear protein bands observed at approximately (1) 90 kDa and (2) 49kDa and the cytoplasmic bands observed at approximately (3) 49kDa and (4) 42kDa in Figure 5.11. Three gel slices were excised from each area in MCF10A and MDA-MB-321 protein lanes (n=3) with one experimental repeat shown.

5.2.8 LC-MS/MS analysis

Label-free LC-MS/MS and subsequent Mascot peptide database searches resulted in the identification of a total of 2346 proteins based on 95% peptide probability and 99% protein probability settings in Scaffold (Searle, 2010). The exclusive spectral count data for each excised band, which was the number of spectra associated with a given protein, was then assessed to determine differences in protein abundance in MCF10A and MDA-MB-231 cells.

Shotgun proteomic analysis was firstly conducted using the unique normalised spectral abundance factor (uNSAF) method described by Zhang *et al.,* (Zhang et al., 2010). This method calculated the abundance of a given protein across multiple independent samples by dividing the unique spectral count data acquired for a given

protein by the protein length (SAF). The individual SAF values were also normalised to account for any count discrepancies between experimental runs (see 2.6.9). As polyacrylamide gel slices were excised to encapsulate patient IgG identified nuclear bands at approximately 90kDa (Figure 5.13, gel slice 1) and 49kDa (Figure 5.13, gel slice 2) and cytoplasmic bands at approximately 49kDa (Figure 5.13, gel slice 3) and 42kDa (Figure 5.13, gel slice 4), only LC-MS/MS identified proteins with similar molecular weights (± 20% kDa) were considered. Significant differences in NSAF values for each protein when comparing MCF10A and MDA-MB-231 samples were then determined by t-tests with 1% FDR (Figure 5.14).

The semi-quantitative analysis identified two statistically significant proteins with higher spectral abundances in MDA-MB-231 80kDa nuclear band samples than the corresponding MCF10A samples (Figure 5.14A). The proteins, targeting protein for TPX2 Microtubule Nucleation Factor (TPX2) and SATB Homeobox 2 (SATB2), were 72kDa and 83kDa, respectively. Excised polyacrylamide gel slices at 49kDa in nuclear fractions led to the identification of two potential antigens, KH RNA Binding Domain Containing, Signal Transduction-associated protein 3 (KHDRBS3) and Secernin-1 (SCRN1) (Figure 5.14B). NSAF proteomic analysis of cytoplasmic bands also cut at approximately 49kDa identified Enolase 2 (ENO2), Solute Carrier Family 29 Member 1 (SLC29A1) and Tumour Protein P53 (TP53) as proteins with significantly higher abundances in MDA-MB-231 samples compared to MCF10A samples (Figure 5.13C). In addition, statistical analysis of the proteins identified in the 42kDa MCF10A and MDA-MB-231 cytoplasmic gel slices found that five proteins were found to have statistically higher NSAF values in MDA-MB-231 samples compared to MCF10A samples. These proteins were Alpha-2-HS-Glycoprotein (AHSG), Dual Specificity Phosphatase 4 (DUSP4), Pelota mRNA Surveillance And Ribosome Rescue Factor (PELO), Calumenin (CALU) and DNAJ Heat Shock Protein Family Member B6 (DNAJB6).

Other statistical methods have also been applied for comparative analysis of label-free proteomics data. Langley and Mayr performed seven different statistical methods for differential expression analyses of label-free proteomics spectral counts and found that DESeq2 led to true positive results being obtained with a limited number of false positives (Langley and Mayr, 2015). Therefore, DESeq2 was implemented to expand the list of potential TNBC antigens identified from the LC-MS/MS proteomic data.

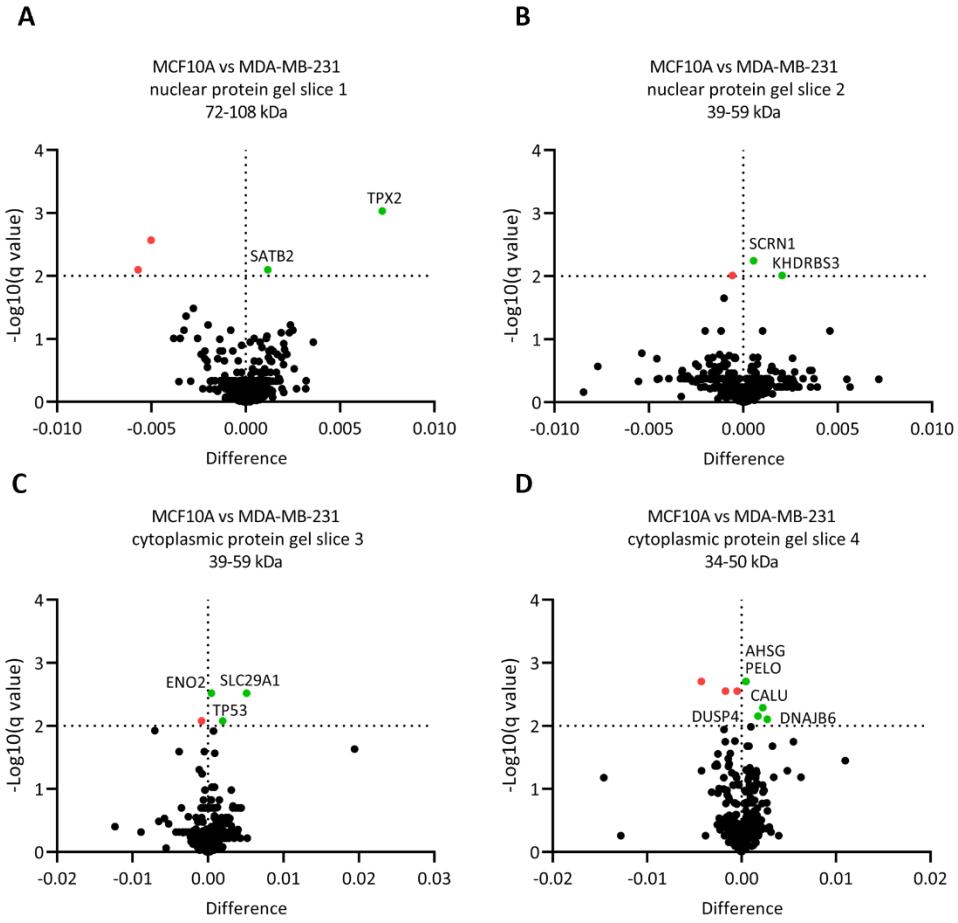


Figure 5.14 Overexpressed proteins in MDA-MB-231 cells identified by NSAF

NSAF values for LC-MS/MS identified proteins in MCF10A and MDA-MB-231 protein samples were compared by multiple unpaired t-tests with Benjamini, Krieger and Yekutieli correction (n=3). (A) Proteins identified from gel slice area 1 samples where MCF10A and MDA-MB-231 polyacrylamide gel bands were excised to encapsulate 90kDa nuclear proteins. (B) Proteins identified from gel slice area 2 samples where MCF10A and MDA-MB-231 polyacrylamide gel bands were excised to encapsulate 49kDa nuclear proteins. (C) Proteins identified from gel slice area 3 samples where MCF10A and MDA-MB-231 polyacrylamide gel bands were excised to encapsulate 49kDa cytoplasmic proteins. (D) Proteins identified from gel slice area 4 samples where MCF10A and MDA-MB-231 polyacrylamide gel bands were excised to encapsulate 42kDa cytoplasmic proteins. Statistically significant proteins (< 0.01 FDR) with decreased protein levels in MDA-MB-231 samples compared to MCF10A samples are shown in red. Statistically significant proteins (< 0.01 FDR) with increased protein levels in MDA-MB-231 samples compared to MCF10A samples are shown in green and labelled with gene names. Non-significant proteins (> 0.01 FDR) are shown in black.

As DESeq2 is designed to model high count RNA-seq data, Benjamini-Hochberg FDR correction was applied as described by Langley and Mayr (Langley and Mayr, 2015). The DESeq2 with BH correction analysis results are shown in Figure 5.15. Unlike NSAF and t-test calculations, the protein lists were not filtered based on molecular weight prior to analysis. The negative binomial distribution model resulted in the identification of three proteins with significantly higher abundances in MDA-MB-231 than MCF10A gel slices which were excised at approximately 90kDa, with TPX2, SATB2 and WRN Helicase Interacting Protein 1 (WRNIP1) resulting in fold changes (log_2FC) of 2.15, 4.39 and 5.68, respectively (Figure 5.15A). DESeq2 comparative analysis of MCF10A and MDA-MB-231 protein abundances in gel slices 2 (nuclear 49kDa band) and 3 (cytoplasmic 49kDa band) resulted in no proteins with statistically higher protein abundances in MDA-MB-231 protein samples than MCF10A protein samples being observed (Figure 5.15B and Figure 5.15C). In contrast, six proteins, Keratin 19 (KRT19), Serine Family B Member 9 (SERPINB9), Vimentin (VIM), Calponin-3 (CNN3), Major Histocompatibility Complex, Class I, A (HLA-A) and DNAJB6 were found to be significantly upregulated in MDA-MB-231 in protein samples which were cut at 42kDa (gel slices 4) in cytoplasmic lanes (Figure 5.15D).

5.2.9 Expression of candidate TNBC antigens in TNBC patient datasets

NSAF and DESeq2 statistical analyses resulted in a total of 12 proteins being identified as upregulated in MDA-MB-231 compared to MCF10A protein gel slices. As protein gel slices were extracted where more intense immunoreactive bands were observed in MDA-MB-231 protein fractions compared to MCF10A when probed with patient IgG, the proteins identified by LC-MS/MS proteomics may be neoantigens recognised by a patient specific autoantibody present in TNBC patient IgG. However, the potential TNBC antigens were identified in an *in vitro* TNBC cell line and not in patient samples, therefore their clinical relevance was evaluated next. To this end, the mRNA expression levels of each potential TNBC antigen were investigated in patient datasets (Figure 5.16).

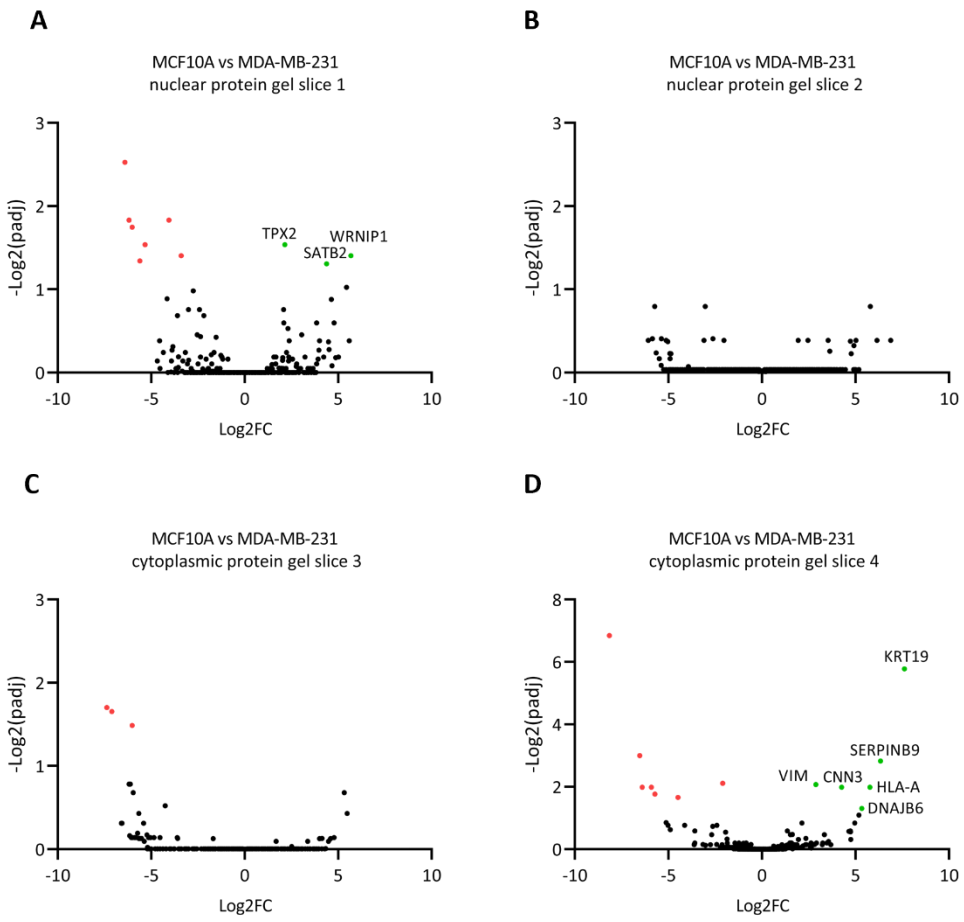


Figure 5.15 Overexpressed proteins in MDA-MB-231 cells identified by DESeq2

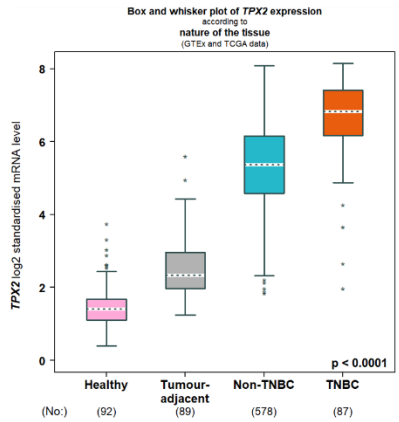
Differences in exclusive spectrum count data for LC-MS/MS identified proteins in MCF10A and MDA-MB-231 protein samples were identified by DESeq2 with Benjamini-Hochberg FDR correction (n=3). (A) Proteins identified from gel slice area 1 samples where MCF10A and MDA-MB-231 polyacrylamide gel bands were excised to encapsulate 90kDa nuclear proteins (± 20%). (B) Proteins identified from gel slice area 2 samples where MCF10A and MDA-MB-231 polyacrylamide gel bands were excised to encapsulate 49kDa nuclear proteins (± 20%). (C) Proteins identified from gel slice area 3 samples where MCF10A and MDA-MB-231 polyacrylamide gel bands were excised to encapsulate 49kDa cytoplasmic proteins (± 20%). (D) Proteins identified from gel slice area 4 samples where MCF10A and MDA-MB-231 polyacrylamide gel bands were excised to encapsulate 42kDa cytoplasmic proteins (± 20%). Statistically significant proteins (padj ≤ 0.05) with decreased protein levels in MDA-MB-231 samples compared to MCF10A samples are shown in red. Statistically significant proteins (padj ≤ 0.05) with increased protein levels in MDA-MB-231 samples compared to MCF10A samples are shown in green and labelled with gene names. Non-significant proteins (padj > 0.05) are shown in black. FC = fold change and padj = adjusted p-value.

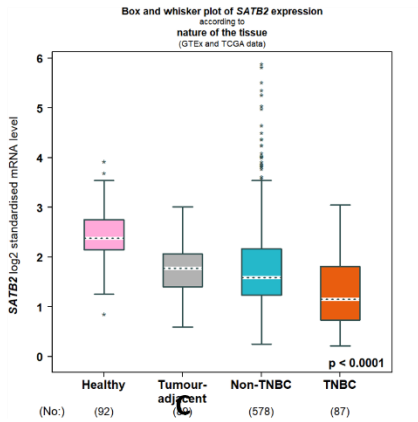
Differential gene expression analyses were conducted using the statistical mining tool BC-GenExMiner (Jézéquel et al., 2021) and by comparing normal (Genotype-Tissue Expression project, GTEx) and BCa (The Cancer Genome Atlas network, TCGA) transcriptomic datasets. The analysis compared the targeted gene expression levels for each LC-MS/MS identified protein depending on the nature of tissue, with mRNA expression levels from healthy (n=92, GTEx), tumour-adjacent (n=89, TCGA), non-TNBC (n=578, TCGA) and TNBC (n=87, TCGA) samples shown in Figure 5.16.

The mRNA expression levels of TPX2, SATB2 and WRNIP1 were firstly analysed, with expression levels of only TPX2 ($p < 0.0001$) and WRNIP1 ($p < 0.0001$) being significantly higher in TNBC than in both healthy and tumour adjacent tissue (Figure 5.16A). The nuclear protein KHDRBS3 ($p < 0.05$) also had higher expression levels in TNBC than in healthy samples (Figure 5.16B), but SCRN1 expression was not significantly higher in TNBC. None of the cytoplasmic proteins ENO2, SLC29A1 and TP53 were overexpressed in TNBC samples (Figure 5.16C). The proteins with significantly higher spectral abundances in MDA-MB-231 cytoplasmic gel slices at 42kDa compared to the relevant MCF10A samples are shown in Figure 5.16D. Out of the ten proteins identified, four proteins, CALU ($p < 0.0001$), DNAJB6 ($p < 0.01$), KRT19 ($p < 0.01$) and HLA-A ($p < 0.0001$) had significantly higher expression levels in TNBC samples compared to healthy samples.

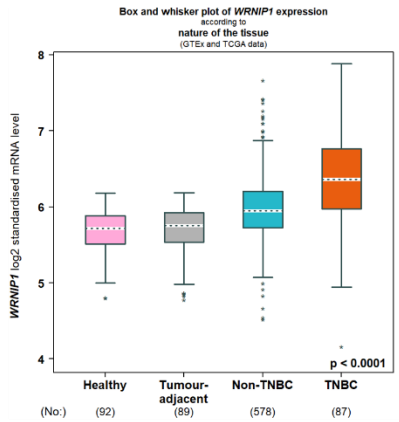
In summary, a total of seven proteins, TPX2, WRNIP1, KHDRBS3, CALU, DNAJB6, KRT19 and HLA-A were found to have higher expression levels in MDA-MB-231 excised protein bands compared to the corresponding MCF10A bands (Figure 5.14 and Figure 5.15), as well as being overexpressed in TNBC from transcriptomic datasets (Figure 5.16). Thus, TPX2, WRNIP1, KHDRBS3, CALU, DNAJB6, KRT19 and HLA-A were taken forward as potential TNBC autoantibody antigens.

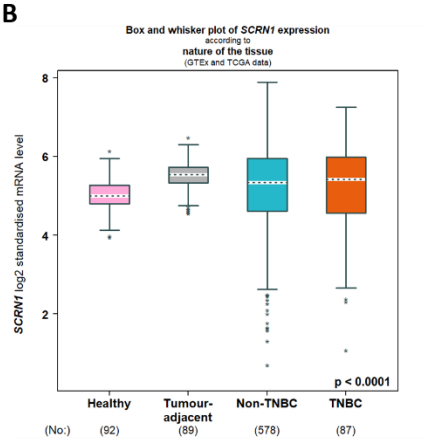
A


Box and whisker plot of TPX2 expression
according to
nature of the tissue
(GTEx and TCGA data)
TPX2 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

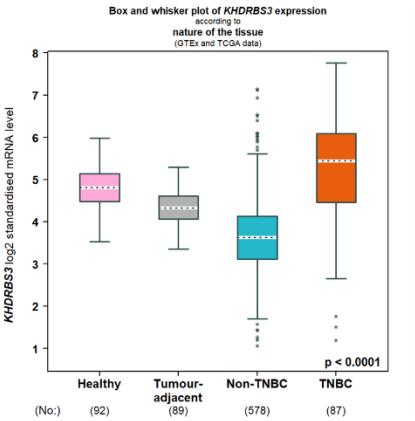

Box and whisker plot of SATB2 expression
according to
nature of the tissue
(GTEx and TCGA data)
SATB2 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

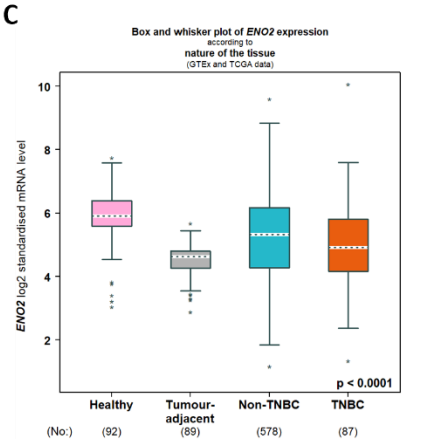
B

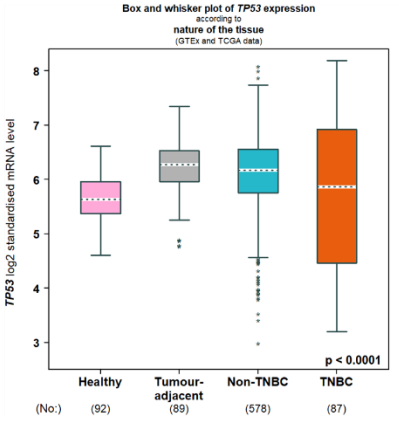

Box and whisker plot of WRNIP1 expression
according to
nature of the tissue
(GTEx and TCGA data)
WRNIP1 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

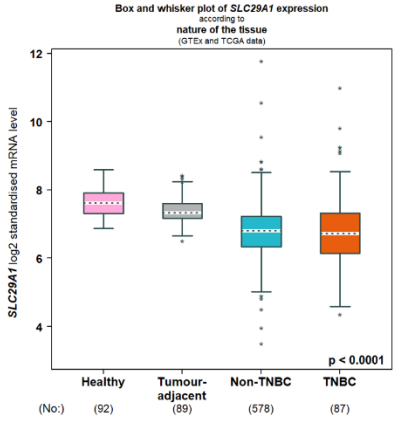

Box and whisker plot of SCRN1 expression
according to
nature of the tissue
(GTEx and TCGA data)
SCRN1 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

C

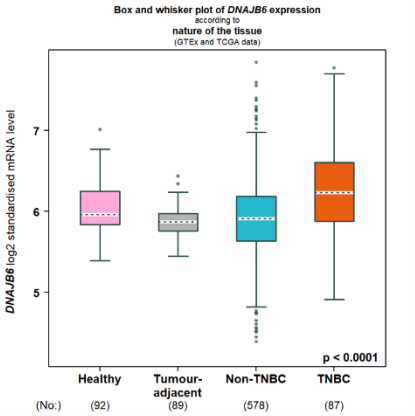

Box and whisker plot of KHDRBS3 expression
according to
nature of the tissue
(GTEx and TCGA data)
KHDRBS3 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

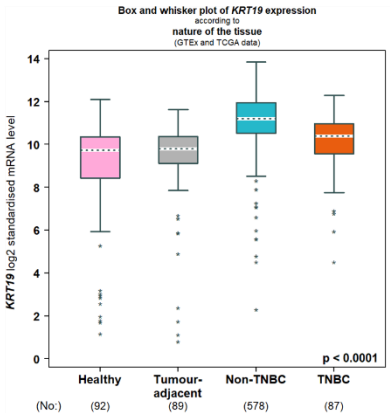

Box and whisker plot of ENO2 expression
according to
nature of the tissue
(GTEx and TCGA data)
ENO2 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

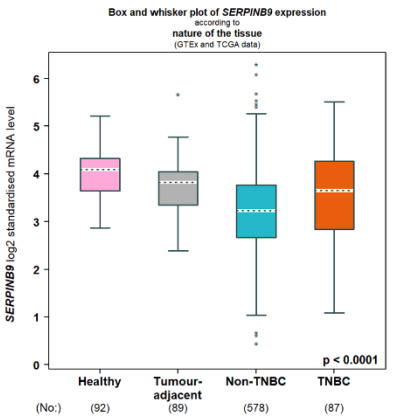

Box and whisker plot of TP53 expression
according to
nature of the tissue
(GTEx and TCGA data)
TP53 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

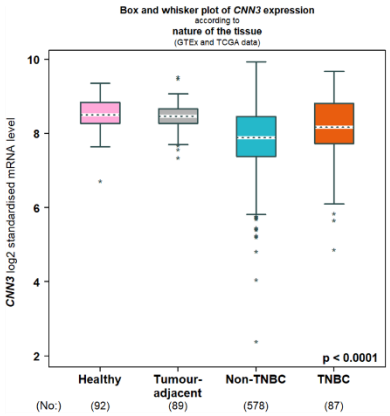

Box and whisker plot of SLC29A1 expression
according to
nature of the tissue
(GTEx and TCGA data)
SLC29A1 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

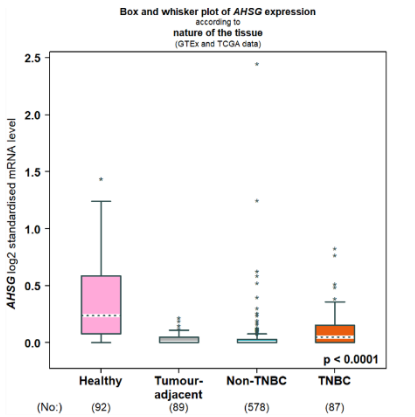
D

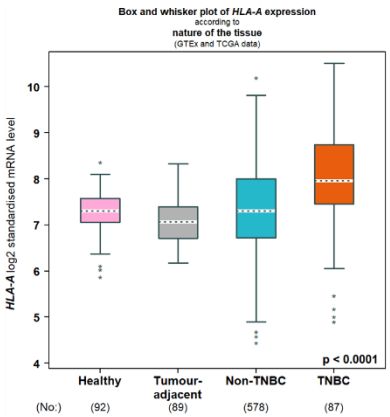

Box and whisker plot of DNAJB6 expression
according to
nature of the tissue
(GTEx and TCGA data)
DNAJB6 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

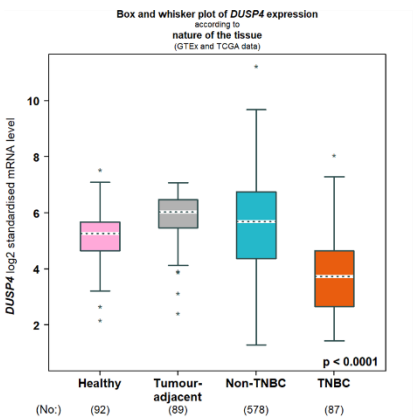

Box and whisker plot of KRT19 expression
according to
nature of the tissue
(GTEx and TCGA data)
KRT19 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

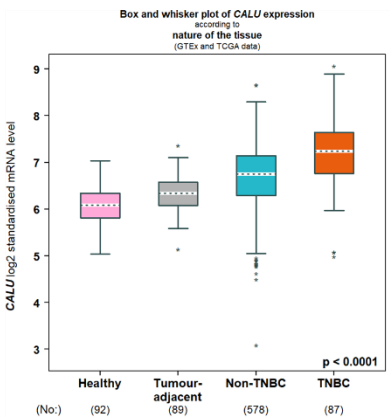

Box and whisker plot of SERPINB9 expression
according to
nature of the tissue
(GTEx and TCGA data)
SERPINB9 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)


Box and whisker plot of CNN3 expression
according to
nature of the tissue
(GTEx and TCGA data)
CNN3 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)


Box and whisker plot of AHSG expression
according to
nature of the tissue
(GTEx and TCGA data)
AHSG log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)


Box and whisker plot of HLA-A expression
according to
nature of the tissue
(GTEx and TCGA data)
HLA-A log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)


Box and whisker plot of DUSP4 expression
according to
nature of the tissue
(GTEx and TCGA data)
DUSP4 log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)


Box and whisker plot of CALU expression
according to
nature of the tissue
(GTEx and TCGA data)
CALU log2 standardised mRNA level
p < 0.0001
Healthy
Tumour-adjacent
Non-TNBC
TNBC
(No:) (92) (89) (578) (87)

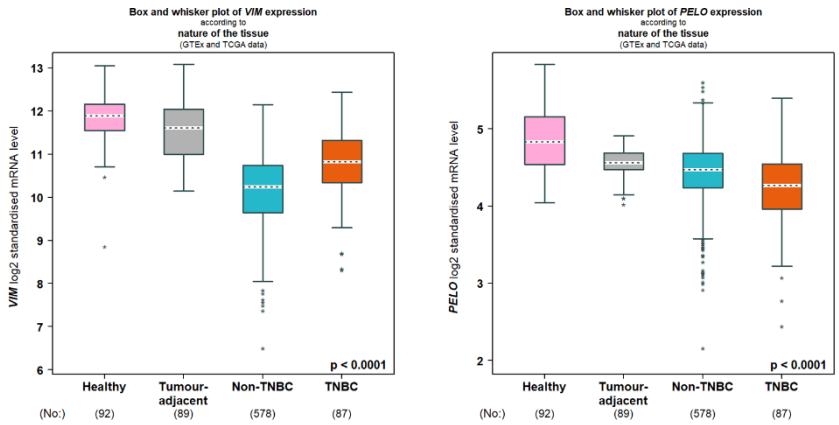


Figure 5.16 mRNA expression levels for NSAF and DESeq2 identified proteins in healthy, tumour adjacent, non-TNBC and TNBC tissue

The expression level of proteins identified to have significantly higher exclusive spectral counts in MDA-MB-231 gel slices compared to MCF10A gel slices from NSAF (Figure 5.14) and DESeq2 (Figure 5.15) analyses were assessed. Standardised mRNA expression levels were compared in healthy (n=92, GTEx), tumour-adjacent (n=89, TCGA), non-TNBC (n=578, TCGA) and TNBC (n=87, TCGA) transcriptomic data from the Genotype-Tissue Expression (GTEx) project (Lonsdale et al., 2013) and The Cancer Genome Atlas (TCGA) Network (Koboldt et al., 2012) using BC-GenExMiner (Jézéquel et al., 2021). Gene expression comparisons were conducted between clinical groups by Welch's t-tests. Multiple pairwise comparisons between each clinical group were performed by Dunnett-Tukey-Kramer's test if Welch p-values were significant ($p < 0.0001$ for all genes, bottom right).

5.2.10 Prognostic values of TPX2, WRNIP1, KHDRBS3, CALU, DNAJB6, KRT19 and HLA-A in TNBC

In addition to mRNA expression levels, the prognostic value of each of the identified autoantibody antigens was investigated to determine whether the presence of TPX2, WRNIP1, KHDRBS3, CALU, DNAJB6, KRT19 or HLA-A autoantibodies may be used as prognostic markers of TNBC. To investigate whether the expression of the genes encoding for these proteins was correlated with TNBC patient prognosis, survival analyses were conducted using the Kaplan-Meier Plotter (https://kmplot.com/analysis) with TNBC data from the TCGA patient cohort (ER negative, PGR negative and HER2 negative breast cancer patients).

Survival analysis for each gene is shown in Figure 5.17. The Kaplan-Meier plots

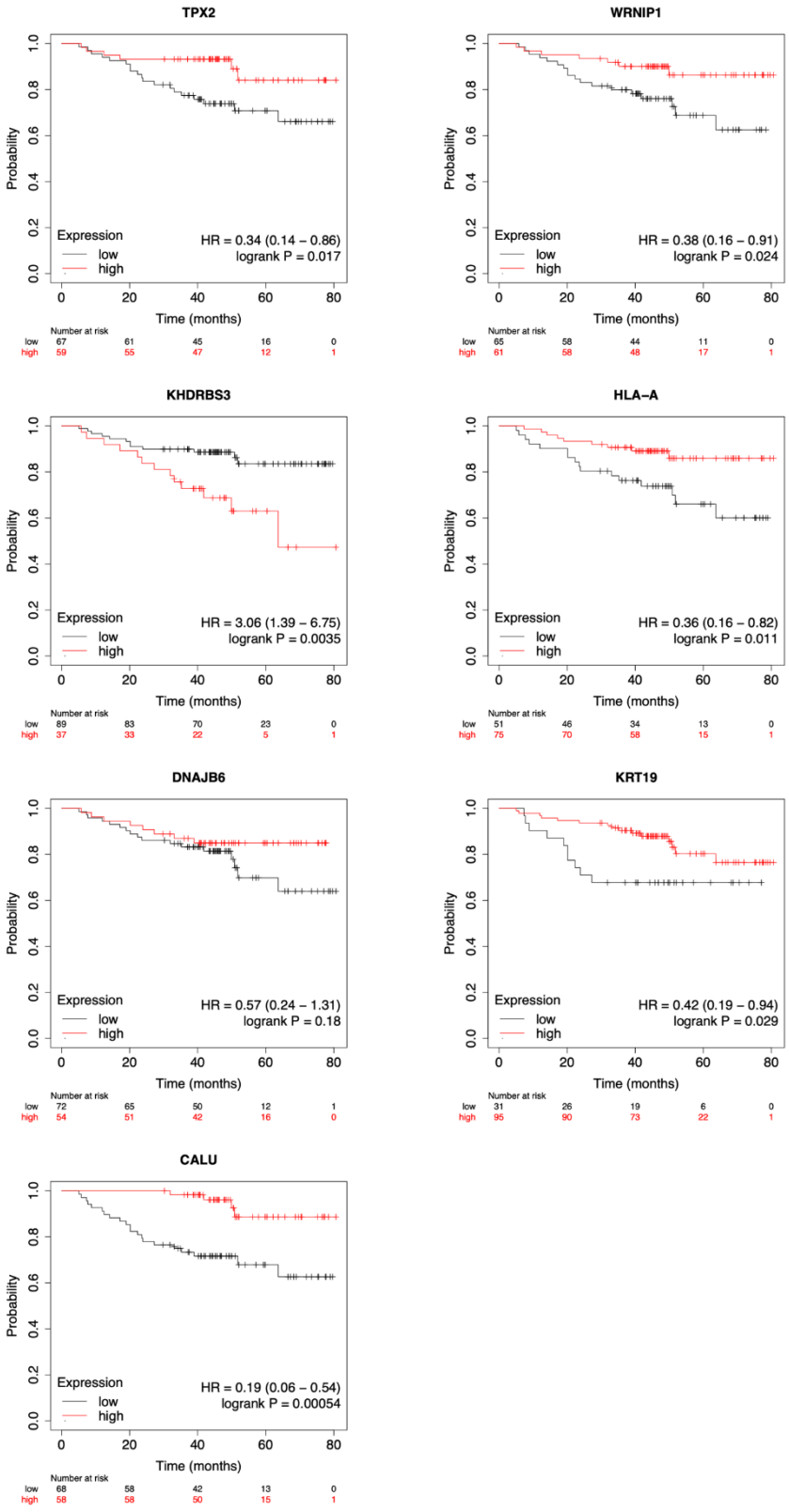


Figure 5.17 Kaplan-Meier survival plots for potential TNBC autoantibody antigens

The prognostic value of TPX2, WRNIP1, KHDRBS3, HLA-A, DNAJB6, KRT19 and CALU in ER negative, PGR negative and HER2 negative breast cancers (n = 126). Probability = overall survival. Plots produced using Kaplan-Meier plotter (Lánczky and Győrffy, 2021).

revealed that CALU expression was associated with better overall survival in TNBC patients (HR = 0.19 (CI = 0.06 – 0.54); logrank P = 0.00054). In contrast, high expression of KHDRBS3 in TNBC patients was associated with lower overall survival (HR = 3.06 (CI = 1.39 – 6.75); logrank P = 0.0032). Therefore, the presence of anti-KHDRBS3 autoantibodies in patient serum samples may be associated with a worse prognosis.

5.2.11 Validation of control and patient IgG binding to TPX2 and KHDRBS3 overexpression lysate

Following the identification of seven potential tumour associated antigens (TPX2, WRNIP1, KHDRBS3, CALU, DNAJB6, KRT19 and HLA-A) from LC-MS/MS and subsequent bioinformatic analysis, two of the proteins were selected from the candidate list to assess whether antibody binding could be confirmed to the antigens. The two proteins selected were TPX2 and KHDRBS3. TPX2 was selected as the protein was identified from both differential expression analyses (Figure 5.14 and Figure 5.15) and overexpression of TPX2 has been correlated with genomic instability and high grade disease in TNBC (Jiang et al., 2019, Matson et al., 2021). KHDRBS3 was also selected for further analysis as the protein has been linked with TNBC stemness and KHDRBS3 was correlated with worse overall survival in the survival analysis shown in Figure 5.17 (Matsumoto et al., 2018).

To determine whether anti-TPX2 or anti-KHDRBS3 autoantibodies were present in the TNBC patient sera used in this study, HEK293T TPX2 and KHDRBS3 transient overexpression lysates were probed with control and TNBC patient-derived IgG. Prior to this, the binding of control and patient-derived IgG to HEK293T protein extracted from HEK293T cells with no transient overexpression was assessed to identify immunoreactive bands that are not TPX2 and KHDRBS3 specific (Figure 5.18). The blot revealed that the control IgG bound to a HEK293T protein at around 50kDa in a similar manner to control IgG blots with MCF10A and MDA-MB-231 extracted protein samples (Figure 5.7). In contrast, probing HEK293T extracted proteins with patient IgG led to an intense band being observed below 80kDa, which was also observed with the patient IgG in Figure 5.7. Patient IgG binding also led to a faint band being observed at around 50kDa.

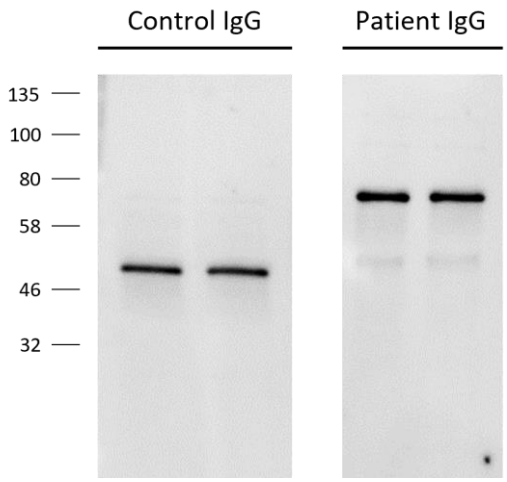


Figure 5.18 Control and Patient IgG binding to HEK293T cell lysate

Proteins extracted from HEK293T cell lysate were separated by SDS-PAGE and probed with control and patient IgG. Protein samples were loaded in duplicate.

Following the identification of non-specific bands, the interaction of control and patient-derived IgG with TPX2 overexpression lysate was investigated (Figure 5.19). Probing the TPX2 overexpression lysate with control IgG (Figure 5.19, Control IgG) resulted in a single band being observed at the same molecular weight as the band observed when TPX2 was not overexpressed in HEK293T cells (Figure 5.18, Control IgG).

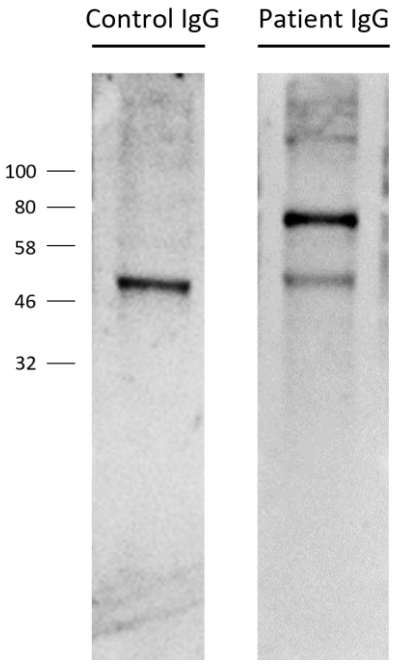


Figure 5.19 Control and patient IgG binding to TPX2 overexpression HEK293T cell lysate

TPX2 overexpression cell lysate was separated by SDS-PAGE and probed with control and patient IgG. TPX2 molecular weight = 85.7 kDa.

In addition, the bands observed with the patient IgG (Figure 5.19, Patient IgG) were also observed in the HEK293T samples with no TPX2 overexpression (Figure 5.18, Patient IgG).

For KHDRBS3, no additional bands were observed when control IgG was incubated with KHDRBS3 overexpression lysate, with a single band being observed at approximately 50kDa (Figure 5.20, Control IgG). In contrast, multiple bands were observed in the patient IgG blot with KHDRBS3 overexpression lysate (Figure 5.20, Patient IgG) compared with HEK293T cell lysate with no KHDRBS3 expression (Figure 5.18, Patient IgG). Firstly, the band present at approximately 50kDa was more intense than the band which was observed at the same molecular weight in both the TPX2 (Figure 5.19) and the HEK293T (Figure 5.18) patient blots. A faint band was also observed at a slightly higher molecular weight which is not observed in the other control or patient IgG blots. Additionally, two clear but faint bands are observed at lower molecular weights, with one band observed at approximately 34kDa and the other at around 40kDa (Figure 5.20, Patient IgG). The 34kDa and 40kDa bands were not present in Figure 5.18 or Figure 5.19, indicating that KHDRBS3 overexpression resulted in KHDRBS3 specific patient IgG binding being observed.

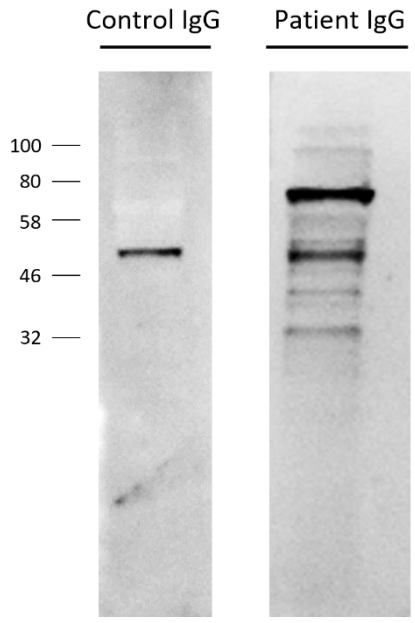


Figure 5.20 Control and patient IgG binding to proteins KHDRBS3 overexpression HEK293T cell lysate

KHDRBS3 overexpression cell lysate was separated by SDS-PAGE and probed with control and patient IgG. KHDRBS3 molecular weight = 38.7 kDa.

5.3 Discussion

The presence of serum autoantibodies, which are produced in response to circulating tumour antigens, have been shown in a number of studies to be indicative of cancers such as BCa (Nesterova et al., 2006, Bassaro et al., 2017, Kathrikolly et al., 2022). In particular, the increase of IgG autoantibodies present in serum from women with BCa is being investigated as a potential biomarker for the early detection of the disease (Fernández Madrid et al., 2015, Stanton et al., 2021). Therefore, in this study, IgG antibodies were purified from 40 TNBC patient serum samples to determine whether autoantibodies were present for use in the early detection of TNBC, a molecular subtype which is often not detected in early-stage disease in women with dense breast tissue or who have not reached the NHS mammographic screening programme age. As there was only a small quantity of serum available for each patient, the patient samples were pooled prior to IgG purification.

Clinical data was provided with each TNBC patient sample and purified IgG were pooled according to the clinical information of blood serum samples, including the age of each woman at the time of sample collection and whether the individual experienced recurrence or had received prior treatment (Table 5.1). The first clinical criteria studied was age, as TNBC mainly affects women who are premenopausal, therefore the presence of autoantibodies in women younger than 50 may be used as an early diagnostic tool for individuals in this age group. Additionally, studies have also demonstrated that treatment and BCa stage can impact autoantibody levels, with decreased levels of autoantibodies being detected in patients who had received combinatorial treatment regimens including chemotherapy and an increased number of BCSC derived autoantibodies being found in later stage BCa, respectively (Evans et al., 2018, Stanton et al., 2018).

