# University of Southampton

FACULTY OF MEDICINE

Division of Cancer Sciences


# The Immune Microenvironment in Young Patients with Triple Negative Breast Cancer

by


**Hayley Suzanne McKenzie**

**MA (Cantab) MBBS (Lond) MRCP (UK)**


Thesis for the degree of Doctor of Philosophy

January 2021

# University of Southampton

## Abstract

Faculty of Medicine
Division of Cancer Sciences
Thesis for the degree of Doctor of Philosophy
The Immune Microenvironment in Young Patients with Triple Negative Breast Cancer
By Hayley Suzanne McKenzie

Introduction: Triple negative breast cancers (TNBC) are high-grade tumours, with poorer disease-specific survival. Following diagnosis of metastatic disease, survival is typically short and novel treatment strategies are required. The immune system is increasingly recognised to play an important role in TNBC biology; the tumour microenvironment (TME) is central to immune evasion. I explored the interaction between the TME, immunogenicity and prognosis in a cohort of young women with TNBC.

Methods: The POSH study is a prospective observational study of almost 3000 patients aged 40 years or younger with a first diagnosis of invasive breast cancer. Associations between clinicopathological features and molecular markers (using tissue microarrays) and prognosis were assessed in those with TNBC. These factors were integrated into a prognostic score using ROC analysis. Correlates of immune activity were identified and digital pathology analyses utilised. Murine breast fibroblasts were differentiated into the myofibroblastic phenotype *in vitro*, with qtPCR, western blot, cell count, immunofluorescence and metabolism assays performed. The effect of myofibroblast co-injection on tumour growth using the 4T1 model *in vivo* was studied. Finally, AZD0156, a novel cancer-associated fibroblast (CAF) inhibitor, was assessed for efficacy and compared to immunotherapy *in vivo*.

Results: Tumour-infiltrating lymphocyte (TIL) density was highly prognostic and was significantly higher in those with a germline *BRCA* mutation. Bioinformatics of publically available datasets revealed that extracellular matrix pathways were upregulated in lymphocyte-deplete TNBC cases. An inverse correlation between smooth muscle actin (SMA) and TIL was found. Murine breast fibroblasts were transformed *in vitro* using TGF-β; myofibroblasts augmented tumour growth in a TNBC murine model. An ATM inhibitor (AZD0156) inhibited the myofibroblastic phenotype and was effective at reducing tumour growth using the CAF-rich 4T1 model *in vivo*. This effect was not mediated via CD8 infiltration but there was a reduction in CAF density.

Conclusion: The TME is an important factor in immune evasion in TNBC. CAF inhibition may be an effective treatment strategy and should be investigated in combination with other agents.

# Table of Contents

# Table of Tables

# Table of Figures

# Research Thesis: Declaration of Authorship

| | |
|---|---|
| Print name: | Hayley McKenzie |

| | |
|---|---|
| Title of thesis: | The Immune Microenvironment in Young Patients with Triple Negative Breast Cancer |

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

McKenzie HS, Maishman T, Simmonds P, Durcan L, Group PS, Eccles D *et al.* Survival and disease characteristics of de novo versus recurrent metastatic breast cancer in a cohort of young patients. Br J Cancer, (2020) doi: 10.1038/s41416-020-0784-z.

| | | | |
|---|---|---|---|
| Signature: | | Date: | 11th January 2021 |

# Acknowledgements

I would like to thank Cancer Research UK for investing in me and generously supporting this work. I would like to thank my supervisors Ellen Copson, Gareth Thomas and Anthony Williams for their guidance over the years. They were always approachable and happy to tackle any of the numerous questions I raised and problems I encountered during this time. I am also indebted to Diana Eccles, who has been an endless inspiration and mentor to me. Both she and Ellen met with me regularly in the early days to brainstorm ideas and I wouldn't have managed this without them. I'd also like to thank the POSH Steering Group for generously allowing me to use data and materials from this important cohort study.

I owe a large amount of gratitude to Max Mellone, Christopher Hanley and Kirsty Ford for their support during the laboratory components of this project. Sara Waise kept my spirits up in the final 18 months as I approached the finishing line. I'd also like to thank others who have contributed to the project and have attributed elements of this work to them on the following page. I'd particularly like to thank Guy Martland for the hours we spent scoring TMAs.

I'd like to thank my son who was born during my PhD and has overtaken it to become my most challenging work in progress and my partner for his patience and support.

Attributions

Metastatic cohort statistical design (2.2.3, Appendix B) – HM with input from Ellen Copson and Tom Maishman

Metastatic statistical analysis (3.2.1.1-3.2.1.4 including Figures, Figure 3.6) – Tom Maishman

Tissue microarray construction and staining and multiplex staining (2.2.6) – Monette Maria Lopez

Evaluation of TILs and expression scoring of TMAs (2.2.7) – Guy Martland

Initial backwards elimination in a Cox regression model for OS and proportional hazards test (3.2.2.3) – Laura Day

Bioinformatics (2.3, 5.2.2) – Steve Thirdborough

ROC analysis (2.2.5.2, Table 5.4, Figure 5.11) – Scott Harris

Digital imaging analysis using Definiens (2.2.10, 5.2.4; all figures and calculations by HM) – Matt Ellis

RNA extraction and western blot (Figure 7.14; analysis/graphs performed by HM) – Max Mellone

Design and primary investigator of pilot AZD0156 *in vivo* experiment (7.2.3) – Max Mellone

# Definitions and Abbreviations

| | |
|---|---|
| 95% CI | 95% confidence interval |
| α-SMA | α-smooth muscle actin |
| ANOVA | analysis of variance |
| APC | antigen presenting cell |
| ATM | ataxia telangiectasia mutated protein |
| BMI | body mass index |
| BSA | bovine serum albumin |
| CAF | cancer-associated fibroblast |
| CR | complete response |
| CTLA4 | cytotoxic T-lymphocyte-associated protein 4 |
| DDR | DNA damage response |
| DFS | disease-free survival |
| DMEM | Dubecco's Modified Eagle's medium |
| dnMBC | *de novo* metastatic breast cancer |
| ECM | extra-cellular matrix |
| EGFR | epidermal growth factor receptor |
| ER | oestrogen receptor |
| FAP-α | fibroblast activation protein-α |
| FBS | fetal bovine serum |
| FFPE | formalin fixed and paraffin embedded |
| FISH | fluorescent in-situ hybridisation |
| FOXP3 | forkhead box P3 |
| GWAS | genome-wide association study |
| H&E | haematoxylin and eosin |
| HER2 | human epidermal growth factor receptor 2 |
| HLA | human leukocyte antigen |
| HR | hazard ratio |
| HRR | homologous recombination repair |
| ICB | immune checkpoint blockade |
| IHC | immunohistochemistry |
| IM | immunomodulatory |
| LPBC | lymphocyte-predominant breast cancer |
| Mb | millions of base pairs |
| MBC | metastatic breast cancer |
| MHC | major histocompatibility complex |
| mRNA | messenger RNA |

| | |
|---|---|
| MVA | multivariable analysis |
| OS | overall survival |
| p53 | tumour suppressor protein 53 |
| PARP | poly (ADP-ribose) polymerase |
| pCR | pathological complete response |
| PCR | polymerase chain reaction |
| PD-1 | programmed death receptor 1 |
| PD-L1 | programmed death ligand-1 |
| PDRS | post distant relapse survival |
| PFS | progression free survival |
| PR | progesterone receptor |
| RCT | randomised controlled trial |
| rMBC | recurrent MBC |
| RNA | ribonucleic acid |
| ROS | reactive oxygen species |
| rtPCR | reverse transcription polymerase chain reaction |
| SD | standard deviation |
| SEM | standard error of the means |
| SNP | single nucleotide polymorphism |
| TAMs | tumour-associated macrophages |
| TCGA | the Cancer Genome Atlas |
| TCR | T-cell receptor |
| TGFβ1 | transforming growth factor β type 1 |
| TILs | tumour-infiltrating lymphocytes |
| TME | tumour microenvironment |
| TMA | tissue microarray |
| TNBC | triple negative breast cancer |
| Treg | Regulatory T-cell |
| UVA | univariable analysis |
| VEGF | vascular endothelial growth factor |

# Chapter 1 Introduction

## 1.1 Background – breast cancer

Breast cancer is the most common neoplasm in women; there are over 55,000 new diagnoses per year in the UK(1). The vast majority (approximately 94%) of patients with breast cancer present with disease localised to the breast and axillary lymph nodes and are treated with the aim of cure(2). Overall survival rates are continuing to improve; the five-year survival rate in the UK has increased from 5% in 1972 to 87% in 2011 (Figure 1.1) (3).

Breast cancer is a heterogenous disease, both clinically and at a molecular level. It has traditionally been divided into subtypes based on immunohistochemistry (IHC) and the presence or absence of oestrogen (ER), progesterone (PR) and HER2 receptor staining. The proportion of cells which stain positively for ER and PR ranges from 0 to 100%, although ≥1% (approximately 75% of breast cancers) is considered ER or PR positive for clinical purposes(4). HER2 expression is scored in the range of 0-3+, with 0/1+ considered negative and 3+ positive. A score of 2+ is regarded as borderline and mandates further assessment with *in* situ hybridisation(5). Triple negative breast cancers (TNBC) are defined by lack of ER and PR staining by IHC and absence of HER2 overexpression by IHC or gene amplification by *in-situ* hybridisation. This classification has been primarily driven by the ability to treat patients with targeted agents depending on their receptor status. However, the true diversity of breast cancer has become apparent in recent years with the advances of genomic and transcriptonomic technology.

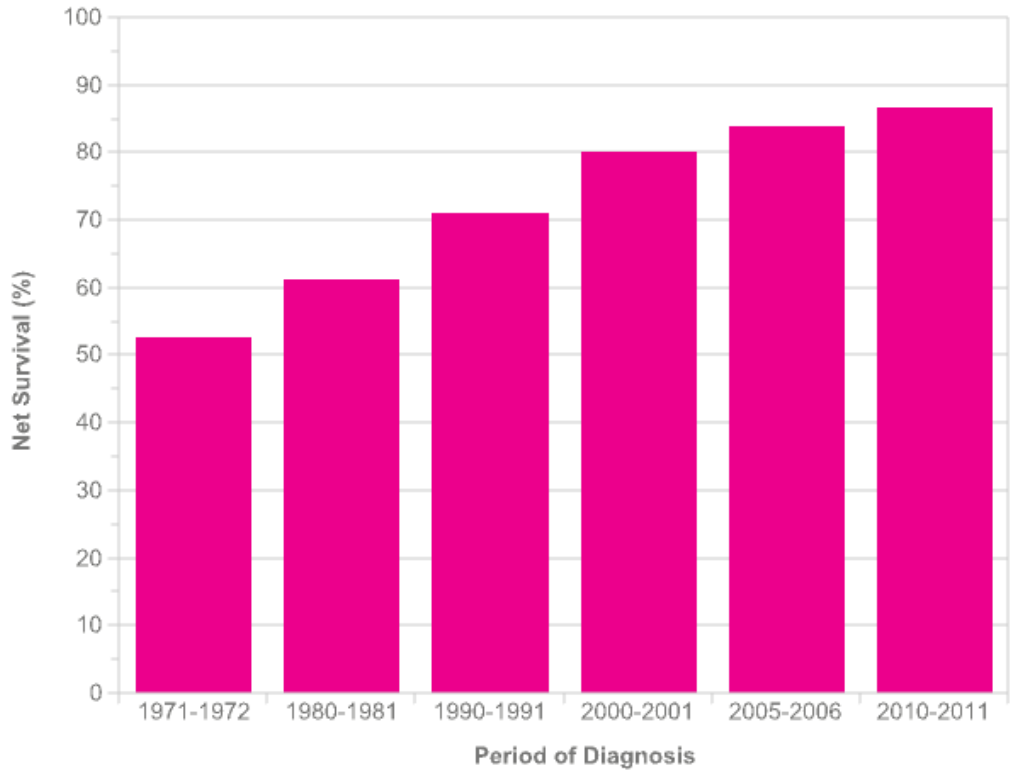



**Figure 1-1. Age-Standardised Five-Year Net Survival over time.**
**For women with a diagnosis of breast cancer (Aged 15-99), England and Wales 1971-2011.**
**Five-year survival has increased from 53% in 1971-72 to 87% in 2010-2011. Ten year survival has increased even further: from 40% to 78%. Survival for 2010-2011 is predicted using an excess hazard statistical model.**
**Image from Cancer Research UK(3).**

## 1.2 Metastatic breast cancer

Although most patients diagnosed with breast cancer are now cured, unfortunately there is still a significant risk of recurrence. The risk of recurrence for a patient undergoing treatment for localised breast cancer is difficult to estimate as breast cancer recurrence risk continues even 24 years after treatment, particularly for ER+ disease(6). With on-going evolution of increasingly effective therapies, recurrence rates for the current population of patients will never be precise. In one study involving 4105 patients enrolled in international trials from 1978-1985, the distant recurrence rate was 24.4%(6). This rate

is likely to be somewhat lower in the modern era and the recurrence risk for an individual patient clearly depends on multiple factors, including staging and breast cancer subtype.

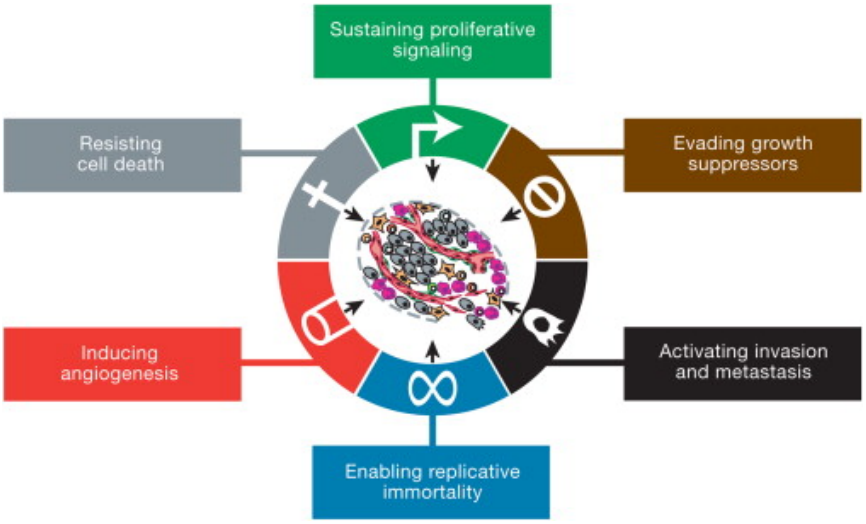
Unfortunately, the prognosis for those with metastatic breast cancer (MBC) remains poor and treatment is usually with palliative intent(7, 8). The median survival is 2 to 3 years. However, some patients with ER+ or HER2+ disease live much longer (more than 10 years in some HER2+ cases) (9, 10). Most patients with MBC develop secondary metastatic disease later, following a diagnosis of a primary cancer. However, 6-7% of patients with MBC have a diagnosis of metastatic disease at first presentation (known as "*de novo* metastatic disease"; dnMBC). It is not clear how the pathology, presentation and outcome for patients who present with dnMBC compare to those who later develop distant metastases following a diagnosed primary (recurrent MBC; rMBC).