To identify non-specific IgG binding and cancer autoantibody binding events, a human isotype control IgG was employed as a negative control in this study. The human isotype control IgG used in this work was a commercially available IgG so it is important to note that a different IgG purification method may have been used compared to the method applied here for TNBC patient IgG purification. In addition, the number of samples included in the human sera pool was not provided by the company and the

age and gender of participants was also unknown. However, the human IgG was purified from serum samples like the TNBC patient derived IgG and was therefore deemed a suitable negative control for distinguishing between non-specific background binding events and autoantibody specific signals.

No significant differences in binding between the patient IgG pools were observed when binding events to TNBC derived MDA-MB-231 cells were analysed by flow cytometry (Figure 5.1). Categorising patients based on their age did lead to pool 2 (0.6%, patients < 50 years old) resulting in more IgG binding events above the control gate than pool 1 (0.31%, patients ≤ 50 years old), but this difference was not significant. The increase may have also been due to a greater number of samples being included in pool 2 (n = 25) than pool 1 (n = 15). This finding is supported by a study by Chapman et al., who found no significant differences in the sensitivity of autoantibody detection when patients with primary BCa or DCIS were categorised based on age, tumour grade or lymph node status (Chapman et al., 2007). In addition, BCa autoantibodies have been found in women which are younger than and older than 50 years old (Lourenco et al., 2017, Henderson et al., 2019). There were also no differences in binding observed between pools based on recurrence or treatment status. This indicated that recurrence or treatment status did not influence autoantibody production, which correlates with research suggesting that cancer autoantibodies are produced in early tumourigenesis (Jiang et al., 2021, Stanton et al., 2021). It is interesting to note that pools 3, 4, 5 and 6 resulted in less autoantibody binding than the global pool and pools 1 and 2, however differences in IgG binding were not significant.

Although there were no differences in IgG binding observed between the patient IgG pools, the global pool and pool 2 resulted in a significant increase in binding events being observed above the control IgG gate, inferring that an antibody present in TNBC patient sera was binding to an MDA-MB-231 antigen. The subpopulation of cells which the global IgG pool bound to equated to 0.3-0.5% of the total cell population, inferring that the corresponding TNBC associated antigens may only be expressed by a small number of cells. Targeting of such a small population of TNBC cells may influence the efficacy of an antigen-directed treatment, however, this population of cells is not bound by control human IgG, which indicates that any treatment directed against the

unique subpopulation of TNBC cells would discriminate between normal and cancerous tissue.

Interestingly, the percentage of patient IgG bound cells is also consistent with estimates of the size of the BCSC population in breast tumours, which range from 0.1-5% (De Angelis et al., 2019). TNBC is characterised by CSC enrichment, therefore a number of cancer autoantibodies may be produced in response to cancer stem cell associated antigens, as suggested to by a clinical trial conducted by Stanton *et al.,* who identified the presence of eight autoantibodies produced in response to BCSC antigens (Stanton et al., 2018). Therefore, the binding of patient IgG to MDA-MB-231 expressing BCSC markers was investigated (Figure 5.6). FACS analysis revealed that a large percentage of MDA-MB-231 cells are $CD44^+/CD24^-$, which is considerably higher than expected. However, other studies have also reported a high percentage of $CD44^+/CD24^-$ cells in MDA-MB-231 cell lines, with $CD44^+/CD24^-$ being associated with cell proliferation and tumourigenesis in basal-like BCas and subsequent poor survival in TNBC (Yin and Glass, 2011, Li et al., 2017). Patient IgG was found to bind to 2-3% of the $CD44^+/CD24^-$ subset of cells, with the percentage of TNBC IgG bound $CD44^+/CD24^-$ cells increasing to 6-9% when gating for the brightest $CD44^+$ cells (Figure 5.6G). This infers that there may be enrichment of patient IgG binding in the $CD44^+/CD24^-$ subpopulation compared to the whole cell population, indicating that a cancer autoantibody present in TNBC patient derived IgG may specifically target an antigen expressed by $CD44^+/CD24^-$ cells. It is also important to note that the patient IgG only bound to a small minority of cells, which may be a result of only a small number of $CD44^+/CD24^-$ MDA-MB-231 cells expressing the corresponding antigen, or due to there being only a small number of TNBC associated autoantibodies in the patient IgG pool.

Exposure of both patient-derived IgG and control IgG to the TNBC cell line MDA-MB-231 and non-tumourigenic MCF10A cells revealed that only the patient-derived IgG appeared to bind to a subpopulation of MDA-MB-231 cells (Figure 5.4). The patient IgG and control IgG interacted in a similar manner with MCF10A cells with no significant difference in binding events being observed, implying that no cancer specific IgG antibodies are interacting with non-tumourigenic epitopes (Figure 5.4A). Regarding MDA-MB-231 binding, the patient IgG resulted in more antibody binding events recorded by flow cytometry than with control IgG (Figure 5.4B). This implies

that autoantibodies may be produced as part of the humoral immune response to TNBC specific antigens. Preliminary work conducted identified that more patient IgG antibody binding events were seen when ice cold acetone was used as a fixative agent rather than PFA. As organic solvents fix and permeabilise mammalian cells, intracellular antigens are able to be detected by flow cytometry (Carbonari et al., 2008). Contrastingly, formaldehyde based solvents such as PFA do not permeabilise cells, therefore indicating that potential autoantibodies present in the patient IgG pool are binding to an intracellular antigen and not cell surface antigens (Hobro and Smith, 2017).

Immunofluorescence studies were then employed to confirm the binding of control and patient IgG to MCF10A and MDA-MB-231 cells and to determine the cellular localisation of patient IgG antibody epitopes in MDA-MB-231 cells. Fluorescence microscopy did confirm the binding of control and patient IgG to both cell lines, as fluorescently labelled anti-human IgG was observed in control and patient IgG wells (Figure 5.5). However, the immunocytochemistry approach was not able to confirm the differences in antibody binding observed by flow cytometry, as differences in control and patient IgG binding to MDA-MB-231 cells could not be visualised (Figure 5.5B). In addition, no clear differences in the level of fluorescence were seen when MCF10A cells were incubated with either the control or patient IgG (Figure 5.5B). As the patient IgG appeared to bind to a small MDA-MB-231 population from flow cytometric analysis, the small number of cells seeded on the coverslip for immunofluorescence may not have covered this population. Furthermore, most of the binding events were below the control IgG gate when patient IgG was incubated with MDA-MB-231 cells (Figure 5.4B), so the small number of patient IgG specific binding events may have been difficult to observe by immunocytochemistry. It also appears that the control IgG bound to nuclear antigens in both MCF10A and MDA-MB-231 cells, with more fluorescent dots observed in MCF10A nuclei than MDA-MB-231 nuclei. However, the staining observed may be an artifact of the fixation method used.

Following the detection of potential TNBC autoantibodies, western blotting and mass spectrometry were employed to identify TNBC specific epitopes by serological proteome analysis (SERPA). Other technologies have also been applied for the discovery of tumour associated antibodies such as serological screening of cDNA

expression libraries (SEREX), phage display libraries and protein microarrays (Li et al., 2004, Lin et al., 2007, Anderson et al., 2011). In contrast to the other serological screening methods, SERPA allows for the detection of post translationally modified antigens and protein isoforms, as well as SERPA not requiring the time-consuming construction of cDNA libraries. There are also some limitations to the SERPA, including difficulties in detecting proteins of low abundance, high molecular weight or membrane-associated proteins (Rauf et al., 2020). However, other studies have successfully identified autoantibodies present in human sera (Nagele et al., 2013, Bazhin et al., 2004, Klade et al., 2001). For instance, Hamrita *et al.,* and Desmetz *et al.,* identified different panels of BCa serum autoantibodies by 2D electrophoresis coupled with western blotting from both cell line and BCa tissue derived proteins, respectively *(*Hamrita et al., 2008*,* Desmetz et al., 2009*)*. Therefore, this study hypothesised that the TNBC patient derived IgG may bind to MDA-MB-231 isolated proteins and lead to immunoreactive bands that may be detected by western blot. Moreover, patient IgG protein bands that are not observed in non-tumourigenic cell lysate, or when MDA-MB-231 proteins are probed with control IgG, may indicate the presence of TNBC specific autoantigens.

1D electrophoresis of MCF10A and MDA-MB-231 whole cell lysate coupled with western blotting with control and patient IgG found that TNBC antibodies led to different immunoreactive bands being observed (Figure 5.7). The patient IgG resulted in a more intense band observed at around 50kDa in MDA-MB-231 cell lysate lanes than MCF10A cell lysate lanes. The increased intensity of this band suggests that the antigen may be aberrantly expressed in the TNBC cell line compared to MCF10A cells, or that the antigen is post translationally modified in TNBC, leading to the production of antibodies against the modified protein (Volkov et al., 2021). A band at 50kDa was also observed in the control IgG blot with both MCF10A and MDA-MB-231 whole cell lysate, however the bands were not as intense as those observed when the patient IgG pool was used as the primary antibody with MDA-MB-231 lysate. A study by Bassaro *et al.,* also detected autoantibodies in healthy donor plasma, therefore quantifying autoantibody levels will be important in determining disease status (Bassaro et al., 2017). Two other bands of interest were also established, with an 80-100kDa band and approximately 40kDa band exclusively present in the MDA-MB-231 protein lanes of the patient IgG blot (Figure 5.7, indicated by top and bottom arrows).

Antibody binding observed exclusively in TNBC protein lanes again may be in response to the aberrant expression of a cancer antigen. Unlike the 50kDa protein, the antigens are only recognised by a TNBC patient antibody, indicating that the interacting autoantibodies may be able to detect TNBC specifically. The two bands were also present when BT-549 cell lysate was probed with patient IgG, suggesting that the neoantigens are not just expressed in one TNBC cell line (Figure 5.8B). Conversely, the 80-110kDa antigen does not appear to be present in CAL-51 cells, highlighting how intertumoural heterogeneity is observed in TNBC (Chiu et al., 2018).

Further immunoblotting with fractionated MCF10A and MDA-MB-231 protein found that the patient IgG immunoreactive 80-110kDa band arose from a cancer autoantibody binding to a nuclear antigen (Figure 5.9, top arrow). This correlates with research which has shown that the release of nuclear antigens during malignant disease progression subsequently leads to the production of anti-nuclear antibodies in BCa patients, including c-MYC, BRCA1 and BRCA2 autoantibodies (Mohammed and Abdelhafiz, 2015, Nisihara et al., 2018, Chapman et al., 2007). The lower molecular weight 40kDa protein was also identified to be localised to the cytoplasm, as no band was observed in the nuclear fractions. In contrast, the approximately 50kDa protein was found to be enriched in both nuclear and cytoplasmic MDA-MB-231 fractions, indicating that the protein is present in both subcellular compartments (Figure 5.9, middle arrow). The separation of membrane and cytoplasmic protein also revealed that the potential autoantibody epitopes were not cell surface proteins in Figure 5.10, corresponding with the increase in patient IgG binding events when cells were permeabilised with acetone for flow cytometry experiments.

The characterisation of potential TNBC autoantibody binding antigens by label-free LC-MS/MS led to the identification of 2346 proteins across all gel slices excised based on the blot areas highlighted in Figure 5.11. The overexpression of antigens has been shown to lead to autoantibody production in a number of studies, including a study by *Hwang et al.,* where high expression of splicing factor 3b subunit 1, SF3B1, was identified in hepatocellular carcinoma (HCC) tissue samples and anti-SF3B1 autoantibodies were detected in HCC patients vs healthy subjects by ELISA with 74% sensitivity and 92% specificity (Hwang et al., 2018). As autoantibodies are produced in response to the overexpression of antigens in malignant disease progression, we

speculated that the potential TNBC antigens have higher expression levels in MDA-MB-231 cells than MCF10A cells and that the differences in protein expression may be detected by mass spectrometry. LC-MS/MS spectral counts were therefore used to estimate relative protein abundances in each sample, as spectral counting can be used as a sensitive, semi-quantitative method to define differential protein expression when combined with statistical analysis (Old et al., 2005, Fu et al., 2008).

In this study, two independent statistical analyses were applied, NSAF with multiple t-tests (Figure 5.14) and DESeq2 (Figure 5.15). Other methods have also been implemented to detect differential expression from spectral count data, including spectral index (spI), normalised spectral index, QSpec and NSAF coupled with power law global error modelling (NSAF-PLGEM) (Griffin et al., 2010, Fu et al., 2008, Choi et al., 2008, Pavelka et al., 2008). However, NSAF based methods have been shown to produce more reproducible results than spectral index calculations (McIlwain et al., 2012). When coupled with t-tests, NSAF can also correctly identify differential expression whilst controlling for the number of false positives. In addition, Langley and Mayr found that DESeq2 and Qspec performed well when compared to the other analyses mentioned, resulting in the detection of true positives whilst controlling for false negatives (Langley and Mayr, 2015). Thus, differential protein expression in MCF10A cells compared to MDA-MB-231 cells was investigated by both methods to increase the likelihood of false positive proteins being identified, with FDR corrections applied in both statistical analyses to control for type I errors.

NSAF and DESeq2 analysis resulted in 18 proteins being identified as overexpressed in MDA-MB-231 cells compared to MCF10A cells, with more proteins being identified by NSAF and t-test than DESeq2 (Figure 5.14 and Figure 5.15). TPX2, SATB2 and DNAJB6 were found to have increased protein levels in MDA-MB-231 cell lysate compared to MCF10A cell lysate from both statistical analysis methods, inferring that these proteins have a higher likelihood of being true positives. In contrast, SCRN1 and KHDRBS3 were identified to be differentially expressed in gel slice 2 (Figure 5.14B) and ENO2, SLC29A1 and TP53 (Figure 5.14C) were identified as differentially expressed in gel slice 3 by NSAF and t-test, whereas no proteins were found to have significantly higher protein expression levels in MDA-MB-231 samples when the count data was analysed by DESeq2 (Figure 5.15B and Figure 5.15C). This highlights how the two methods differ

statistically and how NSAF based methods such as NSAF-PLGEM may lead to the identification of a larger number of true and false positives (Langley and Mayr, 2015). To try and limit the number of false positive results associated with NSAF calculations, a 1% FDR was implemented in the NSAF and t-test analysis instead of the 5% FDR applied in the DESeq2 analysis.

As potential autoantibodies present in the blood serum samples were produced in response to triple negative breast tumours in patients, the expression of the tumour associated antigen candidates identified from LC-MS/MS proteomic analysis with MDA-MB-231 protein was assessed in patient datasets using bc-GenExMiner (Figure 5.16). Although relatively high levels of correlation have been demonstrated between transcriptional and protein expression profiles in cell lines compared to tumour samples, some studies have demonstrated that cell lines may not be fully representative models of tumours (Domcke et al., 2013, Vincent et al., 2015, Jarnuczak et al., 2021). Therefore, protein expression profiles in TCGA TNBC datasets were used to confirm the upregulation of proteins observed in MDA-MB-231 cells. To identify differential expressed proteins, the mRNA expression levels of each potential tumour associated antigen was compared to the mRNA expression level of a given protein in healthy tissue.

The potential TNBC antigens identified from LC-MS/MS and patient dataset overexpression analyses were TPX2, WRNIP1, KHDRBS3, CALU, DNAJB6, KRT19 and HLA-A. To our knowledge, autoantibodies produced in response to TPX2, WRNIP1, KHDRBS3, DNAJB6 and HLA-A overexpression have not yet been identified in studies focusing on the presence of autoantibodies in BCa patient sera (Rauf et al., 2020). However, Katayama *et al.,* identified the presence of anti-CALU and anti-KRT19 autoantibodies in prediagnostic TNBC plasma samples when the samples were individually hybridised with MDA-MB-231 or C3(1)-T mouse model derived protein arrays, respectively (Katayama et al., 2015).

Studies have also demonstrated the role of the TNBC antigen candidates in tumourigenesis. TPX2 regulates the formation of the mitotic spindle through activation of Aurora-A and is correlated with genomic instability, high grade disease and poor clinical outcome in TNBC (Jiang et al., 2019, Matson et al., 2021).

Overexpression of the protein has been identified in TNBC and has also been associated with *BRCA1/2* mutated cancers and sensitivity of TNBC tumours to PARP inhibitors such as Olaparib (Mosler et al., 2022). WRNIP1 also plays a role in with genomic integrity, however loss of WRNIP1 is associated with genomic instability through R-loop accumulation (Leuzzi et al., 2016, Marabitti et al., 2020). Interestingly, upregulation of the KH domain RNA binding protein, KHDRBS3, has been linked to BCa stemness in basal-like BCas through alteration of CD44 isoform expression (Matsumoto et al., 2018). In addition, CALU has been shown to be a drive of EMT and a subsequent marker of poor survival in glioma patients, as well as being a key protein in lung cancer cell invasiveness (Nagano et al., 2015, Yang et al., 2021). Conversely, Mitra *et al.,* demonstrated that DNAJB6 activity suppresses tumour progression, leading to the reversal of a mesenchymal phenotype. Research by Mitra *et al.,* also reported depleted DNAJB6 in advanced grade IDC, contradicting the increased mRNA expression levels observed in the TCGA dataset presented in Figure 5.17D (Mitra et al., 2008, Mitra et al., 2010). Keratin 19 (KRT19) expression has also been linked to the suppression of invasive BCa through regulation of the NOTCH signalling pathway (Saha et al., 2017). Lastly, HLA-A expression is associated with a more favourable prognosis in basal-like BCas (Noblejas-López et al., 2019, Kaneko et al., 2011).

In addition to the potential diagnostic role of autoantibodies in BCa, the presence of cancer autoantibodies has also been correlated with BCa patient prognosis. For example, *Gu et al.,* found that high serum anti-HSPA4 IgG levels, produced in response to HSPA4 overexpression, was correlated with poor prognosis in IDC patients (Gu et al., 2019). To investigate whether the autoantibodies identified in this study are prognostic markers for TNBC, Kaplan-Meier curves of overall survival and the expression of each tumour associated antigen gene were produced. The survival analysis shown in Figure 5.17 revealed that only KHDRBS3 expression was correlated with worse overall survival in TNBC patients. This relates to studies showing that the overexpression of WRNIP1, DNAJB6, KRT19 or HLA-A is associated with better patient prognosis, as the increased expression of these genes did not lead to worse overall survival in ER^-, PGR^- and $HER2^-$ BCa patients. The prognostic analysis also found that CALU expression was correlated with a better overall survival in ER^-, PGR^- and $HER2^-$ BCa patients, which contradicts findings in glioma patients (Nagano et al., 2015, Yang et al., 2021).

Based on the LC-MS/MS proteomics analysis and following TNBC overexpression and survival analyses, the antigens selected for further study were TPX2 and KHDRBS3. As previously mentioned, these proteins were selected as TPX2 was identified from both proteomics analyses methods (Figure 5.14 and Figure 5.15) and overexpression of TPX2 has been linked poor clinical outcome in TNBC, whereas KHDRBS3 was identified in the NSAF and t-test analysis (Figure 5.14) and was correlated with poor survival in TNBC patients (Figure 5.17). All the patient IgG reactive proteins identified from the serological proteome analysis could be leading to TNBC specific autoantibody production, therefore future work would involve the validation of all seven proteins.

To confirm whether anti-TPX2 and anti-KHDRBS3 autoantibodies are present in the TNBC patient-derived IgG pool, HEK293T lysate from cells over expressing TPX2 or KHDRBS3 was probed with control and patient IgG (Figure 5.19 and Figure 5.20). The overexpression western blots revealed that KHDRBS3 overexpression led to differential reactive bands being observed with the patient IgG, suggesting that ant-KHDRBS3 antibodies are produced in response to increased expression. The expected molecular weight of the RNA binding protein is 39kDa. In Figure 5.20, there is a faint band present at this molecular weight, suggesting that a patient-derived IgG antibody may be binding to KHDRBS3. However, KHDRBS3 was highlighted in the NSAF and t-test statistical analysis of proteins isolated from the band excised at 39-59kDa, as this band was extracted to encapsulate a potential antigen observed at around 50kDa in MDA-MB-231 nuclear fractions not 39kDa (Figure 5.11). However, the band which is observed at approximately 50kDa in the overexpression blots did appear to be more intense in the KHDRBS3 patient blot (Figure 5.20) when compared to the HEK293T patient IgG (Figure 5.18) and TPX2 patient IgG blots (Figure 5.19), with a fainter band observed at a marginally higher molecular weight. This correlates with research by Ehrmann *et al.,* which identified the major isoform of T-STAR (KHDRBS3) at 55kDa and a minor isoform of the protein at a slightly higher kDa by western blot (Ehrmann et al., 2013). Additional bands were also observed in the KHDRBS3 patient IgG blot, including a 34kDa band (Figure 5.20). As KHDRBS3 is involved in the regulation of alternative mRNA splicing, the generation of splice variants via KHDRBS3 activity may result in TNBC cancer-specific immunogenic neoantigens being produced, thus leading to the production of autoantibodies against splice variants (Alsarkhi and Weber, 2018, Bessa

et al., 2020) . Therefore, the patient IgG reactive proteins in Figure 5.20 may be due to the generation of TNBC protein isoforms through increased KHDRBS3 activity.

As previously mentioned, KHDRBS3 has been linked to cancer stemness in TNBC through the modulation of CD44 splicing (Matsumoto et al., 2018). This correlates with the data presented in Figure 5.6, where patient IgG binding events were observed in the $CD44^{++}$ MDA-MB-231 cell population (Figure 5.6G), indicating that KHDRBS3 is expressed by BCSCs. In addition, overexpression of the protein has been correlated with alternative splicing, the acquisition of stem cell-like properties and poor patient prognosis in other malignancies, including colorectal and ovarian cancer (Ukai et al., 2021, Gov, 2020, Gao et al., 2020). The survival analysis conducted in this study found that KHDRBS3 expression is also linked with decreased overall survival in HER2 negative and hormone receptor negative BCa patients (Figure 5.17). Therefore, the results in this chapter indicate the presence of anti-KHDRBS3 antibodies in TNBC patient serum samples, thus highlighting the clinical potential of anti-KHDRBS3 as a diagnostic and prognostic marker of TNBC.

Chapter 6: Discussion

This study aimed to identify tumour associated antigens, their corresponding autoantibodies and targeting small peptides for the development of early diagnostic methods and targeted therapeutics for TNBC. TNBC currently has the worse prognosis among the molecular subtypes of BCa, with early diagnosis resulting in improved patient outcome (Chen et al., 2020). However, TNBC can be difficult to detect in young women through imaging methods, which highlights the need for complimentary diagnostic techniques. A patient's prognosis is also influenced by the lack of target therapies available for TNBC treatment compared to the other BCa subtypes. The development of targeted therapeutics for the treatment of TNBC would therefore address a currently unmet clinical need and improve the treatment landscape for this aggressive subtype of BCa (Bianchini et al., 2022).

Regarding TNBC early detection, we hypothesised that the identification of TNBC-specific autoantibodies in patient blood sera could be used for the early detection of TNBC, as autoantibodies are produced in early carcinogenesis. Blood-based liquid biopsies are gaining interest in BCa diagnostics as they are minimally invasive and can provide information on tumour heterogeneity and disease stage, with tumour associated autoantibodies panels showing encouraging clinical results (Henderson et al., 2019, Setayesh et al., 2022). The work conducted in this project investigated whether cancer autoantibodies were present in IgG antibody fractions purified from TNBC patient blood serum sample for the early detection of the disease. Autoantibodies were present in TNBC patient purified IgG fractions compared to control fractions (Figure 5.3), indicating that autoantibodies are produced in response to TNBC and be utilised as discriminatory protein biomarkers of the disease. Furthermore, the patient purified IgG fraction bound to permeabilised $CD44^{+}/CD24^{-}$ MDA-MB-231 cells, suggesting that autoantibodies may be produced in response to TNBC BCSC intracellular antigens (Figure 5.6). As mentioned, BCSCs are a therapy resistant subpopulation of cells associated with tumour progression and metastasis, therefore the detection of BCSC autoantibodies in serum samples may be indicative of aggressive disease and be used to inform treatment decisions (Stanton et al., 2018).
Autoantibody recognition of TNBC antigens led to the identification of a novel panel of potential tumour associated antigens using a SERPA approach. Validation of the

most promising candidates confirmed the presence of anti-KHDRBS3 in TNBC patient sera and not in sera extracted from healthy individuals, confirming that the approach taken in this study was able to successfully identify TNBC specific autoantibodies (Figure 5.20). KHDRBS3 overexpression resulted in decreased overall survival in TNBC patients, therefore indicating that KHDRBS3 may be a prognostic marker of TNBC. Additionally, increased expression of KHDRBS3 was only observed in TNBC tumours, which may indicate that KHDRBS3 autoantibodies can be used to discriminate between TNBC and non-TNBC breast tumours, as well as between TNBC versus non-cancerous cases (Figure 5.16). Thus, KHDRBS3 warrants further investigation as a potential diagnostic and prognostic marker for TNBC.

As well as cancer autoantibodies acting as diagnostic markers of TNBC, their corresponding antigens may also be utilised as targets for potential TNBC therapeutics. Currently, there are a limited number of targeted therapies available for the treatment of the disease, therefore the identification of molecular targets of TNBC are needed to improve treatment options available for patients. In this study, KHDRBS3 has been identified as a TNBC neoantigen. The involvement of KHDRBS3 in cancer stemness pathways indicates that targeting the splicing factor may hinder cancer stemness in TNBC and target the chemotherapy and radiotherapy resistant cell population which is thought to drive tumour invasion and metastasis (Matsumoto et al., 2018). Therefore, this work highlights how KHDRBS3 may be a novel molecular target for TNBC treatments and how further investigations into the role of KHDRBS3 in TNBC through functional studies are required.

Targeting moieties, such as antibodies or peptides, are also being utilised in novel cancer therapeutic approaches to deliver cytotoxic ligands to tumour cells in a site-specific manner, such as via ADCs and PDCs, respectively. When compared to antibodies, small peptides display low immunogenicity and are easily synthesised and modifiable. The versatility of small peptides also allows for a wide variety of therapeutic ligands to be conjugated or complexed to the targeting moiety, such as different chemotherapeutic agents and nanocarrier systems (Samec et al., 2022b). Therefore, a tumour targeting peptide could be applied to a range of therapeutic approaches to combat TNBC.

To identify a TNBC targeting peptide for peptide-based therapeutics, we conducted a non-bias high-throughput screen of small peptides against TNBC cells by NGPD. This approach allowed for cancer targeting ligands to be selected without a known target, as there is currently a lack of known TNBC molecular targets. The potential TNBC targeting candidates were then validated by *in vitro* studies, which confirmed that TNBC targeting peptides had been identified. Therefore, this study demonstrated how NGPD can be applied to successfully isolate and identify cancer cell targeting ligands. Two potential TNBC targeting ligands were prioritised from flow cytometric analysis, TLYRMVVTS and LMQDMALRL. To our knowledge, the peptides have not been identified in other studies and may therefore be novel targeting agents for TNBC therapeutics. The *in vitro* validation studies found that TLYRMVVTS and LMQDMALRL delivered more mRFP into TNBC cells compared to non-tumourigenic cell lines, indicating that the targeting peptides may lead to tumour-specific delivery of cytotoxic agents and subsequently increase the cytotoxic payload delivered to a TNBC tumour through efficient uptake into cancerous cells. In addition, the selective targeting of TNBC cells may limit off-target toxicity and side effects in patients, which are often experienced with other systemic therapies (Hoppenz et al., 2020). The peptides were also able to efficiently deliver nanoparticles into TNBC cell lines, a molecular cargo which has utilised in cancer diagnostics and drug delivery studies (Estelrich et al., 2015, Avasthi et al., 2020). This further highlights the potential for each peptide to be incorporated into peptide-based drug delivery systems for TNBC targeted treatment, however further work is required to quantify MNP uptake.

A major limitation to this work is the use of TNBC cancer lines compared to primary tissue. The TNBC cell lines used in this study were BT-540, CAL-51 and MDA-MB-231. Regarding TNBC intrinsic subtypes, Lehmann *et al.,* assigned BT-549 and CAL-51 to the TNBC M subtype and identified MDA-MB-231 cells to be MSL (Lehmann et al., 2011, Lehmann et al., 2021). A recent study by Espinosa Fernandez *et al.,* confirmed that BT-549 cells maintain their M subtype properties *in vitro* and in tumour xenografts, whereas the MDA-MB-231 cell line was classified as unstable, with data suggesting that this cell line is both MSL and BL2 (Espinosa Fernandez et al., 2020). This indicates that the cell lines used in this study do not account for all the different TNBC intrinsic subtypes and intertumoural heterogeneity observed between patients. However, MSL and BL2 have been linked to poor patient prognosis, with the subtypes being linked to

higher rates of metastatic disease and worse OS rates than the other intrinsic subtypes (Yin et al., 2020, Ortiz Valdez et al., 2023). Thus, despite the cell lines representing only two intrinsic subtypes, the three TNBC cell lines used were representative of aggressive TNBC intrinsic subtypes that are difficult to treat. The non-tumourigenic and TNBC cell lines were also maintained in 2D monolayer cultures, which are not representative of solid tumours and do not consider the TME and extracellular factors. Therefore, 3D *in vitro* models such as spheroid and organoid systems may have been a better representation of *in vivo* tumours and could enable tumour penetration studies to be conducted (van den Brand et al., 2017).

Future work aims to validate the remaining autoantibody candidates, WRNIP1, CALU, DNAJB6, KRT19 and HLA-A. Research has shown that using a panel of autoantibodies leads to increased sensitivity and specificity, therefore the ultimate aim is to establish a novel panel of serum autoantibodies, including anti-KHDRBS3, for the detection of TNBC (Henderson et al., 2019, Pagaza-Straffon et al., 2020). Current blood-based autoantibody tests are performed by indirect ELISA (Lam et al., 2011, Lourenco et al., 2017). Therefore, once a TNBC autoantibody panel has been established, future work aims to develop an indirect ELISA test and assess the performance of the TAA panel in detecting serum autoantibodies using a larger TNBC patient cohort. As aforementioned, this work has highlighted KHDRBS3 as a potential TNBC molecular target, therefore future studies will also be conducted to further investigate the role of KHDRBS3 in TNBC stemness. Regarding the TNBC peptides, *in vivo* biodistribution studies are going to be conducted to assess the TNBC targeting ability of peptides TLYRMVVTS and LMQDMALRL in an *in vivo* model. Depending on the results obtained from the biodistribution studies, future work also aims to develop a peptide-drug conjugate which could be used to deliver anti-cancer agents directly to TNBC tumours for TNBC targeted treatment.

In conclusion, the work conducted in this project has led to the discovery of novel TNBC associated antigens and targeting small peptides for the development of autoantibody-based early diagnostic tools and peptide-based targeted treatments for TNBC, with the hope of improving the outcome for women with the disease.

Bibliography

2005. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). *The Lancet,* 365**,** 1769-1778.

'T HOEN, P. A. C., JIRKA, S. M. G., TEN BROEKE, B. R., SCHULTES, E. A., AGUILERA, B., PANG, K. H., HEEMSKERK, H., AARTSMA-RUS, A., VAN OMMEN, G. J. & DEN DUNNEN, J. T. 2012. Phage display screening without repetitious selection rounds. *Analytical Biochemistry,* 421**,** 622-631.

ABDOU, Y., ATTWOOD, K., CHENG, T.-Y. D., YAO, S., BANDERA, E. V., ZIRPOLI, G. R., ONDRACEK, R. P., STEIN, L., BSHARA, W., KHOURY, T., AMBROSONE, C. B. & OMILIAN, A. R. 2020. Racial differences in CD8+ T cell infiltration in breast tumors from Black and White women. *Breast Cancer Research,* 22**,** 62.

ABUBAKAR, M., GUO, C., KOKA, H., SUNG, H., SHAO, N., GUIDA, J., DENG, J., LI, M., HU, N., ZHOU, B., LU, N. & YANG, X. R. 2019. Clinicopathological and epidemiological significance of breast cancer subtype reclassification based on p53 immunohistochemical expression. *npj Breast Cancer,* 5**,** 20.

ADAMS, D., GONZALEZ-DUARTE, A., O'RIORDAN, W. D., YANG, C. C., UEDA, M., KRISTEN, A. V., TOURNEV, I., SCHMIDT, H. H., COELHO, T., BERK, J. L., LIN, K. P., VITA, G., ATTARIAN, S., PLANTÉ-BORDENEUVE, V., MEZEI, M. M., CAMPISTOL, J. M., BUADES, J., BRANNAGAN, T. H., 3RD, KIM, B. J., OH, J., PARMAN, Y., SEKIJIMA, Y., HAWKINS, P. N., SOLOMON, S. D., POLYDEFKIS, M., DYCK, P. J., GANDHI, P. J., GOYAL, S., CHEN, J., STRAHS, A. L., NOCHUR, S. V., SWEETSER, M. T., GARG, P. P., VAISHNAW, A. K., GOLLOB, J. A. & SUHR, O. B. 2018. Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. *N Engl J Med,* 379**,** 11-21.

AHERN, C. H., SHIH, Y. C. T., DONG, W., PARMIGIANI, G. & SHEN, Y. 2014. Cost-effectiveness of alternative strategies for integrating MRI into breast cancer screening for women at high risk. *British Journal Of Cancer,* 111**,** 1542.

AL-HAJJ, M., WICHA, M. S., BENITO-HERNANDEZ, A., MORRISON, S. J. & CLARKE, M. F. 2003. Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 3983-3988.

ALESKANDARANY, M. A., VANDENBERGHE, M. E., MARCHIÒ, C., ELLIS, I. O., SAPINO, A. & RAKHA, E. A. 2018. Tumour Heterogeneity of Breast Cancer: From Morphology to Personalised Medicine. *Pathobiology,* 85**,** 23-34.

ALISTAIR, R. & MITCH, D. 2004. Mechanisms of tamoxifen resistance. *Endocrine-Related Cancer Endocr Relat Cancer,* 11**,** 643-658.

ALLRED, C. A., GORMLEY, C., VENUGOPAL, I., LI, S., MCGUIRE, M. J. & BROWN, K. C. 2023. Tumor-specific intracellular delivery: peptide-guided transport of a catalytic toxin. *Communications Biology,* 6**,** 60.

ALSARKHI, L. K. & WEBER, G. F. 2018. Anti-osteopontin autoantibodies in various types of cancer. *Oncol Rep,* 40**,** 3879-3889.

ALTHOBITI, M., ALESKANDARANY, M. A., JOSEPH, C., TOSS, M., MONGAN, N., DIEZ-RODRIGUEZ, M., NOLAN, C. C., ASHANKYTY, I., ELLIS, I. O., GREEN, A. R. & RAKHA, E. A. 2018. Heterogeneity of tumour-infiltrating lymphocytes in breast cancer and its prognostic significance. *Histopathology,* 73**,** 887-896.

AMIRI-KORDESTANI, L., BLUMENTHAL, G. M., XU, Q. C., ZHANG, L., TANG, S. W., HA, L., WEINBERG, W. C., CHI, B., CANDAU-CHACON, R., HUGHES, P., RUSSELL, A. M., MIKSINSKI, S. P., CHEN, X. H., MCGUINN, W. D., PALMBY, T., SCHRIEBER, S.

J., LIU, Q., WANG, J., SONG, P., MEHROTRA, N., SKARUPA, L., CLOUSE, K., AL-HAKIM, A., SRIDHARA, R., IBRAHIM, A., JUSTICE, R., PAZDUR, R. & CORTAZAR, P. 2014. FDA Approval: Ado-Trastuzumab Emtansine for the Treatment of Patients with HER2-Positive Metastatic Breast Cancer. *Clinical Cancer Research,* 20**,** 4436.

ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biology,* 11**,** R106.

ANDERSON, K. S., CRAMER, D. W., SIBANI, S., WALLSTROM, G., WONG, J., PARK, J., QIU, J., VITONIS, A. & LABAER, J. 2015. Autoantibody signature for the serologic detection of ovarian cancer. *J Proteome Res,* 14**,** 578-86.

ANDERSON, K. S., SIBANI, S., WALLSTROM, G., QIU, J., MENDOZA, E. A., RAPHAEL, J., HAINSWORTH, E., MONTOR, W. R., WONG, J., PARK, J. G., LOKKO, N., LOGVINENKO, T., RAMACHANDRAN, N., GODWIN, A. K., MARKS, J., ENGSTROM, P. & LABAER, J. 2011. Protein microarray signature of autoantibody biomarkers for the early detection of breast cancer. *Journal of proteome research,* 10**,** 85-96.

ANTONIOU, A., PHAROAH, P. D. P., NAROD, S., RISCH, H. A., EYFJORD, J. E., HOPPER, J. L., LOMAN, N., OLSSON, H., JOHANNSSON, O., BORG, A., PASINI, B., RADICE, P., MANOUKIAN, S., ECCLES, D. M., TANG, N., OLAH, E., ANTON-CULVER, H., WARNER, E., LUBINSKI, J., GRONWALD, J., GORSKI, B., TULINIUS, H., THORLACIUS, S., EEROLA, H., NEVANLINNA, H., SYRJÄKOSKI, K., KALLIONIEMI, O. P., THOMPSON, D., EVANS, C., PETO, J., LALLOO, F., EVANS, D. G. & EASTON, D. F. 2003. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *American journal of human genetics,* 72**,** 1117-1130.

APPLETON, D. C., HACKNEY, L. & NARAYANAN, S. 2014. Ultrasonography alone for diagnosis of breast cancer in women under 40. *Annals of the Royal College of Surgeons of England,* 96**,** 202-206.

ARAP, W., PASQUALINI, R. & RUOSLAHTI, E. 1998. Cancer Treatment by Targeted Drug Delivery to Tumor Vasculature in a Mouse Model. *Science,* 279**,** 377-380.

ASSELAIN, B., BARLOW, W., BARTLETT, J., BERGH, J., BERGSTEN-NORDSTRÖM, E., BLISS, J., BOCCARDO, F., BODDINGTON, C., BOGAERTS, J., BONADONNA, G., BRADLEY, R., BRAIN, E., BRAYBROOKE, J., BROET, P., BRYANT, J., BURRETT, J., CAMERON, D., CLARKE, M., COATES, A., COLEMAN, R., COOMBES, R. C., CORREA, C., COSTANTINO, J., CUZICK, J., DANFORTH, D., DAVIDSON, N., DAVIES, C., DAVIES, L., DI LEO, A., DODWELL, D., DOWSETT, M., DUANE, F., EVANS, V., EWERTZ, M., FISHER, B., FORBES, J., FORD, L., GAZET, J.-C., GELBER, R., GETTINS, L., GIANNI, L., GNANT, M., GODWIN, J., GOLDHIRSCH, A., GOODWIN, P., GRAY, R., HAYES, D., HILL, C., INGLE, J., JAGSI, R., JAKESZ, R., JAMES, S., JANNI, W., LIU, H., LIU, Z., LOHRISCH, C., LOIBL, S., MACKINNON, L., MAKRIS, A., MAMOUNAS, E., MANNU, G., MARTÍN, M., MATHOULIN, S., MAURIAC, L., MCGALE, P., MCHUGH, T., MORRIS, P., MUKAI, H., NORTON, L., OHASHI, Y., OLIVOTTO, I., PAIK, S., PAN, H., PETO, R., PICCART, M., PIERCE, L., POORTMANS, P., POWLES, T., PRITCHARD, K., RAGAZ, J., RAINA, V., RAVDIN, P., READ, S., REGAN, M., ROBERTSON, J., RUTGERS, E., SCHOLL, S., SLAMON, D., SÖLKNER, L., SPARANO, J., STEINBERG, S., SUTCLIFFE, R., SWAIN, S., TAYLOR, C., TUTT, A., VALAGUSSA, P., VAN DE VELDE, C., VAN DER HAGE, J., VIALE, G., VON MINCKWITZ, G., et al. 2018. Long-term outcomes for neoadjuvant versus adjuvant chemotherapy in early breast cancer: meta-

analysis of individual patient data from ten randomised trials. *The Lancet Oncology,* 19**,** 27-39.

AVASTHI, A., CARO, C., POZO-TORRES, E., LEAL, M. P. & GARCÍA-MARTÍN, M. L. 2020. Magnetic Nanoparticles as MRI Contrast Agents. *Top Curr Chem (Cham),* 378**,** 40.

AYSOLA, K., DESAI, A., WELCH, C., XU, J., QIN, Y., REDDY, V., MATTHEWS, R., OWENS, C., OKOLI, J., BEECH, D. J., PIYATHILAKE, C. J., REDDY, S. P. & RAO, V. N. 2013. Triple Negative Breast Cancer - An Overview. *Hereditary Genet,* 2013.

BADEN, L. R., EL SAHLY, H. M., ESSINK, B., KOTLOFF, K., FREY, S., NOVAK, R., DIEMERT, D., SPECTOR, S. A., ROUPHAEL, N., CREECH, C. B., MCGETTIGAN, J., KHETAN, S., SEGALL, N., SOLIS, J., BROSZ, A., FIERRO, C., SCHWARTZ, H., NEUZIL, K., COREY, L., GILBERT, P., JANES, H., FOLLMANN, D., MAROVICH, M., MASCOLA, J., POLAKOWSKI, L., LEDGERWOOD, J., GRAHAM, B. S., BENNETT, H., PAJON, R., KNIGHTLY, C., LEAV, B., DENG, W., ZHOU, H., HAN, S., IVARSSON, M., MILLER, J. & ZAKS, T. 2020. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine,* 384**,** 403-416.

BANNISTER, N. & BROGGIO, J. 2016. Cancer survival by stage at diagnosis for England (experimental statistics): adults diagnosed 2012, 2013 and 2014 and followed up to 2015. *Produced in collaboration with Public Health England*.

BARDIA, A., MAYER, I. A., DIAMOND, J. R., MOROOSE, R. L., ISAKOFF, S. J., STARODUB, A. N., SHAH, N. C., O'SHAUGHNESSY, J., KALINSKY, K., GUARINO, M., ABRAMSON, V., JURIC, D., TOLANEY, S. M., BERLIN, J., MESSERSMITH, W. A., OCEAN, A. J., WEGENER, W. A., MALIAKAL, P., SHARKEY, R. M., GOVINDAN, S. V., GOLDENBERG, D. M. & VAHDAT, L. T. 2017. Efficacy and Safety of Anti-Trop-2 Antibody Drug Conjugate Sacituzumab Govitecan (IMMU-132) in Heavily Pretreated Patients With Metastatic Triple-Negative Breast Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* 35**,** 2141-2148.