Some studies have shown a longer survival time for those presenting with dnMBC compared to those with rMBC (11-14). In one multicentre retrospective study evaluating 815 consecutive patients with MBC in the Netherlands from 2007 to 2009, with a median follow-up of 37.1 months, this was only true for rMBC patients with a metastasis-free interval (MFI) of less than 24 months(15). In common with other studies on this subject, the follow-time is insufficient as late ER+ recurrences are omitted from the data. Additionally, most MBC survival analyses are retrospective, with a median age of 53-65 (with less than 15% of participants being aged under 40)(11, 15, 16). Baseline imaging investigations for metastatic disease are not mandatory in early breast cancer (unlike in some other solid tumours) and are generally done at the discretion of the treating physician. In general, imaging (such as a CT scan of the chest, abdomen and pelvis) is only performed for clinically positive axillary nodes, large tumours (≥5cm) or clinical signs or symptoms suggestive of metastatic disease or aggressive biology(17).

## 1.3 Background – Immunology In Cancer

It has long been known that the immune system is important in tumour biology. In 1893 William Coley published his data using live bacteria to stimulate the immune system in the treatment of cancer. Using a preparation derived from streptococcal cultures he injected over 900 cancer patients with an alleged cure rate of 10%(18). It is now established that the host immune system can recognize and eliminate malignant cells, with tumours only developing when these cells are able to evade immune recognition. This was acknowledged as one of the emerging hallmarks of cancer, in addition to the six original hallmarks in the seminal paper by Hanahan *et al* (Figure 1.2)(19).

**A.**




**B.**

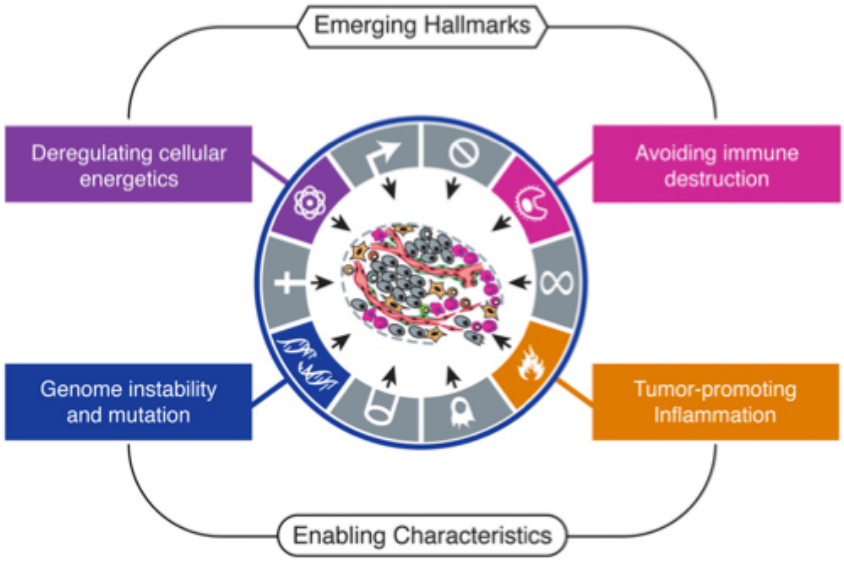


**Figure 1-2. Hallmarks of Cancer**

**A. In their seminal article published in 2000, Hanahan et al(20) attempted to simplify cancer biology into a number of key principles. They argued that all cancers can be defined by six hallmarks that govern neoplastic development. These six hallmarks were: self-sufficiency in growth signals, insensitivity to anti-growth signals, evasion of apoptosis, replicative immortality, sustained angiogenesis and tissue invasion/metastasis.**

**B. In 2011, the authors published an updated paper, outlining two emerging hallmarks and two emerging "characteristics" (defined as features which lead to the development of hallmarks). The new hallmarks were deregulated metabolism and evasion of the immune system. The emerging characteristics were genomic instability and inflammation(19).**

**Therefore the study of immunology is now recognised to be of vital importance in cancer biology and therapeutics.**

**Images from Hanahan *et al*(19).**

Cancer cells are potentially immunogenic; they can produce antigens recognized as foreign by the immune system and tumour-specific T-cell clones can be grown from animals bearing tumours(21). There are several ways in which a tumour can become antigenic, including the generation of mutated neoantigens, expression of oncofetal antigens or viral proteins and the overexpression of normal self antigens(21). The primary factor in determining tumour antigenicity is the degree to which it expresses epitopes selectively recognized by T-cells. An epitope within a mutated protein has to be sufficiently non-self to be immunogenic. The number of tumour-specific antigens differs between tumours and the immunogenicity of these antigens is also variable; therefore some tumours are highly immunogenic while others are relatively immune-silent(22). The overall immunogenicity of a tumour is also dependent on other immunomodulatory mechanisms, including the ability of cytotoxic T-cells to effectively infiltrate the tumour(23). It is clear that immunogenicity varies between tumours and this is reflected in responsiveness to immunotherapy; melanoma, renal cell carcinoma and non-small cell lung cancer are examples of the most immunogenic tumours, whereas uterine carcinoma and glioblastoma are considered poorly immunogenic(24).

The host response commences with processing and presentation of the antigen in association with the major histocompatibility complex (MHC) on the surface of an

antigen-presenting cell (APC). Naïve T-cells interact with the APC and are activated in a secondary lymphoid organ. Activation requires both T-cell receptor (TCR)/MHC interaction and co-stimulatory signals. Differentiation and clonal expansion of specific T-cells then occurs, resulting in the production of effector and memory T-cells(25). The innate and adaptive immune systems coordinate to eliminate the foreign cells.

Mononuclear cells localising within a tumour are thought to reflect the adaptive host response and are known as tumour-infiltrating lymphocytes (TILs). They can be considered a surrogate marker of immunogenicity(22). In solid tumours they are largely composed of CD3+ T-cells; CD20+ B-cells are seen infrequently. The CD3+ T-cells are further subdivided into CD8+ cytotoxic T-cells, CD4+ regulatory T-cells and CD4+ helper T-cells (26). CD8+ T-cells recognise antigens presented by MHC class I molecules and CD4+ helper T-cells recognise those presented by MHC class II molecules. The anti-tumour response is primarily cytotoxic (mediated by CD8+ cells) as tumour cells are usually negative for class II and positive for class I MHC molecules. However, the CD4+ T-cells also play an important role, with the Th1 cells assisting in cellular immunity and the Th2 cells in humoral immunity(25).

The process whereby the immune system and tumour interact, resulting in elimination, equilibrium and/or escape of malignant cells is known as immunoediting(27). Tumours may be weakly immunogenic or exhibit tumour evasion strategies which result in an inadequate immune response. Tumours use a multitude of mechanisms to evade immune destruction, including promotion of immunosuppressive cells in the microenvironment (e.g. tumour-derived regulatory T-cells), downregulation of antigen processing/presentation, production of immunosuppressive cytokines, expression of inhibitory molecules (e.g. PD-L1) and apoptosis of cytotoxic T-cells(28). In addition, intratumoural heterogeneity and clonal evolution significantly contribute to the challenge the immune system has in successfully eliminating tumour cells.

## 1.4 Tumour-infiltrating Lymphocytes

### 1.4.1 The assessment of TILs in breast cancer

Although not classically considered an immunogenic tumour, there is emerging evidence that the immune system is important in the pathogenesis of some types of breast cancer, in particular HER2-positive and TNBC. As outlined above, in every study using transcriptonomic and/or genomic data to subclassify breast cancer, an "immune-active" subtype has been identified, which confers a good prognosis.

In contrast to normal breast tissue, in which only very low numbers of CD4+ and CD8+ T-cells are detected, breast tumours and their associated stroma have high numbers of immune cells(29). The presence of TILs in a solid tumour is generally associated with a positive prognosis and they have long been known to have prognostic implications in breast cancer(26, 30). TILs are detected using light microscopy with haematoxylin and eosin (H&E) stained histological slides. TILs can be divided into those seen within the stroma (sTILs) and within the tumour itself (intratumoural). The majority of TILs in breast cancer are located in the stroma. Stromal and intratumoural TILs are highly correlated, but sTILs are seen in higher numbers and display larger variability between tumours(29). Therefore, sTILs are now considered to be a better biomarker in predicting patient response to treatment and subsequent outcome and when quantifying TILs, it is now advised to report the stromal component only(31).

TILs are a broad marker of lymphocyte infiltration but do not tell us which lymphocyte subpopulations are present. In general, TILs are primarily CD8 cells, along with variable numbers of CD4 (helper T cells), CD19 (B cells), macrophages and NK cells(29). Immunohistochemical staining can be used to identify which subtypes of immune cells are present.

Clearly if TILs are going to be used as a biomarker, it is important that their evaluation must be standardised and reproducible with clear methodology. In order to facilitate this, international guidelines for the evaluation of TILs in breast cancer were published in 2014 and the guidelines are reproduced in Table 1.1.

**Table 1-1. Recommendations for assessing TILs in breast cancer.**

**Reproduced from recommendations by an International TILs Working Group 2014(31).**

| | |
|---|---|
| 1 | TILs should be reported for the stromal compartment. The denominator used to determine the % stromal TILs is the area of stromal tissue (i.e. area occupied by mononuclear inflammatory cells over total intratumoral stromal area), not the number of stromal cells (i.e. fraction of total stromal nuclei that represent mononuclear inflammatory cell nuclei). |
| 2 | TILs should be evaluated within the borders of the invasive tumor. |
| 3 | Exclude TILs outside of the tumor border and around DCIS and normal lobules. |
| 4 | Exclude TILs in tumor zones with crush artifacts, necrosis, regressive hyalinization as well as in the previous core biopsy site. |
| 5 | All mononuclear cells (including lymphocytes and plasma cells) should be scored, but polymorphonuclear leukocytes are excluded. |
| 6 | One section (4–5 µm, magnification x200–400) per patient is currently considered to be sufficient. |
| 7 | Full sections are preferred over biopsies whenever possible. Cores can be used in the pre-therapeutic neoadjuvant setting; currently no validated methodology has been developed to score TILs after neoadjuvant treatment. |
| 8 | A full assessment of average TILs in the tumor area by the pathologist should be used. Do not focus on hotspots. |
| 9 | The working group's consensus is that TILs may provide more biological relevant information when scored as a continuous variable, since this will |

| | |
|---|---|
| | allow more accurate statistical analyses, which can later be categorized around different thresholds. However, in daily practice, most pathologists will rarely report for example 13.5% and will round up to the nearest 5%–10%, in this example thus 15%. Pathologist should report their scores in as much detail as the pathologist feels comfortable with. |
| 10 | TILs should be assessed as a continuous parameter. The percentage of stromal TILs is a semi-quantitative parameter for this assessment, for example, 80% stromal TILs means that 80% of the stromal area shows a dense mononuclear infiltrate. For assessment of percentage values, the dissociated growth pattern of lymphocytes needs to be taken into account. Lymphocytes typically do not form solid cellular aggregates; therefore, the designation '100% stromal TILs' would still allow some empty tissue space between the individual lymphocytes. |
| 11 | No formal recommendation for a clinically relevant TIL threshold(s) can be given at this stage. The consensus was that a valid methodology is currently more important than issues of thresholds for clinical use, which will be determined once a solid methodology is in place. Lymphocyte-predominant breast cancer can be used as a descriptive term for tumors that contain 'more lymphocytes than tumor cells'. However, the thresholds vary between 50% and 60% stromal lymphocytes. |

These guidelines were updated in 2017, to include guidance on TIL assessment in the setting of post-neoadjuvant chemotherapy and carcinoma *in situ*(32*)*.

### 1.4.2 TILs and survival in breast cancer

Increased density of tumour-infiltrating CD8+ lymphocytes has long been associated with increased survival in breast cancer(30, 33) although in many studies, the association has

been limited to ER-negative (34) and/or basal-like TNBC (35). This reflects the heterogeneity of breast cancer and the importance of immunology in the treatment of this subtype. Tumours that have a high degree of TIL infiltration are named lymphocyte-predominant breast cancers (LPBC) and accordingly have the best prognosis. The threshold is usually defined as greater than 50 or 60% TIL infiltrate. Studies have reported LPBC rates of between 4.4 and 28.3%(36, 37). Unfortunately, this probably reflects the heterogeneity of TIL scoring, although the International Working Guidelines(31) should go some way towards standardising this.

### 1.4.3 TILs as a biomarker

TILs can be used as a predictive biomarker of chemotherapy response in TNBC; there is a strong and consistent association between TIL density or expression of immune markers and the likelihood of achieving a pCR after neoadjuvant chemotherapy (37, 38) (Table 1.2). In a prospective study of 580 patients with HER2-positive and triple-negative breast cancers, both greater TILs and elevated expression of 12 mRNA immune markers were significantly associated with pCR to neoadjuvant chemotherapy(37). LPBC tumours had a pCR rate of 59.9% vs. 33.8% for non-LPBC. This was confirmed in a large pooled analysis of 906 patients treated neoadjuvantly for TNBC(39). They found increased TILs were associated with longer PFS and OS, even when adjusted for baseline parameters within a multivariable analysis (MVA). The implication being that the immune system is necessary for chemotherapy efficacy in TNBC; the exact mechanisms by which this interaction leads to improved outcomes are unknown. It is presumed that immunogenic cell death allows cross-presentation of antigens, activation of dendritic cells and induction of cytotoxic T-cells(40). Multiple agents, including anthracyclines, gemcitabine and hormonal therapy, have been shown to be capable of inducing immunogenic cell death(41). It is likely that chemotherapy can have a modulatory effect on the immune system – neoadjuvant chemotherapy can convert a low-TIL tumour to a high-TIL tumour in a number of patients(42).

TILs can also be used as a prognostic biomarker at baseline in patients receiving adjuvant therapy. In this setting, increasing TIL density has again been found to have a robust and consistent correlation with better outcomes(36, 41, 43-45) (Table 1.2). In the BIG02-09 trial consisting of 256 ER-negative HER2-negative node-positive patients treated with adjuvant chemotherapy, the 5-year OS was 92% for those with the LPBC phenotype (10.6% of all patients) vs. 71% for those without(41). In a large study combining 481 TNBC patients from two phase III adjuvant trials (Eastern Cooperative Oncology Group [ECOG] 2197 and ECOG 1199), for every 10% increase in sTILs, a 19% reduction of risk of death was observed(36). sTILs were found to be an independent prognostic marker of DFS and OS.

TILs can also be used as a biomarker of residual disease. Early work compared 25 tumour samples from unselected breast cancer patients pre and post neoadjuvant chemotherapy(46). They found that of six patients without TILs at baseline but with a CR, four showed development of TILs following chemotherapy (67%). Only three of the 12 patients with a partial response developed TILs and none of the three with stable disease did. Another group evaluated T cell infiltration before and after neoadjuvant chemotherapy in 111 consecutive HER2-negative tumours and found that high CD8 infiltrate after chemotherapy was significantly associated with DFS and OS(47). Although overall absence of pCR indicates a poor outcome, they found a subgroup of patients with minor residual tumour but a high CD8/FOXP3 ratio with a long-term DFS and OS of almost 100%. More recently, a retrospective series of 278 TNBC patients treated with neoadjuvant chemotherapy found that those with high-TIL residual disease had significantly better five-year DFS (81.5% vs. 46%) and OS (91% vs. 55%) than the low-TIL group(40). Of the 19 cases where a pre-treatment biopsy was available for review, only one patient had a lower sTIL score after chemotherapy. Therefore in the vast majority of patients, chemotherapy appeared to induce lymphocyte activation and infiltration into the tumour.

Little is known about the molecular characteristics of TIL-dense breast cancers – such as the T-cell subtype repertoire and the genomic characteristics of the tumour. Given that

chemotherapy can convert a TIL-negative tumour into a TIL-positive state, it is rational to postulate whether we can augment TIL infiltration by other mechanisms. It may be possible to alter the balance of chemokines to promote cytotoxic activity locally. In one neoadjuvant study using lymphocyte mRNA expression, two immune markers that correlated with pCR and TILs were CXCL9 and CXCL13(37). These chemokines may be released by tumour cells and attract CD8+ lymphocytes. Toll-like receptor agonists have been shown to enhance antigen-specific CD8+ T-cell responses in pre-clinical models(48). Histone deacetylase inhibitors, which can upregulate MHC class I and II molecules, have been used *in vitro* to alter the immunogenicity of cancer cells(49, 50).