BARDIA, A., MAYER, I. A., VAHDAT, L. T., TOLANEY, S. M., ISAKOFF, S. J., DIAMOND, J. R., O'SHAUGHNESSY, J., MOROOSE, R. L., SANTIN, A. D., ABRAMSON, V. G., SHAH, N. C., RUGO, H. S., GOLDENBERG, D. M., SWEIDAN, A. M., IANNONE, R., WASHKOWITZ, S., SHARKEY, R. M., WEGENER, W. A. & KALINSKY, K. 2019. Sacituzumab Govitecan-hziy in Refractory Metastatic Triple-Negative Breast Cancer. *N Engl J Med,* 380**,** 741-751.

BARECHE, Y., VENET, D., IGNATIADIS, M., AFTIMOS, P., PICCART, M., ROTHE, F. & SOTIRIOU, C. 2018. Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Ann Oncol,* 29**,** 895-902.

BARROSO-SOUSA, R., KEENAN, T. E., PERNAS, S., EXMAN, P., JAIN, E., GARRIDO-CASTRO, A. C., HUGHES, M., BYCHKOVSKY, B., UMETON, R., FILES, J. L., LINDEMAN, N. I., MACCONAILL, L. E., HODI, F. S., KROP, I. E., DILLON, D., WINER, E. P., WAGLE, N., LIN, N. U., MITTENDORF, E. A., VAN ALLEN, E. M. & TOLANEY, S. M. 2020. Tumor Mutational Burden and PTEN Alterations as Molecular Correlates of Response to PD-1/L1 Blockade in Metastatic Triple-Negative Breast Cancer. *Clinical Cancer Research,* 26**,** 2565-2572.

BARROSO-SOUSA, R. & METZGER-FILHO, O. 2016. Differences between invasive lobular and invasive ductal carcinoma of the breast: results and therapeutic implications. *Therapeutic advances in medical oncology,* 8**,** 261-266.

BASSARO, L., RUSSELL, S. J., PASTWA, E., SOMIARI, S. A. & SOMIARI, R. I. 2017. Screening for Multiple Autoantibodies in Plasma of Patients with Breast Cancer. *Cancer genomics & proteomics,* 14**,** 427-435.

BATLLE, E. & CLEVERS, H. 2017. Cancer stem cells revisited. *Nature Medicine,* 23**,** 1124-1134.

BAZAN, J., CAŁKOSIŃSKI, I. & GAMIAN, A. 2012. Phage display--a powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications. *Human vaccines & immunotherapeutics,* 8**,** 1817-1828.

BAZHIN, A. V., SAVCHENKO, M. S., SHIFRINA, O. N., DEMOURA, S. A., CHIKINA, S. Y., JAQUES, G., KOGAN, E. A., CHUCHALIN, A. G. & PHILIPPOV, P. P. 2004. Recoverin as a paraneoplastic antigen in lung cancer: the occurrence of anti-recoverin autoantibodies in sera and recoverin in tumors. *Lung Cancer,* 44**,** 193-198.

BEGICEVIC, R. R. & FALASCA, M. 2017. ABC Transporters in Cancer Stem Cells: Beyond Chemoresistance. *Int J Mol Sci,* 18.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological),* 57**,** 289-300.

BENJAMINI, Y., KRIEGER, A. M. & YEKUTIELI, D. 2006. Adaptive Linear Step-up Procedures That Control the False Discovery Rate. *Biometrika,* 93**,** 491-507.

BENNETT, R. L., BLANKS, R. G. & MOSS, S. M. 2009. Evaluation of extension of breast screening to women aged 65-70 in England using screening performance measures. *British journal of cancer,* 100**,** 1043-1047.

BERRY, D. A., CRONIN, K. A., PLEVRITIS, S. K., FRYBACK, D. G., CLARKE, L., ZELEN, M., MANDELBLATT, J. S., YAKOVLEV, A. Y., HABBEMA, J. D. F. & FEUER, E. J. 2005. Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer. *New England Journal of Medicine,* 353**,** 1784-1792.

BERTUCCI, F., FINETTI, P., GONCALVES, A. & BIRNBAUM, D. 2020. The therapeutic response of ER+/HER2− breast cancers differs according to the molecular Basal or Luminal subtype. *npj Breast Cancer,* 6**,** 8.

BESSA, C., MATOS, P., JORDAN, P. & GONÇALVES, V. 2020. Alternative Splicing: Expanding the Landscape of Cancer Biomarkers and Therapeutics. *International Journal of Molecular Sciences,* 21**,** 9032.

BEZU, L., KEPP, O., CERRATO, G., POL, J., FUCIKOVA, J., SPISEK, R., ZITVOGEL, L., KROEMER, G. & GALLUZZI, L. 2018. Trial watch: Peptide-based vaccines in anticancer therapy. *Oncoimmunology,* 7**,** e1511506-e1511506.

BIANCHINI, G., DE ANGELIS, C., LICATA, L. & GIANNI, L. 2022. Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. *Nat Rev Clin Oncol,* 19**,** 91-113.

BINLADEN, J., GILBERT, M. T. P., BOLLBACK, J. P., PANITZ, F., BENDIXEN, C., NIELSEN, R. & WILLERSLEV, E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PloS one,* 2**,** e197-e197.

BISWAS, S., DODWADKAR, N. S., DESHPANDE, P. P., PARAB, S. & TORCHILIN, V. P. 2013. Surface functionalization of doxorubicin-loaded liposomes with octa-arginine for enhanced anticancer activity. *Eur J Pharm Biopharm,* 84**,** 517-25.

BLANKS, R. G., MOSS, S. M., MCGAHAN, C. E., QUINN, M. J. & BABB, P. J. 2000. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: comparison of observed with predicted mortality. *BMJ (Clinical research ed.),* 321**,** 665-669.

BLOKPOEL FERRERAS, L. A., SCOTT, D., VAZQUEZ REINA, S., ROACH, P., TORRES, T. E., GOYA, G. F., SHAKESHEFF, K. M. & DIXON, J. E. 2020. Enhanced Cellular Transduction of Nanoparticles Resistant to Rapidly Forming Plasma Protein Coronas. *Advanced Biosystems,* 4**,** 2000162.

BOND, M., PAVEY, T., WELCH, K., COOPER, C., GARSIDE, R., DEAN, S. & HYDE, C. J. 2013. Psychological consequences of false-positive screening mammograms in the UK. *Evidence Based Medicine,* 18**,** 54-61.

BONNEFOI, H., GRELLETY, T., TREDAN, O., SAGHATCHIAN, M., DALENC, F., MAILLIEZ, A., L'HARIDON, T., COTTU, P., ABADIE-LACOURTOISIE, S., YOU, B., MOUSSEAU, M., DAUBA, J., DEL PIANO, F., DESMOULINS, I., COUSSY, F., MADRANGES, N., GRENIER, J., BIDARD, F. C., PROUDHON, C., MACGROGAN, G., ORSINI, C., PULIDO, M. & GONCALVES, A. 2016. A phase II trial of abiraterone acetate plus prednisone in patients with triple-negative androgen receptor positive locally advanced or metastatic breast cancer (UCBG 12-1). *Ann Oncol,* 27**,** 812-8.

BONNET, D. & DICK, J. E. 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine,* 3**,** 730-737.

BOOHAKER, R. J., LEE, M. W., VISHNUBHOTLA, P., PEREZ, J. M. & KHALED, A. R. 2012. The use of therapeutic peptides to target and to kill cancer cells. *Current medicinal chemistry,* 19**,** 3794-3804.

BUIST, D. S. M., PORTER, P. L., LEHMAN, C., TAPLIN, S. H. & WHITE, E. 2004. Factors Contributing to Mammography Failure in Women Aged 40–49 Years. *JNCI: Journal of the National Cancer Institute,* 96**,** 1432-1440.

BURSTEIN, H. J., PRESTRUD, A. A., SEIDENFELD, J., ANDERSON, H., BUCHHOLZ, T. A., DAVIDSON, N. E., GELMON, K. E., GIORDANO, S. H., HUDIS, C. A., MALIN, J., MAMOUNAS, E. P., ROWDEN, D., SOLKY, A. J., SOWERS, M. R., STEARNS, V., WINER, E. P., SOMERFIELD, M. R., GRIGGS, J. J. & AMERICAN SOCIETY OF CLINICAL, O. 2010. American Society of Clinical Oncology clinical practice guideline: update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* 28**,** 3784-3796.

BURSTEIN, M. D., TSIMELZON, A., POAGE, G. M., COVINGTON, K. R., CONTRERAS, A., FUQUA, S. A., SAVAGE, M. I., OSBORNE, C. K., HILSENBECK, S. G., CHANG, J. C., MILLS, G. B., LAU, C. C. & BROWN, P. H. 2015. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin Cancer Res,* 21**,** 1688-98.

BYBEE, S. M., BRACKEN-GRISSOM, H., HAYNES, B. D., HERMANSEN, R. A., BYERS, R. L., CLEMENT, M. J., UDALL, J. A., WILCOX, E. R. & CRANDALL, K. A. 2011. Targeted Amplicon Sequencing (TAS): A Scalable Next-Gen Approach to Multilocus, Multitaxa Phylogenetics. *Genome Biology and Evolution,* 3**,** 1312-1323.

CAILLEAU, R., YOUNG, R., OLIVÉ, M. & REEVES, W. J., JR. 1974. Breast tumor cell lines from pleural effusions. *J Natl Cancer Inst,* 53**,** 661-74.

CANTINI, L., ATTAWAY, C. C., BUTLER, B., ANDINO, L. M., SOKOLOSKY, M. L. & JAKYMIW, A. 2013. Fusogenic-Oligoarginine Peptide-Mediated Delivery of siRNAs Targeting the CIP2A Oncogene into Oral Cancer Cells. *PLOS ONE,* 8**,** e73348.

CARBONARI, M., MANCANIELLO, D., TEDESCO, T. & FIORILLI, M. 2008. Flow acetone-staining technique: A highly efficient procedure for the simultaneous analysis of DNA content, cell morphology, and immunophenotype by flow cytometry. *Cytometry Part A,* 73A**,** 168-174.

CARTER, P., PRESTA, L., GORMAN, C. M., RIDGWAY, J. B., HENNER, D., WONG, W. L., ROWLAND, A. M., KOTTS, C., CARVER, M. E. & SHEPARD, H. M. 1992. Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 4285-4289.

CHANGAVI, A. A., SHASHIKALA, A. & RAMJI, A. S. 2015. Epidermal Growth Factor Receptor Expression in Triple Negative and Nontriple Negative Breast Carcinomas. *Journal of laboratory physicians,* 7**,** 79-83.

CHAPMAN, C., MURRAY, A., CHAKRABARTI, J., THORPE, A., WOOLSTON, C., SAHIN, U., BARNES, A. & ROBERTSON, J. 2007. Autoantibodies in breast cancer: their use as an aid to early diagnosis. *Annals of Oncology,* 18**,** 868-873.

CHAPMAN, C. J., MURRAY, A., MCELVEEN, J. E., SAHIN, U., LUXEMBURGER, U., TÜRECI, Ö., WIEWRODT, R., BARNES, A. C. & ROBERTSON, J. F. 2008. Autoantibodies in lung cancer: possibilities for early detection and subsequent cure. *Thorax,* 63**,** 228-233.

CHARFI, C., DEMEULE, M., CURRIE, J.-C., LAROCQUE, A., ZGHEIB, A., DANALACHE, B. A., OUANOUKI, A., BÉLIVEAU, R., MARSOLAIS, C. & ANNABI, B. 2021. New Peptide-Drug Conjugates for Precise Targeting of SORT1-Mediated Vasculogenic Mimicry in the Tumor Microenvironment of TNBC-Derived MDA-MB-231 Breast and Ovarian ES-2 Clear Cell Carcinoma Cells. *Frontiers in Oncology,* 11.

CHAVEZ, K. J., GARIMELLA, S. V. & LIPKOWITZ, S. 2010. Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer. *Breast Dis,* 32**,** 35-48.

CHECKA, C. M., CHUN, J. E., SCHNABEL, F. R., LEE, J. & TOTH, H. 2012. The Relationship of Mammographic Density and Age: Implications for Breast Cancer Screening. *American Journal of Roentgenology,* 198**,** W292-W295.

CHEN, C., POOLE, L., SIZER, L., CARTER, W. B. & FRAZIER, T. G. 2019. Diagnostic accuracy of the Videssa protein-based liquid biopsy for breast cancer in suspicious mammography. *Journal of Clinical Oncology,* 37**,** e13034-e13034.

CHEN, H., WU, J., ZHANG, Z., TANG, Y., LI, X., LIU, S., CAO, S. & LI, X. 2018. Association Between BRCA Status and Triple-Negative Breast Cancer: A Meta-Analysis. *Frontiers in pharmacology,* 9**,** 909-909.

CHEN, J., LI, Y., YU, T. S., MCKAY, R. M., BURNS, D. K., KERNIE, S. G. & PARADA, L. F. 2012. A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature,* 488**,** 522-6.

CHEN, K., ZHANG, C., LING, S., WEI, R., WANG, J. & XU, X. 2021. The metabolic flexibility of quiescent CSC: implications for chemotherapy resistance. *Cell Death & Disease,* 12**,** 835.

CHEN, Y., SUSICK, L., DAVIS, M., BENSENHAVER, J., NATHANSON, S. D., BURNS, J. & NEWMAN, L. A. 2020. Evaluation of Triple-Negative Breast Cancer Early Detection via Mammography Screening and Outcomes in African American and White American Patients. *JAMA Surgery,* 155**,** 440-442.

CHIU, A. M., MITRA, M., BOYMOUSHAKIAN, L. & COLLER, H. A. 2018. Integrative analysis of the inter-tumoral heterogeneity of triple-negative breast cancer. *Sci Rep,* 8**,** 11807.

CHOI, H., FERMIN, D. & NESVIZHSKII, A. I. 2008. Significance Analysis of Spectral Count Data in Label-free Shotgun Proteomics. *Molecular & Cellular Proteomics,* 7**,** 2373-2385.

CLAESSENS, A. K. M., IBRAGIMOVA, K. I. E., GEURTS, S. M. E., BOS, M. E. M. M., ERDKAMP, F. L. G. & TJAN-HEIJNEN, V. C. G. 2020. The role of chemotherapy in treatment of advanced breast cancer: an overview for clinical practice. *Critical Reviews in Oncology/Hematology,* 153**,** 102988.

CLINICALTRIALS.GOV/NCT03623945. 2018. *Autoantibodies in Breast Cancer Detection* [Online]. https://ClinicalTrials.gov/show/NCT03623945. [Accessed 19th December 2022].

COCHRANE, C. R., SZCZEPNY, A., WATKINS, D. N. & CAIN, J. E. 2015. Hedgehog Signaling in the Maintenance of Cancer Stem Cells. *Cancers,* 7**,** 1554-1585.

COLLETT, K., STEFANSSON, I. M., EIDE, J., BRAATEN, A., WANG, H., EIDE, G. E., THORESEN, S. Ø., FOULKES, W. D. & AKSLEN, L. A. 2005. A Basal Epithelial Phenotype Is More Frequent in Interval Breast Cancers Compared with Screen Detected Tumors. *Cancer Epidemiology Biomarkers & Prevention,* 14**,** 1108-1112.

CORSETTI, V., HOUSSAMI, N., FERRARI, A., GHIRARDI, M., BELLAROSA, S., ANGELINI, O., BANI, C., SARDO, P., REMIDA, G., GALLIGIONI, E. & CIATTO, S. 2008. Breast screening with ultrasound in women with mammography-negative dense breasts: Evidence on incremental cancer detection and false positives, and associated cost. *European Journal of Cancer,* 44**,** 539-544.

COSSETTI, R. J. D., TYLDESLEY, S. K., SPEERS, C. H., ZHENG, Y. & GELMON, K. A. 2015. Comparison of Breast Cancer Recurrence and Outcome Patterns Between Patients Treated From 1986 to 1992 and From 2004 to 2008. *Journal of Clinical Oncology,* 33**,** 65-73.

CRISCITIELLO, C., ESPOSITO, A., TRAPANI, D. & CURIGLIANO, G. 2016. Prognostic and predictive value of tumor infiltrating lymphocytes in early breast cancer. *Cancer Treatment Reviews,* 50**,** 205-207.

CROWN, J., DIÉRAS, V., KAUFMANN, M., VON MINCKWITZ, G., KAYE, S., LEONARD, R., MARTY, M., MISSET, J.-L., OSTERWALDER, B. & PICCART, M. 2002. Chemotherapy for metastatic breast cancer–report of a European expert panel. *The Lancet Oncology,* 3**,** 719-727.

CUZICK, J., POWLES, T., VERONESI, U., FORBES, J., EDWARDS, R., ASHLEY, S. & BOYLE, P. 2003. Overview of the main outcomes in breast-cancer prevention trials. *The Lancet,* 361**,** 296-300.

DAI, X., LI, T., BAI, Z., YANG, Y., LIU, X., ZHAN, J. & SHI, B. 2015. Breast cancer intrinsic subtype classification, clinical use and future trends. *American journal of cancer research,* 5**,** 2929-2943.

DASS, S. A., TAN, K. L., SELVA RAJAN, R., MOKHTAR, N. F., MOHD ADZMI, E. R., WAN ABDUL RAHMAN, W. F., TENGKU DIN, T. & BALAKRISHNAN, V. 2021. Triple Negative Breast Cancer: A Review of Present and Future Diagnostic Modalities. *Medicina (Kaunas),* 57.

DAWOOD, S. 2010. Triple-negative breast cancer: epidemiology and management options. *Drugs,* 70**,** 2247-58.

DE ANGELIS, M. L., FRANCESCANGELI, F. & ZEUNER, A. 2019. Breast Cancer Stem Cells as Drivers of Tumor Chemoresistance, Dormancy and Relapse: New Challenges and Therapeutic Opportunities. *Cancers,* 11**,** 1569.

DE OLIVEIRA, E. C. L., SANTANA, K., JOSINO, L., LIMA, E. L. A. H. & DE SOUZA DE SALES JÚNIOR, C. 2021. Predicting cell-penetrating peptides using machine learning algorithms and navigating in their chemical space. *Sci Rep,* 11**,** 7628.

DEMEULE, M., CHARFI, C., CURRIE, J. C., LAROCQUE, A., ZGHEIB, A., KOZELKO, S., BÉLIVEAU, R., MARSOLAIS, C. & ANNABI, B. 2021. TH1902, a new docetaxel-peptide conjugate for the treatment of sortilin-positive triple-negative breast cancer. *Cancer Sci,* 112**,** 4317-4334.

DENKERT, C., VON MINCKWITZ, G., DARB-ESFAHANI, S., LEDERER, B., HEPPNER, B. I., WEBER, K. E., BUDCZIES, J., HUOBER, J., KLAUSCHEN, F., FURLANETTO, J., SCHMITT, W. D., BLOHMER, J. U., KARN, T., PFITZNER, B. M., KUMMEL, S., ENGELS, K., SCHNEEWEISS, A., HARTMANN, A., NOSKE, A., FASCHING, P. A., JACKISCH, C., VAN MACKELENBERGH, M., SINN, P., SCHEM, C., HANUSCH, C., UNTCH, M. & LOIBL, S. 2018. Tumour-infiltrating lymphocytes and prognosis

in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol,* 19**,** 40-50.

DENT, R., TRUDEAU, M., PRITCHARD, K. I., HANNA, W. M., KAHN, H. K., SAWKA, C. A., LICKLEY, L. A., RAWLINSON, E., SUN, P. & NAROD, S. A. 2007. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res,* 13**,** 4429-34.

DESMETZ, C., BASCOUL-MOLLEVI, C., ROCHAIX, P., LAMY, P.-J., KRAMAR, A., ROUANET, P., MAUDELONDE, T., MANGÉ, A. & SOLASSOL, J. 2009. Identification of a New Panel of Serum Autoantibodies Associated with the Presence of In situ Carcinoma of the Breast in Younger Women. *Clinical Cancer Research,* 15**,** 4733-4741.

DEVLIN, J. J., PANGANIBAN, L. C. & DEVLIN, P. E. 1990. Random peptide libraries: a source of specific protein binding molecules. *Science,* 249**,** 404-6.

DEY, N., DE, P. & LEYLAND-JONES, B. 2017. PI3K-AKT-mTOR inhibitors in breast cancers: From tumor cell signaling to clinical trials. *Pharmacology & Therapeutics,* 175**,** 91-106.

DIAS-NETO, E., NUNES, D. N., GIORDANO, R. J., SUN, J., BOTZ, G. H., YANG, K., SETUBAL, J. C., PASQUALINI, R. & ARAP, W. 2009. Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis. *PloS one,* 4**,** e8338-e8338.

DIECI, M. V., MATHIEU, M. C., GUARNERI, V., CONTE, P., DELALOGE, S., ANDRE, F. & GOUBAR, A. 2015. Prognostic and predictive value of tumor-infiltrating lymphocytes in two phase III randomized adjuvant breast cancer trials. *Ann Oncol,* 26**,** 1698-704.

DILL, E. A., GRU, A. A., ATKINS, K. A., FRIEDMAN, L. A., MOORE, M. E., BULLOCK, T. N., CROSS, J. V., DILLON, P. M. & MILLS, A. M. 2017. PD-L1 Expression and Intratumoral Heterogeneity Across Breast Cancer Subtypes and Stages: An Assessment of 245 Primary and 40 Metastatic Tumors. *The American Journal of Surgical Pathology,* 41.

DISSANAYAKE, S., DENNY, W. A., GAMAGE, S. & SAROJINI, V. 2017. Recent developments in anticancer drug delivery using cell penetrating and tumor targeting peptides. *J Control Release,* 250**,** 62-76.

DIXON, J. E., OSMAN, G., MORRIS, G. E., MARKIDES, H., ROTHERHAM, M., BAYOUSSEF, Z., EL HAJ, A. J., DENNING, C. & SHAKESHEFF, K. M. 2016. Highly efficient delivery of functional cargoes by the synergistic effect of GAG binding motifs and cell-penetrating peptides. *Proceedings of the National Academy of Sciences,* 113**,** E291-E299.

DOMCKE, S., SINHA, R., LEVINE, D. A., SANDER, C. & SCHULTZ, N. 2013. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat Commun,* 4**,** 2126.

DRIN, G., COTTIN, S., BLANC, E., REES, A. R. & TEMSAMANI, J. 2003. Studies on the Internalization Mechanism of Cationic Cell-penetrating Peptides*. *Journal of Biological Chemistry,* 278**,** 31192-31201.

DUBIKOVSKAYA, E. A., THORNE, S. H., PILLOW, T. H., CONTAG, C. H. & WENDER, P. A. 2008. Overcoming multidrug resistance of small-molecule therapeutics through conjugation with releasable octaarginine transporters. *Proceedings of the National Academy of Sciences,* 105**,** 12128-12133.

DUFFY, S. W., AGBAJE, O., TABAR, L., VITAK, B., BJURSTAM, N., BJÖRNELD, L., MYLES, J. P. & WARWICK, J. 2005. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast cancer research : BCR,* 7**,** 258-265.

DUFFY, S. W., VULKAN, D., CUCKLE, H., PARMAR, D., SHEIKH, S., SMITH, R. A., EVANS, A., BLYUSS, O., JOHNS, L., ELLIS, I. O., MYLES, J., SASIENI, P. D. & MOSS, S. M. 2020. Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial. *The Lancet Oncology,* 21**,** 1165-1172.

DURANTI, S., FABI, A., FILETTI, M., FALCONE, R., LOMBARDI, P., DANIELE, G., FRANCESCHINI, G., CARBOGNIN, L., PALAZZO, A., GARGANESE, G., PARIS, I., SCAMBIA, G. & PIETRAGALLA, A. 2021. Breast Cancer Drug Approvals Issued by EMA: A Review of Clinical Trials. *Cancers (Basel),* 13.

EARLY BREAST CANCER TRIALISTS' COLLABORATIVE, G. 1998. Tamoxifen for early breast cancer: an overview of the randomised trials. *The Lancet,* 351**,** 1451-1467.

EHRMANN, I., DALGLIESH, C., LIU, Y., DANILENKO, M., CROSIER, M., OVERMAN, L., ARTHUR, H. M., LINDSAY, S., CLOWRY, G. J., VENABLES, J. P., FORT, P. & ELLIOTT, D. J. 2013. The tissue-specific RNA binding protein T-STAR controls regional splicing patterns of neurexin pre-mRNAs in the brain. *PLoS Genet,* 9**,** e1003474.

ELLIS, M. J. & PEROU, C. M. 2013. The genomic landscape of breast cancer as a therapeutic roadmap. *Cancer discovery,* 3**,** 27-34.

ELSTON, C. W. & ELLIS, I. O. 1991. 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,* 19**,** 403-410.

ENGEL, C., RHIEM, K., HAHNEN, E., LOIBL, S., WEBER, K. E., SEILER, S., ZACHARIAE, S., HAUKE, J., WAPPENSCHMIDT, B., WAHA, A., BLUMCKE, B., KIECHLE, M., MEINDL, A., NIEDERACHER, D., BARTRAM, C. R., SPEISER, D., SCHLEGELBERGER, B., ARNOLD, N., WIEACKER, P., LEINERT, E., GEHRIG, A., BRIEST, S., KAST, K., RIESS, O., EMONS, G., WEBER, B. H. F., ENGEL, J., SCHMUTZLER, R. K., GERMAN CONSORTIUM FOR HEREDITARY, B. & OVARIAN, C. 2018. Prevalence of pathogenic BRCA1/2 germline mutations among 802 women with unilateral triple-negative breast cancer without family cancer history. *BMC Cancer,* 18**,** 265.

EROLES, P., BOSCH, A., PEREZ-FIDALGO, J. A. & LLUCH, A. 2012. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev,* 38**,** 698-707.

ESPINOSA FERNANDEZ, J. R., ECKHARDT, B. L., LEE, J., LIM, B., PEARSON, T., SEITZ, R. S., HOUT, D. R., SCHWEITZER, B. L., NIELSEN, T. J., LAWRENCE, O. R., WANG, Y., RAO, A. & UENO, N. T. 2020. Identification of triple-negative breast cancer cell lines classified under the same molecular subtype using different molecular characterization techniques: Implications for translational research. *PLoS One,* 15**,** e0231953.

ESTELRICH, J., ESCRIBANO, E., QUERALT, J. & BUSQUETS, M. A. 2015. Iron Oxide Nanoparticles for Magnetically-Guided and Magnetically-Responsive Drug Delivery. *International Journal of Molecular Sciences* [Online], 16.

EVANS, R. L., POTTALA, J. V., NAGATA, S. & EGLAND, K. A. 2018. Longitudinal autoantibody responses against tumor-associated antigens decrease in breast cancer patients according to treatment modality. *BMC Cancer,* 18**,** 119.

FAGERLUND, A., MYRSET, A. H. & KULSETH, M. A. 2008. Construction and characterization of a 9-mer phage display pVIII-library with regulated peptide density. *Applied Microbiology and Biotechnology,* 80**,** 925-936.

FALCIANI, C., LOZZI, L., PINI, A. & BRACCI, L. 2005. Bioactive Peptides from Libraries. *Chemistry & Biology,* 12**,** 417-426.

FALLAHPOUR, S., NAVANEELAN, T., DE, P. & BORGO, A. 2017. Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data. *CMAJ Open,* 5**,** E734-E739.

FAN, C., OH, D. S., WESSELS, L., WEIGELT, B., NUYTEN, D. S. A., NOBEL, A. B., VAN'T VEER, L. J. & PEROU, C. M. 2006. Concordance among Gene-Expression–Based Predictors for Breast Cancer. *New England Journal of Medicine,* 355**,** 560-569.

FARINA, N. H., RAMSEY, J. E., CUKE, M. E., AHERN, T. P., SHIRLEY, D. J., STEIN, J. L., STEIN, G. S., LIAN, J. B. & WOOD, M. E. 2017. Development of a predictive miRNA signature for breast cancer risk among high-risk women. *Oncotarget,* 8.

FARZIN, A., ETESAMI, S. A., QUINT, J., MEMIC, A. & TAMAYOL, A. 2020. Magnetic Nanoparticles in Cancer Therapy and Diagnosis. *Adv Healthc Mater,* 9**,** e1901058.

FENG, Y., SPEZIA, M., HUANG, S., YUAN, C., ZENG, Z., ZHANG, L., JI, X., LIU, W., HUANG, B., LUO, W., LIU, B., LEI, Y., DU, S., VUPPALAPATI, A., LUU, H. H., HAYDON, R. C., HE, T. C. & REN, G. 2018. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes Dis,* 5**,** 77-106.

FEOLA, S., CHIARO, J., MARTINS, B. & CERULLO, V. 2020. Uncovering the Tumor Antigen Landscape: What to Know about the Discovery Process. *Cancers (Basel),* 12.

FERNÁNDEZ MADRID, F., MAROUN, M.-C., OLIVERO, O. A., LONG, M., STARK, A., GROSSMAN, L. I., BINDER, W., DONG, J., BURKE, M., NATHANSON, S. D., ZARBO, R., CHITALE, D., ZEBALLOS-CHÁVEZ, R. & PEEBLES, C. 2015. Autoantibodies in breast cancer sera are not epiphenomena and may participate in carcinogenesis. *BMC Cancer,* 15**,** 407.

FIGUEROA-MAGALHÃES, M. C., JELOVAC, D., CONNOLLY, R. & WOLFF, A. C. 2014. Treatment of HER2-positive breast cancer. *Breast (Edinburgh, Scotland),* 23**,** 128-136.

FIORILLO, M., SOTGIA, F. & LISANTI, M. P. 2019. "Energetic" Cancer Stem Cells (e-CSCs): A New Hyper-Metabolic and Proliferative Tumor Cell Phenotype, Driven by Mitochondrial Energy. *Frontiers in Oncology,* 8.

FISHER, B., ANDERSON, S., BRYANT, J., MARGOLESE, R. G., DEUTSCH, M., FISHER, E. R., JEONG, J.-H. & WOLMARK, N. 2002. Twenty-Year Follow-up of a Randomized Trial Comparing Total Mastectomy, Lumpectomy, and Lumpectomy plus Irradiation for the Treatment of Invasive Breast Cancer. *New England Journal of Medicine,* 347**,** 1233-1241.

FOOTE, F. W. & STEWART, F. W. 1941. Lobular carcinoma in situ: A rare form of mammary cancer. *The American journal of pathology,* 17**,** 491-496.3.

FOUGNER, C., BERGHOLTZ, H., NORUM, J. H. & SØRLIE, T. 2019. Re-definition of claudin-low as a breast cancer phenotype. *bioRxiv***,** 756411.

FOULKES, W. D., SMITH, I. E. & REIS-FILHO, J. S. 2010. Triple-negative breast cancer. *N Engl J Med,* 363**,** 1938-48.

FRANKEL, A. D. & PABO, C. O. 1988. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell,* 55**,** 1189-93.

FROKJAER, S. & OTZEN, D. E. 2005. Protein drug stability: a formulation challenge. *Nature Reviews Drug Discovery,* 4**,** 298-306.

FU, X., GHARIB, S. A., GREEN, P. S., AITKEN, M. L., FRAZER, D. A., PARK, D. R., VAISAR, T. & HEINECKE, J. W. 2008. Spectral Index for Assessment of Differential

Protein Expression in Shotgun Proteomics. *Journal of Proteome Research,* 7**,** 845-854.

FUTAKI, S., SUZUKI, T., OHASHI, W., YAGAMI, T., TANAKA, S., UEDA, K. & SUGIURA, Y. 2001. Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J Biol Chem,* 276**,** 5836-40.

GALAN, M., GUIVIER, E., CARAUX, G., CHARBONNEL, N. & COSSON, J.-F. 2010. A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics,* 11**,** 296.

GALEA, M. H., BLAMEY, R. W., ELSTON, C. E. & ELLIS, I. O. 1992. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat,* 22**,** 207-19.

GALLO, E., KELIL, A., BAYLISS, P. E., JEGANATHAN, A., EGOROVA, O., PLODER, L., ADAMS, J. A., GIBLIN, P. & SIDHU, S. S. 2020. In situ antibody phage display yields optimal inhibitors of integrin α11/β1. *MAbs,* 12**,** 1717265.

GANDIOSO, A., MASSAGUER, A., VILLEGAS, N., SALVANS, C., SÁNCHEZ, D., BRUN-HEATH, I., MARCHÁN, V., OROZCO, M. & TERRAZAS, M. 2017. Efficient siRNA–peptide conjugation for specific targeted delivery into tumor cells. *Chemical Communications,* 53**,** 2870-2873.

GAO, L., HE, R.-Q., HUANG, Z.-G., DANG, Y.-W., GU, Y.-Y., YAN, H.-B., LI, S.-H. & CHEN, G. 2020. Genome-wide Analysis of the Alternative Splicing Profiles Revealed Novel Prognostic Index for Kidney Renal Cell Clear Cell Carcinoma. *Journal of Cancer,* 11**,** 1542-1554.

GASTON, J., MAESTRALI, N., LALLE, G., GAGNAIRE, M., MASIERO, A., DUMAS, B., DABDOUBI, T., RADOŠEVIĆ, K. & BERNE, P.-F. 2019. Intracellular delivery of therapeutic antibodies into specific cells using antibody-peptide fusions. *Scientific Reports,* 9**,** 18688.

GATTI-MAYS, M. E., BALKO, J. M., GAMEIRO, S. R., BEAR, H. D., PRABHAKARAN, S., FUKUI, J., DISIS, M. L., NANDA, R., GULLEY, J. L., KALINSKY, K., ABDUL SATER, H., SPARANO, J. A., CESCON, D., PAGE, D. B., MCARTHUR, H., ADAMS, S. & MITTENDORF, E. A. 2019. If we build it they will come: targeting the immune response to breast cancer. *npj Breast Cancer,* 5**,** 37.

GHEBEH, H., SLEIMAN, G. M., MANOGARAN, P. S., AL-MAZROU, A., BARHOUSH, E., AL-MOHANNA, F. H., TULBAH, A., AL-FAQEEH, K. & ADRA, C. N. 2013. Profiling of normal and malignant breast tissue show CD44high/CD24low phenotype as a predominant stem/progenitor marker when used in combination with Ep-CAM/CD49f markers. *BMC Cancer,* 13**,** 289.

GIOANNI, J., LE FRANÇOIS, D., ZANGHELLINI, E., MAZEAU, C., ETTORE, F., LAMBERT, J. C., SCHNEIDER, M. & DUTRILLAUX, B. 1990. Establishment and characterisation of a new tumorigenic cell line with a normal karyotype derived from a human breast adenocarcinoma. *Br J Cancer,* 62**,** 8-13.

GOGAS, H. & FOUNTZILAS, G. 2003. The role of taxanes as a component of neoadjuvant chemotherapy for breast cancer. *Annals of Oncology,* 14**,** 667-674.

GONZALEZ-ANGULO, A. M., MORALES-VASQUEZ, F. & HORTOBAGYI, G. N. 2007. Overview of Resistance to Systemic Therapy in Patients with Breast Cancer. *In:* YU, D. & HUNG, M.-C. (eds.) *Breast Cancer Chemosensitivity.* New York, NY: Springer New York.

GOODELL, M. A., NGUYEN, H. & SHROYER, N. 2015. Somatic stem cell heterogeneity: diversity in the blood, skin and intestinal stem cell compartments. *Nature Reviews Molecular Cell Biology,* 16**,** 299.

GOV, E. 2020. Co-expressed functional module-related genes in ovarian cancer stem cells represent novel prognostic biomarkers in ovarian cancer. *Systems Biology in Reproductive Medicine,* 66**,** 255-266.

GREENBLATT, M. S., CHAPPUIS, P. O., BOND, J. P., HAMEL, N. & FOULKES, W. D. 2001. TP53 Mutations in Breast Cancer Associated with BRCA1 or BRCA2 Germ-line Mutations. *Cancer Research,* 61**,** 4092.

GREENUP, R., BUCHANAN, A., LORIZIO, W., RHOADS, K., CHAN, S., LEEDOM, T., KING, R., MCLENNAN, J., CRAWFORD, B., KELLY MARCOM, P. & SHELLEY HWANG, E. 2013. Prevalence of BRCA mutations among women with triple-negative breast cancer (TNBC) in a genetic counseling cohort. *Ann Surg Oncol,* 20**,** 3254-8.

GRIFFIN, N. M., YU, J., LONG, F., OH, P., SHORE, S., LI, Y., KOZIOL, J. A. & SCHNITZER, J. E. 2010. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol,* 28**,** 83-9.

GU, Y., LIU, Y., FU, L., ZHAI, L., ZHU, J., HAN, Y., JIANG, Y., ZHANG, Y., ZHANG, P., JIANG, Z., ZHANG, X. & CAO, X. 2019. Tumor-educated B cells selectively promote breast cancer lymph node metastasis by HSPA4-targeting IgG. *Nature Medicine,* 25**,** 312-322.

GUCALP, A., TOLANEY, S., ISAKOFF, S. J., INGLE, J. N., LIU, M. C., CAREY, L. A., BLACKWELL, K., RUGO, H., NABELL, L., FORERO, A., STEARNS, V., DOANE, A. S., DANSO, M., MOYNAHAN, M. E., MOMEN, L. F., GONZALEZ, J. M., AKHTAR, A., GIRI, D. D., PATIL, S., FEIGIN, K. N., HUDIS, C. A., TRAINA, T. A. & TRANSLATIONAL BREAST CANCER RESEARCH, C. 2013. Phase II trial of bicalutamide in patients with androgen receptor-positive, estrogen receptor-negative metastatic Breast Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research,* 19**,** 5505-5512.

GUO, J., CHENG, T., XU, H., LI, Y. & ZENG, J. 2019. An efficient and cost-effective method for primer-induced nucleotide labeling for massive sequencing on next-generation sequencing platforms. *Scientific Reports,* 9**,** 3125.

GUO, Z., DING, Y., WANG, M., LIU, J., ZHAI, Q. & DU, Q. 2022. Safety of trastuzumab deruxtecan: A meta-analysis and pharmacovigilance study. *Journal of Clinical Pharmacy and Therapeutics,* 47**,** 1837-1844.

GURNEY, A., AXELROD, F., BOND, C. J., CAIN, J., CHARTIER, C., DONIGAN, L., FISCHER, M., CHAUDHARI, A., JI, M., KAPOUN, A. M., LAM, A., LAZETIC, S., MA, S., MITRA, S., PARK, I.-K., PICKELL, K., SATO, A., SATYAL, S., STROUD, M., TRAN, H., YEN, W.-C., LEWICKI, J. & HOEY, T. 2012. Wnt pathway inhibition via the targeting of Frizzled receptors results in decreased growth and tumorigenicity of human tumors. *Proceedings of the National Academy of Sciences,* 109**,** 11717-11722.

HAFFTY, B. G., YANG, Q., REISS, M., KEARNEY, T., HIGGINS, S. A., WEIDHAAS, J., HARRIS, L., HAIT, W. & TOPPMEYER, D. 2006. Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. *J Clin Oncol,* 24**,** 5652-7.

HAMRITA, B., CHAHED, K., KABBAGE, M., GUILLIER, C. L., TRIMECHE, M., CHAÏEB, A. & CHOUCHANE, L. 2008. Identification of tumor antigens that elicit a humoral immune response in breast cancer patients' sera by serological proteome analysis (SERPA). *Clinica Chimica Acta,* 393**,** 95-102.

HARDY-WERBIN, M., QUIROGA, V., CIRAUQUI, B., ROMEO, M., FELIP, E., TERUEL, I., GARCIA, J. J., ERASUN, C., ESPAÑA, S., CUCURULL, M., MONTPRADE, E., PARDO, J. C., CARBALLO, D., VELARDE, J. M. & MARGELI, M. 2019. Real-world data on T-DM1 efficacy – results of a single-center retrospective study of HER2-positive breast cancer patients. *Scientific Reports,* 9**,** 12760.

HATHAWAY, H. J., BUTLER, K. S., ADOLPHI, N. L., LOVATO, D. M., BELFON, R., FEGAN, D., MONSON, T. C., TRUJILLO, J. E., TESSIER, T. E., BRYANT, H. C., HUBER, D. L., LARSON, R. S. & FLYNN, E. R. 2011. Detection of breast cancer cells using targeted magnetic nanoparticles and ultra-sensitive magnetic field sensors. *Breast Cancer Research,* 13**,** R108.

HENDERSON, I. C., BERRY, D. A., DEMETRI, G. D., CIRRINCIONE, C. T., GOLDSTEIN, L. J., MARTINO, S., INGLE, J. N., COOPER, M. R., HAYES, D. F., TKACZUK, K. H., FLEMING, G., HOLLAND, J. F., DUGGAN, D. B., CARPENTER, J. T., III, E. F., SCHILSKY, R. L., WOOD, W. C., MUSS, H. B. & NORTON, L. 2003. Improved Outcomes From Adding Sequential Paclitaxel but Not From Escalating Doxorubicin Dose in an Adjuvant Chemotherapy Regimen for Patients With Node-Positive Primary Breast Cancer. *Journal of Clinical Oncology,* 21**,** 976-983.

HENDERSON, M. C., SILVER, M., TRAN, Q., LETSIOS, E. E., MULPURI, R., REESE, D. E., LOURENCO, A. P., LABAER, J., ANDERSON, K. S., ALPERS, J., COSTANTINI, C., ROHATGI, N., ALI, H., BAKER, K., NORTHFELT, D. W., GHOSH, K., GROBMYER, S. R., POLEN, W. & WOLF, J. K. 2019. A Noninvasive Blood-based Combinatorial Proteomic Biomarker Assay to Detect Breast Cancer in Women over age 50 with BI-RADS 3, 4, or 5 Assessment. *Clinical Cancer Research,* 25**,** 142-149.

HERSCHKOWITZ, J. I., SIMIN, K., WEIGMAN, V. J., MIKAELIAN, I., USARY, J., HU, Z., RASMUSSEN, K. E., JONES, L. P., ASSEFNIA, S., CHANDRASEKHARAN, S., BACKLUND, M. G., YIN, Y., KHRAMTSOV, A. I., BASTEIN, R., QUACKENBUSH, J., GLAZER, R. I., BROWN, P. H., GREEN, J. E., KOPELOVICH, L., FURTH, P. A., PALAZZO, J. P., OLOPADE, O. I., BERNARD, P. S., CHURCHILL, G. A., VAN DYKE, T. & PEROU, C. M. 2007. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome biology,* 8**,** R76-R76.

HESS, G. T., CRAGNOLINI, J. J., POPP, M. W., ALLEN, M. A., DOUGAN, S. K., SPOONER, E., PLOEGH, H. L., BELCHER, A. M. & GUIMARAES, C. P. 2012. M13 Bacteriophage Display Framework That Allows Sortase-Mediated Modification of Surface-Accessible Phage Proteins. *Bioconjugate Chemistry,* 23**,** 1478-1487.