### 1.4.4 CD4+ TILs

While CD8+ lymphocyte infiltration is associated with positive outcomes in TNBC, there is evidence that the opposite may be true for CD4+ infiltration. In other cancers, the accumulation of FOXP3+ immunosuppressive regulatory T-cells (Tregs) correlates with shorter survival, presumed to be due to the suppression of endogenous anti-tumour activity (51). When the impact of Tregs on prognosis in breast cancer has been studied, the results have been conflicting. In one study, evaluation of 677 treatment-naïve breast cancers revealed that a low-CD8/high-CD4/high-macrophage immune signature was associated with significantly reduced OS and DFS, controlling for other clinicopathologic variables (52). A pCR to neoadjuvant chemotherapy has been associated with the disappearance of FOXP3+ cells by the end of treatment(47). Tregs may promote metastasis in breast cancer by acting in a paracrine fashion via RANK ligand, in addition to their usual role as mediators of immune tolerance (53).

However, the negative association between the presence of Tregs and outcome is not a universal finding and may depend on the subtypes being analysed. In a series of 237 breast tumours, higher numbers of FOXP3+ cells were correlated with shorter DFS and OS (54). When the ER-negative tumours were evaluated alone however, this correlation was

not found. In another series of 175 ER-negative tumours, high FOXP3+ infiltration was actually associated with *longer* recurrence-free survival, particularly amongst basal tumours(55). In common with the previously mentioned neoadjuvant study(47), generally there was a decrease in the number of FOXP3+ cells by the end of chemotherapy. The difference in association between FOXP3+ cells and outcome may be partly explained by the higher number of FOXP3+ cells in ER-negative tumours and the differing cut-offs used for defining the amount of cells infiltrated. In the latter study that showed FOXP3+ infiltration to have a positive impact on survival in ER-negative tumours, they used a cut-off that they felt was more relevant to this group. It should be noted that the positive impact of FOXP3+ cells was dependent on the simultaneous presence of large numbers of CD8+ lymphocytes. The authors hypothesised that CD8+ and FOXP3+ T-cells accumulate in tumours by the same mechanisms. To explain the benefit in prognosis they found, it was suggested that either the regulatory T-cells can exert pro-inflammatory functions in ER-negative tumours (e.g. through suppression of tumour-associated macrophages) or they may be exerting their usual suppressive function but at an insufficient level to combat the activity of the CD8+ cells(55). In a much larger series of 1445 breast tumours, the number of FOXP3+ cells was not an independent prognostic factor, in either ER-positive or ER-negative tumours(56).

Clearly further evaluation of the role of Tregs in ER-negative tumours is required, including assessment of their phenotype and function, in addition to the mechanisms by which they accumulate, before strategies to target them as a source of anti-tumour immunity can be employed. Blockade of CD25 (a receptor on FOXP3+ cells) has been shown to reprogram Tregs and enable them to produce IFNγ *in vitro* and when used *in vivo* in patients with metastatic breast cancer resulted in depletion of circulating Tregs (57).

**Table 1-2. Summary of studies evaluating TILs in ER-ve/HER2-ve or Triple Negative Breast Cancer[1].**

| First Author | Year | Study type (R/P[2]) | PR status | No. of patients[3] | 2nd pathologist | Median follow up (m/y[4]) | Median age (y) | Median % sTIL score | Threshold % TIL-high | % TIL high | % TIL low | Outcomes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **No-chemotherapy studies** | | | | | | | | | | | | **OS UVA – HR for death** |
| Park(58) | 2019 | R | ✔ | 476 | NK | 8y | 65 | 10 | 30 | 29.0 | 71.0 | 10%↑TILs: 0.91 (0.82-1.00,p=0.06) |
| **Adjuvant studies** | | | | | | | | | | | | |
| Loi(41) | 2013 | R | ✖ | 256 | Yes | 8y | 49 | 20 | 50 | 10.6 | 89.4 | 10%↑TILs: 0.82 (0.70-0.96, p=0.016)<br>LPBC: 0.30 (0.094-0.95, p=0.040) |
| Adams(36) | 2014 | R | ✔ | 481 | Yes | 10.6y | NK (22.7%≤ 40) | 10 | 50 | 4.4 | 95.6 | 10%↑TILs: 0.81 (0.69-0.95, p=0.01) |
| Loi(43) | 2014 | R | ✔ | 134 | Yes | 62m | 50.9 | 25 | 50 | - | - | 10%↑TILs: 0.80 (0.62-1.03, p=0.08) |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dieci(44) | 2015 | R | ✖ | 199 | No | 12.7y | 54 | 15 | 50 | 5.0 | 95.0 | 10%↑TILs: 0.89 (0.78-1.02, p= 0.1) |
| Pruneri(45) | 2016 | R | ✔ | 647 | Yes | 6.9y | 52 | 18 | 50 | 18.4 | 81.6 | 10%↑TILs: 0.83 (0.74-0.92, p<0.001)<br>LPBC: 0.42 (0.22-0.78, p=0.006) |
| Leon-Ferre(59) | 2018 | R | ✔ | 605 | No | 10.6y | 56.3 | 20 | 50 | 27.0 | 73.0 | 10%↑TILs: 0.90 (0.84-0.96, p=0.001)<br>LPBC: 0.60 (0.43-0.83, p=0.001) |

| **Neoadjuvant studies** | | | | | | | | | | | | **Outcomes as reported** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Denkert(37) | 2015 | P | ✔ | 314 | Yes | NK | 47 | NK | 60 | 28.3 | 71.7 | <u>pCR rate UVA</u><br>LPBC: HR 2.01 (1.22-1.31, p=0.006)<br>10%↑TILs: HR 1.15 (1.05-1.26, p=0.004) |
| Denkert(39) | 2017 | R/P (pooled analysis) | ✔ | 906 | No | 62.8m | NK (19% ≤ 40) | NK | 60 | 30.0 | 70.0 | <u>pCR rate</u><br>31% low, 31% intermediate, 50% high (p<0.0001) |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | pCR rate UVA<br>1%↑TILs: HR 1.16 (1.10-1.22, p<0.0001)<br>OS UVA – HR for death<br>10%↑TILs: 0.92 (0.86-0.99, p=0.032) |
| Mori(60) | 2017 | R | ✔ | 248 | No | 68m | 58.4-61.8[5] | NK | 50 | 47.6 | 53.4 | OS UVA – HR for death<br>TILs high vs. low: 0.4 (0.2-0.8, p=0.015) |
| Ruan(61) | 2018 | R | ✔ | 166 | Yes | NK | 50 | 15 | 50 | 4.2 | 95.8 | pCR rate UVA<br>10%↑TILs: HR 1.07 (1.03-1.10, p=0.0001) |
| Telli(62) | 2019 | R | ✔ | 161 | No | NK | 49.4 | NK | 50 | 8.1 | 91.9 | pCR rate UVA<br>10%↑TILs: HR 1.16 (0.97-1.40, p=0.107) |
| Ochi(63) | 2019 | R | ✔ | 80 | Yes | 98m | 52 | NK | 50 | NK | NK | pCR rate<br>4% low & 43.6% intermediate/high (p<0.001) |

[1]Between January 2010 and December 2019, I identified studies of interest by conducting a literature search using the PubMed database. I used the keywords: “breast cancer”, “triple negative”, “neoadjuvant”, “lymphocytes”, “TILs”, “adjuvant”, “TILs”, “TIL” and “infiltrate”. I also reviewed references from review articles to identify additional studies not found in my search. I included studies that were: 1) published in English between the above dates, 2) included patients with early stage triple negative or ER-negative breast cancer, 3) included at least 50 cases of ER-negative breast cancer or TNBC, 4) investigated the prognostic role of TILs in the neoadjuvant/adjuvant/no-chemotherapy settings, prior to the administration of chemotherapy 5) reported stromal TILs as a percentage infiltrate, 6) reported hazard ratios for overall survival (OS), or pCR rate for neoadjuvant studies, 7) used H&E slides, not tissue microarrays or specific markers such as CD8, 8) Published as original articles. Studies were manually reviewed for quality and excluded if not meeting this threshold.

[2]R=retrospective; P=prospective

[3]In studies without progesterone status, number consists of ER-ve HER2-ve patients

[4]m/y = months or years, depending on what was reported

[5]Median age for entire cohort not reported; median age was 58.4 years for patients with PD-L1+ve cancer, 61.8 years for patients with PD-L1-ve cancer

PR = Progesterone Receptor, HR = hazard ratio, NK = not known, UVA = univariable analysis, pCR = pathological complete response, LPBC = lymphocyte predominant breast cancer (threshold reported in Table)

## 1.5 Triple Negative Breast Cancer – heterogeneity and genomics

### 1.5.1 Background

Triple Negative Breast cancers (TNBC) are defined by lack of ER and PR staining by IHC and absence of HER2 overexpression by IHC or gene amplification by *in-situ* hybridisation. The majority are classified histologically as invasive ductal carcinoma, not otherwise specified (IDC-NOS)(64). TNBCs comprise 15-20% of breast cancers and are typically high-grade tumours, with poorer disease-specific survival(65).

Unlike the other subtypes of breast cancer, neither endocrine nor HER2-directed therapies are of clinical benefit in TNBC. PARP inhibitors are now approved by the Food and Drug Administration (USA) for the treatment of metastatic breast cancer in patients with a germline or somatic *BRCA1/2* mutation(66). There are no NICE-approved targeted therapies and novel strategies are required. Treatment of early-stage TNBC is comprised of surgery, radiotherapy and cytotoxic chemotherapy. Chemotherapy can be neoadjuvant (prior to surgery), adjuvant (after surgery) or palliative (for metastatic disease). A number of regimens, consisting of a combination of agents, are used, as the optimum combination has not yet been determined.

Despite its poorer prognosis relative to other breast cancers, patients with TNBC are more likely to respond to chemotherapy, particularly when given in the neoadjuvant setting. This is sometimes referred to as the TNBC paradox. In fact, 30-40% of early stage TNBC patients treated with standard neoadjuvant anthracycline and taxane based chemotherapy regimens achieve a pathological complete response (pCR) to treatment(42). pCR is considered a surrogate marker of improved survival in TNBC(67);

those with residual disease have a 6-fold higher risk of relapse and a 12-fold higher risk of death(68).

Neoadjuvant or adjuvant chemotherapy is considered for stage I to III (early-stage) TNBC. The chemotherapy recommended for early-stage disease is not unique to this subtype; in general this is anthracycline and taxane based treatment(69). However, there is growing evidence that platinum compounds may be particularly active in TNBC. A meta-analysis of nine randomised controlled trials (RCTs) incorporating platinums into neoadjuvant chemotherapy for TNBC indicates an improvement in pCR rate from 37.0% to 52.1% (OR 1.96, 95% CI 1.46–2.62,$p<0.001$) but at the expense of increased haematological toxicity(70). Of the two RCTs included that reported survival outcomes, there was no improvement in overall survival (OS). Therefore, platinum therapy is currently considered an option in the management of early-stage TNBC, but whether it translates into a meaningful survival benefit, possibly in a subset of patients, remains unclear.

The lack of specific targets is a major challenge in the treatment of TNBC. In recent years there has been a drive to understand its biology at a molecular level in order to further define the heterogeneity of this disease and discover potential novel targets.

### 1.5.2 Heterogeneity

Beyond the traditional IHC-based classification of breast cancer, with advances in genomic technology it is now evident that the disease is far more heterogenous than these basic subtypes. This may account for the differences in biological characteristics, clinical outcomes and response to treatment within the same IHC subtype(71). Attempts have been made to further delineate breast cancer into molecular subgroups with biological and prognostic relevance. In their seminal paper in 2000, Perou *et al*(72) used gene expression profiles to define five subgroups of breast cancer with distinct biological

and clinical features: Luminal A, Luminal B, HER2-enriched, normal-like and basal-like. Later, analysis of genomic data across multiple platforms from 825 breast cancer patients from the Cancer Genome Atlas (TCGA) confirmed these five intrinsic subtypes(73), although they found normal-like to be very infrequent.

One of the intrinsic subtypes of breast cancer, basal-like, has significant overlap with TNBC, although the two are not synonymous. There are no accepted international criteria for identification of basal-like tumours and so its use remains controversial(74). In two studies using mRNA expression array analysis, 8-29% of TNBCs did not have a basal-like subtype(74). Overall the overlap appears to be in the order of 60-90%(75, 76). Basal-like cancers are characterised by overexpression of basal markers (such as cytokeratins 8 and 18) and enrichment of genes involved in cell cycle checkpoints, proliferation and response to DNA damage(74). There is also overlap with the BRCAness phenotype, which is characterised by homologous recombination repair (HRR) defects and an inability to repair double strand DNA breaks; it has not been defined at a genomic level(77). Indeed, three quarters of tumours in germline *BRCA1* mutation carriers are basal-like(75) (Figure 1.3). A *BRCA1* mutation can be inherited or the gene can be mutated somatically. Dysfunction of the BRCA pathways can also occur by other mechanisms such as defects in other HRR genes or *BRCA1* promoter hypermethylation(77).

Over the last ten years, genomic and transcriptonomic analyses have been integrated with clinical outcome data in an attempt to define novel biological subgroups and elucidate common molecular themes. Some early work using a pooled analysis of gene expression data in over 2000 cases of breast cancer showed only the immune response gene module had an association with clinical outcome in both univariate and multivariate analysis in the ER-ve HER2-ve subtype(78). Similarly, a seven-gene immune response module was able to identify a subgroup of patients with ER-ve disease with reduced risk of distant relapse(79). Another group refined a six-gene T-cell related metagene and found that in patients with ER-ve HER2-ve disease, there was a higher rate of pCR and a higher rate of metastasis-free survival at five years (85.4% vs. 43.9% in the highest vs lowest tertile of expression)(80).

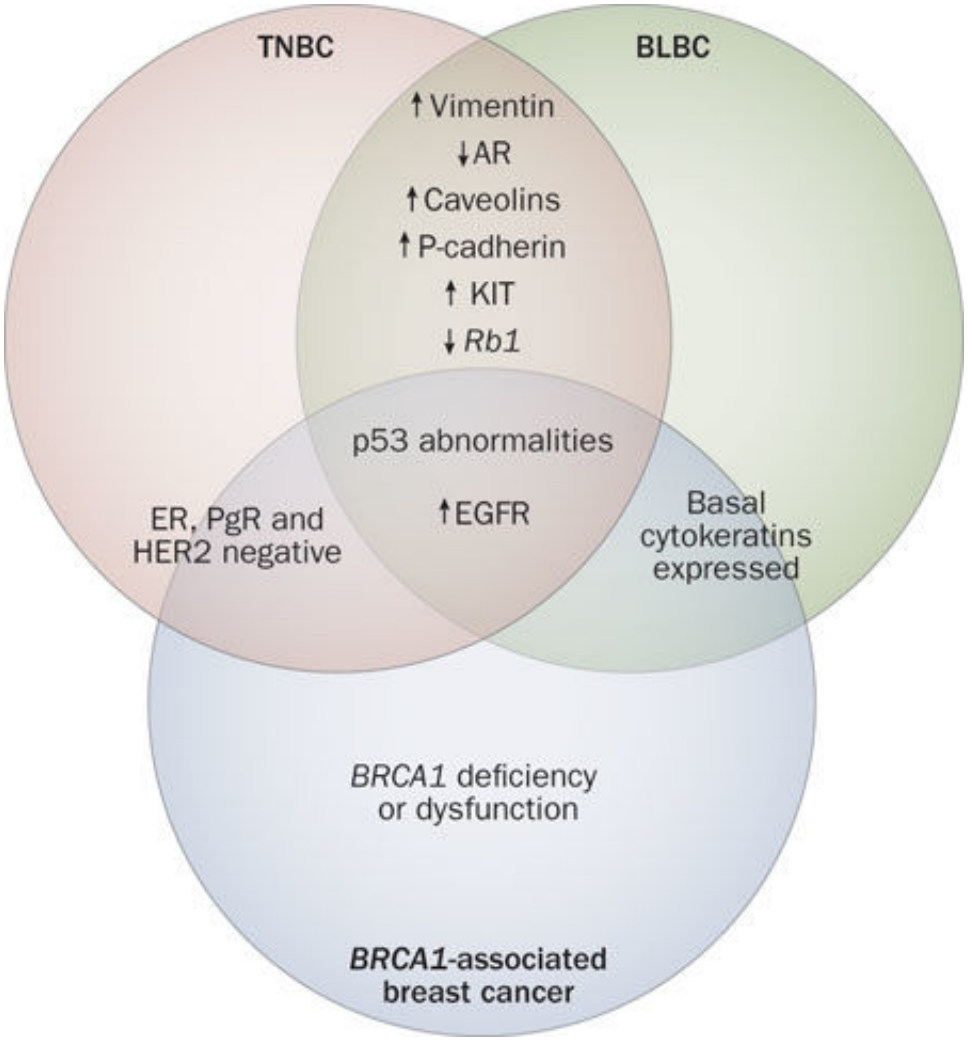



**Figure 1-3. Overlap between triple-negative (TNBC), basal-like (BLBC) and BRCA1-associated breast cancers.**

**Although there is significant overlap between these classifications, they are based on different technologies and are distinct. TNBC is an immunohistochemical classification while BLBC is a molecular subtype based on gene expression profiling. BLBC typically express genes such as cytokeratins 5, 14, 17, smooth muscle actin and P-cadherin (usually expressed in cells of the normal, non-luminal, myoepithelial layer of the breast duct(81). *BRCA1*-associated breast cancers may occur as a result of a germline or somatic mutation; alternatively they can occur due to epigenetic modification of the gene or abnormalities in other members of the BRCA1 pathway(77).**

**Image from Carey *et al(81).***

In the largest cohort using integrated genomic/transcriptonomic and clinical data in breast cancer, Curtis *et al*(82) used discovery and validation sets of 997 and 995 tumours,

respectively, to identify 10 novel subgroups based on joint clustering of copy number and gene expression data. One subgroup called “IntClust4” or “CNA-devoid” had a flat copy number landscape with a high proportion in this group showing heavy lymphocytic infiltration. The IntClust4 group was heterogenous, composed of both ER-negative and –positive cases, with a good clinical outcome.

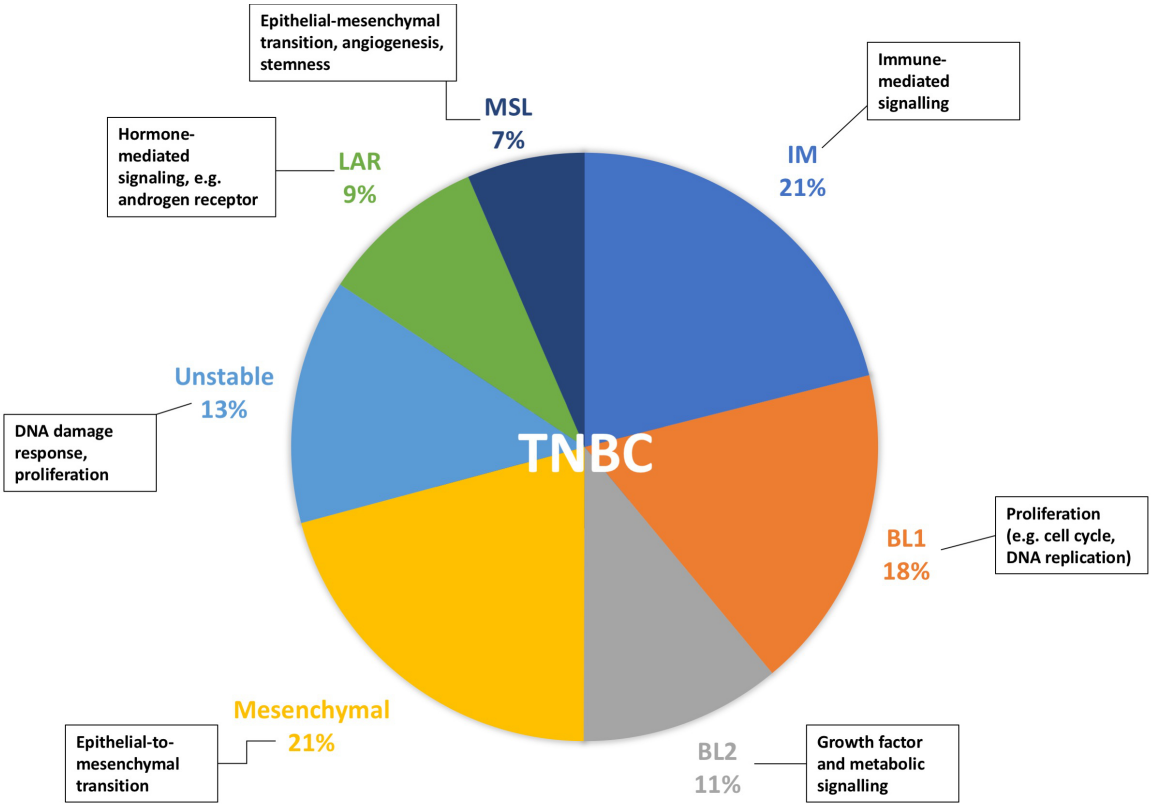



**Figure 1-4. TNBC subtypes with characteristic gene ontologies.**

**TNBCs subdivided into six subtypes using gene expression clustering from 587 cases from public data sets – includes basal-like-1 and -2, immunomodulatory (IM), mesenchymal, mesenchymal-stem-like and luminal androgen receptor subtypes. Data from Lehmann *et al(83).***

TNBC has been further subclassified into six subtypes using gene expression clustering from 587 cases from public data sets – basal-like-1 and -2, immunomodulatory (IM), mesenchymal, mesenchymal-stem-like and luminal androgen receptor subtype (83). These subtypes had characteristic gene ontologies and distinct clinical outcomes (additionally there was one unstable cluster) (Figure 1.4). The IM subtype was enriched

for genes involved in immune signalling and antigen presentation. They acknowledged the potential for stromal contamination but in microdissection the same proportion of IM subtype was found. Using RNA gene expression clustering, 198 fresh frozen TNBC samples were subdivided into four subtypes (84) and a “basal-like immune activated” subtype was identified. This subtype had upregulation of genes involved in B cell, T cell and natural killer (NK) cell function and had the best prognosis.

Therefore, TNBC is treated as a single disease entity, but is clearly heterogenous at a molecular level. This may reflect why it does not respond to treatment in a uniform manner. Gradually, treatments are emerging which target these subtypes, or molecular alterations occurring across subtypes. For example, bicalutamide and enzalutamide are undergoing phase II/III clinical testing for androgen-receptor-positive TNBC(85). Activation of immune gene pathways appears to confer a good prognosis in a subset of TNBCs. It may be that immune-targeted therapies would be most efficacious in this subset. It is likely that future trials in TNBC will utilise biomarker stratification. In the PARTNERING study, patients with residual TNBC disease after neoadjuvant chemotherapy were stratified by TIL score and then allocated to receive either: a combination of a cell cycle checkpoint inhibitor, a PARP inhibitor or an immune checkpoint inhibitor(86).

### 1.5.3 Genomic changes in TNBC

#### 1.5.3.1 Genomic instability

Recent developments in high-throughput sequencing technology have facilitated greater understanding of the mutational landscape of cancers. Somatic alterations can include copy number or structural aberrations, mutations, transcriptional modification and epigenetic changes. It is hoped that elucidating the driver signalling pathways underlying the pathogenesis of specific subtypes will enable the development of specific therapies.

Tumour mutational burden is defined as the total number of somatic mutations per coding area of a tumour genome. Mutations can be clonal (present in every tumour cell) or subclonal(87). On average, TNBCs have a somatic mutation rate of 1.68 per Mb (approximately 60 mutations in each tumour). The mutation rate is variable, with some tumours having a mutation burden of more than 4.68 per Mb(42). Basal-like tumours have a more heterogenous genomic landscape than non-basal, with variable transcriptional and DNA copy number activity(88). Genomic rearrangements occur at a much higher rate in basal tumours compared to the luminal A subtype (237 vs. 30 rearrangements)(89). TNBCs are more likely to have nonsense, frameshift and complex mutations compared to other subtypes. They are characterised by frequent DNA copy number losses and gains (71) and have a higher expression of a chromosomal instability signature compared to ER-positive tumours(90).

A minority of somatic mutations in tumours can produce neoantigens, defined as antigens derived from mutations that can potentially be recognised and targeted by the immune system. More highly mutated tumours can produce a greater number of neoantigens, hypothesised to enhance immune reactivity and checkpoint blockade responses. Indeed high tumour mutational burden and neoantigen load have been associated with response to checkpoint blockade (see section 1.6.3).

Overall, TNBC exhibits greater genomic instability and a higher mutational burden than the other subtypes, resulting in the generation of a greater number of neoantigens (42). This may be why triple negative tumours appear to be more immunogenic.

#### 1.5.3.2 Driver mutations

It was hoped that the identification of driver mutations would lead to the development of novel and effective targets in cancer. Unfortunately, intratumoural heterogeneity, clonal evolution and development of resistance mechanisms often render these therapies ineffective. For instance, in the treatment of non-small cell lung cancer with epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs), many tumours develop a point mutation of EGFR that alters the conformation of the binding domain, reducing the drug's potency through altered binding affinity(91). Using whole exome sequencing of chronic lymphocytic leukaemia cells at two time points, drug treatment has been shown to select for resistant subclones, possibly by removing competition between subclone populations(92).

Nevertheless, a better understanding of the neoplastic mechanisms should deepen our understanding of TNBC at a molecular level. In a recent large study, whole genome sequencing was performed on 560 breast cancers (93). They identified 93 genes carrying probable driver mutations, all in exons. At least one driver was identified in 95% of cases. The 10 most frequently mutated genes accounted for 62% of drivers (Figure 1.5). Although they found high mutation frequencies in non-coding regions, it was felt that this was secondary to structural features in these areas and they were unlikely to contain drivers. They identified at least 12 base substitution mutational signatures and 6 rearrangement signatures.

Analysis of 466 breast cancers from the TCGA dataset identified only three genes (*TP53*, *PIK3CA* and *GATA3*) that had somatic mutations at an incidence of >10% across all breast cancers, with the majority of mutations occurring in only a very small proportion of cases. Consistently, basal-like and TNBCs have shown a high frequency of *TP53* mutations (80%) and loss of *RB1* and *BRCA1* are also found. Activation of the PI3K/Akt pathway is also seen, by mechanisms that include loss of *PTEN* or amplification of *PIK3CA*. However, *PIK3CA* was also one of the more commonly mutated in basal-like tumours (9%) in the TCGA cohort(73).

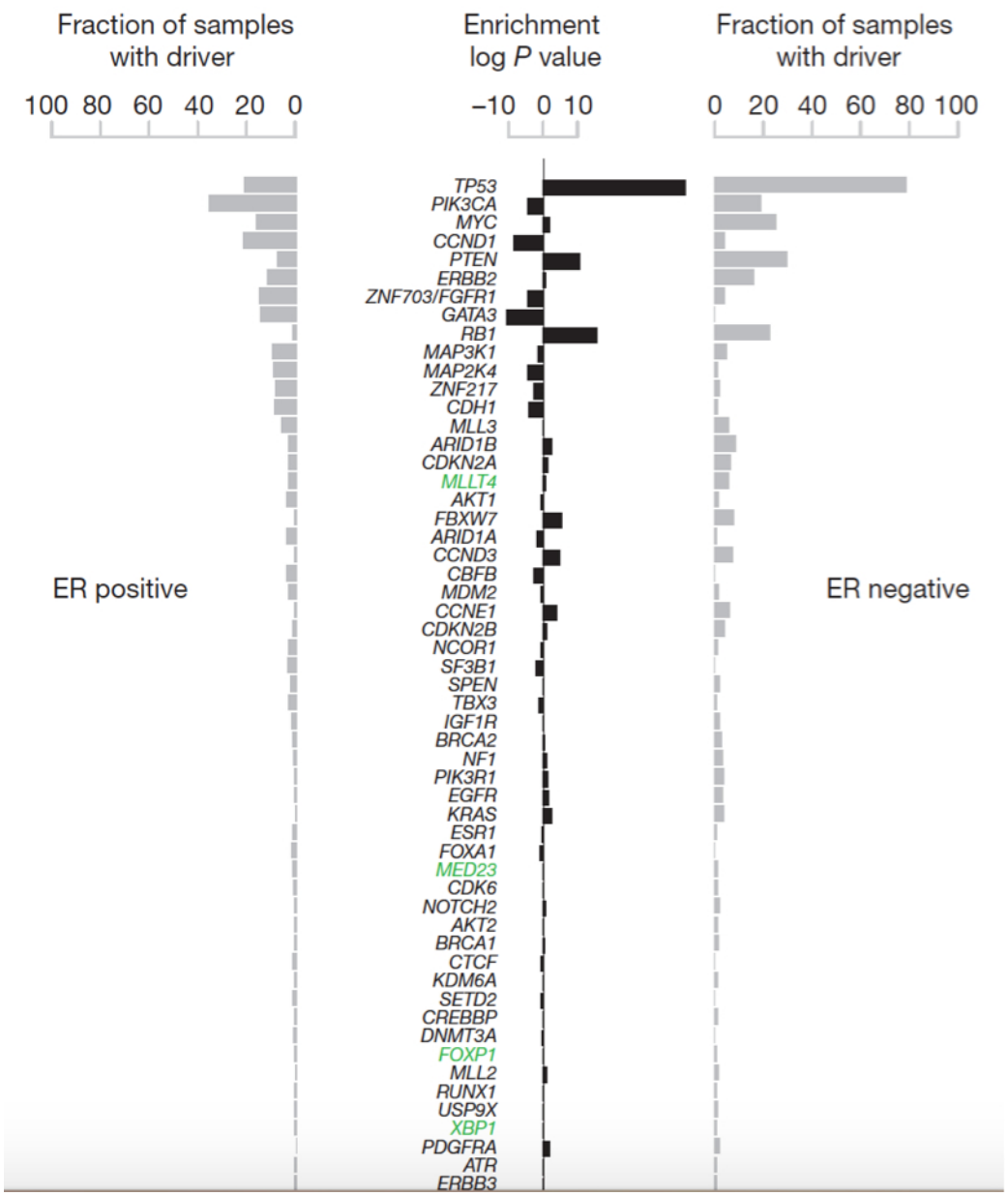



**Figure 1-5. Somatic mutation rates in ER-positive and ER-negative tumours.**

**Image from Nik-Zainal *et al*(93).**

**This constitutes a list of curated driver genes, sorted in descending order of frequency. The total ER-positive samples were 366 and ER-negative samples were 194; the fraction of samples found to have a mutation in the driver gene is shown in grey. The genes highlighted in green have new or additional evidence supporting that they are likely to be novel breast cancer genes.**

Using whole-exome sequencing of 104 early TNBC cases, *TP53* was again shown to be the most frequently mutated gene, in 62% of basal TNBCs and 43% of non-basal TNBCs. The next five most commonly mutated genes were *PIK3CA* (10.2%), *USHA2A* (9.2%), *MYO3A*

(9.2%), *PTEN* (7.7%) and *RB1* (7.7%)(88). Excluding *MYO3A*, these genes were mutated at a similar frequency in the TCGA study(73). Other more common mutations in other studies have included *MLL3*, *GATA3* and amplification of *CCNE1*(71).