HOBRO, A. J. & SMITH, N. I. 2017. An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging. *Vibrational Spectroscopy,* 91**,** 31-45.

HOLSTEGE, H., JOOSSE, S. A., VAN OOSTROM, C. T. M., NEDERLOF, P. M., DE VRIES, A. & JONKERS, J. 2009. High Incidence of Protein-Truncating TP53 Mutations in BRCA1-Related Breast Cancer. *Cancer Research,* 69**,** 3625.

HOPPENZ, P., ELS-HEINDL, S. & BECK-SICKINGER, A. G. 2020. Peptide-Drug Conjugates and Their Targets in Advanced Cancer Therapies. *Frontiers in Chemistry,* 8.

HORTOBAGYI, G. N., STEMMER, S. M., BURRIS, H. A., YAP, Y.-S., SONKE, G. S., PALUCH-SHIMON, S., CAMPONE, M., BLACKWELL, K. L., ANDRÉ, F., WINER, E. P., JANNI, W., VERMA, S., CONTE, P., ARTEAGA, C. L., CAMERON, D. A., PETRAKOVA, K., HART, L. L., VILLANUEVA, C., CHAN, A., JAKOBSEN, E., NUSCH, A., BURDAEVA, O., GRISCHKE, E.-M., ALBA, E., WIST, E., MARSCHNER, N., FAVRET, A. M.,

YARDLEY, D., BACHELOT, T., TSENG, L.-M., BLAU, S., XUAN, F., SOUAMI, F., MILLER, M., GERMA, C., HIRAWAT, S. & O'SHAUGHNESSY, J. 2016. Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. *New England Journal of Medicine,* 375**,** 1738-1748.

HOSHINO, I., NAGATA, M., TAKIGUCHI, N., NABEYA, Y., IKEDA, A., YOKOI, S., KUWAJIMA, A., TAGAWA, M., MATSUSHITA, K., SATOSHI, Y. & HIDEAKI, S. 2017. Panel of autoantibodies against multiple tumor-associated antigens for detecting gastric cancer. *Cancer science,* 108**,** 308-315.

HOSSEIN-NEJAD-ARIANI, H., ALTHAGAFI, E. & KAUR, K. 2019. Small Peptide Ligands for Targeting EGFR in Triple Negative Breast Cancer Cells. *Sci Rep,* 9**,** 2723.

HOUSSAMI, N., CIATTO, S., MARTINELLI, F., BONARDI, R. & DUFFY, S. W. 2009. Early detection of second breast cancers improves prognosis in breast cancer survivors. *Annals of Oncology,* 20**,** 1505-1510.

HUANG, H., GROTH, J., SOSSEY-ALAOUI, K., HAWTHORN, L., BEALL, S. & GERADTS, J. 2005. Aberrant Expression of Novel and Previously Described Cell Membrane Markers in Human Breast Cancer Cell Lines and Tumors. *Clinical Cancer Research,* 11**,** 4357.

HUDIS, C., SWANTON, C., JANJIGIAN, Y. Y., LEE, R., SUTHERLAND, S., LEHMAN, R., CHANDARLAPATY, S., HAMILTON, N., GAJRIA, D., KNOWLES, J., SHAH, J., SHANNON, K., TETTEH, E., SULLIVAN, D. M., MORENO, C., YAN, L. & HAN, H. S. 2013. A phase 1 study evaluating the combination of an allosteric AKT inhibitor (MK-2206) and trastuzumab in patients with HER2-positive solid tumors. *Breast cancer research : BCR,* 15**,** R110-R110.

HURVITZ, S. A., HEGG, R., CHUNG, W.-P., IM, S.-A., JACOT, W., GANJU, V., CHIU, J. W. Y., XU, B., HAMILTON, E., MADHUSUDAN, S., IWATA, H., ALTINTAS, S., HENNING, J.-W., CURIGLIANO, G., PEREZ-GARCIA, J. M., KIM, S.-B., PETRY, V., HUANG, C.-S., LI, W., FRENEL, J.-S., ANTOLIN, S., YEO, W., BIANCHINI, G., LOI, S., TSURUTANI, J., EGOROV, A., LIU, Y., CATHCART, J., ASHFAQUE, S. & CORTÉS, J. 2023. Trastuzumab deruxtecan versus trastuzumab emtansine in patients with HER2-positive metastatic breast cancer: updated results from DESTINY-Breast03, a randomised, open-label, phase 3 trial. *The Lancet,* 401**,** 105-117.

HWANG, H.-M., HEO, C.-K., LEE, H. J., KWAK, S.-S., LIM, W.-H., YOO, J.-S., YU, D.-Y., LIM, K. J., KIM, J.-Y. & CHO, E.-W. 2018. Identification of anti-SF3B1 autoantibody as a diagnostic marker in patients with hepatocellular carcinoma. *Journal of Translational Medicine,* 16**,** 177.

IANNOLO, G., MINENKOVA, O., PETRUZZELLI, R. & CESARENI, G. 1995. Modifying Filamentous Phage Capsid: Limits in the Size of the Major Capsid Protein. *Journal of Molecular Biology,* 248**,** 835-844.

JÄGER, S., JAHNKE, A., WILMES, T., ADEBAHR, S., VÖGTLE, F. N., DELIMA-HAHN, E., PFEIFER, D., BERG, T., LÜBBERT, M. & TREPEL, M. 2007. Leukemia targeting ligands isolated from phage display peptide libraries. *Leukemia,* 21**,** 411-420.

JARNUCZAK, A. F., NAJGEBAUER, H., BARZINE, M., KUNDU, D. J., GHAVIDEL, F., PEREZ-RIVEROL, Y., PAPATHEODOROU, I., BRAZMA, A. & VIZCAÍNO, J. A. 2021. An integrated landscape of protein expression in human cancer. *Scientific Data,* 8**,** 115.

JAROSZEWICZ, W., MORCINEK-ORŁOWSKA, J., PIERZYNOWSKA, K., GAFFKE, L. & WĘGRZYN, G. 2022. Phage display and other peptide display technologies. *FEMS Microbiol Rev,* 46.

JENKINS, B. D., MARTINI, R. N., HIRE, R., BROWN, A., BENNETT, B., BROWN, I., HOWERTH, E. W., EGAN, M., HODGSON, J., YATES, C., KITTLES, R., CHITALE, D., ALI, H., NATHANSON, D., NIKOLINAKOS, P., NEWMAN, L., MONTEIL, M. &

DAVIS, M. B. 2019. Atypical Chemokine Receptor 1 (DARC/ACKR1) in Breast Tumors Is Associated with Survival, Circulating Chemokines, Tumor-Infiltrating Immune Cells, and African Ancestry. *Cancer Epidemiol Biomarkers Prev,* 28**,** 690-700.

JÉZÉQUEL, P., GOURAUD, W., BEN AZZOUZ, F., GUÉRIN-CHARBONNEL, C., JUIN, P. P., LASLA, H. & CAMPONE, M. 2021. bc-GenExMiner 4.5: new mining module computes breast cancer differential gene expression analyses. *Database (Oxford),* 2021.

JIANG, D., ZHANG, X., LIU, M., WANG, Y., WANG, T., PEI, L., WANG, P., YE, H., SHI, J., SONG, C., WANG, K., WANG, X., DAI, L. & ZHANG, J. 2021. Discovering Panel of Autoantibodies for Early Detection of Lung Cancer Based on Focused Protein Array. *Front Immunol,* 12**,** 658922.

JIANG, Y., LIU, Y., TAN, X., YU, S. & LUO, J. 2019. TPX2 as a Novel Prognostic Indicator and Promising Therapeutic Target in Triple-negative Breast Cancer. *Clin Breast Cancer,* 19**,** 450-455.

JOHNS, L. E., MOSS, S. M. & AGE TRIAL MANAGEMENT, G. 2010. False-positive results in the randomized controlled trial of mammographic screening from age 40 ("Age" trial). *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology,* 19**,** 2758-2764.

JOHNSTON, S. R. D. & DOWSETT, M. 2003. Aromatase inhibitors for breast cancer: lessons from the laboratory. *Nature Reviews Cancer,* 3**,** 821-831.

JONES, K. M., KARANAM, B., JONES-TRICHE, J., SANDEY, M., HENDERSON, H. J., SAMANT, R. S., TEMESGEN, S., YATES, C. & BEDI, D. 2018. Phage Ligands for Identification of Mesenchymal-Like Breast Cancer Cells and Cancer-Associated Fibroblasts. *Frontiers in oncology,* 8**,** 625-625.

KALOYANNIDIS, P., HERTZBERG, M., WEBB, K., ZOMAS, A., SCHROVER, R., HURST, M., JACOB, I., NIKOGLOU, T. & CONNORS, J. M. 2020. Brentuximab vedotin for the treatment of patients with relapsed or refractory Hodgkin lymphoma after autologous stem cell transplantation. *British Journal of Haematology,* 188**,** 540-549.

KANEKO, K., ISHIGAMI, S., KIJIMA, Y., FUNASAKO, Y., HIRATA, M., OKUMURA, H., SHINCHI, H., KORIYAMA, C., UENO, S., YOSHINAKA, H. & NATSUGOE, S. 2011. Clinical implication of HLA class I expression in breast cancer. *BMC Cancer,* 11**,** 454.

KATAYAMA, H., BOLDT, C., LADD, J. J., JOHNSON, M. M., CHAO, T., CAPELLO, M., SUO, J., MAO, J., MANSON, J. E., PRENTICE, R., ESTEVA, F., WANG, H., DISIS, M. L. & HANASH, S. 2015. An Autoimmune Response Signature Associated with the Development of Triple-Negative Breast Cancer Reflects Disease Pathogenesis. *Cancer Res,* 75**,** 3246-54.

KATHRIKOLLY, T., NAIR, S. N., MATHEW, A., SAXENA, P. P. U. & NAIR, S. 2022. Can serum autoantibodies be a potential early detection biomarker for breast cancer in women? A diagnostic test accuracy review and meta-analysis. *Systematic Reviews,* 11**,** 215.

KELLER, A., NESVIZHSKII, A. I., KOLKER, E. & AEBERSOLD, R. 2002. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Analytical Chemistry,* 74**,** 5383-5392.

KELLY, K. A. & JONES, D. A. 2003. Isolation of a colon tumor specific binding peptide using phage display selection. *Neoplasia (New York, N.Y.),* 5**,** 437-444.

KESKIN, D. B., ANANDAPPA, A. J., SUN, J., TIROSH, I., MATHEWSON, N. D., LI, S., OLIVEIRA, G., GIOBBIE-HURDER, A., FELT, K., GJINI, E., SHUKLA, S. A., HU, Z., LI,

L., LE, P. M., ALLESØE, R. L., RICHMAN, A. R., KOWALCZYK, M. S., ABDELRAHMAN, S., GEDULDIG, J. E., CHARBONNEAU, S., PELTON, K., IORGULESCU, J. B., ELAGINA, L., ZHANG, W., OLIVE, O., MCCLUSKEY, C., OLSEN, L. R., STEVENS, J., LANE, W. J., SALAZAR, A. M., DALEY, H., WEN, P. Y., CHIOCCA, E. A., HARDEN, M., LENNON, N. J., GABRIEL, S., GETZ, G., LANDER, E. S., REGEV, A., RITZ, J., NEUBERG, D., RODIG, S. J., LIGON, K. L., SUVÀ, M. L., WUCHERPFENNIG, K. W., HACOHEN, N., FRITSCH, E. F., LIVAK, K. J., OTT, P. A., WU, C. J. & REARDON, D. A. 2019. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature,* 565**,** 234-239.

KHOSRAVI-SHAHI, P., CABEZON-GUTIERREZ, L. & APARICIO SALCEDO, M. I. 2019. State of art of advanced triple negative breast cancer. *Breast J*.

KIM, S. B., DENT, R., IM, S. A., ESPIÉ, M., BLAU, S., TAN, A. R., ISAKOFF, S. J., OLIVEIRA, M., SAURA, C., WONGCHENKO, M. J., KAPP, A. V., CHAN, W. Y., SINGEL, S. M., MASLYAR, D. J. & BASELGA, J. 2017. Ipatasertib plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer (LOTUS): a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Oncol,* 18**,** 1360-1372.

KING, M.-C., MARKS, J. H. & MANDELL, J. B. 2003. Breast and Ovarian Cancer Risks Due to Inherited Mutations in BRCA1 and <em>BRCA2</em&gt. *Science,* 302**,** 643.

KIYONO, T., FOSTER, S. A., KOOP, J. I., MCDOUGALL, J. K., GALLOWAY, D. A. & KLINGELHUTZ, A. J. 1998. Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature,* 396**,** 84-8.

KLADE, C. S., VOSS, T., KRYSTEK, E., AHORN, H., ZATLOUKAL, K., PUMMER, K. & ADOLF, G. R. 2001. Identification of tumor antigens in renal cell carcinoma by serological proteome analysis. *PROTEOMICS,* 1**,** 890-898.

KLEIN, J., TRAN, W., WATKINS, E., VESPRINI, D., WRIGHT, F. C., LOOK HONG, N. J., GHANDI, S., KISS, A. & CZARNOTA, G. J. 2019. Locally advanced breast cancer treated with neoadjuvant chemotherapy and adjuvant radiotherapy: a retrospective cohort analysis. *BMC Cancer,* 19**,** 306.

KO, Y. S., JIN, H., LEE, J. S., PARK, S. W., CHANG, K. C., KANG, K. M., JEONG, B. K. & KIM, H. J. 2018. Radioresistant breast cancer cells exhibit increased resistance to chemotherapy and enhanced invasive properties due to cancer stem cells. *Oncol Rep,* 40**,** 3752-3762.

KOBOLDT, D. C., FULTON, R. S., MCLELLAN, M. D., SCHMIDT, H., KALICKI-VEIZER, J., MCMICHAEL, J. F., FULTON, L. L., DOOLING, D. J., DING, L., MARDIS, E. R., WILSON, R. K., ALLY, A., BALASUNDARAM, M., BUTTERFIELD, Y. S. N., CARLSEN, R., CARTER, C., CHU, A., CHUAH, E., CHUN, H.-J. E., COOPE, R. J. N., DHALLA, N., GUIN, R., HIRST, C., HIRST, M., HOLT, R. A., LEE, D., LI, H. I., MAYO, M., MOORE, R. A., MUNGALL, A. J., PLEASANCE, E., GORDON ROBERTSON, A., SCHEIN, J. E., SHAFIEI, A., SIPAHIMALANI, P., SLOBODAN, J. R., STOLL, D., TAM, A., THIESSEN, N., VARHOL, R. J., WYE, N., ZENG, T., ZHAO, Y., BIROL, I., JONES, S. J. M., MARRA, M. A., CHERNIACK, A. D., SAKSENA, G., ONOFRIO, R. C., PHO, N. H., CARTER, S. L., SCHUMACHER, S. E., TABAK, B., HERNANDEZ, B., GENTRY, J., NGUYEN, H., CRENSHAW, A., ARDLIE, K., BEROUKHIM, R., WINCKLER, W., GETZ, G., GABRIEL, S. B., MEYERSON, M., CHIN, L., PARK, P. J., KUCHERLAPATI, R., HOADLEY, K. A., TODD AUMAN, J., FAN, C., TURMAN, Y. J., SHI, Y., LI, L., TOPAL, M. D., HE, X., CHAO, H.-H., PRAT, A., SILVA, G. O., IGLESIA, M. D., ZHAO, W., USARY, J., BERG, J. S., ADAMS, M., BOOKER, J., WU, J., GULABANI, A., BODENHEIMER, T., HOYLE, A. P., SIMONS, J. V., SOLOWAY, M. G., MOSE, L. E., JEFFERYS, S. R., BALU, S., PARKER, J. S., NEIL HAYES, D., PEROU, C. M., MALIK,

S., MAHURKAR, S., SHEN, H., WEISENBERGER, D. J., TRICHE JR, T., et al. 2012. Comprehensive molecular portraits of human breast tumours. *Nature,* 490**,** 61-70.

KOOREMAN, N. G., KIM, Y., DE ALMEIDA, P. E., TERMGLINCHAN, V., DIECKE, S., SHAO, N. Y., WEI, T. T., YI, H., DEY, D., NELAKANTI, R., BROUWER, T. P., PAIK, D. T., SAGIV-BARFI, I., HAN, A., QUAX, P. H. A., HAMMING, J. F., LEVY, R., DAVIS, M. M. & WU, J. C. 2018. Autologous iPSC-Based Vaccines Elicit Anti-tumor Responses In Vivo. *Cell Stem Cell,* 22**,** 501-513 e7.

KOSTIANETS, O., SHYYAN, M., ANTONIUK, S. V., FILONENKO, V. & KIYAMOVA, R. 2017. Panel of SEREX-defined antigens for breast cancer autoantibodies profile detection. *Biomarkers,* 22**,** 149-156.

KOZIOL, J. A., IMAI, H., DAI, L., ZHANG, J.-Y. & TAN, E. M. 2018. Early detection of hepatocellular carcinoma using autoantibody profiles from a panel of tumor-associated antigens. *Cancer immunology, immunotherapy : CII,* 67**,** 835-841.

KRIJGSMAN, O., ROEPMAN, P., ZWART, W., CARROLL, J. S., TIAN, S., DE SNOO, F. A., BENDER, R. A., BERNARDS, R. & GLAS, A. M. 2012. A diagnostic gene profile for molecular subtyping of breast cancer associated with treatment response. *Breast Cancer Research and Treatment,* 133**,** 37-47.

KROP, I. E., KIM, S.-B., MARTIN, A. G., LORUSSO, P. M., FERRERO, J.-M., BADOVINAC-CRNJEVIC, T., HOERSCH, S., SMITT, M. & WILDIERS, H. 2017. Trastuzumab emtansine versus treatment of physician's choice in patients with previously treated HER2-positive metastatic breast cancer (TH3RESA): final overall survival results from a randomised open-label phase 3 trial. *The Lancet Oncology,* 18**,** 743-754.

KROP, I. E., LIN, N. U., BLACKWELL, K., GUARDINO, E., HUOBER, J., LU, M., MILES, D., SAMANT, M., WELSLAU, M. & DIÉRAS, V. 2015. Trastuzumab emtansine (T-DM1) versus lapatinib plus capecitabine in patients with HER2-positive metastatic breast cancer and central nervous system metastases: a retrospective, exploratory analysis in EMILIA†. *Annals of Oncology,* 26**,** 113-119.

KRUMPE, L. R. & MORI, T. 2007. Potential of phage-displayed peptide library technology to identify functional targeting peptides. *Expert opinion on drug discovery,* 2**,** 525-525.

KUERER, H. M., ALBARRACIN, C. T., YANG, W. T., CARDIFF, R. D., BREWSTER, A. M., SYMMANS, W. F., HYLTON, N. M., MIDDLETON, L. P., KRISHNAMURTHY, S., PERKINS, G. H., BABIERA, G., EDGERTON, M. E., CZERNIECKI, B. J., ARUN, B. K. & HORTOBAGYI, G. N. 2009. Ductal Carcinoma in Situ: State of the Science and Roadmap to Advance the Field. *Journal of Clinical Oncology,* 27**,** 279-288.

LACOMBE, J., MANGÉ, A., BOUGNOUX, A. C., PRASSAS, I. & SOLASSOL, J. 2014. A multiparametric serum marker panel as a complementary test to mammography for the diagnosis of node-negative early-stage breast cancer and DCIS in young women. *Cancer Epidemiol Biomarkers Prev,* 23**,** 1834-42.

LAM, S., BOYLE, P., HEALEY, G. F., MADDISON, P., PEEK, L., MURRAY, A., CHAPMAN, C. J., ALLEN, J., WOOD, W. C., SEWELL, H. F. & ROBERTSON, J. F. R. 2011. EarlyCDT-Lung: An Immunobiomarker Test as an Aid to Early Detection of Lung Cancer. *Cancer Prevention Research,* 4**,** 1126-1134.

LÁNCZKY, A. & GYŐRFFY, B. 2021. Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. *J Med Internet Res,* 23**,** e27633.

LANGLEY, S. R. & MAYR, M. 2015. Comparative analysis of statistical methods used for detecting differential expression in label-free mass spectrometry proteomics. *Journal of Proteomics,* 129**,** 83-92.

LAPIDOT, T., SIRARD, C., VORMOOR, J., MURDOCH, B., HOANG, T., CACERES-CORTES, J., MINDEN, M., PATERSON, B., CALIGIURI, M. A. & DICK, J. E. 1994. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature,* 367**,** 645-648.

LE JONCOUR, V. & LAAKKONEN, P. 2018. Seek & Destroy, use of targeting peptides for cancer detection and drug delivery. *Bioorganic & Medicinal Chemistry,* 26**,** 2797-2806.

LEDSGAARD, L., KILSTRUP, M., KARATT-VELLATT, A., MCCAFFERTY, J. & LAUSTSEN, A. H. 2018. Basics of Antibody Phage Display Technology. *Toxins (Basel),* 10.

LEE, R. J., VALLOW, L. A., MCLAUGHLIN, S. A., TZOU, K. S., HINES, S. L. & PETERSON, J. L. 2012. Ductal carcinoma in situ of the breast. *International journal of surgical oncology,* 2012**,** 123549-123549.

LEHMANN, B. D., ABRAMSON, V. G., SANDERS, M. E., MAYER, E. L., HADDAD, T. C., NANDA, R., VAN POZNAK, C., STORNIOLO, A. M., NANGIA, J. R., GONZALEZ-ERICSSON, P. I., SANCHEZ, V., JOHNSON, K. N., ABRAMSON, R. G., CHEN, S. C., SHYR, Y., ARTEAGA, C. L., WOLFF, A. C. & PIETENPOL, J. A. 2020. TBCRC 032 IB/II Multicenter Study: Molecular Insights to AR Antagonist and PI3K Inhibitor Efficacy in Patients with AR(+) Metastatic Triple-Negative Breast Cancer. *Clin Cancer Res,* 26**,** 2111-2123.

LEHMANN, B. D., BAUER, J. A., CHEN, X., SANDERS, M. E., CHAKRAVARTHY, A. B., SHYR, Y. & PIETENPOL, J. A. 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest,* 121**,** 2750-67.

LEHMANN, B. D., COLAPRICO, A., SILVA, T. C., CHEN, J., AN, H., BAN, Y., HUANG, H., WANG, L., JAMES, J. L., BALKO, J. M., GONZALEZ-ERICSSON, P. I., SANDERS, M. E., ZHANG, B., PIETENPOL, J. A. & CHEN, X. S. 2021. Multi-omics analysis identifies therapeutic vulnerabilities in triple-negative breast cancer subtypes. *Nature Communications,* 12**,** 6276.

LEHMANN, B. D., JOVANOVIC, B., CHEN, X., ESTRADA, M. V., JOHNSON, K. N., SHYR, Y., MOSES, H. L., SANDERS, M. E. & PIETENPOL, J. A. 2016. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. *PLoS One,* 11**,** e0157368.

LEUZZI, G., MARABITTI, V., PICHIERRI, P. & FRANCHITTO, A. 2016. WRNIP1: A new guardian of genome integrity at stalled replication forks. *Mol Cell Oncol,* 3**,** e1215777.

LEWIS, E. E. L., CHILD, H. W., HURSTHOUSE, A., STIRLING, D., MCCULLY, M., PATERSON, D., MULLIN, M. & BERRY, C. C. 2015. The influence of particle size and static magnetic fields on the uptake of magnetic nanoparticles into three dimensional cell-seeded collagen gel cultures. *Journal of Biomedical Materials Research Part B: Applied Biomaterials,* 103**,** 1294-1301.

LI, G., MILES, A., LINE, A. & REES, R. C. 2004. Identification of tumour antigens by serological analysis of cDNA expression cloning. *Cancer Immunology, Immunotherapy,* 53**,** 139-143.

LI, S. Y., SUN, R., WANG, H. X., SHEN, S., LIU, Y., DU, X. J., ZHU, Y. H. & JUN, W. 2015. Combination therapy with epigenetic-targeted and chemotherapeutic drugs delivered by nanoparticles to enhance the chemotherapy response and overcome resistance by breast cancer stem cells. *J Control Release,* 205**,** 7-14.

LI, W., MA, H., ZHANG, J., ZHU, L., WANG, C. & YANG, Y. 2017. Unraveling the roles of CD44/CD24 and ALDH1 as cancer stem cell markers in tumorigenesis and metastasis. *Sci Rep,* 7**,** 13856.

LI, Z., LIU, C. P., HE, Y. L., TIAN, Y. & HUANG, T. 2008. [Analysis on clone in vitro and tumorigenic capacity in vivo of different subsets cells from the MCF-7 human breast cancer cell line]. *Sichuan Da Xue Xue Bao Yi Xue Ban,* 39**,** 547-9.

LIANG, Y., LÜ, W., ZHANG, X. & LÜ, B. 2018. Tumor-infiltrating CD8+ and FOXP3+ lymphocytes before and after neoadjuvant chemotherapy in cervical cancer. *Diagnostic pathology,* 13**,** 93-93.

LIAO, G., JIANG, Z., YANG, Y., ZHANG, C., JIANG, M., ZHU, J., XU, L., XIE, A., YAN, M., ZHANG, Y., XIAO, Y. & LI, X. 2021. Combined homologous recombination repair deficiency and immune activation analysis for predicting intensified responses of anthracycline, cyclophosphamide and taxane chemotherapy in triple-negative breast cancer. *BMC Medicine,* 19**,** 190.

LIN, H.-S., TALWAR, H. S., TARCA, A. L., IONAN, A., CHATTERJEE, M., YE, B., WOJCIECHOWSKI, J., MOHAPATRA, S., BASSON, M. D., YOO, G. H., PESHEK, B., LONARDO, F., PAN, C.-J. G., FOLBE, A. J., DRAGHICI, S., ABRAMS, J. & TAINSKY, M. A. 2007. Autoantibody Approach for Serum-Based Detection of Head and Neck Cancer. *Cancer Epidemiology, Biomarkers & Prevention,* 16**,** 2396-2405.

LITIÈRE, S., WERUTSKY, G., FENTIMAN, I. S., RUTGERS, E., CHRISTIAENS, M.-R., VAN LIMBERGEN, E., BAAIJENS, M. H. A., BOGAERTS, J. & BARTELINK, H. 2012. Breast conserving therapy versus mastectomy for stage I–II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. *The Lancet Oncology,* 13**,** 412-419.

LITTON, J. K., HURVITZ, S. A., MINA, L. A., RUGO, H. S., LEE, K. H., GONÇALVES, A., DIAB, S., WOODWARD, N., GOODWIN, A., YERUSHALMI, R., ROCHÉ, H., IM, Y. H., EIERMANN, W., QUEK, R. G. W., USARI, T., LANZALONE, S., CZIBERE, A., BLUM, J. L., MARTIN, M. & ETTL, J. 2020. Talazoparib versus chemotherapy in patients with germline BRCA1/2-mutated HER2-negative advanced breast cancer: final overall survival results from the EMBRACA trial. *Ann Oncol,* 31**,** 1526-1535.

LITTON, J. K., RUGO, H. S., ETTL, J., HURVITZ, S. A., GONÇALVES, A., LEE, K.-H., FEHRENBACHER, L., YERUSHALMI, R., MINA, L. A., MARTIN, M., ROCHÉ, H., IM, Y.-H., QUEK, R. G. W., MARKOVA, D., TUDOR, I. C., HANNAH, A. L., EIERMANN, W. & BLUM, J. L. 2018. Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. *New England Journal of Medicine,* 379**,** 753-763.

LIU, F., QI, C. L., KONG, M., LIU, T. T., LI, L. & LI, B. J. 2016. Screening specific polypeptides of breast cancer stem cells from a phage display random peptide library. *Oncol Lett,* 12**,** 4727-4731.

LIU, H. L., HUA, M. Y., YANG, H. W., HUANG, C. Y., CHU, P. C., WU, J. S., TSENG, I. C., WANG, J. J., YEN, T. C., CHEN, P. Y. & WEI, K. C. 2010. Magnetic resonance monitoring of focused ultrasound/magnetic nanoparticle targeting delivery of therapeutic agents to the brain. *Proc Natl Acad Sci U S A,* 107**,** 15205-10.

LIU, J. C., VOISIN, V., BADER, G. D., DENG, T., PUSZTAI, L., SYMMANS, W. F., ESTEVA, F. J., EGAN, S. E. & ZACKSENHAUS, E. 2012. Seventeen-gene signature from enriched Her2/Neu mammary tumor-initiating cells predicts clinical outcome for human HER2+:ERalpha- breast cancer. *Proc Natl Acad Sci U S A,* 109**,** 5832-7.

LIU, S., CONG, Y., WANG, D., SUN, Y., DENG, L., LIU, Y., MARTIN-TREVINO, R., SHANG, L., MCDERMOTT, S. P., LANDIS, M. D., HONG, S., ADAMS, A., D'ANGELO, R., GINESTIER, C., CHARAFE-JAUFFRET, E., CLOUTHIER, S. G., BIRNBAUM, D.,

WONG, S. T., ZHAN, M., CHANG, J. C. & WICHA, M. S. 2014. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports,* 2**,** 78-91.

LOI, S., DRUBAY, D., ADAMS, S., PRUNERI, G., FRANCIS, P. A., LACROIX-TRIKI, M., JOENSUU, H., DIECI, M. V., BADVE, S., DEMARIA, S., GRAY, R., MUNZONE, E., LEMONNIER, J., SOTIRIOU, C., PICCART, M. J., KELLOKUMPU-LEHTINEN, P.-L., VINGIANI, A., GRAY, K., ANDRE, F., DENKERT, C., SALGADO, R. & MICHIELS, S. 2019. Tumor-Infiltrating Lymphocytes and Prognosis: A Pooled Individual Patient Analysis of Early-Stage Triple-Negative Breast Cancers. *Journal of Clinical Oncology,* 37**,** 559-569.

LOMAN, N., JOHANNSSON, O., KRISTOFFERSSON, U., OLSSON, H. K. & BORG, A. K. 2001. Family History of Breast and Ovarian Cancers and BRCA1 and BRCA2 Mutations in a Population-Based Series of Early-Onset Breast Cancer. *JNCI: Journal of the National Cancer Institute,* 93**,** 1215-1223.

LONE, S. N., NISAR, S., MASOODI, T., SINGH, M., RIZWAN, A., HASHEM, S., EL-RIFAI, W., BEDOGNETTI, D., BATRA, S. K., HARIS, M., BHAT, A. A. & MACHA, M. A. 2022. Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. *Molecular Cancer,* 21**,** 79.

LONGLEY, D. & JOHNSTON, P. 2005. Molecular mechanisms of drug resistance. *The Journal of Pathology,* 205**,** 275-292.

LONSDALE, J., THOMAS, J., SALVATORE, M., PHILLIPS, R., LO, E., SHAD, S., HASZ, R., WALTERS, G., GARCIA, F., YOUNG, N., FOSTER, B., MOSER, M., KARASIK, E., GILLARD, B., RAMSEY, K., SULLIVAN, S., BRIDGE, J., MAGAZINE, H., SYRON, J., FLEMING, J., SIMINOFF, L., TRAINO, H., MOSAVEL, M., BARKER, L., JEWELL, S., ROHRER, D., MAXIM, D., FILKINS, D., HARBACH, P., CORTADILLO, E., BERGHUIS, B., TURNER, L., HUDSON, E., FEENSTRA, K., SOBIN, L., ROBB, J., BRANTON, P., KORZENIEWSKI, G., SHIVE, C., TABOR, D., QI, L., GROCH, K., NAMPALLY, S., BUIA, S., ZIMMERMAN, A., SMITH, A., BURGES, R., ROBINSON, K., VALENTINO, K., BRADBURY, D., COSENTINO, M., DIAZ-MAYORAL, N., KENNEDY, M., ENGEL, T., WILLIAMS, P., ERICKSON, K., ARDLIE, K., WINCKLER, W., GETZ, G., DELUCA, D., MACARTHUR, D., KELLIS, M., THOMSON, A., YOUNG, T., GELFAND, E., DONOVAN, M., MENG, Y., GRANT, G., MASH, D., MARCUS, Y., BASILE, M., LIU, J., ZHU, J., TU, Z., COX, N. J., NICOLAE, D. L., GAMAZON, E. R., IM, H. K., KONKASHBAEV, A., PRITCHARD, J., STEVENS, M., FLUTRE, T., WEN, X., DERMITZAKIS, E. T., LAPPALAINEN, T., GUIGO, R., MONLONG, J., SAMMETH, M., KOLLER, D., BATTLE, A., MOSTAFAVI, S., MCCARTHY, M., RIVAS, M., MALLER, J., RUSYN, I., NOBEL, A., WRIGHT, F., SHABALIN, A., FEOLO, M., SHAROPOVA, N., et al. 2013. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics,* 45**,** 580-585.

LÓPEZ DE ANDRÉS, J., GRIÑÁN-LISÓN, C., JIMÉNEZ, G. & MARCHAL, J. A. 2020. Cancer stem cell secretome in the tumor microenvironment: a key point for an effective personalized cancer treatment. *Journal of Hematology & Oncology,* 13**,** 136.

LOURENCO, A. P., BENSON, K. L., HENDERSON, M. C., SILVER, M., LETSIOS, E., TRAN, Q., GORDON, K. J., BORMAN, S., CORN, C., MULPURI, R., SMITH, W., ALPERS, J., COSTANTINI, C., ROHATGI, N., YANG, R., HAYTHEM, A., BIREN, S., MORRIS, M., KASS, F. & REESE, D. E. 2017. A Noninvasive Blood-based Combinatorial Proteomic Biomarker Assay to Detect Breast Cancer in Women Under the Age of 50 Years. *Clinical Breast Cancer,* 17**,** 516-525.e6.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15**,** 550.

LOVRICS, P. J., CORNACCHI, S. D., FARROKHYAR, F., GARNETT, A., CHEN, V., FRANIC, S. & SIMUNOVIC, M. 2009. The relationship between surgical factors and margin status after breast-conservation surgery for early stage breast cancer. *The American Journal of Surgery,* 197**,** 740-746.

LOWMAN, H. B. 1997. BACTERIOPHAGE DISPLAY AND DISCOVERY OF PEPTIDE LEADS FOR DRUG DEVELOPMENT. *Annual Review of Biophysics and Biomolecular Structure,* 26**,** 401-424.

LU, Y., ZI, X., ZHAO, Y., MASCARENHAS, D. & POLLAK, M. 2001. Insulin-Like Growth Factor-I Receptor Signaling and Resistance to Trastuzumab (Herceptin). *JNCI: Journal of the National Cancer Institute,* 93**,** 1852-1857.

MALHOTRA, G. K., ZHAO, X., BAND, H. & BAND, V. 2010. Histological, molecular and functional subtypes of breast cancers. *Cancer Biol Ther,* 10**,** 955-60.

MANDEL-BREHM, C., DUBEY, D., KRYZER, T. J., O'DONOVAN, B. D., TRAN, B., VAZQUEZ, S. E., SAMPLE, H. A., ZORN, K. C., KHAN, L. M., BLEDSOE, I. O., MCKEON, A., PLEASURE, S. J., LENNON, V. A., DERISI, J. L., WILSON, M. R. & PITTOCK, S. J. 2019. Kelch-like Protein 11 Antibodies in Seminoma-Associated Paraneoplastic Encephalitis. *New England Journal of Medicine,* 381**,** 47-54.

MANDELSON, M. T., OESTREICHER, N., PORTER, P. L., WHITE, D., FINDER, C. A., TAPLIN, S. H. & WHITE, E. 2000. Breast Density as a Predictor of Mammographic Detection: Comparison of Interval- and Screen-Detected Cancers. *JNCI: Journal of the National Cancer Institute,* 92**,** 1081-1087.

MANN, R. M., KUHL, C. K. & MOY, L. 2019. Contrast-enhanced MRI for breast cancer screening. *Journal of Magnetic Resonance Imaging,* 50**,** 377-390.

MANNU, G. S., WANG, Z., BROGGIO, J., CHARMAN, J., CHEUNG, S., KEARINS, O., DODWELL, D. & DARBY, S. C. 2020. Invasive breast cancer and breast cancer mortality after ductal carcinoma in situ in women attending for breast screening in England, 1988-2014: population based observational cohort study. *BMJ,* 369**,** m1570.

MANSOUR, S., ADHYA, I., LEBLEU, C., DUMPATI, R., REHAN, A., CHALL, S., DAI, J., ERRASTI, G., DELACROIX, T. & CHAKRABARTI, R. 2022. Identification of a novel peptide ligand for the cancer-specific receptor mutation EGFRvIII using high-throughput sequencing of phage-selected peptides. *Scientific Reports,* 12**,** 20725.

MARABITTI, V., LILLO, G., MALACARIA, E., PALERMO, V., PICHIERRI, P. & FRANCHITTO, A. 2020. Checkpoint Defects Elicit a WRNIP1-Mediated Response to Counteract R-Loop-Associated Genomic Instability. *Cancers,* 12**,** 389.

MARCOS-CAMPOS, I., ASÍN, L., TORRES, T. E., MARQUINA, C., TRES, A., IBARRA, M. R. & GOYA, G. F. 2011. Cell death induced by the application of alternating magnetic fields to nanoparticle-loaded dendritic cells. *Nanotechnology,* 22**,** 205101.

MARKIDES, H., NEWELL, K. J., RUDORF, H., FERRERAS, L. B., DIXON, J. E., MORRIS, R. H., GRAVES, M., KAGGIE, J., HENSON, F. & EL HAJ, A. J. 2019. Ex vivo MRI cell tracking of autologous mesenchymal stromal cells in an ovine osteochondral defect model. *Stem Cell Research & Therapy,* 10**,** 25.

MARKS, J. D., HOOGENBOOM, H. R., GRIFFITHS, A. D. & WINTER, G. 1992. Molecular evolution of proteins on filamentous phage. Mimicking the strategy of the immune system. *Journal of Biological Chemistry,* 267**,** 16007-16010.

MARMOT, M. G., ALTMAN, D. G., CAMERON, D. A., DEWAR, J. A., THOMPSON, S. G. & WILCOX, M. 2013. The benefits and harms of breast cancer screening: an independent review. *British journal of cancer,* 108**,** 2205-2240.

MARQUS, S., PIROGOVA, E. & PIVA, T. J. 2017. Evaluation of the use of therapeutic peptides for cancer treatment. *Journal of Biomedical Science,* 24**,** 21.

MARTINKOVICH, S., SHAH, D., PLANEY, S. L. & ARNOTT, J. A. 2014. Selective estrogen receptor modulators: tissue specificity and clinical utility. *Clinical interventions in aging,* 9**,** 1437-1452.

MASUDA, H., BAGGERLY, K. A., WANG, Y., ZHANG, Y., GONZALEZ-ANGULO, A. M., MERIC-BERNSTAM, F., VALERO, V., LEHMANN, B. D., PIETENPOL, J. A., HORTOBAGYI, G. N., SYMMANS, W. F. & UENO, N. T. 2013. Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. *Clin Cancer Res,* 19**,** 5533-40.

MATOCHKO, W. L., CHU, K., JIN, B., LEE, S. W., WHITESIDES, G. M. & DERDA, R. 2012. Deep sequencing analysis of phage libraries using Illumina platform. *Methods,* 58**,** 47-55.

MATOCHKO, W. L., CORY LI, S., TANG, S. K. Y. & DERDA, R. 2014. Prospective identification of parasitic sequences in phage display screens. *Nucleic acids research,* 42**,** 1784-1798.

MATSON, D. R., DENU, R. A., ZASADIL, L. M., BURKARD, M. E., WEAVER, B. A., FLYNN, C. & STUKENBERG, P. T. 2021. High nuclear TPX2 expression correlates with TP53 mutation and poor clinical behavior in a large breast cancer cohort, but is not an independent predictor of chromosomal instability. *BMC Cancer,* 21**,** 186.

MATSUMOTO, Y., ITOU, J., SATO, F. & TOI, M. 2018. SALL4 - KHDRBS3 network enhances stemness by modulating CD44 splicing in basal-like breast cancer. *Cancer Med,* 7**,** 454-462.

MCCAFFERTY, J., GRIFFITHS, A. D., WINTER, G. & CHISWELL, D. J. 1990. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature,* 348**,** 552-4.

MCCANN, J., STOCKTON, D. & GODWARD, S. 2002. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Research,* 4**,** R11.

MCGOWAN, P. M., SIMEDREA, C., RIBOT, E. J., FOSTER, P. J., PALMIERI, D., STEEG, P. S., ALLAN, A. L. & CHAMBERS, A. F. 2011. Notch1 inhibition alters the CD44hi/CD24lo population and reduces the formation of brain metastases from breast cancer. *Mol Cancer Res,* 9**,** 834-44.

MCGREGOR, D. P. 2008. Discovering and improving novel peptide therapeutics. *Current Opinion in Pharmacology,* 8**,** 616-619.

MCGUIRE, M. J., GRAY, B. P., LI, S., CUPKA, D., BYERS, L. A., WU, L., REZAIE, S., LIU, Y.-H., PATTISAPU, N., ISSAC, J., OYAMA, T., DIAO, L., HEYMACH, J. V., XIE, X.-J., MINNA, J. D. & BROWN, K. C. 2014. Identification and Characterization of a Suite of Tumor Targeting Peptides for Non-Small Cell Lung Cancer. *Scientific Reports,* 4**,** 4480.

MCGUIRE, M. J., LI, S. & BROWN, K. C. 2009. Biopanning of phage displayed peptide libraries for the isolation of cell-specific ligands. *Methods in molecular biology (Clifton, N.J.),* 504**,** 291-321.

MCILWAIN, S., MATHEWS, M., BEREMAN, M. S., RUBEL, E. W., MACCOSS, M. J. & NOBLE, W. S. 2012. Estimating relative abundances of proteins from shotgun proteomics data. *BMC Bioinformatics,* 13**,** 308.

MEATTINI, I., PEZZULLA, D., SAIEVA, C., BERNINI, M., ORZALESI, L., SANCHEZ, L. J., DESIDERI, I., FRANCOLINI, G., BONOMO, P., GRETO, D., LOI, M., MANGONI, M., BRUNI, A., NORI, J., VEZZOSI, V., BIANCHI, S. & LIVI, L. 2018. Triple Negative

Apocrine Carcinomas as a Distinct Subtype of Triple Negative Breast Cancer: A Case-control Study. *Clin Breast Cancer,* 18**,** e773-e780.

MENDES, D., ALVES, C., AFONSO, N., CARDOSO, F., PASSOS-COELHO, J. L., COSTA, L., ANDRADE, S. & BATEL-MARQUES, F. 2015. The benefit of HER2-targeted therapies on overall survival of patients with metastatic HER2-positive breast cancer – a systematic review. *Breast Cancer Research,* 17**,** 140.

METCALFE, K. A., LUBINSKI, J., GRONWALD, J., HUZARSKI, T., MCCUAIG, J., LYNCH, H. T., KARLAN, B., FOULKES, W. D., SINGER, C. F., NEUHAUSEN, S. L., SENTER, L., EISEN, A., SUN, P., NAROD, S. A. & AND THE HEREDITARY BREAST CANCER CLINICAL STUDY, G. 2018. The risk of breast cancer in BRCA1 and BRCA2 mutation carriers without a first-degree relative with breast cancer. *Clinical Genetics,* 93**,** 1063-1068.

MICHIGAMI, M., TAKAHASHI, K., YAMASHITA, H., YE, Z., NAKASE, I. & FUJII, I. 2021. A "ligand-targeting" peptide-drug conjugate: Targeted intracellular drug delivery by VEGF-binding helix-loop-helix peptides via receptor-mediated endocytosis. *PLoS One,* 16**,** e0247045.

MILIOLI, H. H., TISHCHENKO, I., RIVEROS, C., BERRETTA, R. & MOSCATO, P. 2017. Basal-like breast cancer: molecular profiles, clinical features and survival outcomes. *BMC Med Genomics,* 10**,** 19.

MILLER, T. W., HENNESSY, B. T., GONZÁLEZ-ANGULO, A. M., FOX, E. M., MILLS, G. B., CHEN, H., HIGHAM, C., GARCÍA-ECHEVERRÍA, C., SHYR, Y. & ARTEAGA, C. L. 2010. Hyperactivation of phosphatidylinositol-3 kinase promotes escape from hormone dependence in estrogen receptor-positive human breast cancer. *The Journal of clinical investigation,* 120**,** 2406-2413.

MILOSEVIC, M., JANKOVIC, D., MILENKOVIC, A. & STOJANOV, D. 2018. Early diagnosis and detection of breast cancer. *Technol Health Care,* 26**,** 729-759.

MITCHELL, D. J., KIM, D. T., STEINMAN, L., FATHMAN, C. G. & ROTHBARD, J. B. 2000. Polyarginine enters cells more efficiently than other polycationic homopolymers. *J Pept Res,* 56**,** 318-25.

MITRA, A., FILLMORE, R. A., METGE, B. J., RAJESH, M., XI, Y., KING, J., JU, J., PANNELL, L., SHEVDE, L. A. & SAMANT, R. S. 2008. Large isoform of MRJ (DNAJB6) reduces malignant activity of breast cancer. *Breast Cancer Research,* 10**,** R22.

MITRA, A., MENEZES, M. E., SHEVDE, L. A. & SAMANT, R. S. 2010. DNAJB6 Induces Degradation of β-Catenin and Causes Partial Reversal of Mesenchymal Phenotype*. *Journal of Biological Chemistry,* 285**,** 24686-24694.

MITTENDORF, E. A., PHILIPS, A. V., MERIC-BERNSTAM, F., QIAO, N., WU, Y., HARRINGTON, S., SU, X., WANG, Y., GONZALEZ-ANGULO, A. M., AKCAKANAT, A., CHAWLA, A., CURRAN, M., HWU, P., SHARMA, P., LITTON, J. K., MOLLDREM, J. J. & ALATRASH, G. 2014. PD-L1 expression in triple-negative breast cancer. *Cancer Immunol Res,* 2**,** 361-70.

MIYASHITA, M., SASANO, H., TAMAKI, K., HIRAKAWA, H., TAKAHASHI, Y., NAKAGAWA, S., WATANABE, G., TADA, H., SUZUKI, A., OHUCHI, N. & ISHIDA, T. 2015. Prognostic significance of tumor-infiltrating CD8+ and FOXP3+ lymphocytes in residual tumors and alterations in these parameters after neoadjuvant chemotherapy in triple-negative breast cancer: a retrospective multicenter study. *Breast Cancer Res,* 17**,** 124.

MODI, S., JACOT, W., YAMASHITA, T., SOHN, J., VIDAL, M., TOKUNAGA, E., TSURUTANI, J., UENO, N. T., PRAT, A., CHAE, Y. S., LEE, K. S., NIIKURA, N., PARK, Y. H., XU, B., WANG, X., GIL-GIL, M., LI, W., PIERGA, J.-Y., IM, S.-A., MOORE, H. C. F., RUGO, H. S., YERUSHALMI, R., ZAGOURI, F., GOMBOS, A., KIM, S.-B., LIU, Q., LUO, T., SAURA, C., SCHMID, P., SUN, T., GAMBHIRE, D., YUNG, L., WANG, Y.,

SINGH, J., VITAZKA, P., MEINHARDT, G., HARBECK, N. & CAMERON, D. A. 2022. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. *New England Journal of Medicine,* 387**,** 9-20.

MODI, S., PUSZTAI, L., FORERO, A., MITA, M., MILLER, K. D., WEISE, A., KROP, I., BURRIS, H., KALINSKY, K., TSAI, M., LIU, M. C., HURVITZ, S. A., WILKS, S., ADEMUYIWA, F., DIAB, S., HAN, H. S., KATO, G., NANDA, R., SHAUGHNESSY, J., KOSTIC, A., LI, M. & SPECHT, J. 2018. Abstract PD3-14: Phase 1 study of the antibody-drug conjugate SGN-LIV1A in patients with heavily pretreated triple-negative metastatic breast cancer. *Cancer Research,* 78**,** PD3-14.

MODI, S., SAURA, C., YAMASHITA, T., PARK, Y. H., KIM, S.-B., TAMURA, K., ANDRE, F., IWATA, H., ITO, Y., TSURUTANI, J., SOHN, J., DENDULURI, N., PERRIN, C., AOGI, K., TOKUNAGA, E., IM, S.-A., LEE, K. S., HURVITZ, S. A., CORTES, J., LEE, C., CHEN, S., ZHANG, L., SHAHIDI, J., YVER, A. & KROP, I. 2019. Trastuzumab Deruxtecan in Previously Treated HER2-Positive Breast Cancer. *New England Journal of Medicine,* 382**,** 610-621.

MOHAMMED, M. E. & ABDELHAFIZ, K. 2015. Autoantibodies in the sera of breast cancer patients: Antinuclear and anti-double stranded DNA antibodies as example. *J Cancer Res Ther,* 11**,** 341-4.

MORRIS, M. C., DEPOLLIER, J., MERY, J., HEITZ, F. & DIVITA, G. 2001. A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nature Biotechnology,* 19**,** 1173-1176.

MORRISON, S. J. & KIMBLE, J. 2006. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature,* 441**,** 1068-1074.

MOSLER, T., BAYMAZ, H. I., GRÄF, J. F., MIKICIC, I., BLATTNER, G., BARTLETT, E., OSTERMAIER, M., PICCINNO, R., YANG, J., VOIGT, A., GATTI, M., PELLEGRINO, S., ALTMEYER, M., LUCK, K., AHEL, I., ROUKOS, V. & BELI, P. 2022. PARP1 proximity proteomics reveals interaction partners at stressed replication forks. *Nucleic Acids Research,* 50**,** 11600-11618.

MOSS, S. M., BROWN, J., GARVICAN, L., COLEMAN, D. A., JOHNS, L. E., BLANKS, R. G., RUBIN, G., OSWALD, J., PAGE, A., EVANS, A., GAMBLE, P., WILSON, R., LEE, L., LISTON, J., STURDY, L., SUTTON, G., WARDMAN, G., PATNICK, J. & WINDER, R. 2001. Routine breast screening for women aged 65-69: results from evaluation of the demonstration sites. *British journal of cancer,* 85**,** 1289-1294.

MOSS, S. M., WALE, C., SMITH, R., EVANS, A., CUCKLE, H. & DUFFY, S. W. 2015. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. *Lancet Oncol,* 16**,** 1123-1132.

MUSGROVE, E. A. & SUTHERLAND, R. L. 2009. Biological determinants of endocrine resistance in breast cancer. *Nature Reviews Cancer,* 9**,** 631.

NAGANO, K., IMAI, S., ZHAO, X., YAMASHITA, T., YOSHIOKA, Y., ABE, Y., MUKAI, Y., KAMADA, H., NAKAGAWA, S., TSUTSUMI, Y. & TSUNODA, S.-I. 2015. Identification and evaluation of metastasis-related proteins, oxysterol binding protein-like 5 and calumenin, in lung tumors. *Int J Oncol,* 47**,** 195-203.

NAGATA, Y., LAN, K.-H., ZHOU, X., TAN, M., ESTEVA, F. J., SAHIN, A. A., KLOS, K. S., LI, P., MONIA, B. P., NGUYEN, N. T., HORTOBAGYI, G. N., HUNG, M.-C. & YU, D. 2004. PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. *Cancer Cell,* 6**,** 117-127.

NAGAYAMA, A., VIDULA, N., ELLISEN, L. & BARDIA, A. 2020. Novel antibody–drug conjugates for triple negative breast cancer. *Therapeutic Advances in Medical Oncology,* 12**,** 1758835920915980.

NAGELE, E. P., HAN, M., ACHARYA, N. K., DEMARSHALL, C., KOSCIUK, M. C. & NAGELE, R. G. 2013. Natural IgG Autoantibodies Are Abundant and Ubiquitous in Human Sera, and Their Number Is Influenced By Age, Gender, and Disease. *PLOS ONE,* 8**,** e60726.

NAGY, P., FRIEDLÄNDER, E., TANNER, M., KAPANEN, A. I., CARRAWAY, K. L., ISOLA, J. & JOVIN, T. M. 2005. Decreased Accessibility and Lack of Activation of ErbB2 in JIMT-1, a Herceptin-Resistant, MUC4-Expressing Breast Cancer Cell Line. *Cancer Research,* 65**,** 473-482.

NAHTA, R., YU, D., HUNG, M.-C., HORTOBAGYI, G. N. & ESTEVA, F. J. 2006. Mechanisms of Disease: understanding resistance to HER2-targeted therapy in human breast cancer. *Nature Clinical Practice Oncology,* 3**,** 269-280.

NAQID, I. A., OWEN, J. P., MADDISON, B. C., SPILIOTOPOULOS, A., EMES, R. D., WARRY, A., TCHÓRZEWSKA, M. A., MARTELLI, F., GOSLING, R. J., DAVIES, R. H., LA RAGIONE, R. M. & GOUGH, K. C. 2016. Mapping polyclonal antibody responses to bacterial infection using next generation phage display. *Scientific Reports,* 6**,** 24232.

NEDELJKOVIC, M., TANIC, N., DRAMICANIN, T., MILOVANOVIC, Z., SUSNJAR, S., MILINKOVIC, V., VUJOVIC, I., PRVANOVIC, M. & TANIC, N. 2019. Importance of Copy Number Alterations of FGFR1 and C-MYC Genes in Triple Negative Breast Cancer. *J Med Biochem,* 38**,** 63-70.

NEGM, O., AHMED, O., FIGUEREDO, G., HAMED, M., POLLARD, L., GARIBALDI, J., SEWELL, H. & ROBERTSON, J. 2018. Abstract P2-02-08: Validation of an autoantibody blood test for the detection of early breast cancer (BC), particularly hormone receptor positive BC. *Cancer Research,* 78**,** P2-02-08-P2-02-08.

NEGM, O. H., HAMED, M. R., SCHOEN, R. E., WHELAN, R. L., STEELE, R. J., SCHOLEFIELD, J., DILNOT, E. M., SHANTHA KUMARA, H. M. C., ROBERTSON, J. F. R. & SEWELL, H. F. 2016. Human Blood Autoantibodies in the Detection of Colorectal Cancer. *PloS one,* 11**,** e0156971-e0156971.

NELSON, H. D., O'MEARA, E. S., KERLIKOWSKE, K., BALCH, S. & MIGLIORETTI, D. 2016. Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data. *Annals of internal medicine,* 164**,** 226-235.

NELSON, H. D., TYNE, K., NAIK, A., BOUGATSOS, C., CHAN, B. K., HUMPHREY, L. & FORCE, U. S. P. S. T. 2009. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Annals of internal medicine,* 151**,** 727-W242.

NESTEROVA, M., JOHNSON, N., CHEADLE, C. & CHO-CHUNG, Y. S. 2006. Autoantibody biomarker opens a new gateway for cancer diagnosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* 1762**,** 398-403.

NESVIZHSKII, A. I., KELLER, A., KOLKER, E. & AEBERSOLD, R. 2003. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Analytical Chemistry,* 75**,** 4646-4658.

NEVE, R. M., CHIN, K., FRIDLYAND, J., YEH, J., BAEHNER, F. L., FEVR, T., CLARK, L., BAYANI, N., COPPE, J. P., TONG, F., SPEED, T., SPELLMAN, P. T., DEVRIES, S., LAPUK, A., WANG, N. J., KUO, W. L., STILWELL, J. L., PINKEL, D., ALBERTSON, D. G., WALDMAN, F. M., MCCORMICK, F., DICKSON, R. B., JOHNSON, M. D., LIPPMAN, M., ETHIER, S., GAZDAR, A. & GRAY, J. W. 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell,* 10**,** 515-27.

NGUYEN, L. V., VANNER, R., DIRKS, P. & EAVES, C. J. 2012. Cancer stem cells: an evolving concept. *Nature Reviews Cancer,* 12**,** 133.

NISIHARA, R., MACHOSKI, M. C. C., NEPPEL, A., MAESTRI, C. A., MESSIAS-REASON, I. & SKARE, T. L. 2018. Anti-nuclear antibodies in patients with breast cancer. *Clin Exp Immunol,* 193**,** 178-182.

NOBLEJAS-LÓPEZ, M. D. M., NIETO-JIMÉNEZ, C., MORCILLO GARCÍA, S., PÉREZ-PEÑA, J., NUNCIA-CANTARERO, M., ANDRÉS-PRETEL, F., GALÁN-MOYA, E. M., AMIR, E., PANDIELLA, A., GYŐRFFY, B. & OCANA, A. 2019. Expression of MHC class I, HLA-A and HLA-B identifies immune-activated breast tumors with favorable outcome. *Oncoimmunology,* 8**,** e1629780.

NOGUCHI, M., FUJIMOTO, K., ARAI, G., UEMURA, H., HASHINE, K., MATSUMOTO, H., FUKASAWA, S., KOHJIMOTO, Y., NAKATSU, H., TAKENAKA, A., FUJISAWA, M., UEMURA, H., NAITO, S., EGAWA, S., FUJIMOTO, H., HINOTSU, S. & ITOH, K. 2021. A randomized phase III trial of personalized peptide vaccination for castration-resistant prostate cancer progressing after docetaxel. *Oncol Rep,* 45**,** 159-168.

O'MEARA, T. A. & TOLANEY, S. M. 2021. Tumor mutational burden as a predictor of immunotherapy response in breast cancer. *Oncotarget,* 12**,** 394-400.

OEFFINGER, K. C., FONTHAM, E. T. H., ETZIONI, R., HERZIG, A., MICHAELSON, J. S., SHIH, Y.-C. T., WALTER, L. C., CHURCH, T. R., FLOWERS, C. R., LAMONTE, S. J., WOLF, A. M. D., DESANTIS, C., LORTET-TIEULENT, J., ANDREWS, K., MANASSARAM-BAPTISTE, D., SASLOW, D., SMITH, R. A., BRAWLEY, O. W., WENDER, R. & AMERICAN CANCER, S. 2015. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *JAMA,* 314**,** 1599-1614.

OKELLO, J., KISEMBO, H., BUGEZA, S. & GALUKANDE, M. 2014. Breast cancer detection using sonography in women with mammographically dense breasts. *BMC medical imaging,* 14**,** 41-41.

OLD, W. M., MEYER-ARENDT, K., AVELINE-WOLF, L., PIERCE, K. G., MENDOZA, A., SEVINSKY, J. R., RESING, K. A. & AHN, N. G. 2005. Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics*S. *Molecular & Cellular Proteomics,* 4**,** 1487-1502.

OLIVEIRA, M., SAURA, C., NUCIFORO, P., CALVO, I., ANDERSEN, J., PASSOS-COELHO, J. L., GIL GIL, M., BERMEJO, B., PATT, D. A., CIRUELOS, E., DE LA PEÑA, L., XU, N., WONGCHENKO, M., SHI, Z., SINGEL, S. M. & ISAKOFF, S. J. 2019. FAIRLANE, a double-blind placebo-controlled randomized phase II trial of neoadjuvant ipatasertib plus paclitaxel for early triple-negative breast cancer. *Ann Oncol,* 30**,** 1289-1297.

ORTIZ VALDEZ, E., RANGEL-ESCAREÑO, C., MATUS SANTOS, J. A., VÁZQUEZ ROMO, R., GUIJOSA, A., VILLARREAL-GARZA, C., ARRIETA, O., RODRÍGUEZ-BAUTISTA, R., CARO-SÁNCHEZ, C. H. & ORTEGA GÓMEZ, A. 2023. Characterization of triple negative breast cancer gene expression profiles in Mexican patients. *Mol Clin Oncol,* 18**,** 5.

OSBORNE, C. K. 1998. Tamoxifen in the Treatment of Breast Cancer. *New England Journal of Medicine,* 339**,** 1609-1618.

OTT, P. A., HU, Z., KESKIN, D. B., SHUKLA, S. A., SUN, J., BOZYM, D. J., ZHANG, W., LUOMA, A., GIOBBIE-HURDER, A., PETER, L., CHEN, C., OLIVE, O., CARTER, T. A., LI, S., LIEB, D. J., EISENHAURE, T., GJINI, E., STEVENS, J., LANE, W. J., JAVERI, I., NELLAIAPPAN, K., SALAZAR, A. M., DALEY, H., SEAMAN, M., BUCHBINDER, E. I., YOON, C. H., HARDEN, M., LENNON, N., GABRIEL, S., RODIG, S. J., BAROUCH, D. H., ASTER, J. C., GETZ, G., WUCHERPFENNIG, K., NEUBERG, D., RITZ, J., LANDER, E. S., FRITSCH, E. F., HACOHEN, N. & WU, C. J. 2017. An

immunogenic personal neoantigen vaccine for patients with melanoma. *Nature,* 547**,** 217-221.

PACI, E. 2012. Summary of the Evidence of Breast Cancer Service Screening Outcomes in Europe and First Estimate of the Benefit and Harm Balance Sheet. *Journal of Medical Screening,* 19**,** 5-13.

PAGAZA-STRAFFON, C., MARCHAT, L. A., HERRERA, L., DÍAZ-CHÁVEZ, J., AVANTE, M. G., RODRÍGUEZ, Y. P., ARREOLA, M. C. & LÓPEZ-CAMARILLO, C. 2020. Evaluation of a panel of tumor-associated antigens in breast cancer. *Cancer Biomark,* 27**,** 207-211.

PANOWSKI, S., BHAKTA, S., RAAB, H., POLAKIS, P. & JUNUTULA, J. R. 2014. Site-specific antibody drug conjugates for cancer therapy. *mAbs,* 6**,** 34-45.

PARK, H. S., JANG, M. H., KIM, E. J., KIM, H. J., LEE, H. J., KIM, Y. J., KIM, J. H., KANG, E., KIM, S. W., KIM, I. A. & PARK, S. Y. 2014. High EGFR gene copy number predicts poor outcome in triple-negative breast cancer. *Mod Pathol,* 27**,** 1212-22.

PARK, J. H., AHN, J. H. & KIM, S. B. 2018. How shall we treat early triple-negative breast cancer (TNBC): from the current standard to upcoming immuno-molecular strategies. *ESMO Open,* 3**,** e000357.

PARK, S. Y., CHOI, J. H. & NAM, J. S. 2019. Targeting Cancer Stem Cells in Triple-Negative Breast Cancer. *Cancers (Basel),* 11.

PARMLEY, S. F. & SMITH, G. P. 1988. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene,* 73**,** 305-318.

PAUL, A. & PAUL, S. 2014. The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers. *Frontiers in bioscience (Landmark edition),* 19**,** 605-618.

PAVELKA, N., FOURNIER, M. L., SWANSON, S. K., PELIZZOLA, M., RICCIARDI-CASTAGNOLI, P., FLORENS, L. & WASHBURN, M. P. 2008. Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol Cell Proteomics,* 7**,** 631-44.

PEITZSCH, C., TYUTYUNNYKOVA, A., PANTEL, K. & DUBROVSKA, A. 2017. Cancer stem cells: The root of tumor recurrence and metastases. *Semin Cancer Biol,* 44**,** 10-24.

PEREIRA DA SILVA, L., MIGUEL NEVES, B., MOURA, L., CRUZ, M. T. & CARVALHO, E. 2014. Neurotensin Decreases the Proinflammatory Status of Human Skin Fibroblasts and Increases Epidermal Growth Factor Expression. *International Journal of Inflammation,* 2014**,** 248240.

PEREZ, E. A., ROMOND, E. H., SUMAN, V. J., JEONG, J.-H., SLEDGE, G., GEYER, C. E., JR., MARTINO, S., RASTOGI, P., GRALOW, J., SWAIN, S. M., WINER, E. P., COLON-OTERO, G., DAVIDSON, N. E., MAMOUNAS, E., ZUJEWSKI, J. A. & WOLMARK, N. 2014. Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* 32**,** 3744-3752.

PEREZ, E. A. & WEILBAECHER, K. 2006. Aromatase inhibitors and bone loss. *Oncology (Williston Park, N.Y.),* 20**,** 1029-1048.

PEROU, C. M., SØRLIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., REES, C. A., POLLACK, J. R., ROSS, D. T., JOHNSEN, H., AKSLEN, L. A., FLUGE, Ø., PERGAMENSCHIKOV, A., WILLIAMS, C., ZHU, S. X., LØNNING, P. E., BØRRESEN-DALE, A.-L., BROWN, P. O. & BOTSTEIN, D. 2000. Molecular portraits of human breast tumours. *Nature,* 406**,** 747-752.

PIETRAS, R. J., PEGRAM, M. D., FINN, R. S., MANEVAL, D. A. & SLAMON, D. J. 1998. Remission of human breast cancer xenografts on therapy with humanized

monoclonal antibody to HER-2 receptor and DNA-reactive drugs. *Oncogene,* 17**,** 2235-2249.

PING, Z., XIA, Y., SHEN, T., PAREKH, V., SIEGAL, G. P., ELTOUM, I.-E., HE, J., CHEN, D., DENG, M., XI, R. & SHEN, D. 2016. A microscopic landscape of the invasive breast cancer genome. *Scientific reports,* 6**,** 27545-27545.

PLANK, C., OBERHAUSER, B., MECHTLER, K., KOCH, C. & WAGNER, E. 1994. The influence of endosome-disruptive peptides on gene transfer using synthetic virus-like gene transfer systems. *J Biol Chem,* 269**,** 12918-24.

PLANK, C., ZAUNER, W. & WAGNER, E. 1998. Application of membrane-active peptides for drug and gene delivery across cellular membranes. *Adv Drug Deliv Rev,* 34**,** 21-35.

POLAK, P., KIM, J., BRAUNSTEIN, L. Z., KARLIC, R., HARADHAVALA, N. J., TIAO, G., ROSEBROCK, D., LIVITZ, D., KÜBLER, K., MOUW, K. W., KAMBUROV, A., MARUVKA, Y. E., LESHCHINER, I., LANDER, E. S., GOLUB, T. R., ZICK, A., ORTHWEIN, A., LAWRENCE, M. S., BATRA, R. N., CALDAS, C., HABER, D. A., LAIRD, P. W., SHEN, H., ELLISEN, L. W., D'ANDREA, A. D., CHANOCK, S. J., FOULKES, W. D. & GETZ, G. 2017. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat Genet,* 49**,** 1476-1486.

PRAT, A., LLUCH, A., ALBANELL, J., BARRY, W. T., FAN, C., CHACON, J. I., PARKER, J. S., CALVO, L., PLAZAOLA, A., ARCUSA, A., SEGUI-PALMER, M. A., BURGUES, O., RIBELLES, N., RODRIGUEZ-LESCURE, A., GUERRERO, A., RUIZ-BORREGO, M., MUNARRIZ, B., LOPEZ, J. A., ADAMO, B., CHEANG, M. C., LI, Y., HU, Z., GULLEY, M. L., VIDAL, M. J., PITCHER, B. N., LIU, M. C., CITRON, M. L., ELLIS, M. J., MARDIS, E., VICKERY, T., HUDIS, C. A., WINER, E. P., CAREY, L. A., CABALLERO, R., CARRASCO, E., MARTIN, M., PEROU, C. M. & ALBA, E. 2014. Predicting response and survival in chemotherapy-treated triple-negative breast cancer. *Br J Cancer,* 111**,** 1532-41.

PRAT, A., PARKER, J. S., KARGINOVA, O., FAN, C., LIVASY, C., HERSCHKOWITZ, J. I., HE, X. & PEROU, C. M. 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research,* 12**,** R68.

PRESS, M. F., CORDON-CARDO, C. & SLAMON, D. J. 1990. Expression of the HER-2/neu proto-oncogene in normal human adult and fetal tissues. *Oncogene,* 5**,** 953-962.

PULITI, D., DUFFY, S. W., MICCINESI, G., DE KONING, H., LYNGE, E., ZAPPA, M. & PACI, E. 2012. Overdiagnosis in Mammographic Screening for Breast Cancer in Europe: A Literature Review. *Journal of Medical Screening,* 19**,** 42-56.

QIU, J., KEYSER, B., LIN, Z.-T. & WU, T. 2018. Autoantibodies as Potential Biomarkers in Breast Cancer. *Biosensors,* 8**,** 67.

QUINTANA, E., SHACKLETON, M., FOSTER, H. R., FULLEN, D. R., SABEL, M. S., JOHNSON, T. M. & MORRISON, S. J. 2010. Phenotypic Heterogeneity among Tumorigenic Melanoma Cells from Patients that Is Reversible and Not Hierarchically Organized. *Cancer Cell,* 18**,** 510-523.

QUINTANA, E., SHACKLETON, M., SABEL, M. S., FULLEN, D. R., JOHNSON, T. M. & MORRISON, S. J. 2008. Efficient tumour formation by single human melanoma cells. *Nature,* 456**,** 593-8.

RADOVICH, M., JIANG, G., HANCOCK, B. A., CHITAMBAR, C., NANDA, R., FALKSON, C., LYNCE, F. C., GALLAGHER, C., ISAACS, C., BLAYA, M., PAPLOMATA, E., WALLING, R., DAILY, K., MAHTANI, R., THOMPSON, M. A., GRAHAM, R., COOPER, M. E., PAVLICK, D. C., ALBACKER, L. A., GREGG, J., SOLZAK, J. P., CHEN,

Y. H., BALES, C. L., CANTOR, E., SHEN, F., STORNIOLO, A. M. V., BADVE, S., BALLINGER, T. J., CHANG, C. L., ZHONG, Y., SAVRAN, C., MILLER, K. D. & SCHNEIDER, B. P. 2020. Association of Circulating Tumor DNA and Circulating Tumor Cells After Neoadjuvant Chemotherapy With Disease Recurrence in Patients With Triple-Negative Breast Cancer: Preplanned Secondary Analysis of the BRE12-158 Randomized Clinical Trial. *JAMA Oncol,* 6**,** 1410-1415.

RAGHUWANSHI, Y., ETAYASH, H., SOUDY, R., PAIVA, I., LAVASANIFAR, A. & KAUR, K. 2017. Proteolytically Stable Cyclic Decapeptide for Breast Cancer Cell Targeting. *Journal of Medicinal Chemistry,* 60**,** 4893-4903.

RAKHA, E. A., REIS-FILHO, J. S., BAEHNER, F., DABBS, D. J., DECKER, T., EUSEBI, V., FOX, S. B., ICHIHARA, S., JACQUEMIER, J., LAKHANI, S. R., PALACIOS, J., RICHARDSON, A. L., SCHNITT, S. J., SCHMITT, F. C., TAN, P.-H., TSE, G. M., BADVE, S. & ELLIS, I. O. 2010. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast cancer research : BCR,* 12**,** 207-207.

RAUF, F., ANDERSON, K. S. & LABAER, J. 2020. Autoantibodies in Early Detection of Breast Cancer. *Cancer Epidemiology, Biomarkers & Prevention,* 29**,** 2475-2485.

REDDY, L. H., ARIAS, J. L., NICOLAS, J. & COUVREUR, P. 2012. Magnetic Nanoparticles: Design and Characterization, Toxicity and Biocompatibility, Pharmaceutical and Biomedical Applications. *Chemical Reviews,* 112**,** 5818-5878.

REDDY, S. M., CARROLL, E. & NANDA, R. 2020. Atezolizumab for the treatment of breast cancer. *Expert Review of Anticancer Therapy,* 20**,** 151-158.

REESE, D. E., HENDERSON, M. C., SILVER, M., MULPURI, R., LETSIOS, E., TRAN, Q. & WOLF, J. K. 2017. Breast density does not impact the ability of Videssa® Breast to detect breast cancer in women under age 50. *PLoS One,* 12**,** e0186198.

RIGGS, B. L. & HARTMANN, L. C. 2003. Selective Estrogen-Receptor Modulators — Mechanisms of Action and Application to Clinical Practice. *New England Journal of Medicine,* 348**,** 618-629.

RIVAS, S., GOMEZ-ORO, C., ANTON, I. M. & WANDOSELL, F. 2018. Role of Akt Isoforms Controlling Cancer Stem Cell Survival, Phenotype and Self-Renewal. *Biomedicines,* 6.

ROBERTSON, J. F. R., BONDARENKO, I. M., TRISHKINA, E., DVORKIN, M., PANASCI, L., MANIKHAS, A., SHPARYK, Y., CARDONA-HUERTA, S., CHEUNG, K.-L., PHILCO-SALAS, M. J., RUIZ-BORREGO, M., SHAO, Z., NOGUCHI, S., ROWBOTTOM, J., STUART, M., GRINSTED, L. M., FAZAL, M. & ELLIS, M. J. 2016. Fulvestrant 500 mg versus anastrozole 1 mg for hormone receptor-positive advanced breast cancer (FALCON): an international, randomised, double-blind, phase 3 trial. *The Lancet,* 388**,** 2997-3005.

ROBSON, M., IM, S.-A., SENKUS, E., XU, B., DOMCHEK, S. M., MASUDA, N., DELALOGE, S., LI, W., TUNG, N., ARMSTRONG, A., WU, W., GOESSL, C., RUNSWICK, S. & CONTE, P. 2017. Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *New England Journal of Medicine,* 377**,** 523-533.

ROBSON, M. E., TUNG, N., CONTE, P., IM, S. A., SENKUS, E., XU, B., MASUDA, N., DELALOGE, S., LI, W., ARMSTRONG, A., WU, W., GOESSL, C., RUNSWICK, S. & DOMCHEK, S. M. 2019. OlympiAD final overall survival and tolerability results: Olaparib versus chemotherapy treatment of physician's choice in patients with a germline BRCA mutation and HER2-negative metastatic breast cancer. *Ann Oncol,* 30**,** 558-566.

ROCCA, A., MALTONI, R., BRAVACCINI, S., DONATI, C. & ANDREIS, D. 2018. Clinical utility of fulvestrant in the treatment of breast cancer: a report on the emerging clinical evidence. *Cancer management and research,* 10**,** 3083-3099.

ROMOND, E. H., PEREZ, E. A., BRYANT, J., SUMAN, V. J., GEYER, C. E., DAVIDSON, N. E., TAN-CHIU, E., MARTINO, S., PAIK, S., KAUFMAN, P. A., SWAIN, S. M., PISANSKY, T. M., FEHRENBACHER, L., KUTTEH, L. A., VOGEL, V. G., VISSCHER, D. W., YOTHERS, G., JENKINS, R. B., BROWN, A. M., DAKHIL, S. R., MAMOUNAS, E. P., LINGLE, W. L., KLEIN, P. M., INGLE, J. N. & WOLMARK, N. 2005. Trastuzumab plus Adjuvant Chemotherapy for Operable HER2-Positive Breast Cancer. *New England Journal of Medicine,* 353**,** 1673-1684.

RONCKERS, C. M., ERDMANN, C. A. & LAND, C. E. 2005. Radiation and breast cancer: a review of current evidence. *Breast cancer research : BCR,* 7**,** 21-32.

ROSENBERG, S. M., RUDDY, K. J., TAMIMI, R. M., GELBER, S., SCHAPIRA, L., COME, S., BORGES, V. F., LARSEN, B., GARBER, J. E. & PARTRIDGE, A. H. 2016. BRCA1 and BRCA2 Mutation Testing in Young Women With Breast Cancer. *JAMA oncology,* 2**,** 730-736.

ROSSI, L., STEVENS, D., PIERGA, J.-Y., LEREBOURS, F., REYAL, F., ROBAIN, M., ASSELAIN, B. & ROUZIER, R. 2015. Impact of Adjuvant Chemotherapy on Breast Cancer Survival: A Real-World Population. *PloS one,* 10**,** e0132853-e0132853.

RUIZ-BORREGO, M., JIMENEZ, B., ANTOLIN, S., GARCIA-SAENZ, J. A., CORRAL, J., JEREZ, Y., TRIGO, J., URRUTICOECHEA, A., COLOM, H., GONZALO, N., MUNOZ, C., BENITO, S., CABALLERO, R., BEZARES, S., CARRASCO, E., ROJO, F. & MARTIN, M. 2019. A phase Ib study of sonidegib (LDE225), an oral small molecule inhibitor of smoothened or Hedgehog pathway, in combination with docetaxel in triple negative advanced breast cancer patients: GEICAM/2012-12 (EDALINE) study. *Invest New Drugs,* 37**,** 98-108.

SABATIER, R., FINETTI, P., GUILLE, A., ADELAIDE, J., CHAFFANET, M., VIENS, P., BIRNBAUM, D. & BERTUCCI, F. 2014. Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. *Molecular cancer,* 13**,** 228-228.

SADER, M., COULEAUD, P., CARPENTIER, G., GILLES, M.-E., BOUSSERRHINE, N., LIVET, A., CASCONE, I., DESTOUCHES, D., CORTAJARENA, A. & COURTY, J. 2015. Functionalization of Iron Oxide Magnetic Nanoparticles with the Multivalent Pseudopeptide N6l for Breast Tumor Targeting. *Journal of Nanomedicine & Nanotechnology,* 06.

SAEG, F. & ANBALAGAN, M. 2018. Breast cancer stem cells and the challenges of eradication: a review of novel therapies. *Stem Cell Investig,* 5**,** 39.

SAEKI, T., TSURUO, T., SATO, W. & NISHIKAWSA, K. 2005. Drug resistance in chemotherapy for breast cancer. *Cancer Chemotherapy and Pharmacology,* 56**,** 84-89.

SAHA, S. K., CHOI, H. Y., KIM, B. W., DAYEM, A. A., YANG, G. M., KIM, K. S., YIN, Y. F. & CHO, S. G. 2017. KRT19 directly interacts with β-catenin/RAC1 complex to regulate NUMB-dependent NOTCH signaling pathway and breast cancer properties. *Oncogene,* 36**,** 332-349.

SAMEC, T., ALATISE, K. L., BOULOS, J., GILMORE, S., HAZELTON, A., COFFIN, C. & ALEXANDER-BRYANT, A. 2022a. Fusogenic peptide delivery of bioactive siRNAs targeting CSNK2A1 for treatment of ovarian cancer. *Molecular Therapy - Nucleic Acids,* 30**,** 95-111.

SAMEC, T., BOULOS, J., GILMORE, S., HAZELTON, A. & ALEXANDER-BRYANT, A. 2022b. Peptide-based delivery of therapeutics in cancer treatment. *Mater Today Bio,* 14**,** 100248.

SANJOSÉ, S., LÉONÉ, M., BEREZ, V., IZQUIERDO-FONT, A., FONT, R., BRUNET, J., LOUAT, T., VILARDELL, L., BORRAS, J., VILADIU, P., BOSCH, F., LENOIR, G. & SINILNIKOVA, O. 2003. Prevalence ofBRCA1 andBRCA2 germline mutations in

young breast cancer patients: A population-based study. *International journal of cancer. Journal international du cancer,* 106**,** 588-93.

SANTONJA, A., SANCHEZ-MUNOZ, A., LLUCH, A., CHICA-PARRADO, M. R., ALBANELL, J., CHACON, J. I., ANTOLIN, S., JEREZ, J. M., DE LA HABA, J., DE LUQUE, V., FERNANDEZ-DE SOUSA, C. E., VICIOSO, L., PLATA, Y., RAMIREZ-TORTOSA, C. L., ALVAREZ, M., LLACER, C., ZARCOS-PEDRINACI, I., CARRASCO, E., CABALLERO, R., MARTIN, M. & ALBA, E. 2018. Triple negative breast cancer subtypes and pathologic complete response rate to neoadjuvant chemotherapy. *Oncotarget,* 9**,** 26406-26416.

SASADA, T., NOGUCHI, M., YAMADA, A. & ITOH, K. 2012. Personalized peptide vaccination: a novel immunotherapeutic approach for advanced cancer. *Human vaccines & immunotherapeutics,* 8**,** 1309-1313.

SAW, P. E. & SONG, E.-W. 2019. Phage display screening of therapeutic peptide for cancer targeting and therapy. *Protein & Cell*.

SCANLAN, M. J., CHEN, Y.-T., WILLIAMSON, B., GURE, A. O., STOCKERT, E., GORDAN, J. D., TÜRECI, Ö., SAHIN, U., PFREUNDSCHUH, M. & OLD, L. J. 1998. Characterization of human colon cancer antigens recognized by autologous antibodies. *International Journal of Cancer,* 76**,** 652-658.

SCANLAN, M. J., GORDAN, J. D., WILLIAMSON, B., STOCKERT, E., BANDER, N. H., JONGENEEL, V., GURE, A. O., JÄGER, D., JÄGER, E., KNUTH, A., CHEN, Y.-T. & OLD, L. J. 1999. Antigens recognized by autologous antibody in patients with renal-cell carcinoma. *International Journal of Cancer,* 83**,** 456-464.

SCHECH, A., KAZI, A., YU, S., SHAH, P. & SABNIS, G. 2015. Histone Deacetylase Inhibitor Entinostat Inhibits Tumor-Initiating Cells in Triple-Negative Breast Cancer Cells. *Mol Cancer Ther,* 14**,** 1848-57.

SCHIFF, R., REDDY, P., AHOTUPA, M., CORONADO-HEINSOHN, E., GRIM, M., HILSENBECK, S. G., LAWRENCE, R., DENEKE, S., HERRERA, R., CHAMNESS, G. C., FUQUA, S. A. W., BROWN, P. H. & OSBORNE, C. K. 2000. Oxidative Stress and AP-1 Activity in Tamoxifen-Resistant Breast Tumors In Vivo. *JNCI: Journal of the National Cancer Institute,* 92**,** 1926-1934.

SCHMID, P., ADAMS, S., RUGO, H. S., SCHNEEWEISS, A., BARRIOS, C. H., IWATA, H., DIÉRAS, V., HEGG, R., IM, S.-A., SHAW WRIGHT, G., HENSCHEL, V., MOLINERO, L., CHUI, S. Y., FUNKE, R., HUSAIN, A., WINER, E. P., LOI, S. & EMENS, L. A. 2018. Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *New England Journal of Medicine,* 379**,** 2108-2121.

SCHMID, P., CHUI, S. Y. & EMENS, L. A. 2019. Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. Reply. *N Engl J Med,* 380**,** 987-988.

SCHMID, P., SALGADO, R., PARK, Y. H., MUÑOZ-COUSELO, E., KIM, S. B., SOHN, J., IM, S. A., FOUKAKIS, T., KUEMMEL, S., DENT, R., YIN, L., WANG, A., TRYFONIDIS, K., KARANTZA, V., CORTÉS, J. & LOI, S. 2020. Pembrolizumab plus chemotherapy as neoadjuvant treatment of high-risk, early-stage triple-negative breast cancer: results from the phase 1b open-label, multicohort KEYNOTE-173 study. *Annals of Oncology,* 31**,** 569-581.

SCHNEIDER, C. A., RASBAND, W. S. & ELICEIRI, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods,* 9**,** 671-675.

SCHOENMAKER, L., WITZIGMANN, D., KULKARNI, J. A., VERBEKE, R., KERSTEN, G., JISKOOT, W. & CROMMELIN, D. J. A. 2021. mRNA-lipid nanoparticle COVID-19 vaccines: Structure and stability. *International Journal of Pharmaceutics,* 601**,** 120586.

SCOTT, J. & SMITH, G. 1990. Searching for peptide ligands with an epitope library. *Science,* 249**,** 386-390.

SEARLE, B. C. 2010. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics,* 10**,** 1265-9.

SEEBACHER, N. A., STACY, A. E., PORTER, G. M. & MERLOT, A. M. 2019. Clinical development of targeted and immune based anti-cancer therapies. *Journal of Experimental & Clinical Cancer Research,* 38**,** 156.

SEELY, J. M. & ALHASSAN, T. 2018. Screening for breast cancer in 2018-what should we be doing today? *Current oncology (Toronto, Ont.),* 25**,** S115-S124.

SEILER, M., RAY-COQUARD, I., MELICHAR, B., YARDLEY, D. A., WANG, R. X., DODION, P. F. & LEE, M. A. 2015. Oral Ridaforolimus Plus Trastuzumab for Patients With HER2+ Trastuzumab-Refractory Metastatic Breast Cancer. *Clinical Breast Cancer,* 15**,** 60-65.

SETAYESH, S. M., HART, O., NAGHDLOO, A., HIGA, N., NIEVA, J., LU, J., HWANG, S., WILKINSON, K., KIDD, M., ANDERSON, A., VELASCO, C. R., KOLATKAR, A., MATSUMOTO, N., NEVAREZ, R., HICKS, J. B., MASON, J., SHISHIDO, S. N. & KUHN, P. 2022. Multianalyte liquid biopsy to aid the diagnostic workup of breast cancer. *npj Breast Cancer,* 8**,** 112.

SHAH, A. N. & CRISTOFANILLI, M. 2017. The Growing Role of CDK4/6 Inhibitors in Treating Hormone Receptor-Positive Advanced Breast Cancer. *Current Treatment Options in Oncology,* 18**,** 6.

SHANG, Y. & BROWN, M. 2002. Molecular determinants for the tissue specificity of SERMs. *Science,* 295**,** 2465-8.

SHARMA, P., ABRAMSON, V. G., O'DEA, A., NYE, L., MAYER, I., PATHAK, H. B., HOFFMANN, M., STECKLEIN, S. R., ELIA, M., LEWIS, S., SCOTT, J., DE JONG, J. A., WANG, Y. Y., YODER, R., SCHWENSEN, K., FINKE, K., HELDSTAB, J., LAFAVER, S., WILLIAMSON, S. K., PHADNIS, M. A., REED, G. A., KIMLER, B. F., KHAN, Q. J. & GODWIN, A. K. 2021. Clinical and Biomarker Results from Phase I/II Study of PI3K Inhibitor Alpelisib plus Nab-paclitaxel in HER2-Negative Metastatic Breast Cancer. *Clin Cancer Res,* 27**,** 3896-3904.