Overall, TNBCs bear a diverse range of mutational patterns, with relatively few recurrently mutated genes, except for *TP53.* Given the predictability of a mutation in *TP53,* its downstream molecules have been investigated as potential targets, e.g. *CHEK1/2, ATM* and *ATR*(71). The PI3K/Akt signalling pathway is frequently activated in TNBC, via activating mutations in *PIK3CA* or *AKT1* and/or loss of *PTEN(94)*. There are early phase clinical trials underway (in patients with metastatic disease) studying the role of PI3K inhibitors in TNBC(95). One study identified recurrent *MAGI3-AKT3* fusion gene, present in five out of 72 TNBC samples(89). This translocation leads to constitutively active Akt kinase signalling, which can be targeted by a small molecule inhibitor. In the phase II PAKT trial assessing the safety and efficacy of an Akt inhibitor (capivasertib) in 140 patients with untreated metastatic breast cancer, the PFS was 5.9 months in the paclitaxel and capivasertib arm versus 4.2 months in the paclitaxel and placebo arm (p=0.06)(94). Median OS was 19.1 months and 12.6 months respectively (p=0.04). Of those tested, 25% (n=28) had activating *PIK3CA/AKT1* mutations or inactivating *PTEN* alterations and there was a PFS benefit of 9.3 vs. 3.7 months for those who received the Akt inhibitor in this subgroup. Inhibitors of mTOR, a protein kinase downstream effector of the PI3K pathway, are already licensed for clinical use in ER+ disease(96) and are also in early stage clinical trials for TNBC(95). Other potential targets could include amplification of receptor tyrosine kinases such as FGFR1/2, c-KIT and MET(73). Use of targeted therapy in unselected TNBC patients is unlikely to generate high response rates; upfront biomarker selection is likely to yield better results.

Attempts have been made to correlate mutations with clinical features and/or outcome in TNBC. Next generation sequencing of 56 pre-treatment primary TNBC biopsies was performed in order to identify mutations that correlated with clinical features(97). The most common mutations found were in *TP53*, *TTN* and *PIK3CA* (rates of 55%, 14% and 9% respectively). Mutations in *TTN* are however often described as a false-positive

finding(98). The mutation rates were similar between those who responded to chemotherapy and those who didn't (average mutation rate 9 vs. 8 mutations). No recurrent mutations were associated with outcome. Those with a relapse were noted to have a higher copy number alteration rate.

Expression profiling of residual disease following neoadjuvant chemotherapy for TNBC has been used to identify subgroups with prognostic relevance in a chemoresistant subset(99). In one study, molecular profiling of residual disease in 74 TNBC cancers identified a potentially actionable therapeutic target in 90% of cases(100). Alterations more frequently found in this cohort (compared to TCGA cohort) were suggested as having a role in the development of chemoresistance. There was a greater frequency of *MCL1* amplifications, *PTEN* deletions/mutations and *JAK2*, *CDK6* and *CCND1-3* amplifications. As discovered previously, a wide diversity of somatic alterations were found, with most present in less than 5% of lesions.

#### 1.5.3.3 Intratumoural heterogeneity

In addition to extensive inter-patient somatic variability, extensive intratumoural heterogeneity has also been identified(101). Using single nuclei sequencing of one ER+ and one TNBC tumour, aneuploidy was found to occur early in tumour evolution, remaining stable during clonal expansion(102). Point mutations however evolved gradually, contributing to extensive clonal diversity. Again, these mutations frequently occurred at a low frequency in the tumour mass (<10%). It was found that the TNBC cells had an increased mutation rate (13.3x) whereas the ER+ tumour cells did not. Whether this holds true across all tumours within these traditional immunohistochemical subsets (ER-positive, TNBC, etc.) remains to be seen.

In another study using single nuclei sequencing, 1000 cells from primary tumours from 12 patients with treatment-naïve TNBC were used to investigate clonal substructure(101). They found that a large number of copy-number aberrations (CNAs) are acquired at an

early stage in tumour evolution, followed by stable clonal expansion ("clonal stasis"). In each tumour they identified one to three major clonal subpopulations with a common evolutionary precursor. CNAs appeared relatively stable throughout the tumour. This challenges the accepted dogma of gradual tumour evolution and also contradicts previous reports of extensive intratumoural heterogeneity in TNBC. As these studies are only using a small number of tumours, this likely reflects the inter-patient heterogeneity of TNBC, with some tumours having a small number of driver pathways and few mutations and others having extensive mutation burden and a diverse set of pathways involved. Overall, TNBC tends to have higher clonality at diagnosis compared to other breast cancers but this is not absolute(88).

In addition to heterogeneity within a single tumour mass, there can also be variability between lesions, e.g. primary tumour and metastases. In order to investigate this, one study used 10 pre-treatment primary TNBC tumours with matched lymph node metastases to compare CNAs and verified their findings with five additional pairs of tumours (primary and metastasis)(103). There was no increase in CNAs in the lymph node metastases, indicating a high clonal relationship with the primary tumour. Using matched primary tumour and brain metastasis from a patient with basal-like breast cancer, it was found that the mutation frequency range in the primary tumour was wide, indicating significant intratumoural heterogeneity(104). However, the mutation frequency range was much narrower in the metastasis, indicating that the metastasis had arisen from a small subset of cells from the primary tumour. There were two *de novo* mutations and a large deletion not present in the primary tumour, indicating on-going clonal evolution. More recently, whole genome and mRNA sequencing were performed on two primary TNBC/basal-like breast cancers with an aggressive clinical phenotype, in addition to four or five site-specific metastases matched with each tumour(105). They found that overall gene expression signatures, copy number patterns and mutation patterns were similar between the primary tumour and associated metastases. The primary tumour and metastases were closely related by hierarchical clustering and were more closely matched than to primary tumours or metastases from other patients. Almost every mutation in the primary tumour was detected in a metastasis and many were shared across the primary

tumour and all matched metastases. Of note, each metastasis originated from multiclonal seeding rather than a single cell of origin. There were a small number of mutations identified only in the metastases; the majority of these were present in two or more sites. Only a few of the metastasis-specific mutations were expressed in mRNAs, indicating that those arising after seeding were more likely to be passenger mutations.

In a prospective study designed to improve understanding of the biology of metastatic disease, whole exome sequencing and SNP arrays were performed on primary tumour and longitudinal metastatic biopsies, in addition to autopsy biopsies, from a single patient with TNBC(106). The case, which had an aggressive clinical course, again showed that a clone from the primary was present in all the metastases (liver and lung). However, there were also new clones that derived from an earlier progenitor to the primary clone. Some of the metastases showed very similar subclonal structures, indicating possible metastatic cross seeding (rather than dissemination from the primary alone). In addition, there was evidence of recurrent subclonal mixtures, indicating possible cooperation between clones, which enhance each other's survival. Although only one case was sequenced, the depth of information about multiple metastatic sites enables us some insight into the clonal evolution of metastatic TNBC. Whether these findings can be generalized is unclear, and many more cases will need to be sequenced in future studies.

### 1.5.4 BRCA-deficient tumours

#### 1.5.4.1 BRCAness

There is significant overlap between basal-like and triple-negative tumours with the "BRCA-ness" phenotype. Breast cancer in *BRCA1* mutation carriers has a basal-like phenotype in 75-90% of cases and high levels of genomic instability are observed (107). BRCA proteins repair double stranded DNA breaks by HRR and are also involved in cell

growth and cell division regulation. They therefore function to maintain genomic stability(108). Individuals with a germline mutation in *BRCA1* or *BRCA2* are at a high lifetime risk of developing cancer, particularly breast and ovarian. It is generally accepted that oncogenesis in those carrying a pathogenic germline *BRCA* mutation follows the "two-hit" rule with the second hit being somatic inactivation of the wild-type allele.

Although the majority of breast cancers are not associated with a germline *BRCA* mutation, at a somatic level there has been found to be loss of at least one *BRCA1* allele in between 50 and 70% of sporadic breast cancers. Epigenetic modification of *BRCA1* promoters has also been found in sporadic tumours(109) and may lead to a BRCAness phenotype. Somatic *BRCA1* promoter methylation and a germline *BRCA1* mutation appear to be mutually exclusive events(110). The mutation of genes which stabilize or promote the function of *BRCA1/2* can also produce the BRCAness phenotype, e.g. *PALB2* and *BRIP1*(108). Mutations in other genes involved in HRR, such as *EMSY* and *RAD51C* may also be grouped together with these tumours.

#### 1.5.4.2 The molecular biology of *BRCA1*-mutated tumours

*BRCA1*-deficient cells exhibit genomic instability and frequently have chromosomal translocations with large deletions and/or amplifications involving multiple, non-homologous chromosomes(109). The function of BRCA1 is now known to be more complex and varied than HRR alone, by way of its interaction with over 100 different proteins. Downstream effects including transcriptional regulation, apoptosis and DNA damage signaling(109).

It is not clear how different *BRCA1*-mutated basal-like tumours are at a genomic level from their non-*BRCA1*-mutated counterparts; i.e. how loss of BRCA1 alters the mutational landscape of a breast cancer. In order to investigate this, one group interrogated the TCGA database and identified 12 tumours with *BRCA1* germline mutations, seven with somatic mutations and one with both germline and somatic mutations(111). There were only minor molecular differences observed between *BRCA1*-mutated versus wild type

basal cancers, according to gene or protein expression, miRNA expression and DNA methylation. No genes were differentially expressed. They did however identify 250 genes with higher amplification (i.e. DNA copy number aberration) rates in *BRCA1*-mutated tumours compared to basal-like wild-type tumours, including *ID4,* which negatively regulates *BRCA1* and *BRCA2* expression. The *BRCA1*-mutated tumours showed a higher average number of mutations (122.6 versus 80.3, p=0.004) and *TP53* mutations were found in 100% (versus 75.9% of wild-type). *PIK3CA* mutations were not found in the *BRCA1*-mutated tumours, compared to 10% in wild-type(111). The implications of this are obviously limited by the small sample size.

Therefore, despite the increased somatic mutations and genomic instability found in *BRCA1*-mutated tumours in this small cohort, no difference in protein expression (according to the limited protein array expression data using 171 unique proteins/phosphor-proteins) was found, i.e. there was no phenotypic change in *BRCA1*-mutated tumours.(111). Of note, four of the 14 chromosomal regions with higher amplification rates found in the *BRCA1*-mutated tumours are characteristic of high-grade serous ovarian carcinomas. In the original TCGA report, similarities were found between the basal-like (and TNBC) mutation spectrum and that of serous ovarian cancers; both having only one gene mutated at a frequency of >10% (*TP53)*(73). Other features in common included *BRCA1* inactivation, *RB1* loss, *Cyclin E1* amplification, increased expression of *AKT3* and amplification/increased expression of *cMYC.* Both also exhibited widespread genomic instability and frequent gains at 1q, 3q, 8q and 12p and losses at 4q, 5q and 8p. Gene expression analysis also showed high correlation between the two tumour types (in addition to squamous lung carcinomas).

In a large study by Nik-Zainal *et al*(93), whole genome sequencing was performed on 560 breast cancers (all subtypes), of which 90 had germline or somatic *BRCA1/2* mutations (60 were germline). Of these 90 cases, loss of wild-type chromosome 13 or 17 was observed in 80 cases. Inactivating *BRCA1/2* mutations were found to confer many genomic rearrangements and characteristic rearrangement signatures were identified. At least five mutational signatures of *BRCA*-deficient tumours were found. A number of breast cancers without an identifiable *BRCA* mutation segregated with these tumours in hierarchical

clustering; it was suggested that promoter methylation or other mechanisms might be conferring a similar effect in these cases. This group has recently developed a whole genome sequencing-based predictor (HRDetect; HRD) in an attempt to quantitatively define genomic features of *BRCA*-deficiency. Of note, HRD was able to detect *BRCA*-deficiency from formalin fixed paraffin embedded (FFPE) tissues. Of the 560 breast cancer patients enrolled in their study, 124 samples had an HRD score of >70% for BRCA1/2 deficiency. Of these, 55 had a germline *BRCA* mutation and 22 had somatic *BRCA* deficiency (epigenetic or genetic inactivation). Amongst these were 47 samples (of which 42 were ER-negative) that did not have a germline or somatic *BRCA* mutation and either absent promoter hypermethylation of *BRCA1* (n=10) or unavailable methylation information (n=37). Three of these 47 patients had a mutation in another HRR gene. The authors highlighted the fact that nearly a third of patients with a high HRD score did not have an identifiable deficiency of *BRCA* by genetic or epigenetic means and suggested the possibility that these patients may also respond well to treatments found to be beneficial for *BRCA*-deficient tumours.

#### 1.5.4.3 Targeting BRCA-deficient tumours

Using these signatures as a biomarker of defective HRR may widen the net for use of drugs such as platinums and poly (ADP-ribose) polymerase (PARP) inhibitors, which have been identified as potentially useful in *BRCA*-mutated cancers(69). Within the TCGA cohort, they noted that around 20% of the basal-like tumours had a germline or somatic mutation of *BRCA1/2*(73). PARP inhibitors exploit the principle of synthetic lethality, while platinums cause fatal intra-strand crosslinking, the repair of which is deficient in *BRCA*-null cells(112). Given the underlying defect in DNA damage repair, it is not surprising that *BRCA*-deficient tumours appear to be more sensitive to treatment with platinums. Amongst the 43 patients with advanced TNBC enrolled in the TNT trial with a germline *BRCA* mutation, the response rate was 68.0% for single agent carboplatin versus 33.3% for docetaxel(113). The median progression-free survival (PFS) for *BRCA*-mutated patients receiving carboplatin was 6.8 months, versus 3.1 months in those with wild type *BRCA.*

There was no difference in OS. In further exploratory analysis, this group found no improvement in response for those with epigenetic silencing of *BRCA1* (BRCAness subgroup)(114). Unfortunately, only a small number had available tumour sequencing data, limiting investigation of responses in those with a somatic *BRCA* mutation. In the neoadjuvant setting, in the GeparSixto trial, the presence of a germline *BRCA* mutation was predictive of a better pCR rate (compared to those without a mutation) following neoadjuvant chemotherapy; however there was no additive effect from carboplatin(115).

PARP inhibitors have been shown to be effective in patients with platinum-sensitive relapsed ovarian carcinoma and a germline or somatic *BRCA* mutation in a phase III trial(116). Median PFS was significantly longer with the PARP inhibitor (olaparib): 19.1 vs. 5.5 months. Olaparib is now approved as treatment in this setting(117). In a phase III trial of 302 patients with metastatic germline *BRCA*-mutated HER2 negative breast cancer, olaparib monotherapy had a significantly longer PFS of 7.0 vs. 4.2 months for physicians' choice chemotherapy(118). In order to ascertain whether PARP inhibitors may be useful in patients with breast cancer and HRR deficiency at a somatic level, a phase II study is underway evaluating the efficacy of a PARP inhibitor in patients with HER2-negative metastatic breast cancer with a BRCAness phenotype(119).

There has been only limited investigation of the somatic picture in germline *BRCA*-mutated cancers relative to their sporadic counterparts and this warrants further attention. Elucidating molecular markers of BRCAness is of importance now that PARP inhibitors are in use clinically and for stratification within clinical trials, e.g. for immunotherapy. Currently, olaparib is recommended by NICE for the treatment of women with germline *BRCA1/*2 mutations and platinum-sensitive relapsed ovarian, fallopian tube or peritoneal cancer after three or more courses of chemotherapy(117). Niraparib is licensed for use in the UK and NICE has recommended it as a candidate for use in the Cancer Drugs Fund for women with a germline *BRCA* mutation and ovarian cancer after 2 courses of platinum-based chemotherapy(120). In the UK, PARP inhibitors are currently only available for patients with breast cancer in the context of a clinical trial.

#### 1.5.4.4 The immunogenicity of BRCA-deficient tumours

Microsatellite instability (caused by defects in DNA mismatch repair) is frequently found in colorectal cancer and gives rise to the accumulation of repetitive sequences of DNA. This generates immunogenic neoantigens, with enhanced anti-tumour response and CD8 infiltration(121). Given the role of BRCA in the maintenance of genomic integrity, it has been hypothesised that a germline *BRCA* mutation will also give rise to a higher mutational burden and subsequently greater immunogenicity. In one study using a small number of tumours from the TCGA data set, *BRCA-1* mutated TNBCs (n=29) had a higher TIL score (and a higher number of non-silent mutations) compared to wild type (n=64)(122). CD8, PD-1 and CTLA4 also correlated with a *BRCA-1* mutation. It is hoped that this will translate to an immunotherapy benefit in these patients. Indeed, there are a number of Phase I/II trials currently recruiting that are investigating the combination of a PARP inhibitor with checkpoint inhibitor immunotherapy(123). In contrast to microsatellite instability, *BRCA* mutations lead to structural variation in the genome, with resultant genetic damage and impairment of cell growth(124). This does not however necessarily lead to neoantigen formation and may mean that immunotherapy is not as effective in *BRCA*-mutant cases as it has been proven to be in microsatellite unstable tumours.

Recently, in a paper designed to investigate the mutational load, immune profile and response to checkpoint blockade in a *BRCA1*-deficient tumour model, the authors first interrogated TCGA. They found that TNBCs from patients with a germline *BRCA1* mutation had an increased somatic mutational load and greater lymphocyte infiltration, along with increased expression of PD-1 and CTLA4 compared to *BRCA1*-wild type tumours. They postulated that BRCA1-deficient tumours have a greater burden of neoantigens, stimulating the recruitment of tumour-specific effector T-cells. However, they only provided evidence of increased mutational burden (not neoantigens) in *BRCA1*-mutant TNBC tumours (BRCA-1 mutant cases n=7, p=0.05). Using a *BRCA1* homozygous mutant mouse model with TNBC mammary tumours, they found that dual immunotherapy with anti-PD-1 and anti-CTLA4 improved antitumour immunity when the mice were treated with cisplatin. This response was characterised by activation of TILs, enhanced dendritic

cell activation and reduced suppressive FOXP3 regulatory T-cells(122). These effects may be mediated by mechanisms other than neoantigen stimulation, e.g. via components of innate immunity. DNA damage response-deficient TNBC cells have been shown to activate the cyclic GMP-AMP synthase (cGAS)/stimulator of interferon genes (STING) pathway, resulting in the production of proinflammatory cytokines and resultant T-cell recruitment(125).

Very recently, the immunogenicity of BRCA-deficient breast cancers was investigated using genetic/somatic *BRCA*-mutated samples from the TCGA dataset (n=89), in addition to 35 local cases(126). The authors sought to determine why immune infiltration varies between *BRCA*-deficient cases, despite generally having a higher mutational load. They used a homologous recombination repair deficiency (HRD) score, which combined three independent DNA-based measures of genomic instability. They confirmed that mutational burden and neoantigen load were higher in *BRCA1/2*-mutated tumours from the TCGA dataset. The HRD score was also higher in *BRCA1/2*-mutated tumours. Mutational burden correlated with increased HRD score.

High-HRD *BRCA1/2*-mutated cancers also appeared to be more genomically unstable, as evidenced by greater aneuploidy(126). Cytolytic index, a measure of CD8-cytotoxicity, was also higher in *BRCA*-deficient cases. Unexpectedly, HRD-high, *BRCA*-deficient cancers had lower cytolytic and immune ESTIMATE (a measure of lymphocytic infiltration) scores.

Amongst *BRCA1/2*-mutated cases, TNBCs had a higher cytolytic index despite comparable mutational burden(126). Membrane PD-L1 was higher in TNBC cases. The greatest disparity when evaluating cytolytic index and immune ESTIMATE scores in *BRCA1/2*-deficient cases was TNBC HRD-low (where these scores were high) and receptor-positive HRD-high cancers. This was validated by IHC in a small number of local cases. They therefore suggested that HRD and hormone receptor expression independently impacted immunogenicity in *BRCA*-deficient cases. The HRD-low subset was the most highly immunogenic, indicating that there are mechanisms other than mutational burden driving

the immune response, e.g. dysfunction in apparatus involved in the recruitment or activation of immune cells. Of note, they found increased TGF-β signalling to be a prominent feature of the HRD-low microenvironment.

## 1.6 Mechanisms of Immune Evasion

Immune editing explains why tumours can remain present, but dormant, for many years before becoming more proliferative and clinically apparent again. The immune system is able to recognize and destroy some tumour cells that present antigens, but cell turnover and new mutations results in the generation of clones that are able to evade immune destruction. The primary mechanisms by which cancer cells achieve this are loss of antigenicity and loss of immunogenicity. Antigenicity is the ability to be recognised by antibodies, specifically generated by the immune system in response to the foreign agent. Immunogenicity is the ability to activate the immune response. A substance can be antigenic, without being immunogenic (whereas immunogenicity implies antigenicity) (127).

### 1.6.1 Altered Antigenicity

MHC class I proteins are membrane-bound proteins expressed on almost all nucleated cells, encoded by three genes at the human leukocyte antigen (HLA) locus (HLA-A, -B and –C) (128). The binding affinity between MHC-I and peptides is a major component of how "antigenic" these peptides are; MHC-I is consequently vital for an anti-tumour immune response. A potential mechanism for immune evasion in solid tumours is HLA loss, resulting in reduced ability to present neoantigens. HLA mutations have also been described, resulting in disrupted MHC-antigen interactions. This is not however common (129). Recently, HLA loss of heterozygosity has been shown to occur in 40% of non-small cell lung cancers and this was associated with increased neoantigens, PD-L1 upregulation

and greater cytolytic activity(130). This was suggested to be a late event in tumour evolution, as evidenced by their subclonality and enrichment within metastases.

A limited number of studies (using variable techniques) have reported a positive association between MHC-I and -II and TILs, in addition to survival(131-133). Downregulation of MHC class I has been reported in different tumour types and correlated with poor prognosis(134-136). There are no studies evaluating MHC class I IHC staining in TNBC specifically.

Recent work has also focused on a patient's individual genotype and its effect on the tumour mutations present. The combination of six MHC-I alleles carried by an individual determines which peptides can be presented effectively. One group was able to show that patient MHC-I genotypes had an impact on the mutations emerging in their tumours(128). Reduced presentation of a specific mutation was correlated with a higher frequency of expression of that mutation. They therefore hypothesised that a person has an inherited personal cancer susceptibility shaped by MHC-I genotype-restricted immunoediting.

In addition to presentation, the processing of antigens within the endoplastic reticulum is also critical in the generation of the immunopeptidome. This comprises the thousands of peptides presented by MHC class I and II molecules on all of a person's cells (including tumour cells)(137). Endoplastic reticulum aminopeptidases (ERAP1 and 2) are enzymes now known to play a critical role in the generation of antigenic peptides in the MHC class I pathway.

Following the cytosolic degradation of proteins by the proteasome and processing by peptidases, a small proportion of peptides are then transported into the endoplastic reticulum by the Transporter Associated with antigen-Processing (TAP) heterodimer. On arrival in the endoplastic reticulum, peptides are of varying lengths. Proteasomes can

generate peptide fragments ranging from 2-25 residues long, although TAP preferentially transports those of 8-16 residues long(138). As MHC I molecules require a peptide of a specific length (typically 8 to 10 amino acid residues) to fit into their groove, many peptides require further N-terminal trimming prior to loading(139). ERAP1 appears to perform a unique function, acting as a "molecular ruler", preferentially trimming peptides of 9-16 residues and generating optimal lengths for class I presentation(138). It can also destroy peptides by excess processing(140). In this way, the enzyme is well placed to influence the antigenic repertoire presented to T-cells.

It has been observed that absence of ERAAP (the mouse homolog of ERAP1) depletes the repertoire of MHC-peptide complexes in antigen-presenting cells. Additionally, its absence was found to elicit changes in the stability of the MHC-peptide complexes with the production of unique but structurally unstable complexes which paradoxically elicited a potent immune response(141). Therefore ERAP1 has a dynamic and complex role in the generation of the immunopeptidome. It therefore follows that investigation of ERAP1 variation in cancer is important; perhaps ERAP modulation is a mechanism by which tumours evade immune destruction. Using a murine colorectal carcinoma model, it was shown that ERAAP downregulation resulted in increased generation of a protective epitope (the hypothesis being that ERAAP overprocesses this epitope, silencing it to the immune system) and a cytotoxic response, resulting in tumour rejection. They also showed that topical application of an ERAAP inhibitor resulted in arrest or reduction in size of established tumours(142). It has been shown that p53 binds to and upregulates ERAP1 and through this mechanism leads to a corresponding increase in MHC complex I expression(143). There is known to be variable expression of ERAP1 and 2 in tumours(139). Downregulation of ERAP1, in addition to other proteins involved in MHC-I presentation, has been shown in a small number of TNBC cases using mass spectrometry(133).

### 1.6.2 Immunomodulation

Although weak antigenicity is likely to be one of the primary causes of a poor host immune response to a tumour, there are a number of mechanisms that generate a tumour-promoting immunosuppressive environment that opposes any existing host response. The tumour stroma has a pivotal role in the orchestration of a hostile environment. The plasticity of the tumour microenvironment may explain why some tumours are able to effectively evade the immune system and will be discussed in detail in Section 1.8. Additionally, tumours can produce cytokines that attract immunosuppressive cells to the microenvironment, including regulatory T-cells and myeloid-derived suppressor cells. These cytokines can alter the immune balance, e.g. by shifting from a Th1 to a Th2 response. Additionally they can downregulate tumour antigens themselves, in addition to inducing apoptosis in cytotoxic T-lymphocytes(28).

### 1.6.3 Mutational Burden and Immunogenicity in TNBC

The neoepitope load is predicted by the burden of non-synonymous mutations; overall this is low in breast cancer(144). The relationship between genomic aberrations and immunogenicity has been explored in many solid tumours. In a pooled analysis of solid tumours from the TCGA database, there was a correlation found between both the neoantigen load and the total number of somatic mutations and immune infiltration(145). Therefore, one hypothesis is that immune infiltration occurs in tandem with more genomic alterations, as a result of the generation of immunogenic neoantigens.

In melanoma the level of mutational load has been correlated with clinical benefit from anti-CTLA4 therapy (146). However there were patients with a high mutational load that did not respond to immunotherapy, so alone this did not necessarily translate to a clinical benefit. These authors found there were a number of shared tetrapeptide sequences (neoepitope signatures) shared across patients' tumours that demonstrated a stronger association with response and survival than mutational load alone. In another study in

melanoma, comparable levels of potentially immunogenic antigens were found in immune-hot vs. immune-cold tumours(24). Of note, only a low number of mutations were found to be immunogenic, producing reactive T-cell clones, although there are likely to be technical limitations in the detection of mutations present in only a small number of tumour cells. Additionally, immunotherapy is effective in renal cell carcinoma, which does not have a large mutational burden(144). Therefore, it is likely that there are mechanisms other than mutational load alone that drive immune infiltration and improve response to immunotherapy.

Mutations in specific genes (e.g. in the PI3K pathway(147)) have been shown to affect the immune microenvironment in other solid tumours. Recently, in an attempt to identify DNA level changes correlating with either high or low immune gene expression in breast cancer, a group used RNA-seq, DNA copy number, mutation and germline SNP data from over 1000 cases from the TCGA database (including 191 TNBC cases)(148). They were unable to identify any recurrent DNA alterations significantly associated with immune expression. It was found that higher immune metagene expression was significantly associated with *lower* clonal heterogeneity. In TNBC specifically, greater mutation rates, neoantigens and CNV loads were also consistently (but not statistically significantly) associated with lower expression of immune metagenes. The hypothesis presented was that breast cancers with minimal TIL infiltrate have greater clonal diversity due to escape from immune surveillance(148). These cells are therefore no longer undergoing clonal elimination by the immune system. They suggested that high-TIL tumours were in a state of “near-equilibrium” with low clonal diversity. Consistent with previous studies discussed above, a broad spectrum of gene and pathway level alterations were found in breast cancer, each being significant in only a small number of patients within each IHC subtype.

## 1.7 Immunotherapy

### 1.7.1 Checkpoint inhibition

The field of cancer immunotherapy has developed dramatically over the previous decade, with the emergence of effective therapies now used in a wide variety of cancers. One potential mechanism of immune resistance includes immune checkpoints, which function to limit T-cell effector function and mediate immune tolerance. Antibodies to these checkpoint ligands and receptors have now been developed and are licensed for use in a number of cancers. When PD-L1 on a tissue surface binds to the PD-1 receptor on a T-cell, inactivation of the lymphocyte occurs. It is now well established the PD-L1 plays a key role in limiting the cytotoxic anti-tumour response. Agents that target cytotoxic T lymphocyte-associated molecule-4 (CTLA-4) and programmed cell death ligand-1/receptor-1 (PD-L1/PD-1) are the most widely used at present but novel targets continue to emerge(144).

Therefore, one might expect PD-L1 expression to be associated with an immunosuppressive microenvironment and poor prognosis, but results have been conflicting and seem to depend on tumour type. PD-1 expression is associated with T cell exhaustion and in general high expression is associated with a poor prognosis in a range of cancers, e.g. renal cell and non-small cell lung cancer (149, 150). Thirty percent of all breast cancers have unregulated PD-L1 on the tumour cell surface(151), thereby downregulating the lymphocyte response. PD-L1 expression has been correlated with TIL density across all breast cancers(151-153) and in TNBC specifically(60, 154), acting as a surrogate marker for immune response (albeit its downregulation). However it is probably an over-simplification to state that all PD-1 positive lymphocytes are exhausted; some may indeed be activated in breast cancer(155).

The association of PD-1/PD-L1 with survival in TNBC is not clear. In general, PD-1 and PD-L1 expression are greater in TNBC compared to the other subtypes; even in metastatic TNBC, 41% of tumours were found to be positive (≥1%) for PD-L1 expression (152, 156). The combination of high PD-L1 expression with low TIL density has been shown to be a negative prognostic factor in treatment-naïve TNBC cases(60). However, two recent

reports in smaller TNBC cohorts have found a positive correlation between PD-L1 and survival; one of these validated these findings at a transcriptional level and in a separate cohort using publically available datasets (149, 154). Another study showed PD-L1 was adversely prognostic, but only in tumours where TILs were low(60). This may fit in with the developing “adaptive resistance” model in melanoma and colorectal cancer, where the prognostic impact of PD-L1 appears to be influenced by the TIL infiltrate(150).