SHERIDAN, C., KISHIMOTO, H., FUCHS, R. K., MEHROTRA, S., BHAT-NAKSHATRI, P., TURNER, C. H., GOULET, R., JR., BADVE, S. & NAKSHATRI, H. 2006. CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. *Breast Cancer Res,* 8**,** R59.

SHI, P., LIU, W., TALA, WANG, H., LI, F., ZHANG, H., WU, Y., KONG, Y., ZHOU, Z., WANG, C., CHEN, W., LIU, R. & CHEN, C. 2017. Metformin suppresses triple-negative breast cancer stem cells by targeting KLF5 for degradation. *Cell Discov,* 3**,** 17010.

SHIMELIS, H., LADUCA, H., HU, C., HART, S. N., NA, J., THOMAS, A., AKINHANMI, M., MOORE, R. M., BRAUCH, H., COX, A., ECCLES, D. M., EWART-TOLAND, A., FASCHING, P. A., FOSTIRA, F., GARBER, J., GODWIN, A. K., KONSTANTOPOULOU, I., NEVANLINNA, H., SHARMA, P., YANNOUKAKOS, D., YAO, S., FENG, B. J., TIPPIN DAVIS, B., LILYQUIST, J., PESARAN, T., GOLDGAR, D. E., POLLEY, E. C., DOLINSKY, J. S. & COUCH, F. J. 2018. Triple-Negative Breast Cancer Risk Genes Identified by Multigene Hereditary Cancer Panel Testing. *J Natl Cancer Inst*.

SHOKRALLA, S., GIBSON, J. F., NIKBAKHT, H., JANZEN, D. H., HALLWACHS, W. & HAJIBABAEI, M. 2014. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources,* 14**,** 892-901.

SHTATLAND, T., GUETTLER, D., KOSSODO, M., PIVOVAROV, M. & WEISSLEDER, R. 2007. PepBank--a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinformatics,* 8**,** 280.

SIEGEL, R. L., MILLER, K. D., FUCHS, H. E. & JEMAL, A. 2022. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians,* 72**,** 7-33.

SIEVERS, E. L. & SENTER, P. D. 2013. Antibody-Drug Conjugates in Cancer Therapy. *Annual Review of Medicine,* 64**,** 15-29.

SIKORA, M. J., JANKOWITZ, R. C., DABBS, D. J. & OESTERREICH, S. 2013. Invasive lobular carcinoma of the breast: Patient response to systemic endocrine therapy and hormone response in model systems. *Steroids,* 78**,** 568-575.

SIN, W. C. & LIM, C. L. 2017. Breast cancer stem cells-from origins to targeted therapy. *Stem Cell Investig,* 4**,** 96.

SINGH, A. P., SHARMA, S. & SHAH, D. K. 2016. Quantitative characterization of in vitro bystander effect of antibody-drug conjugates. *Journal of pharmacokinetics and pharmacodynamics,* 43**,** 567-582.

SLAMON, D. J., CLARK, G. M., WONG, S. G., LEVIN, W. J., ULLRICH, A. & MCGUIRE, W. L. 1987. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science,* 235**,** 177.

SLAMON, D. J., LEYLAND-JONES, B., SHAK, S., FUCHS, H., PATON, V., BAJAMONDE, A., FLEMING, T., EIERMANN, W., WOLTER, J., PEGRAM, M., BASELGA, J. & NORTON, L. 2001. Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *New England Journal of Medicine,* 344**,** 783-792.

SMITH, G. P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science,* 228**,** 1315-7.

SOPIK, V. & NAROD, S. A. 2018. The relationship between tumour size, nodal status and distant metastases: on the origins of breast cancer. *Breast cancer research and treatment,* 170**,** 647-656.

SØRLIE, T., PEROU, C. M., TIBSHIRANI, R., AAS, T., GEISLER, S., JOHNSEN, H., HASTIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., THORSEN, T., QUIST, H., MATESE, J. C., BROWN, P. O., BOTSTEIN, D., LØNNING, P. E. & BØRRESEN-DALE, A. L. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 10869-10874.

SOUDY, R., AHMED, S. & KAUR, K. 2012. NGR Peptide Ligands for Targeting CD13/APN Identified through Peptide Array Screening Resemble Fibronectin Sequences. *ACS Combinatorial Science,* 14**,** 590-599.

SOUDY, R., ETAYASH, H., BAHADORANI, K., LAVASANIFAR, A. & KAUR, K. 2017. Breast Cancer Targeting Peptide Binds Keratin 1: A New Molecular Marker for Targeted Drug Delivery to Breast Cancer. *Molecular Pharmaceutics,* 14**,** 593-604.

SOULE, H. D., MALONEY, T. M., WOLMAN, S. R., PETERSON, W. D., JR., BRENZ, R., MCGRATH, C. M., RUSSO, J., PAULEY, R. J., JONES, R. F. & BROOKS, S. C. 1990. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res,* 50**,** 6075-86.

STANTON, S. E., CECIL, D. L., O'MEARA, M. M., CHILDS, J. S. & DISIS, M. L. 2018. Breast cancer stem cell autoantibodies to identify women with advanced breast cancer. *Journal of Clinical Oncology,* 36**,** 1079-1079.

STANTON, S. E., GAD, E., RAMOS, E., CORULLI, L., ANNIS, J., CHILDS, J., KATAYAMA, H., HANASH, S., MARKS, J. & DISIS, M. L. 2021. Tumor-associated autoantibodies from mouse breast cancer models are found in serum of breast cancer patients. *npj Breast Cancer,* 7**,** 50.

STEVIC, I., MULLER, V., WEBER, K., FASCHING, P. A., KARN, T., MARME, F., SCHEM, C., STICKELER, E., DENKERT, C., VAN MACKELENBERGH, M., SALAT, C.,

SCHNEEWEISS, A., PANTEL, K., LOIBL, S., UNTCH, M. & SCHWARZENBACH, H. 2018. Specific microRNA signatures in exosomes of triple-negative and HER2-positive breast cancer patients undergoing neoadjuvant therapy within the GeparSixto trial. *BMC Med,* 16**,** 179.

SUGAHARA, K. N., TEESALU, T., KARMALI, P. P., KOTAMRAJU, V. R., AGEMY, L., GREENWALD, D. R. & RUOSLAHTI, E. 2010. Coadministration of a Tumor-Penetrating Peptide Enhances the Efficacy of Cancer Drugs. *Science,* 328**,** 1031.

SUN, C., LEE, J. S. H. & ZHANG, M. 2008. Magnetic nanoparticles in MR imaging and drug delivery. *Adv Drug Deliv Rev,* 60**,** 1252-1265.

SUN, J., ZHANG, C., LIU, G., LIU, H., ZHOU, C., LU, Y., ZHOU, C., YUAN, L. & LI, X. 2012. A novel mouse CD133 binding-peptide screened by phage display inhibits cancer cell motility in vitro. *Clinical & Experimental Metastasis,* 29**,** 185-196.

SUNG, H., FERLAY, J., SIEGEL, R. L., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A. & BRAY, F. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin,* 71**,** 209-249.

SWAIN, S. M., SORACE, R. A., BAGLEY, C. S., DANFORTH, D. N., BADER, J., WESLEY, M. N., STEINBERG, S. M. & LIPPMAN, M. E. 1987. Neoadjuvant Chemotherapy in the Combined Modality Approach of Locally Advanced Nonmetastatic Breast Cancer. *Cancer Research,* 47**,** 3889-3894.

TABÁR, L., GAD, A., HOLMBERG, L. H., LJUNGQUIST, U., FAGERBERG, C. J. G., BALDETORP, L., GRÖNTOFT, O., LUNDSTRÖM, B., MÅNSON, J. C., EKLUND, G., DAY, N. E. & PETTERSSON, F. 1985. REDUCTION IN MORTALITY FROM BREAST CANCER AFTER MASS SCREENING WITH MAMMOGRAPHY: Randomised Trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *The Lancet,* 325**,** 829-832.

THERATECHNOLOGIES & PPD. 2021. *TH1902 in Patients With Advanced Solid Tumors* [Online]. https://ClinicalTrials.gov/show/NCT04706962. [Accessed 03 January 2023].

THIERY, J. P., ACLOQUE, H., HUANG, R. Y. J. & NIETO, M. A. 2009. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell,* 139**,** 871-890.

THIESEN, B. & JORDAN, A. 2008. Clinical applications of magnetic nanoparticles for hyperthermia. *International Journal of Hyperthermia,* 24**,** 467-474.

THOMSEN, M. C. & NIELSEN, M. 2012. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res,* 40**,** W281-7.

TOLANEY, S., BURRIS, H., GARTNER, E., MAYER, I. A., SAURA, C., MAURER, M., CIRUELOS, E., GARCIA, A. A., CAMPANA, F., WU, B., XU, Y., JIANG, J., WINER, E. & KROP, I. 2015. Phase I/II study of pilaralisib (SAR245408) in combination with trastuzumab or trastuzumab plus paclitaxel in trastuzumab-refractory HER2-positive metastatic breast cancer. *Breast Cancer Research and Treatment,* 149**,** 151-161.

TOOGOOD, P. L., HARVEY, P. J., REPINE, J. T., SHEEHAN, D. J., VANDERWEL, S. N., ZHOU, H., KELLER, P. R., MCNAMARA, D. J., SHERRY, D., ZHU, T., BRODFUEHRER, J., CHOI, C., BARVIAN, M. R. & FRY, D. W. 2005. Discovery of a Potent and Selective Inhibitor of Cyclin-Dependent Kinase 4/6. *Journal of Medicinal Chemistry,* 48**,** 2388-2406.

TRAIL, P. A., DUBOWCHIK, G. M. & LOWINGER, T. B. 2018. Antibody drug conjugates for treatment of breast cancer: Novel targets and diverse approaches in ADC design. *Pharmacology & Therapeutics,* 181**,** 126-142.

TSOUMPELI, M. T., GRAY, A., PARSONS, A. L., SPILIOTOPOULOS, A., OWEN, J. P., BISHOP, K., MADDISON, B. C. & GOUGH, K. C. 2022. A Simple Whole-Plasmid PCR Method to Construct High-Diversity Synthetic Phage Display Libraries. *Mol Biotechnol,* 64**,** 791-803.

UKAI, S., SAKAMOTO, N., TANIYAMA, D., HARADA, K., HONMA, R., MARUYAMA, R., NAKA, K., HINOI, T., TAKAKURA, Y., SHIMIZU, W., OHDAN, H. & YASUI, W. 2021. KHDRBS3 promotes multi-drug resistance and anchorage-independent growth in colorectal cancer. *Cancer Sci,* 112**,** 1196-1208.

VAN DEN BRAND, D., MASSUGER, L. F., BROCK, R. & VERDURMEN, W. P. R. 2017. Mimicking Tumors: Toward More Predictive In Vitro Models for Peptide- and Protein-Conjugated Drugs. *Bioconjugate Chemistry,* 28**,** 846-856.

VAN SEIJEN, M., LIPS, E. H., THOMPSON, A. M., NIK-ZAINAL, S., FUTREAL, A., HWANG, E. S., VERSCHUUR, E., LANE, J., JONKERS, J., REA, D. W., WESSELING, J. & ON BEHALF OF THE, P. T. 2019. Ductal carcinoma in situ: to treat or not to treat, that is the question. *British Journal of Cancer*.

VAN WEZENBEEK, P. M. G. F., HULSEBOS, T. J. M. & SCHOENMAKERS, J. G. G. 1980. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. *Gene,* 11**,** 129-148.

VEKRIS, A., PILALIS, E., CHATZIIOANNOU, A. & PETRY, K. G. 2019. A Computational Pipeline for the Extraction of Actionable Biological Information From NGS-Phage Display Experiments. *Frontiers in Physiology,* 10.

VENKATESH, V., NATARAJ, R., THANGARAJ, G. S., KARTHIKEYAN, M., GNANASEKARAN, A., KAGINELLI, S. B., KUPPANNA, G., KALLAPPA, C. G. & BASALINGAPPA, K. M. 2018. Targeting Notch signalling pathway of cancer stem cells. *Stem cell investigation,* 5**,** 5-5.

VERONESI, A., GIACOMI, C. D., MAGRI, M. D., LOMBARDI, D., ZANETTI, M., SCUDERI, C., DOLCETTI, R., VIEL, A., CRIVELLARI, D., BIDOLI, E. & BOIOCCHI, M. 2005. Familial breast cancer: characteristics and outcome of BRCA 1–2 positive and negative cases. *BMC Cancer,* 5**,** 70.

VINCENT, K. M., FINDLAY, S. D. & POSTOVIT, L. M. 2015. Assessing breast cancer cell lines as tumour models by comparison of mRNA expression profiles. *Breast Cancer Research,* 17**,** 114.

VODNIK, M., ZAGER, U., STRUKELJ, B. & LUNDER, M. 2011. Phage display: selecting straws instead of a needle from a haystack. *Molecules,* 16**,** 790-817.

VOLKOV, M., KAMPSTRA, A. S. B., VAN SCHIE, K. A., KAWAKAMI, A., TAMAI, M., KAWASHIRI, S., MAEDA, T., HUIZINGA, T. W. J., TOES, R. E. M. & VAN DER WOUDE, D. 2021. Evolution of anti-modified protein antibody responses can be driven by consecutive exposure to different post-translational modifications. *Arthritis Research & Therapy,* 23**,** 298.

WADA, A., TERASHIMA, T., KAGEYAMA, S., YOSHIDA, T., NARITA, M., KAWAUCHI, A. & KOJIMA, H. 2019. Efficient Prostate Cancer Therapy with Tissue-Specific Homing Peptides Identified by Advanced Phage Display Technology. *Molecular Therapy - Oncolytics,* 12**,** 138-146.

WALSH, H. R., CRUICKSHANK, B. M., BROWN, J. M. & MARCATO, P. 2019. The Flick of a Switch: Conferring Survival Advantage to Breast Cancer Stem Cells Through Metabolic Plasticity. *Frontiers in Oncology,* 9.

WANG, T., FAHRMANN, J. F., LEE, H., LI, Y.-J., TRIPATHI, S. C., YUE, C., ZHANG, C., LIFSHITZ, V., SONG, J., YUAN, Y., SOMLO, G., JANDIAL, R., ANN, D., HANASH,

S., JOVE, R. & YU, H. 2018. JAK/STAT3-Regulated Fatty Acid β-Oxidation Is Critical for Breast Cancer Stem Cell Self-Renewal and Chemoresistance. *Cell metabolism,* 27**,** 136-150.e5.

WARD, H. W. 1973. Anti-oestrogen therapy for breast cancer: a trial of tamoxifen at two dose levels. *British medical journal,* 1**,** 13-14.

WARDEMANN, H. & NUSSENZWEIG, M. C. 2007. B‐Cell Self‐Tolerance in Humans. *Advances in Immunology.* Academic Press.

WEDAM, S., FASHOYIN-AJE, L., GAO, X., BLOOMQUIST, E., TANG, S., SRIDHARA, R., GOLDBERG, K. B., KING-KALLIMANIS, B. L., THEORET, M. R., IBRAHIM, A., AMIRI-KORDESTANI, L., PAZDUR, R. & BEAVER, J. A. 2020. FDA Approval Summary: Ado-trastuzumab emtansine for the adjuvant treatment of HER2-positive early breast cancer. *Clinical Cancer Research***,** clincanres.3980.2019.

WEIGELT, B., PETERSE, J. L. & VAN'T VEER, L. J. 2005. Breast cancer metastasis: markers and models. *Nature Reviews Cancer,* 5**,** 591.

WELCH, H. G. & BLACK, W. C. 2010. Overdiagnosis in Cancer. *JNCI: Journal of the National Cancer Institute,* 102**,** 605-613.

WEN, H. Y. & BROGI, E. 2018. Lobular Carcinoma In Situ. *Surgical pathology clinics,* 11**,** 123-145.

WHITNEY, M., CRISP, J. L., OLSON, E. S., AGUILERA, T. A., GROSS, L. A., ELLIES, L. G. & TSIEN, R. Y. 2010. Parallel in vivo and in vitro selection using phage display identifies protease-dependent tumor-targeting peptides. *The Journal of biological chemistry,* 285**,** 22532-22541.

WHITT, J., HONG, W. S., TELANGE, R. R., LIN, C. P., BIBB, J., BEEBE, D. J., CHEN, H. & JASKULA-SZTUL, R. 2020. Non-toxic fragment of botulinum neurotoxin type A and monomethyl auristatin E conjugate for targeted therapy for neuroendocrine tumors. *Cancer Gene Therapy*.

WORM, D. J., ELS-HEINDL, S. & BECK-SICKINGER, A. G. 2020. Targeting of peptide-binding receptors on cancer cells with peptide-drug conjugates. *Peptide Science,* 112**,** e24171.

WU, C.-P., CALCAGNO, A. M. & AMBUDKAR, S. V. 2008. Reversal of ABC drug transporter-mediated multidrug resistance in cancer cells: evaluation of current strategies. *Current molecular pharmacology,* 1**,** 93-105.

XIAO, W., LI, T., BONONI, F. C., LAC, D., KEKESSIE, I. A., LIU, Y., SANCHEZ, E., MAZLOOM, A., MA, A.-H., LIN, J., TRAN, J., YANG, K., LAM, K. S. & LIU, R. 2016. Discovery and characterization of a high-affinity and high-specificity peptide ligand LXY30 for in vivo targeting of α3 integrin-expressing human tumors. *EJNMMI Research,* 6**,** 18.

YANG, F., HUANG, X., SUN, C., LI, J., WANG, B., YAN, M., JIN, F., WANG, H., ZHANG, J., FU, P., ZENG, T., WANG, J., LI, W., LI, Y., YANG, M., LI, J., WU, H., FU, Z., YIN, Y. & JIANG, Z. 2020. Lapatinib in combination with capecitabine versus continued use of trastuzumab in breast cancer patients with trastuzumab-resistance: a retrospective study of a Chinese population. *BMC cancer,* 20**,** 255-255.

YANG, G., MA, W., ZHANG, B. & XIE, Q. 2016. The labeling of stem cells by superparamagnetic iron oxide nanoparticles modified with PEG/PVP or PEG/PEI. *Materials Science and Engineering: C,* 62**,** 384-390.

YANG, W., YOON, A., LEE, S., KIM, S., HAN, J. & CHUNG, J. 2017. Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. *Experimental & Molecular Medicine,* 49**,** e308-e308.

YANG, Y., WANG, J., XU, S., SHI, F. & SHAN, A. 2021. Calumenin contributes to epithelial-mesenchymal transition and predicts poor survival in glioma. *Transl Neurosci,* 12**,** 67-75.

YEN, M. F., TABÁR, L., VITAK, B., SMITH, R. A., CHEN, H. H. & DUFFY, S. W. 2003. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *European Journal of Cancer,* 39**,** 1746-1754.

YERSAL, O. & BARUTCA, S. 2014. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World journal of clinical oncology,* 5**,** 412-424.

YIN, H. & GLASS, J. 2011. The phenotypic radiation resistance of CD44+/CD24(-or low) breast cancer cells is mediated through the enhanced activation of ATM signaling. *PloS one,* 6**,** e24080-e24080.

YIN, L., DUAN, J.-J., BIAN, X.-W. & YU, S.-C. 2020. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research,* 22**,** 61.

YU, K. D., ZHU, R., ZHAN, M., RODRIGUEZ, A. A., YANG, W., WONG, S., MAKRIS, A., LEHMANN, B. D., CHEN, X., MAYER, I., PIETENPOL, J. A., SHAO, Z. M., SYMMANS, W. F. & CHANG, J. C. 2013. Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer. *Clin Cancer Res,* 19**,** 2723-33.

YUAN, T., MOHAN, D., LASERSON, U., RUCZINSKI, I., BAER, A. N. & LARMAN, H. B. 2018. Improved Analysis of Phage ImmunoPrecipitation Sequencing (PhIP-Seq) Data Using a Z-score Algorithm. *bioRxiv***,** 285916.

ZACKRISSON, S., ANDERSSON, I., JANZON, L., MANJER, J. & GARNE, J. P. 2006. Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ (Clinical research ed.),* 332**,** 689-692.

ZAENKER, P., GRAY, E. S. & ZIMAN, M. R. 2016. Autoantibody Production in Cancer—The Humoral Immune Response toward Autologous Antigens in Cancer Patients. *Autoimmunity Reviews,* 15**,** 477-483.

ZAGAMI, P. & CAREY, L. A. 2022. Triple negative breast cancer: Pitfalls and progress. *npj Breast Cancer,* 8**,** 95.

ZHANG, C., TANG, N., LIU, X., LIANG, W., XU, W. & TORCHILIN, V. P. 2006. siRNA-containing liposomes modified with polyarginine effectively silence the targeted gene. *J Control Release,* 112**,** 229-39.

ZHANG, H., TORKAMANI, A., JONES, T., RUIZ, D., PONS, J. & LERNER, R. 2011. Phenotype-information-phenotype cycle for deconvolution of combinatorial antibody libraries selected against complex systems. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 13456-61.

ZHANG, J.-F., XIONG, H.-L., CAO, J.-L., WANG, S.-J., GUO, X.-R., LIN, B.-Y., ZHANG, Y., ZHAO, J.-H., WANG, Y.-B., ZHANG, T.-Y., YUAN, Q., ZHANG, J. & XIA, N.-S. 2018. A cell-penetrating whole molecule antibody targeting intracellular HBx suppresses hepatitis B virus via TRIM21-dependent pathway. *Theranostics,* 8**,** 549-562.

ZHANG, P., LIU, R., JIANG, H., LIU, Y., ZHANG, Q. & HUANG, Y. 2012. [Antitumor effect of a triple negative breast cancer-dendritic cell fusion vaccine]. *Nan Fang Yi Ke Da Xue Xue Bao,* 32**,** 778-83.

ZHANG, P., LOCK, L. L., CHEETHAM, A. G. & CUI, H. 2014. Enhanced Cellular Entry and Efficacy of Tat Conjugates by Rational Design of the Auxiliary Segment. *Molecular Pharmaceutics,* 11**,** 964-973.

ZHANG, X., POWELL, K. & LI, L. 2020. Breast Cancer Stem Cells: Biomarkers, Identification and Isolation Methods, Regulating Mechanisms, Cellular Origin, and Beyond. *Cancers (Basel),* 12.

ZHANG, Y., WEN, Z., WASHBURN, M. P. & FLORENS, L. 2010. Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins. *Analytical Chemistry,* 82**,** 2272-2281.

ZHONG, L., GE, K., ZU, J.-C., ZHAO, L.-H., SHEN, W.-K., WANG, J.-F., ZHANG, X.-G., GAO, X., HU, W., YEN, Y. & KERNSTINE, K. H. 2008. Autoantibodies as potential biomarkers for breast cancer. *Breast cancer research : BCR,* 10**,** R40-R40.

ZHOU, C., KANG, J., WANG, X., WEI, W. & JIANG, W. 2015a. Phage display screening identifies a novel peptide to suppress ovarian cancer cells in vitro and in vivo in mouse models. *BMC Cancer,* 15**,** 889.

ZHOU, S. M., CHENG, L., GUO, S. J., WANG, Y., CZAJKOWSKY, D. M., GAO, H., HU, X. F. & TAO, S. C. 2015b. Lectin RCA-I specifically binds to metastasis-associated cell surface glycans in triple-negative breast cancer. *Breast Cancer Res,* 17**,** 36.

ZHU, X., LU, N., ZHOU, Y., XUAN, S., ZHANG, J., GIAMPIERI, F., ZHANG, Y., YANG, F., YU, R., BATTINO, M. & WANG, Z. 2019. Targeting Pancreatic Cancer Cells with Peptide-Functionalized Polymeric Magnetic Nanoparticles. *International Journal of Molecular Sciences,* 20**,** 2988.

ZOU, Y., XIA, Y., MENG, F., ZHANG, J. & ZHONG, Z. 2018. GE11-Directed Functional Polymersomal Doxorubicin as an Advanced Alternative to Clinical Liposomal Formulation for Ovarian Cancer Treatment. *Molecular Pharmaceutics,* 15**,** 3664-3671.

Appendices

Appendix 1. Permission to reuse copyright material

Figure 3.1

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Sep 20, 2022

This Agreement between University of Nottingham -- Megan Wilde ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	5393130891804
License date	Sep 20, 2022
Licensed content publisher	Oxford University Press
Licensed content publication	FEMS Microbiology Reviews
Licensed content title	Phage display and other peptide display technologies
Licensed content author	Jaroszewicz, Weronika; Morcinek-Orłowska, Joanna
Licensed content date	Oct 21, 2021
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	PhD thesis
Publisher of your work	University of Nottingham
Expected publication date	Mar 2023
Permissions cost	0.00 GBP
Value added tax	0.00 GBP
Total	0.00 GBP
Title	PhD thesis
Institution name	University of Nottingham
Expected presentation date	Mar 2023
Portions	Figure 5. Adapted phagemid section of the figure
	University of Nottingham Sutton Bonington Campus
Requestor Location	
	Leciestershire, LE12 5RD United Kingdom Attn: University ofNottingham
Publisher Tax ID	GB125506730
Total	0.00 GBP
Terms and Conditions	

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

Figure 3.2

Figure adapted under the terms of the Creative Commons Attribution 4.0 International Licence (http://creativecommons.org/licenses/by/4.0/)

Appendix 2. Materials and methods

Appendix 2.1 NGS barcoded primers

Supplementary Table 1 Ion Torrent sequencing primers

Primer name	Sequence 5'-3' (barcode sequence is highlighted in bold, followed by linker1)
Akey-BC1-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTAAGGTAAC**GTAATCCTTGTGGTATCG
Akey-BC2-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TAAGGAGAAC**GTAATCCTTGTGGTATCG
Akey-BC3-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**AAGAGGATTC**GTAATCCTTGTGGTATCG
Akey-BC4-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TACCAAGATC**GTAATCCTTGTGGTATCG
Akey-BC5-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CAGAAGGAAC**GTAATCCTTGTGGTATCG
Akey-BC6-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTGCAAGTTC**GTAATCCTTGTGGTATCG
Akey-BC7-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTCGTGATTC**GTAATCCTTGTGGTATCG
Akey-BC8-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTCCGATAAC**GTAATCCTTGTGGTATCG
Akey-BC9-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TGAGCGGAAC**GTAATCCTTGTGGTATCG
Akey-BC10-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTGACCGAAC**GTAATCCTTGTGGTATCG
Akey-BC11-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCCTCGAATC**GTAATCCTTGTGGTATCG
Akey-BC12-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TAGGTGGTTC**GTAATCCTTGTGGTATCG
Akey-BC13-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCTAACGGAC**GTAATCCTTGTGGTATCG
Akey-BC14-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTGGAGTGTC**GTAATCCTTGTGGTATCG
Akey-BC15-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCTAGAGGTC**GTAATCCTTGTGGTATCG
Akey-BC16-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCTGGATGAC**GTAATCCTTGTGGTATCG
Akey-BC17-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCTATTCGTC**GTAATCCTTGTGGTATCG
Akey-BC18-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**AGGCAATTGC**GTAATCCTTGTGGTATCG
Akey-BC19-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTAGTCGGAC**GTAATCCTTGTGGTATCG
Akey-BC20-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CAGATCCATC**GTAATCCTTGTGGTATCG
Akey-BC21-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCGCAATTAC**GTAATCCTTGTGGTATCG
Akey-BC22-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTCGAGACGC**GTAATCCTTGTGGTATCG
Akey-BC23-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TGCCACGAAC**GTAATCCTTGTGGTATCG
Akey-BC24-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**AACCTCATTC**GTAATCCTTGTGGTATCG

Akey-BC25-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCTGAGATAC**GTAATCCTTGTGGTATCG
Akey-BC26-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTACAACCTC**GTAATCCTTGTGGTATCG
Akey-BC27-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**AACCATCCGC**GTAATCCTTGTGGTATCG
Akey-BC28-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**ATCCGGAATC**GTAATCCTTGTGGTATCG
Akey-BC29-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCGACCACTC**GTAATCCTTGTGGTATCG
Akey-BC30-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CGAGGTTATC**GTAATCCTTGTGGTATCG
Akey-BC31-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCCAAGCTGC**GTAATCCTTGTGGTATCG
Akey-BC32-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCTTACACAC**GTAATCCTTGTGGTATCG
Akey-BC33-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTCTCATTGAAC**GTAATCCTTGTGGTATCG
Akey-BC34-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCGCATCGTTC**GTAATCCTTGTGGTATCG
Akey-BC35-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TAAGCCATTGTC**GTAATCCTTGTGGTATCG
Akey-BC36-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**AAGGAATCGTC**GTAATCCTTGTGGTATCG
Akey-BC37-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTTGAGAATGTC**GTAATCCTTGTGGTATCG
Akey-BC38-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TGGAGGACGGAC**GTAATCCTTGTGGTATCG
Akey-BC39-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TAACAATCGGC**GTAATCCTTGTGGTATCG
Akey-BC40-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTGACATAATC**GTAATCCTTGTGGTATCG
Akey-BC41-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTCCACTTCGC**GTAATCCTTGTGGTATCG
Akey-BC43-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTTGACACCGC**GTAATCCTTGTGGTATCG
Akey-BC44-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTGGAGGCCAGC**GTAATCCTTGTGGTATCG
Akey-BC45-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TGGAGCTTCCTC**GTAATCCTTGTGGTATCG
Akey-BC46-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCAGTCCGAAC**GTAATCCTTGTGGTATCG
Akey-BC47-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TAAGGCAACCAC**GTAATCCTTGTGGTATCG
Akey-BC48-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTCTAAGAGAC**GTAATCCTTGTGGTATCG
Akey-BC49-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCCTAACATAAC**GTAATCCTTGTGGTATCG
Akey-BC50-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CGGACAATGGC**GTAATCCTTGTGGTATCG
Akey-BC51-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTGAGCCTATTC**GTAATCCTTGTGGTATCG
Akey-BC52-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCGCATGGAAC**GTAATCCTTGTGGTATCG
Akey-BC53-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTGGCAATCCTC**GTAATCCTTGTGGTATCG

Akey-BC54-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCGGAGAATCGC**GTAATCCTTGTGGTATCG
Akey-BC55-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCCACCTCCTC**GTAATCCTTGTGGTATCG
Akey-BC56-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CAGCATTAATTC**GTAATCCTTGTGGTATCG
Akey-BC57-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCTGGCAACGGC**GTAATCCTTGTGGTATCG
Akey-BC58-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCCTAGAACAC**GTAATCCTTGTGGTATCG
Akey-BC59-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCCTTGATGTTC**GTAATCCTTGTGGTATCG
Akey-BC60-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCTAGCTCTTC**GTAATCCTTGTGGTATCG
Akey-BC61-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCACTCGGATC**GTAATCCTTGTGGTATCG
Akey-BC62-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTCCTGCTTCAC**GTAATCCTTGTGGTATCG
Akey-BC63-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCTTAGAGTTC**GTAATCCTTGTGGTATCG
Akey-BC64-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTGAGTTCCGAC**GTAATCCTTGTGGTATCG
Akey-BC65-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCCTGGCACAT**CGTAATCCTTGTGGTATCG
Akey-BC66-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCGCAATCATC**GTAATCCTTGTGGTATCG
Akey-BC67-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTCCTACCAGTC**GTAATCCTTGTGGTATCG
Akey-BC68-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCAAGAAGTTC**GTAATCCTTGTGGTATCG
Akey-BC69-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTCAATTGGC**GTAATCCTTGTGGTATCG
Akey-BC70-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCTACTGGTC**GTAATCCTTGTGGTATCG
Akey-BC71-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TGAGGCTCCGAC**GTAATCCTTGTGGTATCG
Akey-BC72-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CGAAGGCCACAC**GTAATCCTTGTGGTATCG
Akey-BC73-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCTGCCTGTC**GTAATCCTTGTGGTATCG
Akey-BC74-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CGATCGGTTC**GTAATCCTTGTGGTATCG
Akey-BC75-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCAGGAATAC**GTAATCCTTGTGGTATCG
Akey-BC76-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CGGAAGAACCTC**GTAATCCTTGTGGTATCG
Akey-BC77-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CGAAGCGATTC**GTAATCCTTGTGGTATCG
Akey-BC78-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CAGCCAATTCTC**GTAATCCTTGTGGTATCG
Akey-BC79-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCTGGTTGTC**GTAATCCTTGTGGTATCG
Akey-BC80-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCGAAGGCAGGC**GTAATCCTTGTGGTATCG
Akey-BC81-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCTGCCATTCGC**GTAATCCTTGTGGTATCG

Akey-BC82-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTGGCATCTC**GTAATCCTTGTGGTATCG
Akey-BC83-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTAGGACATTC**GTAATCCTTGTGGTATCG
Akey-BC84-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTTCCATAAC**GTAATCCTTGTGGTATCG
Akey-BC85-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCAGCCTCAAC**GTAATCCTTGTGGTATCG
Akey-BC86-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTTGGTTATTC**GTAATCCTTGTGGTATCG
Akey-BC87-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTGGCTGGAC**GTAATCCTTGTGGTATCG
Akey-BC88-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CCGAACACTTC**GTAATCCTTGTGGTATCG
Akey-BC89-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCCTGAATCTC**GTAATCCTTGTGGTATCG
Akey-BC90-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTAACCACGGC**GTAATCCTTGTGGTATCG
Akey-BC91-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CGGAAGGATGC**GTAATCCTTGTGGTATCG
Akey-BC92-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTAGGAACCGC**GTAATCCTTGTGGTATCG
Akey-BC93-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CTTGTCCAATC**GTAATCCTTGTGGTATCG
Akey-BC94-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TCCGACAAGC**GTAATCCTTGTGGTATCG
Akey-BC95-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**CGGACAGATC**GTAATCCTTGTGGTATCG
Akey-BC96-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAG**TTAAGCGGTC**GTAATCCTTGTGGTATCG

N.B Akey-BC93-Ink1 was used to label the first PCR product for sample 42.

All scripts were kindly provided by Dr Jonathan Owen, (ADAS UK group, School of Veterinary Medicine and Science, University of Nottingham). Adaptations to the scripts to fit this analysis are highlighted.

pipeline.1.02 script

The script was used to process fastq files from the University of Pennsylvania for the NGS analysis. The script firstly converted decompressed fastq files to fasta files, which were then used to demultiplex the NGS data using the list of identifier barcodes shown in Appendix 2.1.

The script runs as follows:

1. convert a (.tgz compressed) fastq file to fasta
2. demultiplex the data using a list of supplied barcodes
3. reverse complement (optional)
4. translate all 3 forward frames
5. identify flanking motifs using a list supplied in 2 files

e.g. --left PAMAMSRP.txt --right PRRRR.txt

6. create fasta files containing the peptides between the motifs
7. rank the frequencies of the peptides on a per barcode basis

Compare_collapsed_Zscore_v2.0.pl

The script automatically ranked output peptides based on their Z score. The Z score ranking system was based on the following equation:

$$z = \frac{p1 - p2}{\sqrt{\frac{p1(1-p1)}{n1} + \frac{p2(1-p2)}{n2}}}$$

where:

p1= the number of reads obtained for a specific peptide in the target sample

p2= the number of reads obtained for a specific peptide in the control pool

n1= total peptide read count (target sample)

n2= total peptide read count (control pool)

The following command is an example of how the Z score script was used to determine which BT549 peptides had a Z score greater than 2 in three replicates:

```
# join all the negative control samples to form a control pool

cat MW_panning_BC36.LR.fasta MW_panning_BC37.LR.fasta
MW_panning_BC38.LR.fasta MW_panning_BC39.LR.fasta
MW_panning_BC40.LR.fasta MW_panning_BC41.LR.fasta
MW_panning_BC93.LR.fasta MW_panning_BC43.LR.fasta
MW_panning_BC44.LR.fasta MW_panning_BC45.LR.fasta
MW_panning_BC46.LR.fasta MW_panning_BC47.LR.fasta
MW_panning_BC48.LR.fasta MW_panning_BC49.LR.fasta
MW_panning_BC50.LR.fasta >BT549_controls_36-50.fasta

# compare BT549 samples vs all controls

perl compare_collapsed_Zscore_v2.0_percentages.pl MW_panning_BC31.LR.fasta
BT549_controls_36-50.fasta >BT5491vs36-50.table
perl compare_collapsed_Zscore_v2.0_percentages.pl MW_panning_BC32.LR.fasta
BT549_controls_36-50.fasta >BT5492vs36-50.table
perl compare_collapsed_Zscore_v2.0_percentages.pl MW_panning_BC33.LR.fasta
BT549_controls_36-50.fasta >BT5493vs36-50.table
perl compare_collapsed_Zscore_v2.0_percentages.pl MW_panning_BC34.LR.fasta
BT549_controls_36-50.fasta >BT5494vs36-50.table
perl compare_collapsed_Zscore_v2.0_percentages.pl MW_panning_BC35.LR.fasta
BT549_controls_36-50.fasta >BT5495vs36-50.table
```

```
# get top Z score 2 or more, get sequences # NR>1 && NR<21 to rank by top X Z scores

awk ' NR>1 && $4>=2 (print $1;)' BT5491vs36-50.table > BT5491-BT5495_Z2vs36-50_sequences
awk ' NR>1 && $4>=2 [print $1;]' BT5492vs36-50.table >BT5491-BT5495_Z2vs36-50_sequences
awk ' NR>1 && $4>=2 [print $1;]' BT5493vs36-50.table > BT5491-BT5495_Z2vs36-50_sequences
awk ' NR>1 && $4>=2 [print $1;]' BT5494vs36-50.table > BT5491-BT5495_Z2vs36-50_sequences
awk ' NR>1 && $4>=2 [print $1;]' BT5495vs36-50.table > BT5491-BT5495_Z2vs36-50_sequences

# Determine how many times a given peptide has a Z score > 2 across 5 replicates

cat BT5491-BT5495_Z2vs36-50_sequences | sort | uniq -c | sort -nr | sed 's/^[ \t]*//' > BT5491-BT5495_Z2vs36-50_sequences_time_seen
```

Countpep Perl script

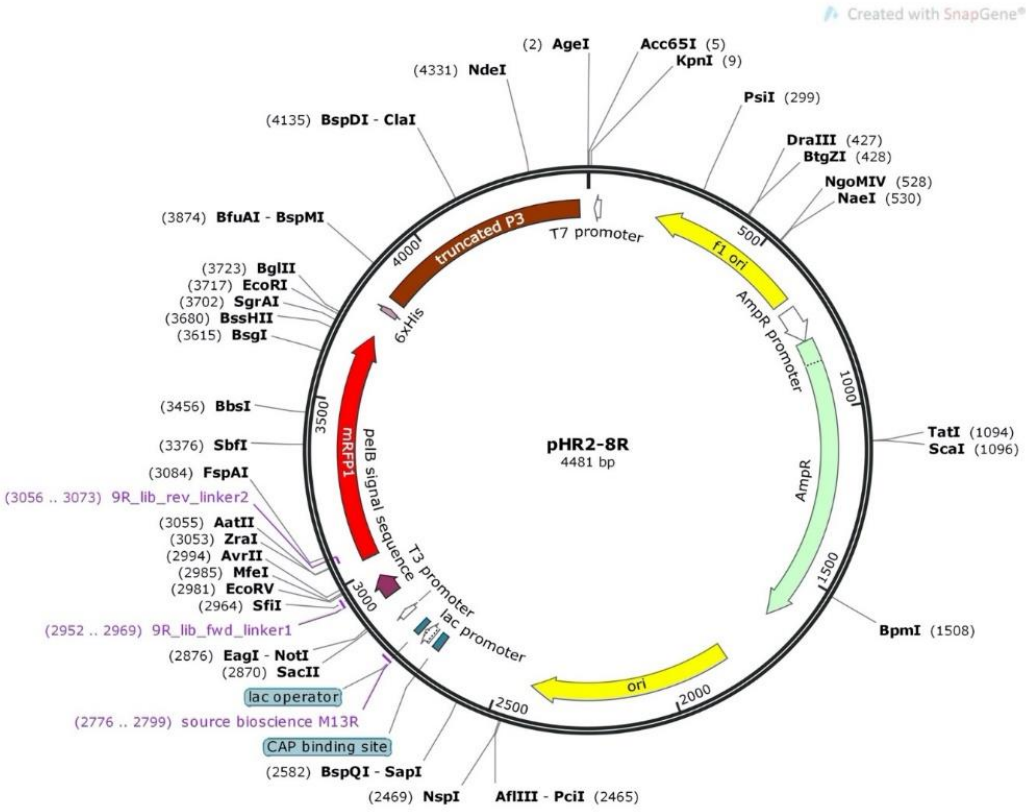
For further analysis of the peptides identified from the compare_collapsed_ Zscore_v2.0_percentages script, the CountPep Perl script was applied to count the frequency of a list of peptides within selected barcodes or a control pool listed in a text file. For example, the following command was used to determine the frequency of peptides with a Z score greater than 2 in all BT549 barcodes and the BT549 control pool:

```
perl countpep.pl BT549_Zscore2_peptides.txt BT549_barcodes.txt
```

This led to two output table files being obtained. The first contained the absolute frequency of each peptide in each barcode and the second contained the percentage of each peptide based on the total number of reads in a particular barcode.

Appendix 3. Identification of potential TNBC targeting small peptides by NGPD

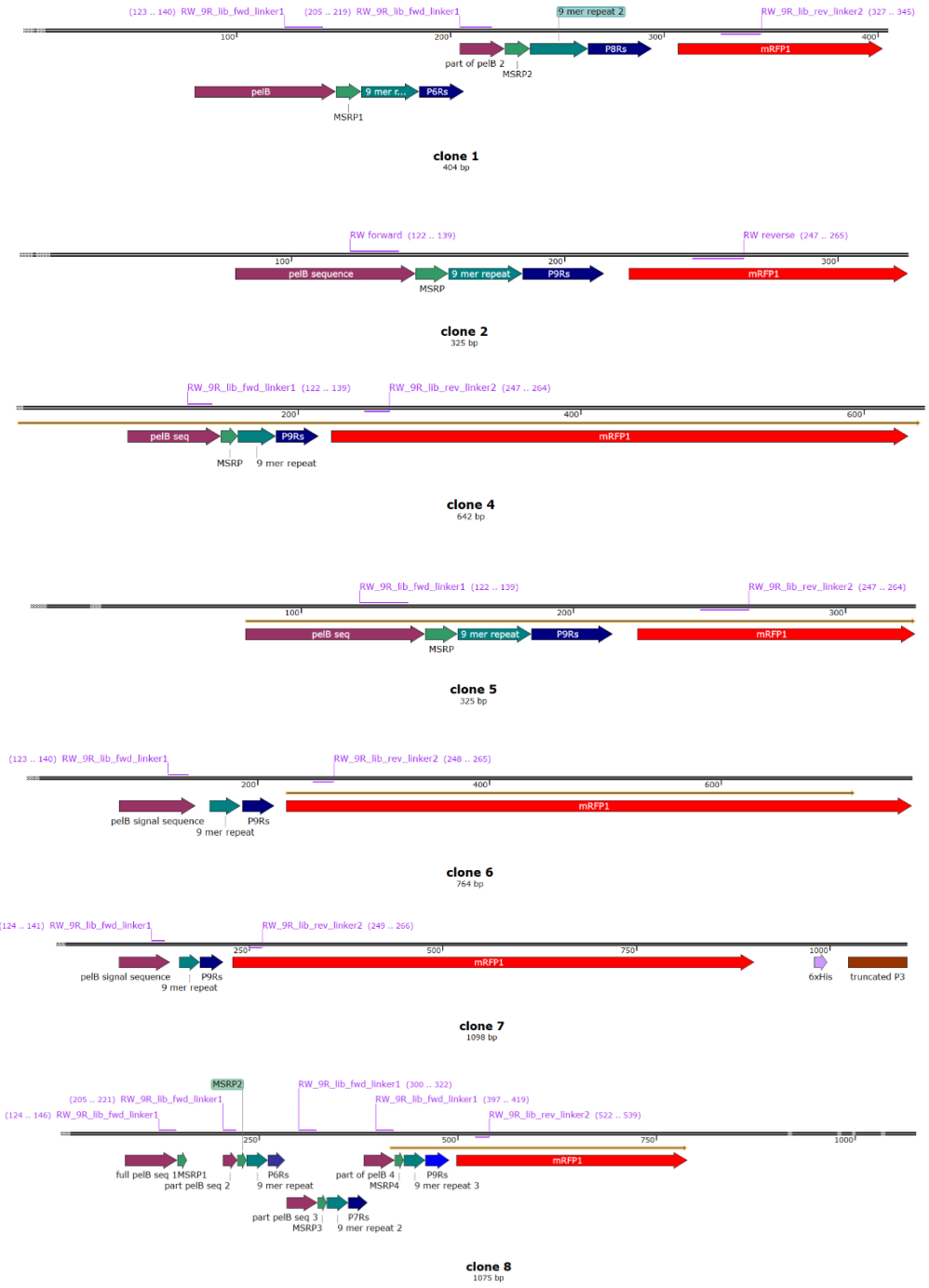
Appendix 3.1 Phagemid map

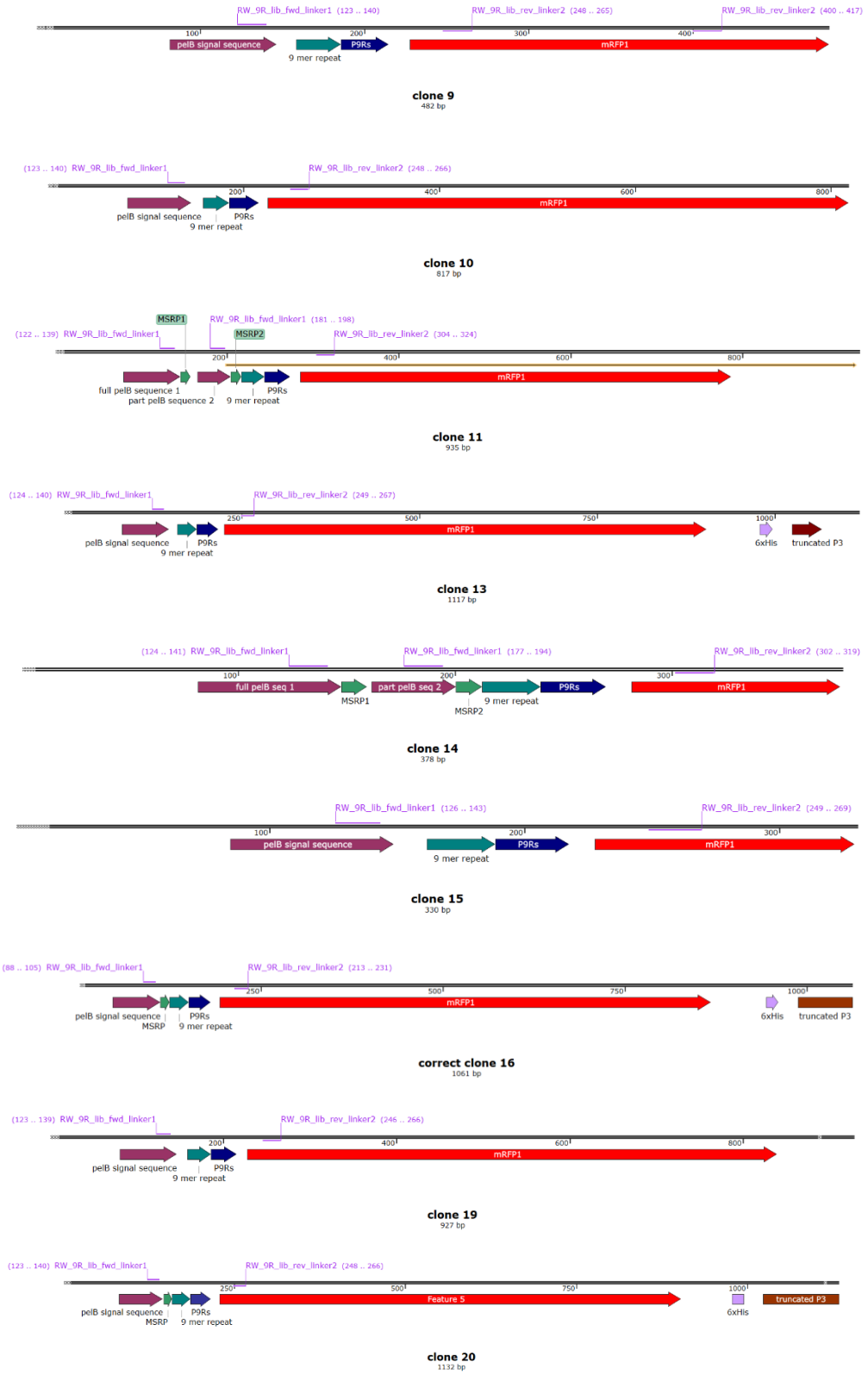


Supplementary Figure 1 pHR2-8R phagemid map

The location of the forward (9R_lib_fwd_linker1) and reverse primer (9R_lib_rev_linker2) annealing sites are highlighted. The annealing site for the M13R source bioscience primer used for Sanger sequencing of individual clones is also shown.

Appendix 3.2 Sanger sequencing of 9-mer library clones

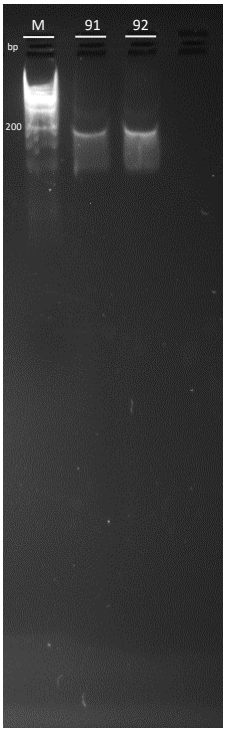




Supplementary Figure 2 Sanger sequencing of bacteriophage library clones

Sequences were annotated in SnapGene Viewer and show annealing sites for the forward and reverse 9R library primers. Clones 1, 8, 11 and 14 show the incorrect insertion of multiple pelB signal sequences.

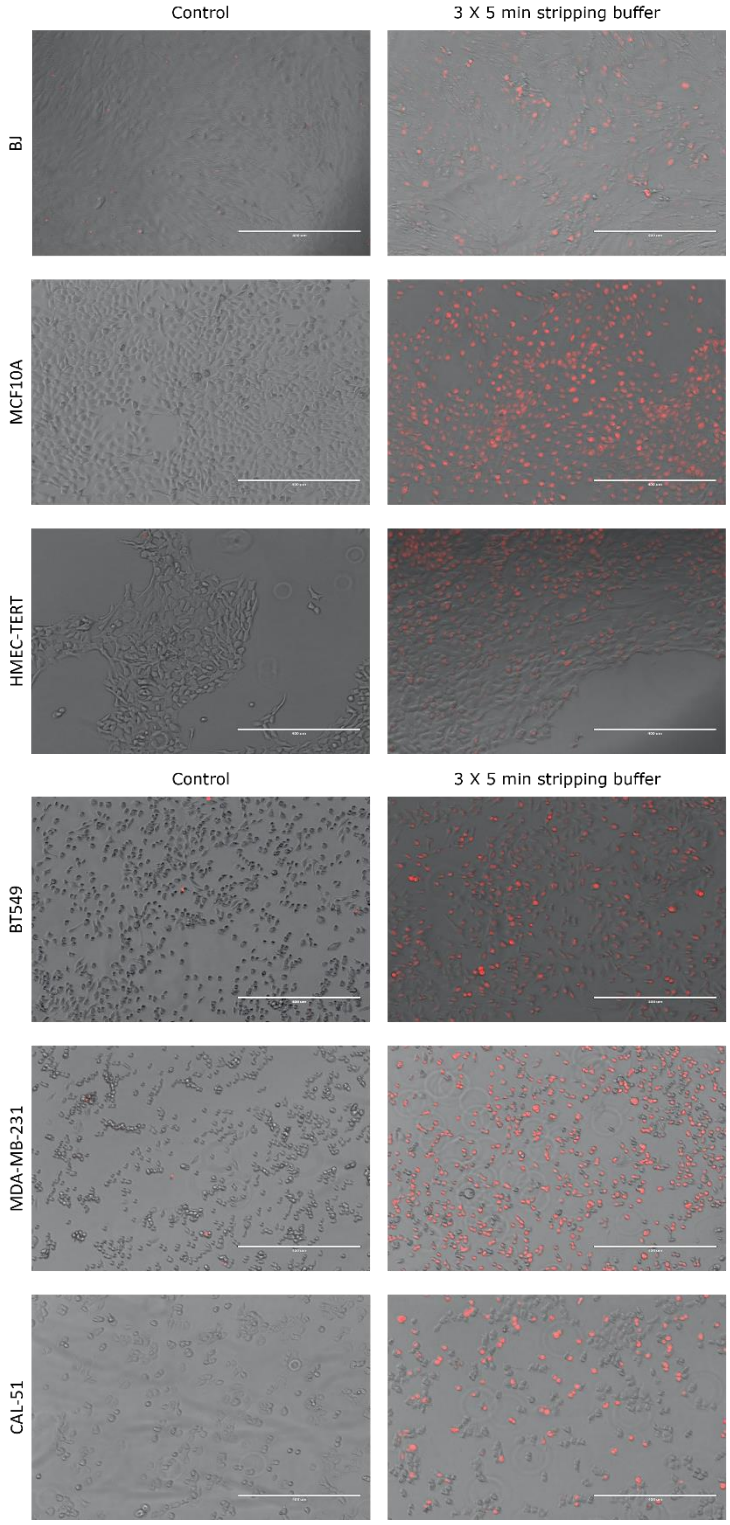
Appendix 3.3 PCR amplification of library clone inserts



Supplementary Figure 3 PCR amplified nonapeptide inserts from library clones

The nonapeptide gene inserts and flanking regions found within the library were amplified using library-specific primers. Amplification of the 176 base pair insert was confirmed across on a 1% agarose gel at 100V. MassRuler™ DNA ladder was used as the molecular marker.

Appendix 3.4 Cell viability with biopanning reagents



Supplementary Figure 4 Cell viability of non-tumourigenic and TNBC cell lines after bacteriophage stripping buffer washes.

Cell permeability as a result of stripping buffer treatment was visualised with propidium iodide staining. Scale bar = 400µM.

Appendix 3.5 NGS barcode reads

Supplementary Table 2 The number of NGS reads for each barcode

Barcode	reads	Percentage	Barcode	reads	Percentage
BC01	674395	1.15	BC47	781952	1.34
BC02	726809	1.24	BC48	771671	1.32
BC03	689805	1.18	BC49	753651	1.29
BC04	726325	1.24	BC50	621435	1.06
BC05	599615	1.02	BC51	684768	1.17
BC06	641265	1.10	BC52	332	0.00
BC07	618280	1.06	BC53	1074648	1.84
BC08	635697	1.09	BC54	608451	1.04
BC09	597526	1.02	BC55	617197	1.05
BC10	597808	1.02	BC56	667406	1.14
BC11	653939	1.12	BC57	581115	0.99
BC12	551371	0.94	BC58	568381	0.97
BC13	657549	1.12	BC59	758073	1.29
BC14	633216	1.08	BC60	784105	1.34
BC15	667561	1.14	BC61	614630	1.05
BC16	610375	1.04	BC62	708676	1.21
BC17	616313	1.05	BC63	450533	0.77
BC18	74601	0.13	BC64	639668	1.09
BC19	616746	1.05	BC65	590064	1.01
BC20	652615	1.11	BC66	614071	1.05
BC21	618188	1.06	BC67	636724	1.09
BC22	541437	0.92	BC68	679284	1.16
BC23	580543	0.99	BC69	627641	1.07
BC24	259912	0.44	BC70	572542	0.98
BC25	540104	0.92	BC71	609729	1.04
BC26	624092	1.07	BC72	542201	0.93
BC27	620944	1.06	BC73	594551	1.02
BC28	635538	1.09	BC74	594108	1.01
BC29	674417	1.15	BC75	648087	1.11
BC30	580240	0.99	BC76	626240	1.07

BC31	569041	0.97	BC77	739438	1.26
BC32	621011	1.06	BC78	651206	1.11
BC33	632317	1.08	BC79	610645	1.04
BC34	622809	1.06	BC80	663475	1.13
BC35	515204	0.88	BC81	669057	1.14
BC36	530515	0.91	BC82	625642	1.07
BC37	136948	0.23	BC83	712354	1.22
BC38	718062	1.23	BC84	706532	1.21
BC39	669961	1.14	BC85	645204	1.10
BC40	723277	1.24	BC86	707939	1.21
BC41	981552	1.68	BC87	664259	1.13
BC93	748798	1.28	BC88	740417	1.26
BC43	677927	1.16	BC89	747423	1.28
BC44	728917	1.25	BC90	744901	1.27
BC45	798452	1.36	BC91	768700	1.31
BC46	716792	1.22	BC92	776862	1.33

Akey-BC93-Ink1 was used to label the first PCR product for sample 42.

total sequences processed: 68379957

total barcodes found: 58539503 (85.61% of total sequences)

sequences containing motifs and at least 1 a.a. insert : 20636955 (35.25% of barcoded sequences)

sequences containing incomplete/absent motif matches: 37902548 (64.75% of barcoded sequences)

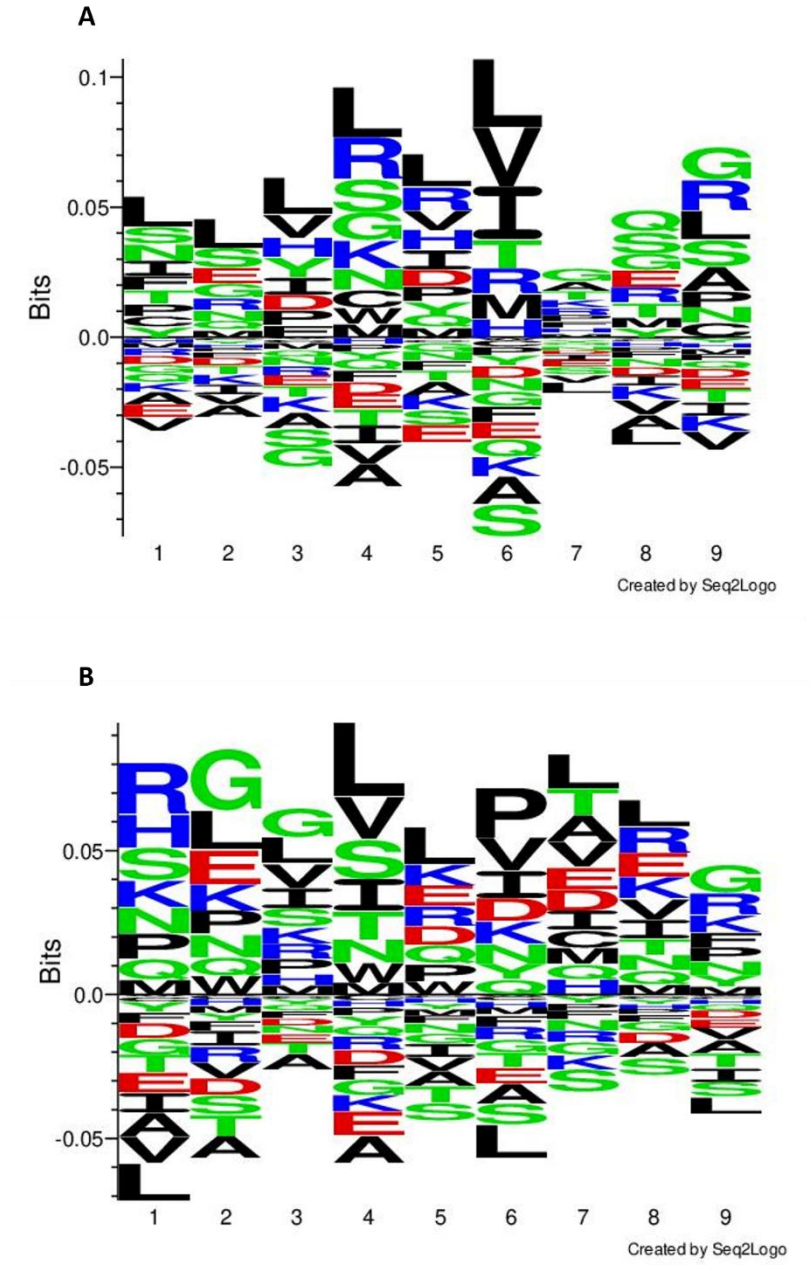
Appendix 3.6 TNBC targeting peptides in naïve bacteriophage library

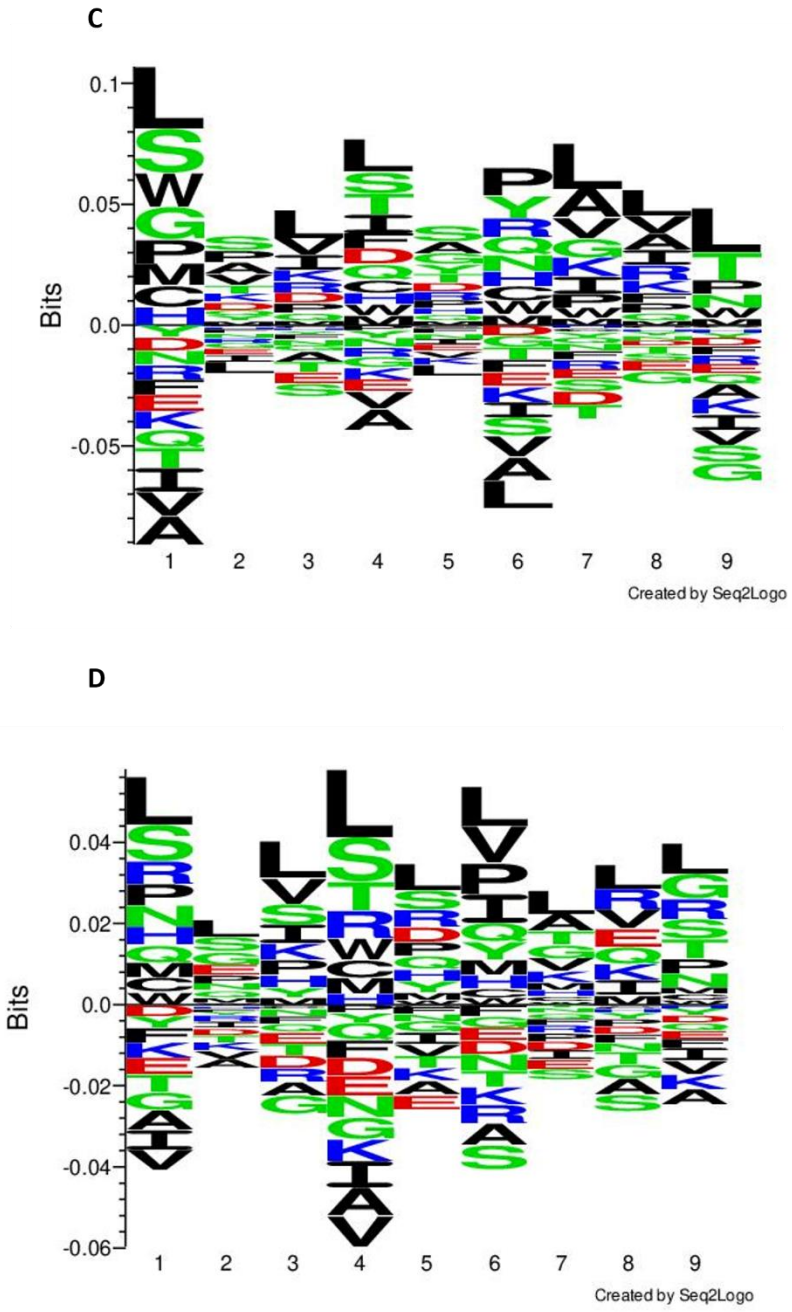
Supplementary Table 3 The number of NGS reads for each TNBC peptide in the original bacteriophage library sequencing data.

TNBC Peptide	Number of times seen in BC91 and BC92 NGS reads
TLYRMVVTS	0
NNVSVIRER	0
LSKKHMPVN	1
LESLSSMSP	5
SPVSSPPIT	20
qYRHSVLqL	3
LSPLqHAPQ	2
LMqDMALRL	0
LASMSYALL	1
CSTTYYTVK	0
MGSVRDTKF	1
LqVSLNSVM	2
HNPNqITMY	1
YGGSDTTMR	0
SLSLHVTqS	0
SLLGLHSSL	0
QSLWRLAqR	0
PEVLqMKMS	0
NSLMRqLGL	0
LQPRLVFQG	0
IMDNPLEqG	0
FSYSYRGRA	0
ELHRESHYV	1
CRHCIIGSC	0
VLHSPVCIG	0
SGSLWPQRN	0
SEGSLPERR	2
RWKVTLLNT	12
RKGLQYMEP	0
RGVTEKDER	0
QERLKPVLS	0
PPIELVMqG	5
NLLWSSASK	1
HTMLDQHTI	1
WVLSSLVRT	0
WLLCDPDAT	0

SVDFANKGP	0
STLLTWKEW	0
SKKqRRALL	0
PDSTYSGKN	0
MQIWLMLVG	3
MPEqGqLLL	0
HSLIPCLAS	1
GSVAHqWFT	0

Appendix 3.7 Seq2Logo results



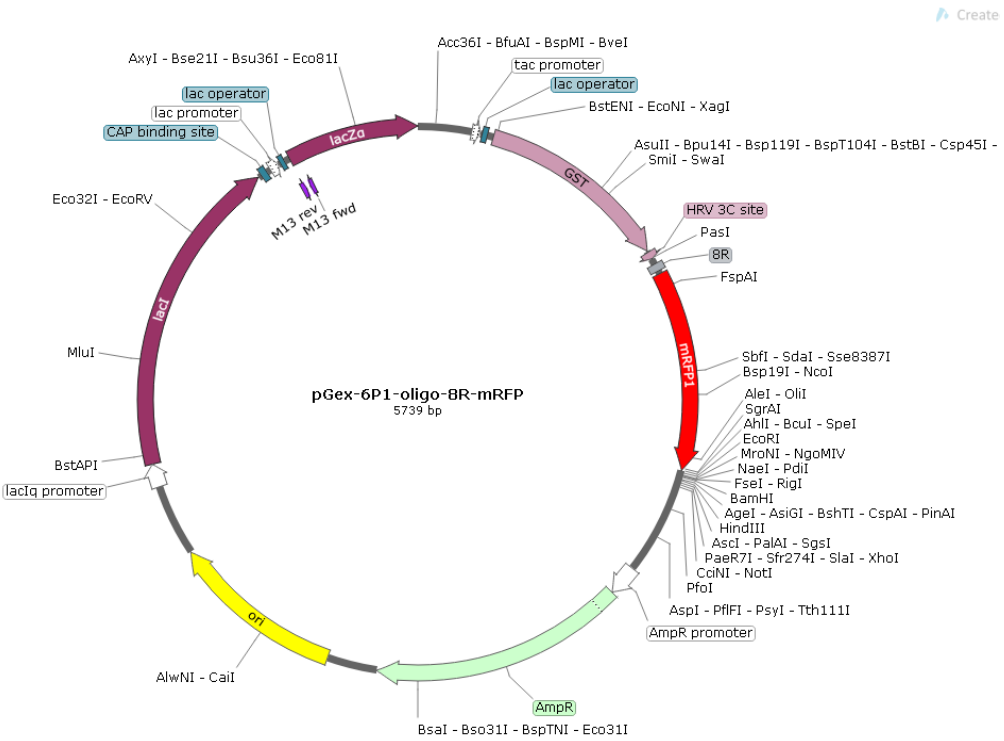


Supplementary Figure 5 Kullback-Leibler Seq2Logos for TNBC targeting 9-mer peptides showing the relative frequency of each amino acid residue at each position.

The logo alignments were generated for (A) MDA-MB-231, (B) BT-549 and (C) CAL-51 targeting peptides, as well as for (D) all identified TNBC targeting peptides. Positively charged, negatively charged, polar and non-polar amino acids are shown in blue, red, green and black, respectively (Thomsen and Nielsen, 2012).

Appendix 4. In vitro validation of potential TNBC targeting peptides

Appendix 4.1 pGEX-6P1-mRFP expression vector



Supplementary Figure 6 pGEX-6P1-mRFP expression vector map

Peptide encoding sequences were inserted by inverse PCR downstream of the HRV 3C site shown. Recovery of the 9-mer peptide encoding sequences into the vector led to the expression of peptide-mRFPs and recombinant protein was purified using the GST tag and PreScission protease cleavage at the HRV 3C site.

Appendix 4.2 pGEX-6P1-mRFP primers for cloning

Inverse PCR primers were designed based on inserting the correct peptide sequence into the pGEX-6P1-mRFP expression vector. The primer set 2 for the insertion of the p21 peptide, KRKKKGKGLGKKRDPCLRKYK, was designed based on using the recombinant plasmid containing peptide 1 as the template DNA.

Supplementary Table 4 Inverse PCR primers designed for the control peptides

Peptide	Forward primer 5'-3'	Reverse primer 5'-3'
mRFP only	ATGGCCTCCTCCGAGGAC	GGGCCCCTGGAACAGAAC
9R-mRFP	GGGCCCCTGGAACAGAAC	GGGCCCCTGGAACAGAAC
P21 set 1	AAGAAGCGCGATCCGTGCCTGCGCAAG TATAAGCCACGCCGAAGA	GCCCAGGCCTTTGCCCTTCTTCTTGCGCT TGGGGCGGCTCAT
P21- set 2	GCCCAGGCCTTTGCCCTTCTTCTTGCGCT TGGGGCGGCTCATGGGCCCCTGGAACA GAAC	AAGAAGCGCGATCCGTGCCTGCGCAAG TATAAGCCCAGACGAAGACGCAGGAGA

Supplementary Table 5 Inverse PCR primers designed for recovery of TNBC targeting peptides

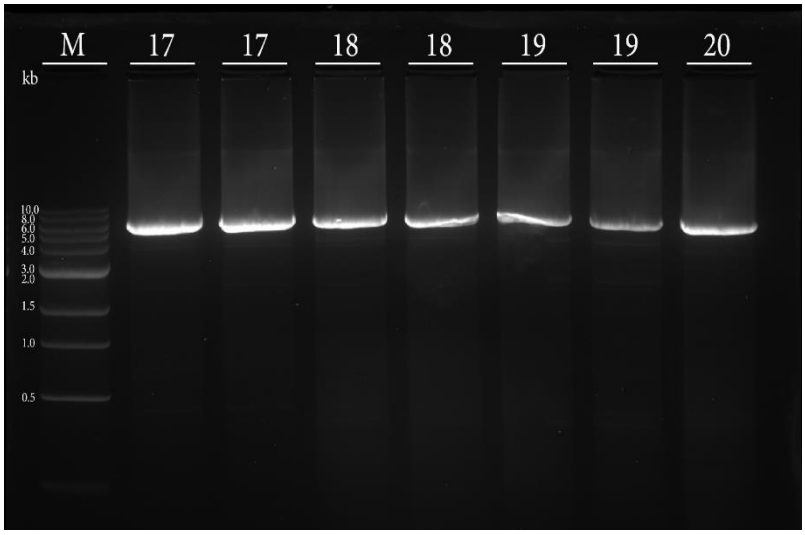
Peptide	Forward primer 5'-3'	Reverse primer 5'-3'
TLYRMVVTS	ATGGTTGTTACCTCG**CCACGC**CG AAGACGCAGGAGACGT	GCGGTACAGCGT**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
NNVSVIRER	GTAATCCGCGAGCGC**CCACGC**C GAAGACGCAGGAGACGT	GCTAACGTTGTT**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
LSKKHMPVN	CATATGCCCGTAAAC**CCACGC**C GAAGACGCAGGAGACGT	CTTTTTCGACAG**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
LESLSSMSP	AGCTCCATGAGCCCC**CCACGC**C GAAGACGCAGGAGACGT	CAGCGATTCCAG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
SPVSSPPIT	AGCCCACCAATCACT**CCACGC**CG AAGACGCAGGAGACGT	GCTAACAGGCGA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
qYRHSVLqL	TCGGTTTTACAACTG**CCACGC**CG AAGACGCAGGAGACGT	GTGGCGGTACTG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
LSPLqHAPQ	CAACATGCGCCACAG**CCACGC**C GAAGACGCAGGAGACGT	CAGAGGCGACAG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
LMqDMALRL	ATGGCGTTACGTTTG**CCACGC**C GAAGACGCAGGAGACGT	ATCCTGCATCAG**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
LASMSYALL	TCGTATGCGTTATTG**CCACGC**CG AAGACGCAGGAGACGT	CATCGAAGCCAG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC

CSTTYYTVK	TACTACACGGTTAAG**CCACGC**C GAAGACGCAGGAGACGT	GGTAGTCGAACA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
MGSVRDTKF	CGTGATACCAAGTTC**CCACGC**C GAAGACGCAGGAGACGT	AACCGACCCCAT**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
LqVSLNSVM	TTGAATTCGGTTATG**CCACGC**CG AAGACGCAGGAGACGT	CGAAACCTGCAG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
HNPNqITMY	CAGATCACCATGTAC**CCACGC**C GAAGACGCAGGAGACGT	ATTAGGATTATG**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
YGGSDTTMR	GACACCACAATGCGC**CCACGC**C GAAGACGCAGGAGACGT	CGAGCCACCGTA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
SLSLHVTqS	CATGTTACGCAGTCG**CCACGC**C GAAGACGCAGGAGACGT	CAGAGACAGCGA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
SLLGLHSSL	CTGCATTCGAGCCTG**CCACGC**C GAAGACGCAGGAGACGT	GCCTAACAGCGA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
QSLWRLAqR	CGTTTGGCGCAGCGT**CCACGC**C GAAGACGCAGGAGACGT	CCACAGCGATTG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
PEVLqMKMS	CAGATGAAAATGTCG**CCACGC**C GAAGACGCAGGAGACGT	CAAAACTTCTGG**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
NSLMRqLGL	CGCCAGCTGGGTCTG**CCACGC**C GAAGACGCAGGAGACGT	CATAAGGCTATT**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
LQPRLVFQG	CTGGTATTTCAGGGT**CCACGC**C GAAGACGCAGGAGACGT	ACGTGGCTGCAG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
IMDNPLEqG	CCTCTGGAGCAGGGT**CCACGC**C GAAGACGCAGGAGACGT	ATTGTCCATAAT**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
FSYSYRGRA	TACCGTGGTCGTGCG**CCACGC**C GAAGACGCAGGAGACGT	CGAGTAAGAGAA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
ELHRESHYV	GAATCGCATTATGTA**CCACGC**C GAAGACGCAGGAGACGT	ACGATGCAGTTC**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
CRHCIIGSC	ATCATTGGCTCGTGT**CCACGC**CG AAGACGCAGGAGACGT	ACAATGACGACA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
VLHSPVCIG	CCTGTATGTATCGGT**CCACGC**CG AAGACGCAGGAGACGT	CGAATGCAGAAC**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
SGSLWPQRN	TGGCCACAGCGTAAT**CCACGC**C GAAGACGCAGGAGACGT	CAGCGAGCCCGA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
SEGSLPERR	CTGCCAGAACGTCGC**CCACGC**C GAAGACGCAGGAGACGT	GCTCCCTTCCGA**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
RWKVTLLNT	ACGCTGTTAAATACG**CCACGC**C GAAGACGCAGGAGACGT	TACTTTCCAACG**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC

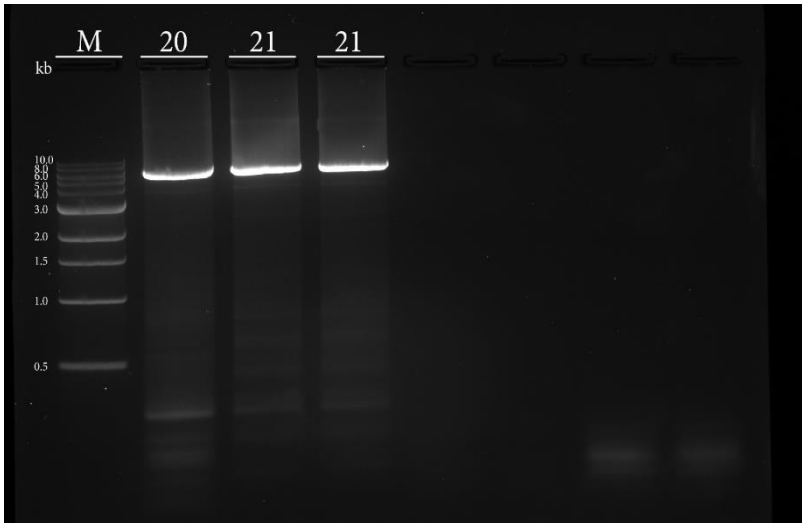
RKGLQYMEP	CAGTATATGGAACCA**CCACGC**C GAAGACGCAGGAGACGT	CAGCCCTTTACG**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
RGVTEKDER	GAGAAAGATGAGCGA**CCACGC**C GAAGACGCAGGAGACGT	GGTAACGCCACG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
QERLKPVLS	AAACCTGTACTGTCG**CCACGC**C GAAGACGCAGGAGACGT	TAAACGTTCTTG**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
PPIELVMqG	TTAGTAATGCAGGGC**CCACGC**C GAAGACGCAGGAGACGT	TTCGATAGGTGG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
NLLWSSASK	TCGTCTGCGTCGAAA**CCACGC**C GAAGACGCAGGAGACGT	CCACAGCAAATT**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
HTMLDQHTI	GATCAACATACAATC**CCACGC**C GAAGACGCAGGAGACGT	CAGCATTGTATG**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
WVLSSLVRT	TCGCTGGTACGTACG**CCACGC**C GAAGACGCAGGAGACGT	CGACAGTACCCA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
WLLCDPDAT	GACCCTGACGCGACA**CCACGC**C GAAGACGCAGGAGACGT	ACATAACAGCCA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
SVDFANKGP	GCGAATAAAGGCCCT**CCACGC**C GAAGACGCAGGAGACGT	GAAATCTACGGA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
STLLTWKEW	ACGTGGAAAGAATGG**CCACGC**C GAAGACGCAGGAGACGT	CAGTAACGTGGA**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
SKKqRRALL	CGTCGCGCGTTGCTT**CCACGC**C GAAGACGCAGGAGACGT	TTGCTTTTTGCT**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
PDSTYSGKN	TACAGCGGCAAAAAC**CCACGC**C GAAGACGCAGGAGACGT	TGTCGAATCAGG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
MQIWLMLVG	CTGATGCTGGTAGGT**CCACGC**C GAAGACGCAGGAGACGT	CCAAATTTGCAT**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
MPEqGqLLL	GGCCAATTGTTACTG**CCACGC**C GAAGACGCAGGAGACGT	CTGTTCAGGCAT**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC
HSLIPCLAS	CCATGTTTAGCGTCG**CCACGC**C GAAGACGCAGGAGACGT	AATCAGCGAATG**GGGGCGGCTCAT**GGGC CCCTGGAACAGAAC
GSVAHqWFT	CATCAGTGGTTCACT**CCACGC**CG AAGACGCAGGAGACGT	AGCTACCGACCC**GGGGCGGCTCAT**GGGCC CCTGGAACAGAAC