So far, clinical trials using PD-L1/PD-1 immune checkpoint blockade (ICB) in advanced TNBC has yielded responses in only a small number of patients (up to 19%)(157). Biomarkers are sought to identify the responders. There are conflicting results regarding the association between PD-L1 expression at baseline and response to anti-PD-L1 therapy; TILs may be a better biomarker of response to immune checkpoint therapy(158). For instance, a subgroup of patients with residual disease and high TILs may be a good target for ICB(159). In the KEYNOTE-086 study of pembrolizumab (anti-PD-1) for metastatic TNBC, PD-L1 positivity on the tumour had no impact on objective response rates (ORR). However those previously untreated for their metastatic disease and with available tumour blocks (n=46) with TIL levels above the median had an ORR of 39% versus 9% in those with low TIL scores(160). They therefore hypothesised that TILs may be used as a biomarker of response to immunotherapy for metastatic TNBC, particularly in the first-line setting. It should be noted that TILs were measured on metastatic biopsies in most cases and on the primary (archival) tumour in others. A trial designed to stratify patients according to TIL status with a standardised method of sampling the tumour would be more robust.

Overall, ICB has yielded only modest responses in TNBC; given that tumour immunogenicity may wane with advancing disease it may be more successful if given earlier in the disease process, e.g. with neoadjuvant therapy. Additionally, ICB is unlikely to be sufficient alone, particularly in those with low TILs and immune escape. Efforts are now focused on combining immunotherapy with chemotherapy in the hope that there will be a synergistic effect.

The Impassion130 trial is a phase III trial in which patients with unresectable/locally advanced or metastatic TNBC were randomised to receive nab-paclitaxel with the PD-L1 antibody atezolizumab or placebo. At a median follow-up of 12.9 months, the median PFS of those in the PD-L1 treatment group was 7.2 months vs. 5.5 months in those who received placebo (p=0.002), with an OS benefit of 21.3 months vs. 17.6 months (p=0.08) at interim analysis(156). The OS benefit for those with PD-L1 positivity (40.9%) was 25.0 vs. 15.5 months (p-value not calculated), indicating this was a useful biomarker in this group of patients as those with PD-L1 expression appeared to derive more benefit.

Two phase II trials, I-SPY2 and KEYNOTE-173, showed a significant improvement in pCR rates with the addition of neoadjuvant pembrolizumab (anti-PD-L1) to chemotherapy(161). More recently, pembrolizumab, was combined with chemotherapy in a phase III trial (KEYNOTE-522) in the neoadjuvant setting(162). At a median follow up of 15.5 months, the PD-L1 arm showed a statistically significant improvement in pCR, from 51.2% to 64.8% (p=0.00055). Whether this translates to an improvement in survival remains to be seen. The efficacy of immunotherapy in the adjuvant setting is also being investigated.

### 1.7.2 Immunotherapy beyond checkpoint inhibition

Cancer vaccines have been developed in the hope that their administration will produce an effective immune response against tumour-specific antigens. The aim is usually to re-establish host immunity to a tumour associated antigen following the development of a cancer(23). Two major limitations of vaccines include the lack of an effective and sustained immune response, and an appropriate antigenic target. Vaccines can be monovalent against a single antigen or polyvalent against multiple antigens(157). Cancer testis (CT) antigens are attractive targets for immunotherapy as they are not found in normal somatic tissue(65). Vaccines are being developed against CT antigens that have

been identified in TNBC, including MAGE-A and NY-ESO-1. Vaccines have also been developed against the glycosylated form of MUC-1, another antigen expressed on breast cancer cells(65).

To date, cancer vaccine trials have been unsuccessful across multiple tumour types(157). This is likely to be related to the failure of vaccination to engage multiple components of the immune system (e.g. cytotoxic cells in addition to B-cells), immune escape mechanisms and the testing of vaccination in patients with established advanced disease(163). Modern vaccination strategy is likely to involve combination regimens that elicit a more complete immune response that is able to overcome immune tolerance. As vaccine therapy may be limited by T-cell exhaustion, checkpoint inhibition is being combined with them in the treatment of TNBC(158). Neoantigen-targeted, personalised vaccines have also been designed, in order to overcome the propensity for tumours to evolve to evade antigenic targeting from simple vaccines(157).

Adoptive T cell transfer involves the administration of lymphocytes to mediate antitumour effects and can include TILs, T-cell receptor cells and chimeric antigen receptors (CAR T-cells)(164). CAR T-cells have the advantage of being able to activate T-cells, independently of MHC-restriction, meaning that tumour-induced loss of MHC can be circumvented(164). CAR T-cells have produced responses in melanoma and haematological malignancies(165). Its use is being investigated in other solid tumours. To date this approach has not been successful in breast cancer despite the development of CAR T-cells targeted to c-Met (for TNBC) and HER2(144). The infusion of HER2-specific T cells in a patient with metastatic colon cancer resulted in a fatal cytokine storm(144). Unfortunately, one limiting factor with CAR T-cells is their lack of specificity to cancer cells; in this case unintended targeting of HER2-positive cells in the lung epithelium triggered a deleterious response(164). The tumour microenvironment is a further limiting factor; the addition of immunotherapy to CAR-T therapy is being investigated as a possible adjunct(164).

## 1.8 Tumour Microenvironment

### 1.8.1 The Stroma In Cancer

Until recently, the field of cancer therapeutics has focused on the tumour-centric model and the role of the stroma, or tumour microenvironment (TME), has been relatively ignored. The TME may play a critical role, due to factors such as the presence of immunosuppressive cells, the metabolic state and the density of the extracellular matrix. Elucidating mechanisms to target these factors is likely to be crucial in the future management of TNBC, in addition to other solid tumours.

The TME is composed of the extracellular matrix (ECM) and non-malignant stromal cells, including immune cells, endothelial cells and fibroblasts(166). This complex environment can alter the behaviour of tumour cells, depending on the balance of cytokines, growth factors and cells present. The restriction of tumour-specific T cells to stromal regions (rather than into tumour islands) implicates a role for the stroma in immunosuppression in cancer(167). As the tumour evolves, the microenvironment also changes in a dynamic, progressive way via paracrine signaling between the tumour and stromal cells(168). Thus, the interaction between neoplastic cells and the TME is now considered to be of pivotal importance in tumour progression. Targeting the TME is also attractive because there may be more mechanistic homogeneity between patients/tumours, compared to the heterogeneity of cancer cells.

One group performed a spatial analysis of the immune cells present in TNBC(169). They classified CD8 staining into Fully Inflamed, Stroma Restricted or Margin Restricted (tumour immune microenvironment classification(170)) using a bioinformatics analysis of publically available transcriptonomic and genomic datasets. Again they found heterogeneity between subtypes, with the immunomodulatory subtype being fully

inflamed, whereas M, MSL and LAR had the highest levels of a margin-restricted signature. Basal-like was predominantly stromal restricted. Of note, stromal restricted tumours were associated with younger patients. This spatial classification was also correlated with TME signatures: the inflamed tumours were negatively associated with stroma/metabolism signatures, while the margin restricted tumours highly expressed stromal signatures. Stromal restricted tumours on the other hand were primarily associated with metabolism signatures. This reflects the diverse and heterogenous stromal-immune microenvironment present within TNBC.

### 1.8.2 CAFs

Fibroblasts are the most abundant cell present in the TME and they have a multifactorial role, including the production of the ECM, regulation of cellular differentiation and alteration of the immune response(171). They are spindle-shaped cells, with a notable lack of specific marker(172). Normal fibroblasts suppress tumour formation, while cancer-associated fibroblasts (CAFs) are generally thought to be pro-tumorigenic(173). CAFs are characterised by expression of α-smooth muscle actin (α-SMA) and fibroblast activation protein-α (FAP-α)(173). Their activated phenotype is analogous to “normal” fibroblasts involved in fibrosis or wound healing(172). These activated cells can also be referred to as myofibroblasts. Key features of myofibroblasts include the ability to newly express α-SMA. α-SMA is an isoform of actin, which increases the ability of myofibroblasts to contract. Following tissue injury, myofibroblasts secrete increased amounts of ECM proteins, such as collagens and fibronectin(174).

There is mounting evidence that CAFs are associated with tumour initiation, progression and metastasis. This effect is mediated by a number of mechanisms, including promotion of proliferation, neoangiogenesis and extracellular matrix remodelling(173). There is also evidence that CAFs contribute to an immunosuppressive microenvironment. The degree to which CAFs mediate the dysfunction of CD8+ T-cells and the mechanisms by which this

may occur are unclear. Mechanisms are likely to include the secretion of immunosuppressive cytokines, T-cell apoptosis and induction of protumoural macrophages(175). CAFs may even present antigen, in an APC-like manner, and kill CD8+ T-cells in an antigen-dependent manner(167).

CAFs do not exhibit a universal phenotype. Indeed, there is not a single universal marker of CAFs; αSMA, FAP-α, integrinβ1, PDGFRβ and caveolin 1 have all been proposed(173). In one study using human breast cancer samples, fluorescent-activated cell sorting was used to distinguish four CAF populations, using expression of various markers. In the same study, they showed that TNBC exhibited the highest expression of CAF markers. TNBC tumours were enriched with the S1 and S4 subsets, both of which showed high expression of SMA. Overall the subsets were enriched in different breast cancer subtypes and had different spatial distributions. They also appeared to be associated with different immune cells; for instance the S1 subtype was associated with more FOXP3 cells. Furthermore, the S1 and S4 subtypes had different transcriptonomic profiles. S1 cells were involved in cell adhesion and ECM organisation, as well as the immune response, and S4 cells behaved more like pericytes, with muscle contraction and oxidative metabolism being features. The S1 subset appeared to orchestrate immunosuppression, through attraction and retention of CD4+ T-lymphocytes and promoting differentiation into FOXP3+ T-cells. In another study using single cell RNA sequencing from breast tumours, CAFs were divided into three major groups based on gene expression patterns. Function appeared to differ between CAF subsets(176).

A group from Sweden has delineated a role for platelet-derived growth factor receptors (PDGFR) in the attraction and behaviour of CAFs in breast cancer. They have shown that one of the platelet-derived growth factor (PDGF) ligand isoforms, expressed by breast cancer epithelial cells, is associated with poor prognosis in breast cancer. They showed that this ligand (PDGF-CC) was correlated with the basal subtype of breast cancer. The corresponding PDGFRs were expressed by CAFs. Genetically PDGF-deficient mice had reduced tumour growth and reduced collagen deposition. Finally, *in vitro,* they showed that PDGF-CC might promote a malignant ER-ve phenotype via expression of growth

factors such as hepatocyte growth factor (HGF)(168). In a separate study, they showed association between PDGF-CC and the triple negative subtype and increased risk of 5-year disease recurrence(177). This indicates one potential mechanism by which paracrine signaling between tumour cells and CAFs promotes tumour growth and perhaps influences the molecular phenotype of breast cancer.

### 1.8.3 Targeting CAFs

Despite the potential role of CAFs in tumour progression and immunosuppression, clinically effective treatments targeting CAFs have not been developed. To date, research has focused on targeting the growth factors produced by the stroma (e.g. matrix metalloproteinase inhibitors), the TGF-β pathway and eliminating CAFs directly(178).

Depletion of FAP-positive cells *in vivo* has been shown to facilitate immunological control of pancreatic adenocarcinoma(179). In a 4T1 *in vivo* model, a FAP vaccine was effective at reducing tumour growth, increasing CD8 influx and suppressing metastases, when given in combination with doxorubicin chemotherapy(180). This effect appeared to be mediated by a shift from Th2 to Th1 polarisation, reduction in tumour-associated macrophages (TAMs) and an increase in CD8+ T-cells. However, a monoclonal antibody to FAP has been ineffective in the treatment of metastatic colorectal cancer in a phase II trial(181).

Given the importance of TGF-β in the differentiation to the myofibroblastic phenotype, drugs targeting this cytokine have been developed. However, TGF-β also acts as a tumour suppressor in early carcinogenesis and TGF-β monoclonal antibodies have resulted in the development of cutaneous keratoacanthomas/squamous cell carcinomas in early-phase human studies(182). It appears that the context is crucial with regards to the effect of TGF-β. Our group has therefore sought a more specific CAF inhibitor and has published

work showing NOX4 is crucial for fibroblast to myofibroblast differentiation. NOX4 inhibition by genetic or pharmacological methods (such as GKT831) was effective in reducing myofibroblastic CAFs and slowing tumour growth *in vivo* (183). A number of other drugs which potentially target CAFs are in preclinical testing, including PDGFR signalling blockade and CXCR4 inhibitors(184). Exploring the interaction between cancer cells and CAFs may also result in success; for instance IL11 is secreted by CAFs and may enhance the metastatic potential of tumour cells(185).

Despite remarkable progress, immunotherapy is only successful in a proportion of patients who receive it. One possible reason for immunotherapy resistance could be the impact of the TME on the function and repertoire of immune cells. The production and alteration of ECM components by CAFs may generate a physical barrier preventing the infiltration of T cells. CAFs have therefore been implicated in resistance to immunotherapy by their possible role in this exclusion effect. In a murine study of pancreatic adenocarcinoma, which is typically resistant to immunotherapy, depletion of FAP+ stromal cells (CAFs) rendered both anti-CTLA-4 and anti-PD-L1 treatment more effective; the authors singled out CXCL12 as the chemokine secreted by FAP+ cells responsible for this immune evasion(186).

In one study, a large cohort of patients with metastatic urothelial cancer who received immunotherapy was studied and correlates of response determined(187). Those who did not respond had greater expression of TGF-β signalling in fibroblasts. The authors noted that patients had exclusion of T-cells from the tumour parenchyma; these CD8 cells were instead resident in the stroma. They then targeted TGF-β and PD-L1 in a CD8-exclusion mouse model and found this allowed the T cells to penetrate into the centre of the tumour.

In a colorectal cancer model, mouse strains were generated bearing four oncogenic mutations, targeted to intestinal stem cells(188). They found that mice bearing all four of these mutations had an immune excluded phenotype; the severity of this phenotype

correlated with the number of mutations introduced. These tumours had high levels of stromal TGF-β and genes induced by TGF-β. Using flow-sorted RNA-sequencing, they confirmed that CAFs are the main source of TGF-β in this model. They then used a TGFBR1-specific inhibitor galunisertib, which was effective in reducing tumour volume and metastases; this appeared to be dependent on CD8 and CD4 cells. However, once metastatic disease was established, galunisertib was not able to induce remission; the T-cells in this case appeared to have increased expression of PD-1 and the metastases (particularly TAMs) had marked PD-L1 expression. With the addition of anti-PD-L1 therapy to galunisertib, lymphocytic infiltration was increased and resulted in the eradication of most metastases (even after cessation of treatment). Therefore, these authors demonstrated that immune cold tumours could be targeted via TGF-β suppression, conferring susceptibility to checkpoint inhibition and resultant immune infiltration.

Therefore, although CAF inhibition is unlikely to be a successful agent used alone in cancer therapy, it could be a useful adjunct when used with other systemic therapy. It could particularly be useful for the large subset of patients who are resistant to immunotherapy.

### 1.8.4 The DNA Damage Response Pathway and Fibroblast Differentiation

The DNA Damage Response (DDR) pathway is triggered by DNA damage or replication stress, mediated by sensors. The DDR pathway is a protein kinase cascade, with transducers such as ataxia-telangiectasia mutated (ATM) protein(189). DDR is crucial as it regulates key events such as apoptosis and differentiation, thereby promoting tumour suppression in the context of DNA damage(190). Given that defective DDR signals are known to occur in cancer (allowing uncontrolled cell proliferation even in the presence of damaged DNA), efforts have been made to target them. It is possible that one could capitalise on existing DDR defects in a tumour, in order to utilize synthetic lethality

(similar to PARP inhibition)(191). Specific DDR inhibitors have been developed, which target proteins including ATR, ATM, CHK1 and WEE1.