Appendix 4.3 Inverse PCR gels for peptides 17-44

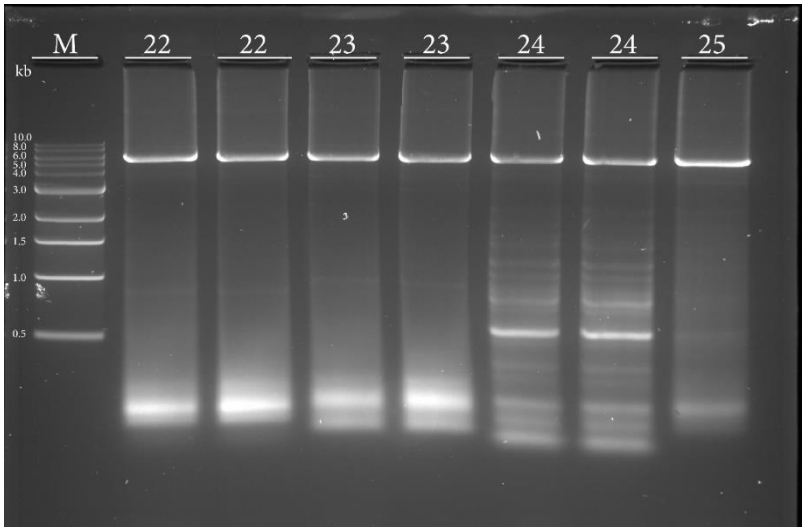
A



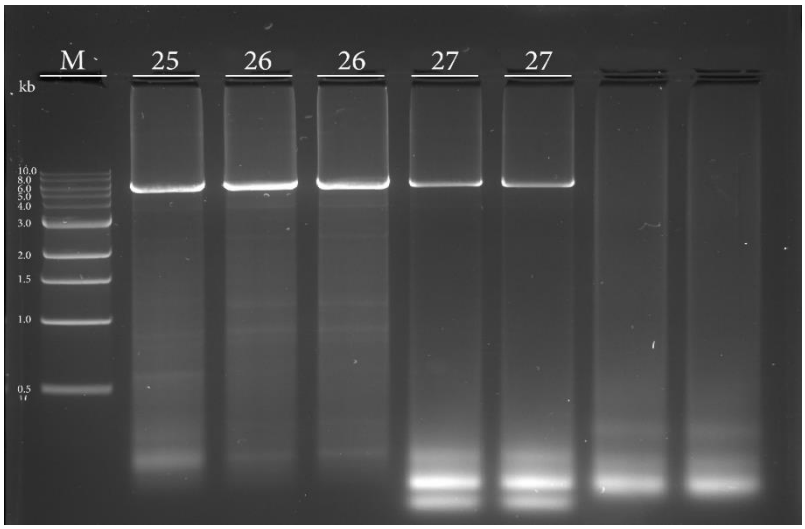
B



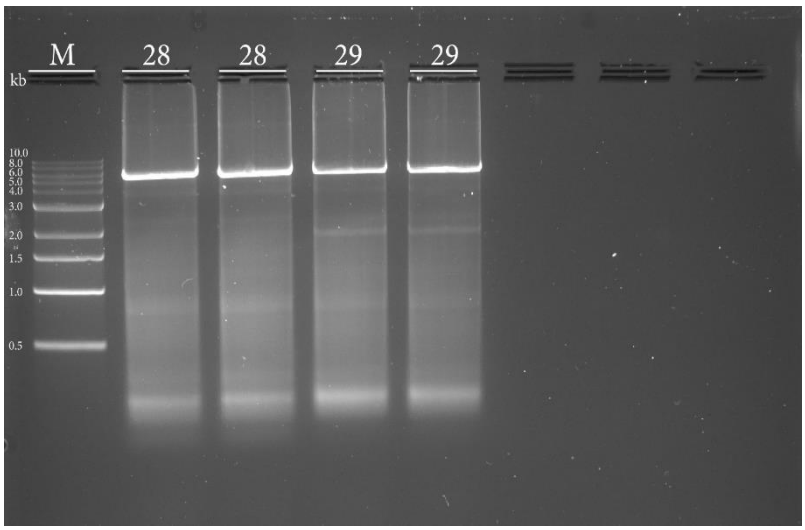
C



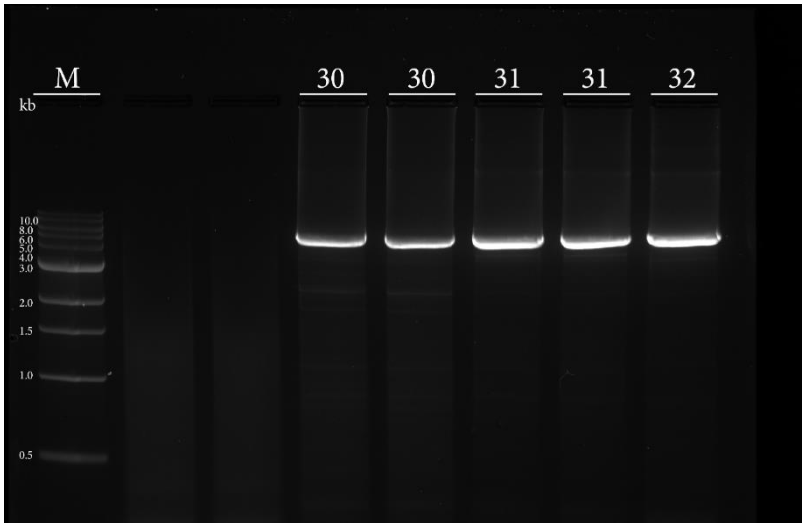
D


M
25
26
26
27
27
kb
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5

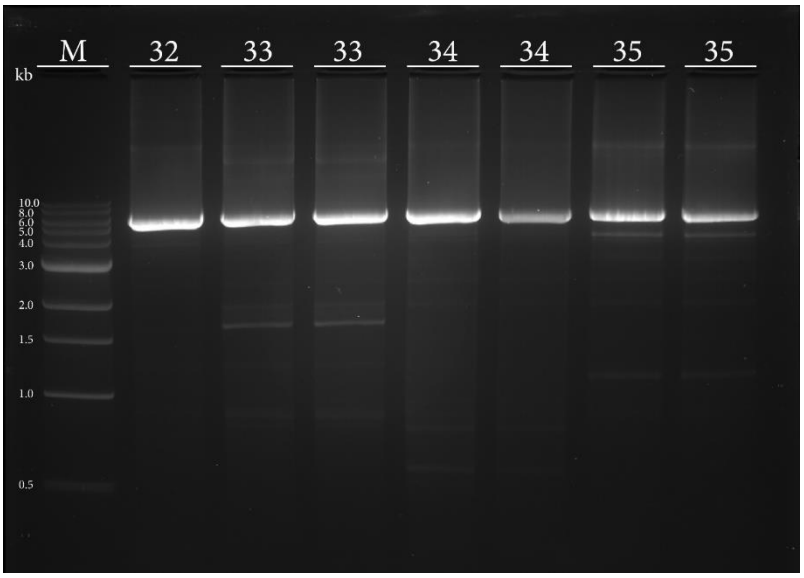
E


M
28
28
29
29
kb
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5

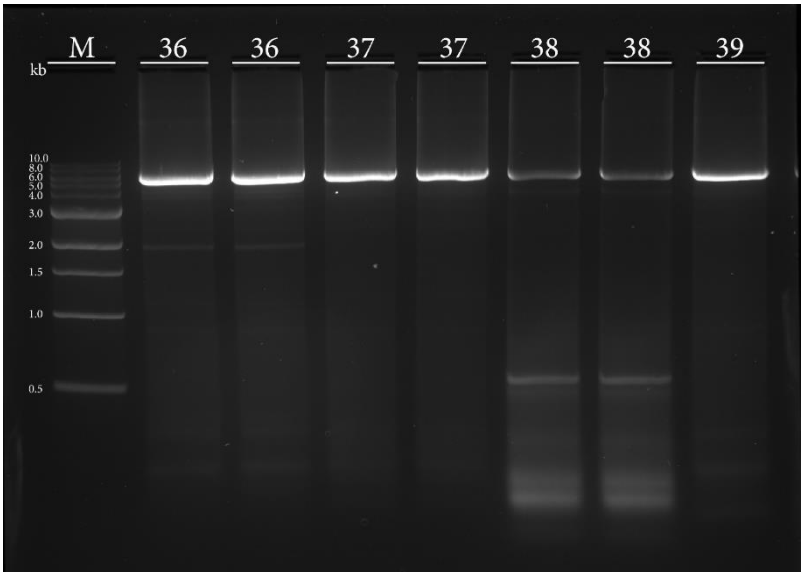
F


M
30
30
31
31
32
kb
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5

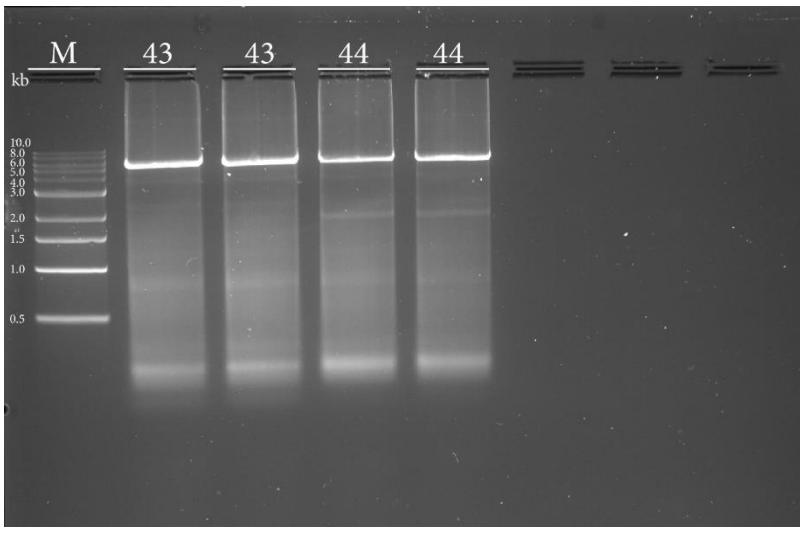
G


M
32
33
33
34
34
35
35
kb
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5

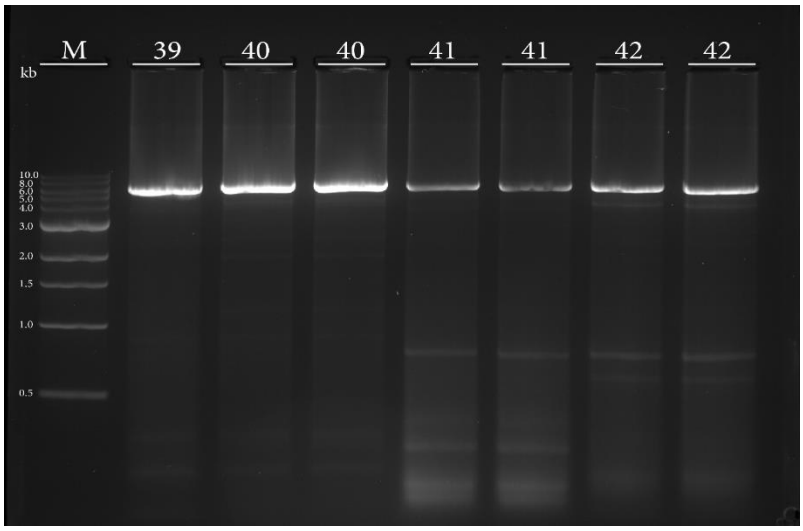
H


M
36
36
37
37
38
38
39
kb
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5

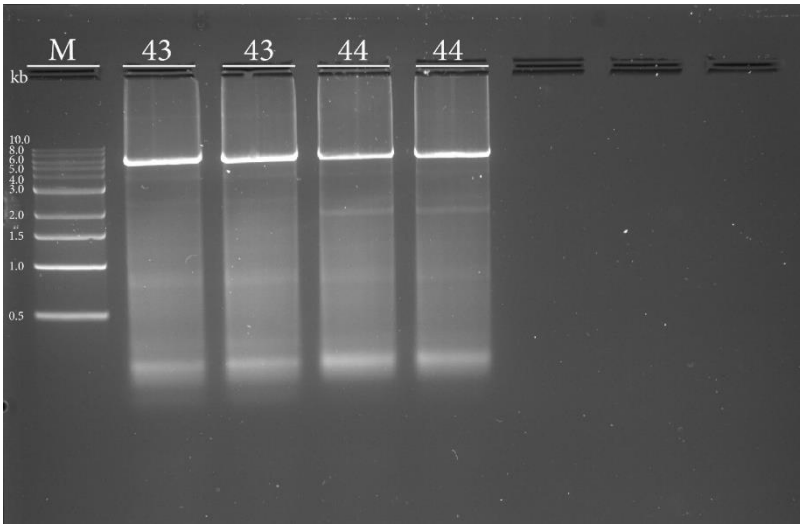
I


M
43
43
44
44
kb
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5

J



K

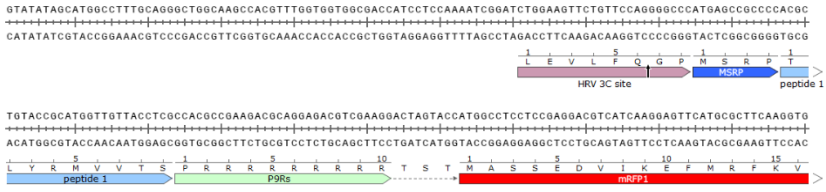


Supplementary Figure 7 Insertion of the 9-mer peptide encoding sequences from the Z>3 analysis into the mRFP expression vector.

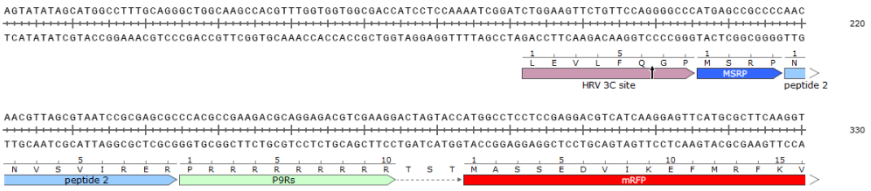
The DNA encoding sequences for the 31 peptides identified from the Z>3 criteria were cloned into the pGEX-6P1-mRFP expression vector by inverse PCR. Subsequent PCR products were observed at around 5.8kb on a 1% agarose gel (A-K). A 1kb ladder was used as the molecular marker.

Appendix 4.4 Sanger sequencing of peptide-mRFP plasmids

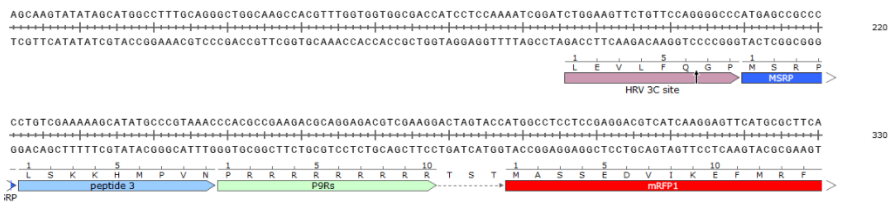
Peptide 1



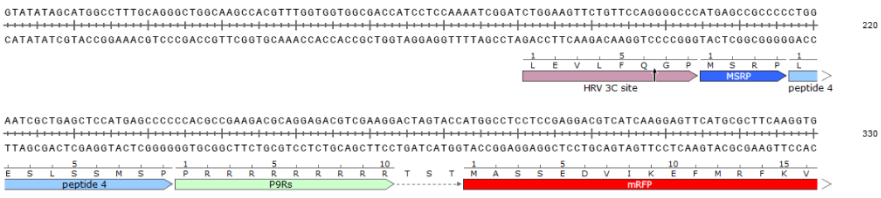
Peptide 2



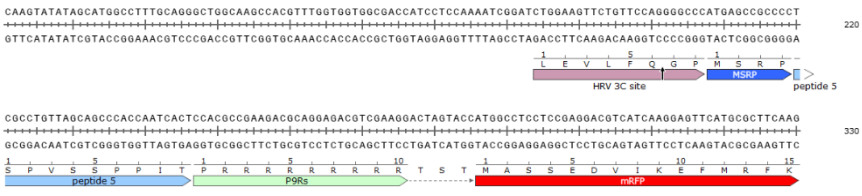
Peptide 3



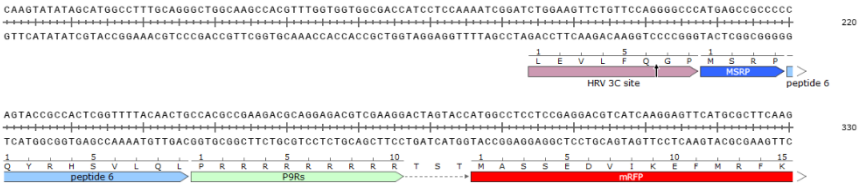
Peptide 4



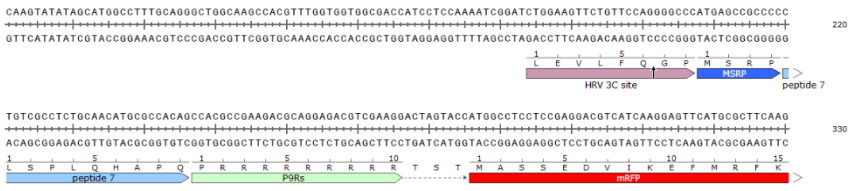
Peptide 5



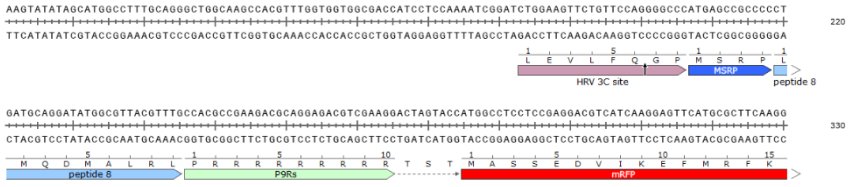
Peptide 6



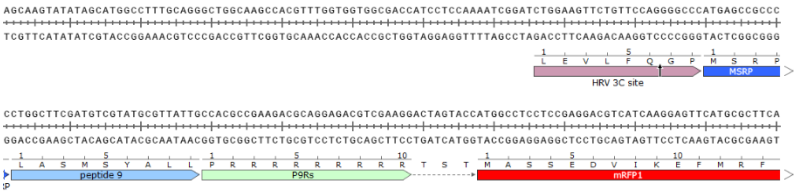
Peptide 7



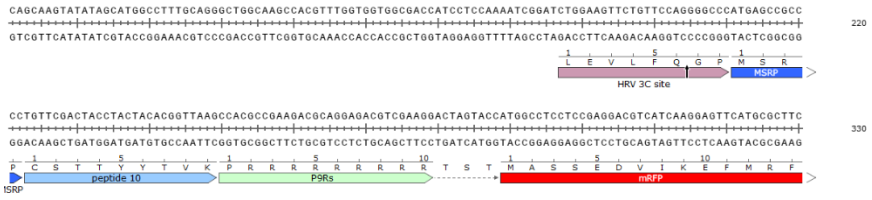
Peptide 8



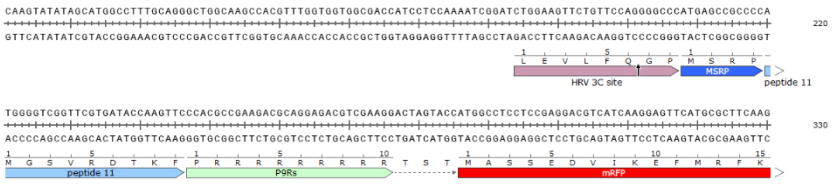
Peptide 9



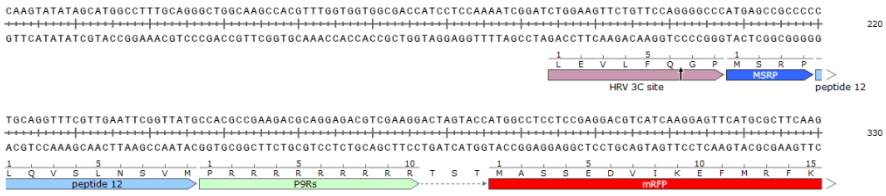
Peptide 10



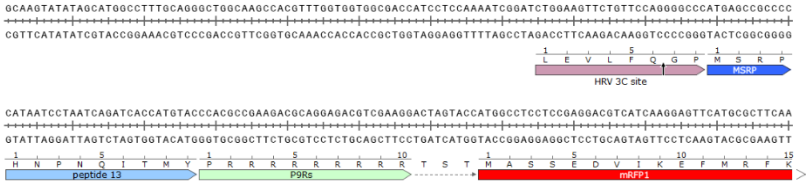
Peptide 11



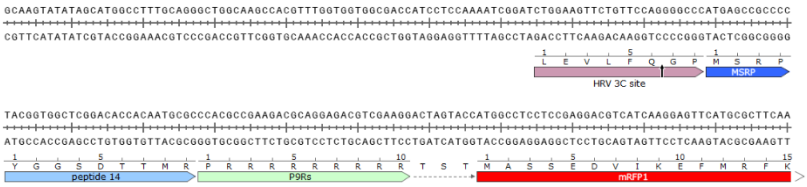
Peptide 12



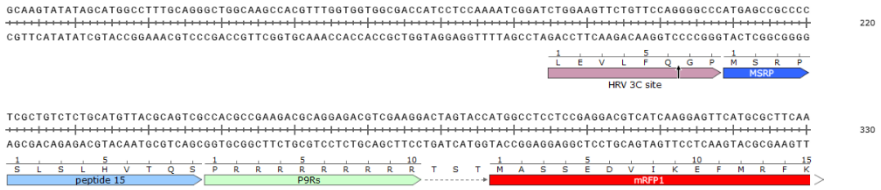
Peptide 13



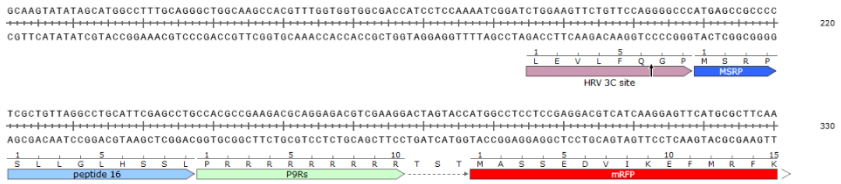
Peptide 14



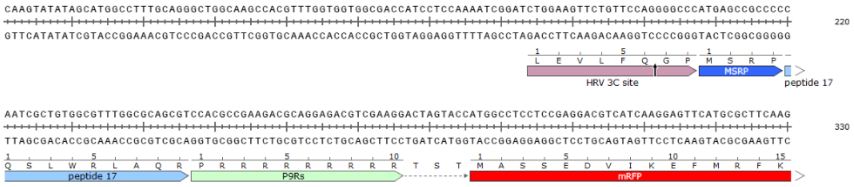
Peptide 15



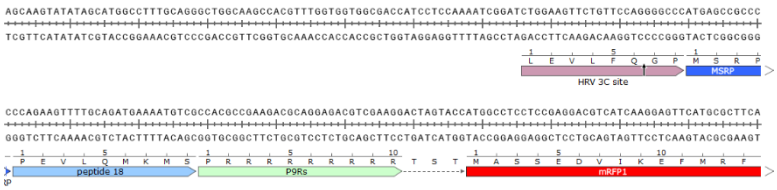
Peptide 16



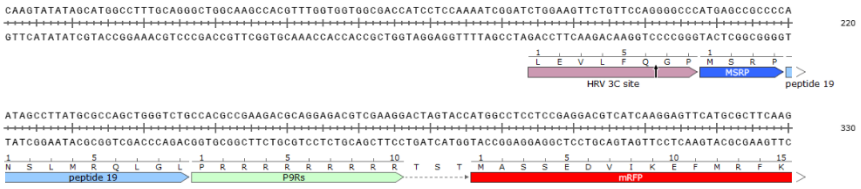
Peptide 17



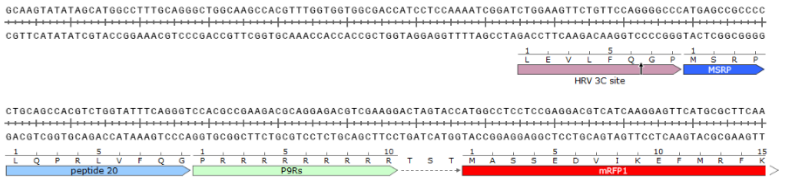
Peptide 18



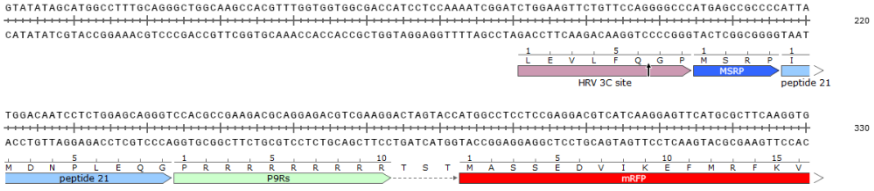
Peptide 19



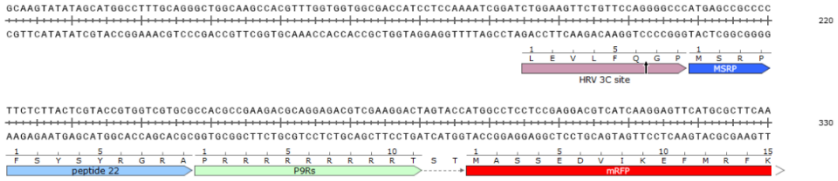
Peptide 20



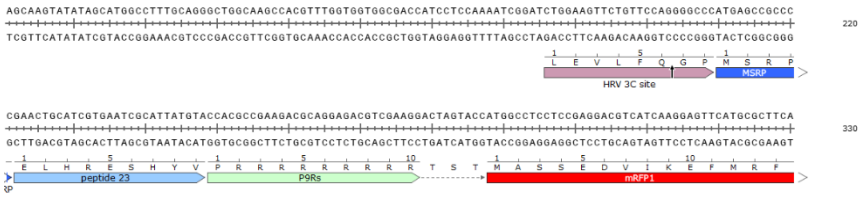
Peptide 21



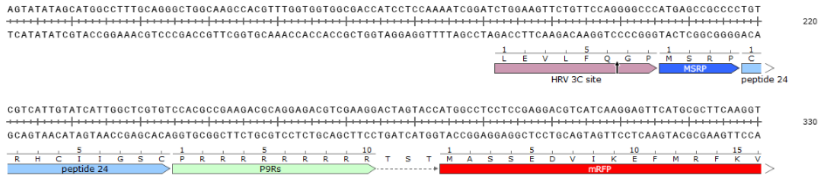
Peptide 22



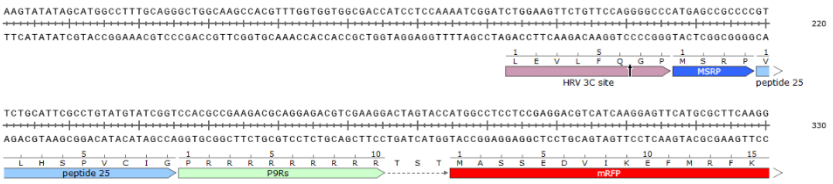
Peptide 23



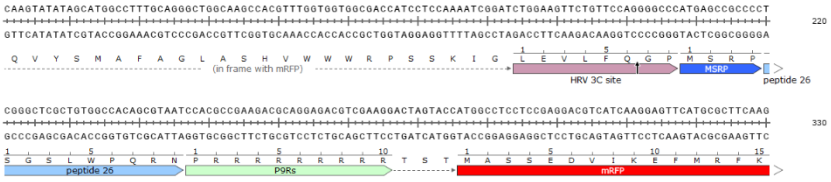
Peptide 24



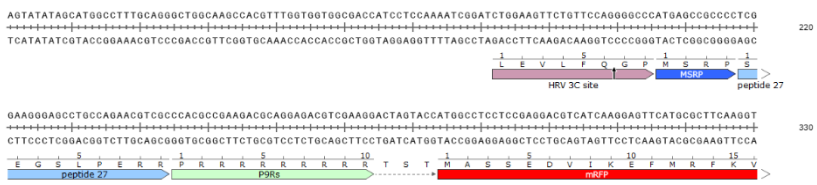
Peptide 25



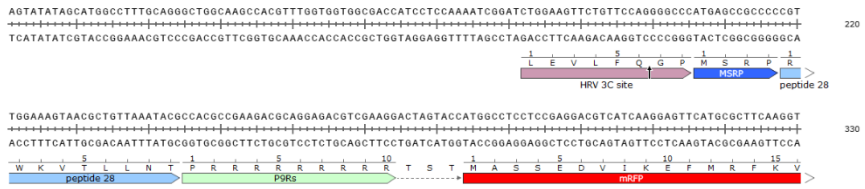
Peptide 26



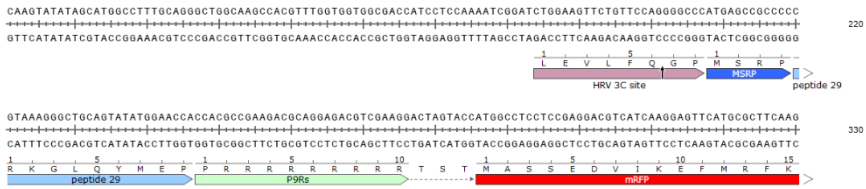
Peptide 27



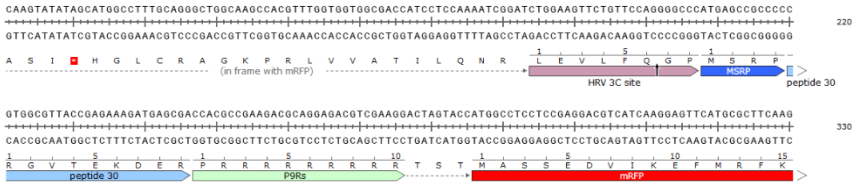
Peptide 28



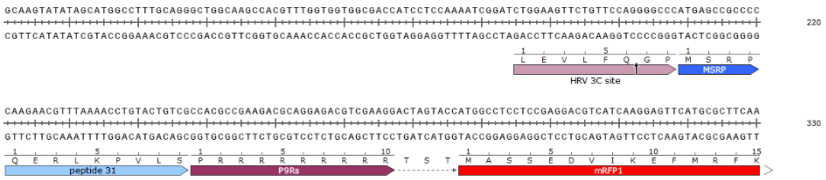
Peptide 29



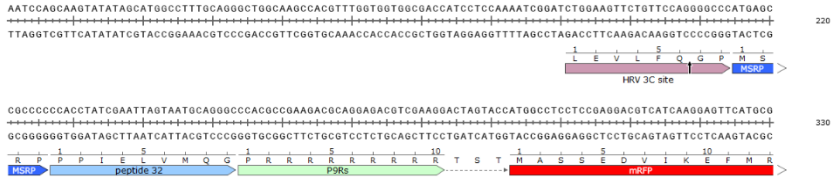
Peptide 30



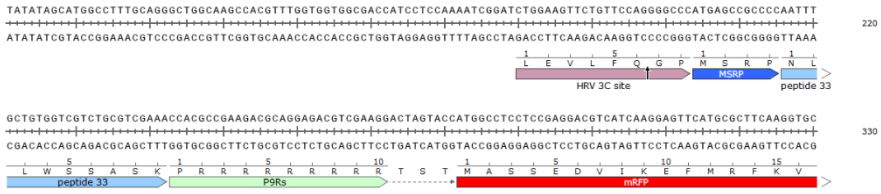
Peptide 31



Peptide 32



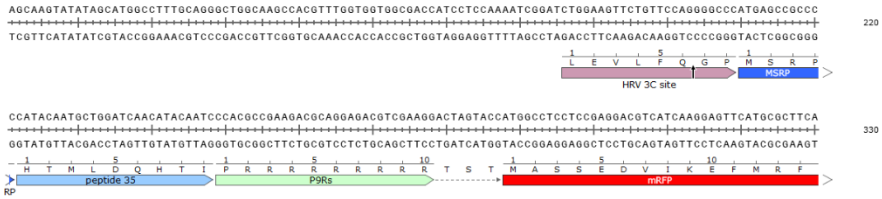
Peptide 33



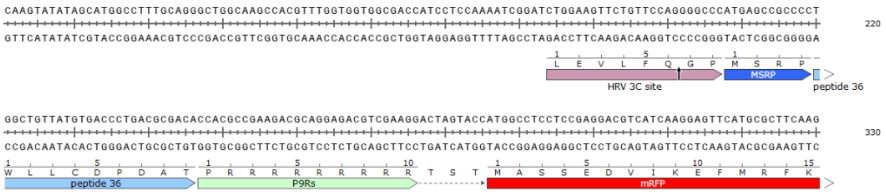
Peptide 34



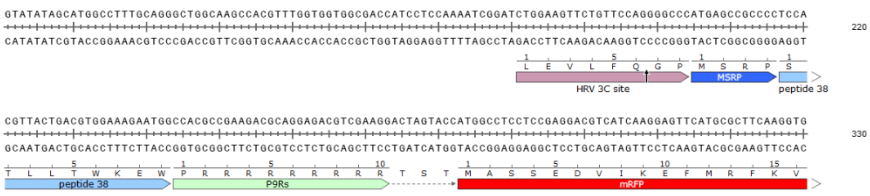
Peptide 35



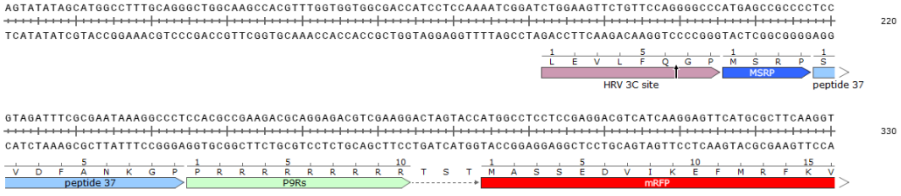
Peptide 36



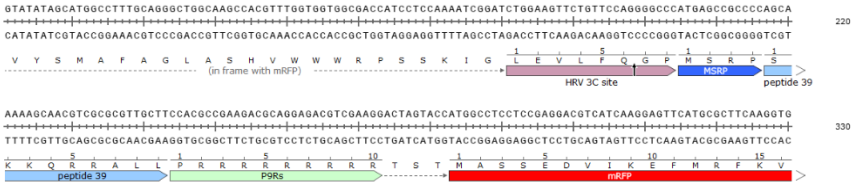
Peptide 37



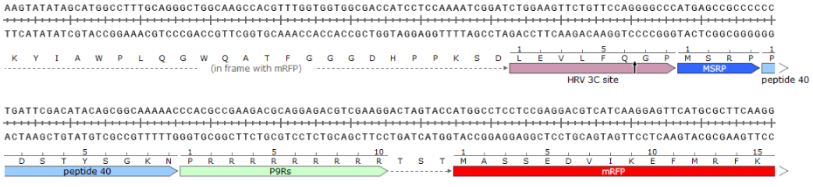
Peptide 38



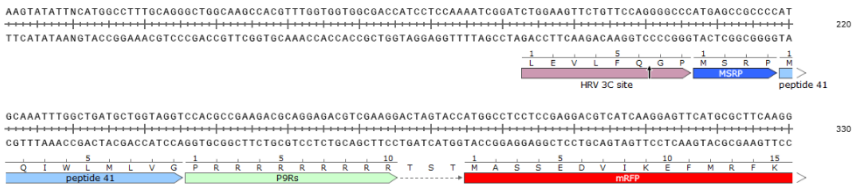
Peptide 39



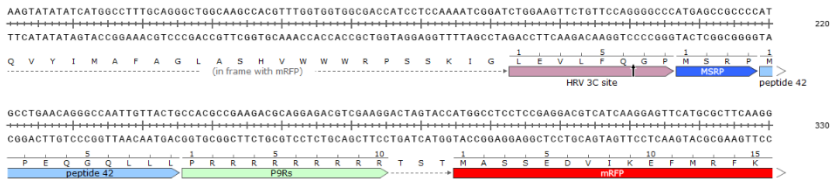
Peptide 40



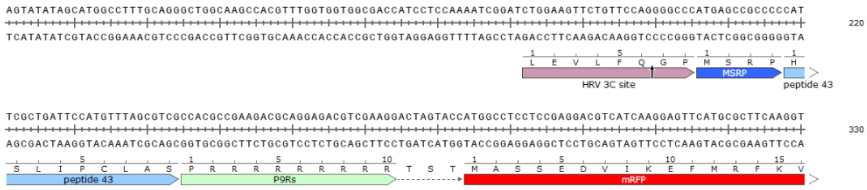
Peptide 41



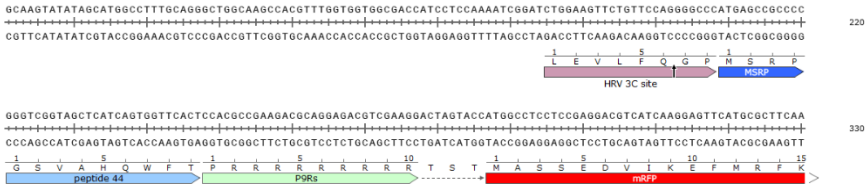
Peptide 42



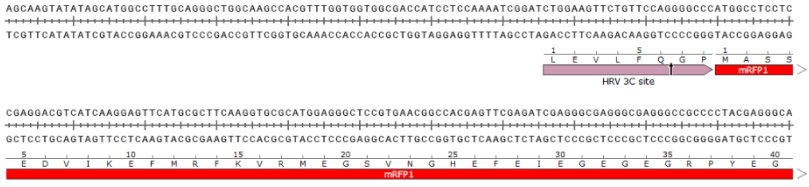
Peptide 43



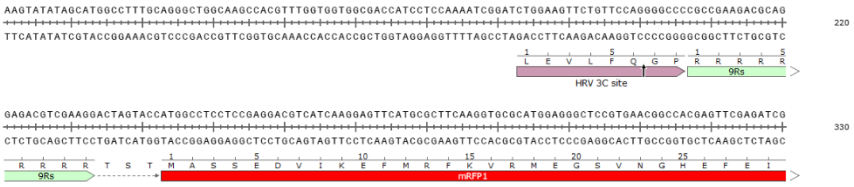
Peptide 44



mRFP only



9R-mRFP



p21-9R-mRFP



Appendix 4.5 CPP prediction for TNBC 9-mer peptides

Supplementary Table 6 BChemRF-CPPred results for 9-mer peptides

	FC-1		FC-2	
FILE	**Classification**	**Probability**	**Classification**	**Probability**
9R	CPP	95.26%	CPP	98.29%
P21-9R	CPP	96.06%	CPP	85.60%
TLYRMVVTS	Non-CPP	53.02%	Non-CPP	60.49%
NNVSVIRER	Non-CPP	50.33%	CPP	64.47%
LSKKHMPVN	Non-CPP	71.85%	CPP	60.11%
LESLSSMSP	Non-CPP	84.92%	Non-CPP	62.76%
SPVSSPPIT	Non-CPP	50.41%	CPP	53.73%
QYRHSVLQL	CPP	52.25%	Non-CPP	59.56%
LSPLQHAPQ	Non-CPP	83.29%	Non-CPP	57.16%
LMQDMALRL	Non-CPP	71.45%	CPP	51.00%
LASMSYALL	Non-CPP	70.14%	Non-CPP	64.07%
CSTTYYTVK	Non-CPP	85.12%	Non-CPP	70.12%
MGSVRDTKF	Non-CPP	82.20%	Non-CPP	64.34%
LQVSLNSVM	Non-CPP	82.69%	Non-CPP	56.49%
HNPNQITMY	Non-CPP	70.75%	Non-CPP	75.54%
YGGSDTTMR	Non-CPP	82.40%	Non-CPP	74.98%
SLSLHVTQS	Non-CPP	73.50%	Non-CPP	67.71%
SLLGLHSSL	Non-CPP	87.52%	Non-CPP	59.82%
QSLWRLAQR	CPP	86.08%	CPP	63.01%
PEVLQMKMS	Non-CPP	88.84%	Non-CPP	54.06%
NSLMRQLGL	Non-CPP	79.49%	CPP	54.30%
LQPRLVFQG	CPP	69.34%	Non-CPP	51.00%
IMDNPLEQG	Non-CPP	92.78%	Non-CPP	66.94%
FSYSYRGRA	CPP	73.08%	CPP	58.42%
ELHRESHYV	Non-CPP	87.73%	Non-CPP	73.06%
CRHCIIGSC	Non-CPP	66.56%	Non-CPP	56.77%
VLHSPVCIG	Non-CPP	89.56%	Non-CPP	50.49%
SGSLWPQRN	Non-CPP	75.73%	Non-CPP	60.36%
SEGSLPERR	CPP	51.05%	CPP	57.23%
RWKVTLLNT	CPP	78.33%	CPP	60.61%
RKGLQYMEP	Non-CPP	87.42%	Non-CPP	56.33%
RGVTEKDER	Non-CPP	54.08%	CPP	65.45%
QERLKPVLS	Non-CPP	67.39%	CPP	66.01%
PPIELVMQG	Non-CPP	91.96%	Non-CPP	50.22%
NLLWSSASK	CPP	58.98%	Non-CPP	60.41%
HTMLDQHTI	Non-CPP	73.61%	Non-CPP	77.30%
WVLSSLVRT	CPP	69.78%	Non-CPP	51.25%
WLLCDPDAT	Non-CPP	63.03%	Non-CPP	70.24%
SVDFANKGP	Non-CPP	93.71%	Non-CPP	68.53%
STLLTWKEW	CPP	70.86%	Non-CPP	58.07%

SKKQRRALL	CPP	89.17%	CPP	85.71%
PDSTYSGKN	Non-CPP	88.35%	Non-CPP	74.84%
MQIWLMLVG	Non-CPP	80.68%	Non-CPP	63.88%
MPEQGQLLL	Non-CPP	88.94%	Non-CPP	58.88%
HSLIPCLAS	Non-CPP	69.26%	Non-CPP	54.55%
GSVAHQWFT	Non-CPP	81.77%	Non-CPP	55.17%

	FC-3		**FC-4**	
FILE	Classification	Probability	Classification	Probability
9R	CPP	96.22%	CPP	93.50%
P21-9R	CPP	92.51%	CPP	92.17%
TLYRMVVTS	Non-CPP	52.72%	CPP	71.21%
NNVSVIRER	CPP	64.39%	CPP	62.64%
LSKKHMPVN	Non-CPP	65.08%	CPP	52.94%
LESLSSMSP	Non-CPP	72.21%	Non-CPP	66.50%
SPVSSPPIT	CPP	65.36%	CPP	60.88%
QYRHSVLQL	CPP	65.32%	CPP	67.83%
LSPLQHAPQ	Non-CPP	65.39%	CPP	61.59%
LMQDMALRL	Non-CPP	65.09%	CPP	68.85%
LASMSYALL	CPP	52.37%	CPP	61.97%
CSTTYYTVK	Non-CPP	72.62%	CPP	75.86%
MGSVRDTKF	Non-CPP	65.74%	Non-CPP	71.75%
LQVSLNSVM	Non-CPP	79.97%	Non-CPP	79.27%
HNPNQITMY	Non-CPP	65.05%	CPP	58.65%
YGGSDTTMR	Non-CPP	57.83%	CPP	67.31%
SLSLHVTQS	Non-CPP	63.64%	CPP	61.82%
SLLGLHSSL	Non-CPP	77.29%	Non-CPP	84.33%
QSLWRLAQR	CPP	77.29%	CPP	87.26%
PEVLQMKMS	Non-CPP	68.57%	Non-CPP	64.62%
NSLMRQLGL	Non-CPP	76.05%	Non-CPP	56.93%
LQPRLVFQG	Non-CPP	72.33%	Non-CPP	73.22%
IMDNPLEQG	Non-CPP	82.89%	Non-CPP	85.39%
FSYSYRGRA	Non-CPP	54.99%	CPP	72.39%
ELHRESHYV	Non-CPP	69.44%	CPP	56.82%
CRHCIIGSC	Non-CPP	66.41%	Non-CPP	72.59%
VLHSPVCIG	Non-CPP	81.18%	Non-CPP	88.10%
SGSLWPQRN	Non-CPP	69.32%	CPP	65.89%
SEGSLPERR	Non-CPP	50.42%	CPP	64.71%
RWKVTLLNT	CPP	83.92%	CPP	84.05%
RKGLQYMEP	Non-CPP	81.94%	Non-CPP	84.68%
RGVTEKDER	CPP	72.35%	CPP	73.00%
QERLKPVLS	Non-CPP	67.53%	CPP	58.96%
PPIELVMQG	Non-CPP	76.23%	Non-CPP	65.78%
NLLWSSASK	CPP	73.22%	CPP	70.45%
HTMLDQHTI	Non-CPP	60.95%	CPP	71.67%
WVLSSLVRT	Non-CPP	59.45%	CPP	73.16%

WLLCDPDAT	Non-CPP	60.73%	Non-CPP	58.91%
SVDFANKGP	Non-CPP	88.48%	Non-CPP	88.48%
STLLTWKEW	CPP	83.61%	CPP	77.65%
SKKQRRALL	CPP	86.02%	CPP	93.84%
PDSTYSGKN	Non-CPP	68.97%	Non-CPP	51.80%
MQIWLMLVG	Non-CPP	71.35%	Non-CPP	66.86%
MPEQGQLLL	Non-CPP	76.96%	Non-CPP	54.23%
HSLIPCLAS	CPP	69.12%	CPP	51.32%
GSVAHQWFT	CPP	56.54%	CPP	54.52%

FC-1 = only sequence-based descriptors (AAC, PseAAC, and DPC)

FC-2 = only structure-based descriptors (structural and physicochemical properties)

FC-3 = the best correlated sequence-based descriptors and structure-based descriptors

FC-4 = optimized selection of structure- and sequence-based descriptors according to Kendall's correlation analysis

Predictions generated using BChemRF-CPPred online tool (de Oliveira et al., 2021).

Appendix 5. Identification of TNBC associated antigens

Appendix 5.1 DESeq2 Script

```
library(DESeq2)
library(apeglm)

#load count matrix
data <- read.csv("/Users/meganwilde/Documents/year
4/mass_spec_analysis/DESeq/1_6.csv", header = T, row.names = 1)
info <- read.table("/Users/meganwilde/Documents/year
4/mass_spec_analysis/DESeq/colData.txt", header = T, sep = '\t')

dds <- DESeqDataSetFromMatrix(data, info, ~condition)

#remove 0 count data across all rows
keep <- rowSums(counts(dds)) > 0
dds <- dds[keep,]

#run DESeq
ddsDE <- DESeq(dds)

#normalised read counts
normCounts <- counts(ddsDE, normalized = TRUE)
write.csv(normCounts, "normal_control_vs_TNBC.csv")

#results of DESeq
res <- results(ddsDE, alpha = 0.05)
summary(res)
resOrdered <- res[order(res$padj),]
write.csv(resOrdered, "deseq_control_vs_TNBC_80_110.csv")
```

Appendix 6. BBSRC DTP Professional Internship for PhD Students (PIPS) reflective statement

The professional internship for PhD students (PIPS) is a mandatory placement that must be undertaken as part of the BBSRC Doctoral Training Partnership (DTP) programme. This reflective statement aims to summarise my PIPS experience and to highlight the skills that I have gained which will be valuable for my future career.

I planned to complete my placement before starting my final year and I was hoping to find a laboratory-based opportunity in industry. However, due to the COVID-19 pandemic, I was unable to obtain an in-person placement, as many laboratories had not returned to full capacity at the time. Therefore, I began enquiring about remote placement opportunities that would enable me to develop skills in bioinformatics or coding, which led to me securing a placement in epidemiological modelling. From July to September 2021, I completed by PIPS placement remotely at Public Health England (PHE) with the Epidemiology and Mathematical Modelling team.

My role at PHE was to carry out an in-depth COVID-19 research project to contribute to the epidemiological investigations being conducted by the emergency response department. During my placement, I developed coding skills in Python and expanded my knowledge of statistical models and their applications. I also gained experience in handling confidential datasets and in presenting project outputs following a project brief. As part of the COVID-19 response team, I was invited to attend weekly meetings where regional updates were discussed and I presented my project outputs at the end of the placement to my team through written report and presentation.

Overall, I really enjoyed my time at PHE and the opportunity to work within an executive agency. The work conducted by the emergency response team highlighted the importance of epidemiological modelling and the crucial role that scientific evidence plays in government decision making. The placement facilitated the development of core skills that will be useful in my future career, such as in communication and teamwork, and has influenced my decision to consider a career outside of academia.