ATM is a serine/threonine protein kinase and is an upstream master regulator of the DDR pathway (189). ATM is therefore involved in a large number of processes that function to maintain genomic stability, for instance through activation of p53 and subsequent apoptosis(191). It is specifically activated by DNA double strand breaks. ATM-deficiency results in increased neoplasm formation, in addition to altered immune responses and metabolism. Many of these outcomes in ATM-deficient mice have been linked to redox homeostasis(190). However, targeting the DDR pathway is hypothesised to render tumour cells sensitive to DNA-damaging therapy.

Elevated Reactive Oxygen Species (ROS) is implicated in the development and progression of pathological fibrosis(192). Our group have previously shown that NADPH Oxidase 4 (NOX4)-generated ROS is crucial for differentiation into myofibroblasts and highlighted NOX4 as a viable potential target in the field of stromal targeting for cancer(183). ATM can be activated following oxidative stress (without DNA damage). Cells with an absence of ATM are also sensitive to insults such as oxidative stress(193). In systemic sclerosis, an autoimmune fibrotic disease, ATM mediates WIF-1 silencing (again in response to oxidative damage) via phosphorylation of the transcription factor c-Jun. Knockdown of WIF-1 in normal fibroblasts resulted in increased Wnt signaling and increased collagen production, postulated to contribute to fibrosis in the setting of systemic sclerosis(194). In a mouse model of renal injury, ATM phosphorylation increased; a concomitant increase in pSMAD3 levels indicated this was TGF-β driven. *In vitro*, ATM inhibition decreased expression of TGF-β target genes including fibronectin. It was shown that ATM was necessary for fibrotic mechanisms, driven by TGF-β, in fibroblasts, in a redox-sensitive manner(192). Our group has previously shown that RNA-sequencing of HFF2 fibroblasts, transdifferentiated by TGF-β, are enriched for the DDR pathway(195). The NOX/ATM/p53 axis is therefore suggested as a possible therapeutic target in fibrotic disorders; cancer can be considered to be amongst these.

## 1.9 The POSH Study

Throughout this work, I will be using data and biological samples from the POSH cohort to explore themes within breast cancer. The POSH cohort is a prospective observational study of almost 3000 patients aged 40 years or younger with a first diagnosis of invasive breast cancer (196). The primary aim of the study was to evaluate the impact of inherited genetic factors on the prognosis of patients with breast cancer(197). Young patients were recruited specifically in order to enrich the cohort for *BRCA* mutation carriers. Patients were recruited between 2000 and 2008 from 127 hospitals in the UK. The main eligibility criteria was a diagnosis of invasive breast cancer at an age of 40 years or younger; patients could be recruited within 12 months of initial diagnosis. This young breast cancer cohort has different baseline pathology characteristics than the general breast cancer population, including lower frequency of T1 tumours and higher grade, absence of hormone receptors, vascular invasion and lymph node involvement. In addition, patients with young onset breast cancer are also recognised to have an increased risk of disease recurrence and death(196).

The POSH cohort is a large study, set up locally, with a wealth of clinicopathological data available and genotyping of germline *BRCA* mutation status for the vast majority (>94%). Published work utilizing this cohort to date has included papers regarding outcomes for those of different ethnicities(198), family history(199), BMI(200) and germline *BRCA* status(201). The latter described similar survival for breast cancer occurring in *BRCA* mutation carriers compared to wild type. A possible survival advantage in the first few years after diagnosis of TNBC for those with a germline *BRCA* mutation was noted. I therefore felt it was an appropriate, and available, dataset to explore outcomes in both metastatic disease and triple negative breast cancer. Both of these constitute an unmet need in the field of breast cancer.

## 1.10 Summary

TNBC is a relatively common disease, with poorer disease-specific survival relative to the other breast cancer subtypes (202). There is a need to develop targeted therapies that are effective in this subtype of breast cancer. There is emerging evidence that the immune system plays an important role in TNBC biology. In integrated genomic/transcriptonomic analyses, an immune-active subtype of TNBC with improved prognosis has been defined(82, 84). Although immunotherapy is undergoing trials in this disease, combination approaches are likely to be required to improve response rates.

TILs are now established as being a useful biomarker in early TNBC. However, the biological mechanisms that render some tumours immune-rich and others immune-cold are unclear. The tumour microenvironment may play a critical role, including factors such as the presence of immunosuppressive cells, the metabolic state and the density of the extracellular matrix. Elucidating mechanisms to target these factors is likely to be crucial in the future management of TNBC

## 1.11 Aims of this research project

The central hypothesis of this project was that an immune correlate could be identified from clinical and immunopathological analysis of breast cancer cases from the POSH cohort and subsequently targeted in an *in vivo* murine model of triple negative breast cancer.

The aims were to:

- Describe the differences in clinical pattern and baseline features of *de novo* versus metastatic breast cancer
- Explore the prognostic impact of baseline clinicopathological features in early-stage TNBC
- Construct tissue microarrays from primary tumour samples from the POSH cohort
- Incorporate bioinformatics of publicly available datasets in the identification of relevant molecular markers
- Explore the prognostic impact of molecular markers in early-stage TNBC
- Identify clinicopathological and molecular correlates of immune activity
- Utilise digital pathology in the scoring of molecular markers
- Utilise digital pathology to undertake a spatial analysis in full-face TNBC tumours
- Undertake further *in vitro* and *in vivo* work using above findings to alter immune infiltration in a murine TNBC model
- Identify a possible target for tumour microenvironment remodelling in TNBC

# Chapter 2 Methods

## 2.1 General materials

| | |
|---|---|
| Agarose (high-resolution) | Sigma Aldrich |
| Amphotericin B | Gibco |
| β-mercaptoethanol | Sigma Aldrich |
| Bovine Serum Albumin (BSA) | Sigma Aldrich |
| Captisol | Glentham |
| Dulbecco's Modified Eagle Medium (DMEM) | Sigma Aldrich |
| Dimethyl sulfoxide (DMSO) | Sigma Aldrich |
| EDTA-trypsin (0.25%) | Gibco |
| Ethanol (99%) | Thermo Fisher Scientific |
| Fetal Calf Serum (FCS) | GE Healthcare |
| Isopropranolol (99%) | Thermo Fisher Scientific |
| Kanamycin | Melford Biolaboratories |
| L-glutamine | Sigma Aldrich |
| Midori green | Nippon Genetics |
| Optimem | Gibco |
| Paraformaldehyde | Thermo Fisher |
| Penicillin/streptomycin | Sigma Aldrich |
| Phosphate-buffered saline (PBS) | Thistle Scientific |
| RIPA buffer | Thermo Fisher Scientific |
| RNase Zap | Sigma Aldrich |
| Standard agarose (high resolution) | Sigma Aldrich |
| TGF-β1 | R&D systems |
| Triton-X | Bio-Rad |
| Trypan blue (0.4%) | Gibco |
| Tween20 | Sigma Aldrich |
| Water (PCR-grade) | Life Technologies |

## 2.2 POSH study

### 2.2.1 Patient population and study design

The Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) was a prospective cohort study of women aged 18 to 40 years old and diagnosed with a first episode of invasive breast cancer. The primary aim of this study was to determine whether the prognosis from breast cancer was changed by inherited genetic factors (197). A detailed report of the presenting characteristics, pathology and treatment of this cohort was published in 2013 and the primary outcome paper was published in 2018(196, 201). Patients were excluded if they had a previous invasive malignancy (excluding non-melanomatous skin cancer). Patients were consented within 12 months of initial diagnosis. All patients received treatment according to local protocols. A detailed protocol has been published(197). In total, 3021 female patients were recruited from 127 UK hospitals between 2000 and 2008 within 12 months of initial diagnosis. The study received approval from the South West Multicenter Research Ethics Committee (MREC 00/6/69). Written informed consent was obtained from all participants.

Information regarding personal characteristics, tumour pathology, stage and treatment received were collected from medical records at study entry. Family history was collected by questionnaire. Pathology and imaging data were verified with copies of original reports. Follow-up data including date and site of disease recurrence were obtained from medical records at 6 months, 12 months and thereafter annually until death or loss to follow-up. Follow-up interval was determined according to local standards; no imaging or other investigation was mandated by this study, as it was observational. Adjuvant chemotherapy was manually categorised into regimens and anthracycline/taxane-containing. Patients were flagged in the National Health Service Medical Research Information Service to facilitate automatic notification of date and cause of death. This work presents analyses conducted on follow-up data received until 26 June 2016.

Oestrogen receptor (ER), progesterone receptor (PR) and HER2 receptor status of primary tumours were determined from routine diagnostic pathology reports including immunohistochemistry (IHC) and/or FISH. Hormone receptor concentrations equivalent to an Allred score of 3 or more were categorised as positive. Tissue microarray (TMA) immunohistochemical staining was used to supplement missing information regarding receptor status. Triple negative breast cancer (TNBC) was defined as tumours that were oestrogen receptor-negative, HER2-negative, and progesterone receptor-negative or unknown.

DNA for genotyping was extracted from whole blood samples collected at recruitment. A multiplex amplicon-based library preparation system, Fluidigm Access Array (Fluidigm UK, Cambridge, UK) was used to sequence a panel of breast cancer susceptibility genes, including *BRCA1/2* and *TP53.* Illumina HiSeq2500 Next Generation Sequencing Platform was utilized (Illumina, Little Chesterford, UK). If patients met current UK guidelines for genetic testing, multiplex ligation probe analysis was used to ensure mutations consisting of large exonic deletions or duplications were not missed. Pathogenic variants were confirmed by Sanger sequencing. Those with variants of unknown significance were classified as *BRCA*-negative.

The study size and power calculations are discussed in the study protocol(196). In brief, the power estimates were dependent on the prevalence of *BRCA*-mutation carriers. The initial power analyses were based on a cohort of 2000 patients; if the prevalence rate of *BRCA1*-mutation carriers was 10%, a study with 200 *BRCA1*-mutation carriers has 97% power to detect a difference in 2-year event rate of 20% in gene-mutation carriers vs. 10% for wild type carriers. Based on preliminary mutation testing from 120 early participants in 2006, the recruitment target was accordingly revised to 3000 patients.

### 2.2.2 Metastatic cohort definitions

For the work included in Section 3.2.1, the dnMBC cohort consisted of patients with confirmed distant metastatic disease at diagnosis (stage M1) according to the local site. Patients who initially had localised disease (stage M0) but developed distant metastatic disease within the follow-up period comprised the rMBC cohort. Tissue diagnosis of metastatic disease was not mandated by the study. Patients without metastatic disease at any time were not included in this analysis.

### 2.2.3 Metastatic cohort statistics

Statistical analyses were performed according to a pre-specified statistical analysis plan (Appendix B) as per STROBE guidelines(203).

The primary objective was the comparison of overall survival (OS) of patients with dnMBC with that of patients with rMBC with a MFI of less than 12 months (early12). OS was defined as the time from date of diagnosis to death from any cause. MFI was defined as the time from date of diagnosis to date of first distant relapse.

Secondary objectives included the comparison of OS and post distant relapse survival (PDRS) of patients with dnMBC with that of patients with rMBC with a MFI of less than 24 months (early24). PDRS was defined as time from date of diagnosis of first distant metastases (date of diagnosis of primary tumour for patients with dnMBC) to death from any cause. Other secondary objectives included the comparison of PDRS of patients with dnMBC vs. early12 patients and the description of clinicopathological features in patients with dnMBC and those with rMBC in four cohorts (recurrent disease within 12 months, within 24 months, between 24 and 60 months, and after 60 months). Patient and tumour

characteristics included ethnicity, body mass index (BMI), germline *BRCA* status, first site of metastasis and primary tumour grading/receptor status.

Time to event outcomes were described using Kaplan-Meier curves and analysed using Cox regression models; stratified Cox models or flexible parametric survival models were used in cases where hazards were time varying. For Kaplan Meier analysis, patients were censored at date of last follow-up or death from any cause. All multivariable analyses were adjusted for age at diagnosis, BMI, grade, tumour size, pathological N stage, ethnicity, and ER and HER2 tumour status. Further objectives included the comparison of OS of dnMBC patients who had surgery (either breast conserving surgery, nodal surgery only or mastectomy) vs. those who had no surgery and assessment of correlation between MFI and PDRS in rMBC patients using the survcorr command in R. Statistical analyses were carried out using Stata v15.1 and RStudio v1.1.456.

All statistical tests were two-sided and a *p*-value less than 0.05 were considered significant.

### 2.2.4 Early-stage TNBC cohort definition

For the work included in Section 3.2.2 and Chapter 4 and 5 (the "early stage TNBC" cohort; see CONSORT diagram in Figure 4.1), only patients with a confirmed diagnosis of TNBC (categorised as per Section 2.2.1) and stage I-III disease were included. Patients who received neoadjuvant chemotherapy were excluded in order to homogenise the cohort and due to the complexity of assessing pre- and post- neoadjuvant tumour samples (there are no TILs to assess in a case of pCR; in many cases both samples were not available for comparison). H&E slides and FFPE tissue was sought for all eligible patients.

### 2.2.5 Statistical analysis of early-stage TNBC cohort

#### 2.2.5.1 Analysis of clinicopathological features and molecular panel

OS was defined as the time from date of diagnosis to death from any cause. Time to event outcomes were described using Kaplan-Meier curves (with log-rank Mantel Cox test) and analysed using Cox regression models. The proportional hazards test was used to identify factors that were not proportional. Backwards elimination in a Cox regression model was used to determine which factors to include in multivariable analyses; these were: pathological nodal stage and stratification by LVI. It was necessary to ensure there was no difference between the baseline clinicopathological features of the TIL-scored cohort and the TNBC cohort as a whole (including those who were not scored because pathology material was not available), in order to rule out selection bias for the TIL-scored cohort. Each category was therefore compared between the two groups. The Chi-square test was used for categorical variables and the Mann-Whitney test for continuous variables. Correlation between TIL score and clinicopathological factors was sought, using Pearson's correlation coefficient (for ordinal variables) and SPSS non-parametric median test (for categorical variables). Statistical analyses were carried out using SPSS v26.

#### 2.2.5.2 ROC analysis for survival prediction model

To evaluate the prognostic impact of various factors, three-year overall survival was used. This survival "cut off" was used, as there were an insufficient number of events occurring by one-year to generate statistical significance. In TNBC the highest risk of recurrence occurs in the first three years(204); by five years there were a larger number of cases censored (51 cases of the total cohort, compared to 14 cases at three years) and the data from these cases would be omitted from the analysis, reducing the power. The detection rate (sensitivity) and false-positive rate (1-specificity) were calculated. The detection rate

is the proportion of patients who died at three years with a positive marker, while the false-positive rate is the proportion of patients who did not die at three years with a positive marker. Likelihood ratios were obtained (detection rate/false-positive rate), indicating the power of an individual marker. A prognostic model was then developed using binary logistic regression, with a backwards selection procedure. The initial variables entered included: TIL percentage, TIL category, CD3, CD4, CD8, FOXP3, CD4:CD8 ratio, FOXP3:CD8 ratio, GLUT1, smooth muscle actin (SMA; a marker of cancer-associated fibroblasts), p53, Ki67, ALDH1, CK5/6, EGFR, MHC, PDL-1 (lymphocytes), age at diagnosis, BMI, clinical T stage, pathological T stage, maximum invasive tumour size, maximum overall tumour size and number of pathological lymph nodes. The regression parameter estimates were used to create a prognostic score. ROC analysis was performed and the area under the curve was calculated.

All statistical tests were two-sided and a *p*-value less than 0.05 were considered significant, unless otherwise specified.

### 2.2.6 TMA construction and scoring

Tissue microarrays (TMAs) were constructed using triplicate, randomly selected within tumour border, paraffin-embedded 1-mm tumour cores (Alphelys Minicore 2, Mitogen, Harpenden, UK). Automated immunostaining (Autostainer Link 48, Agilent Dako, Santa Clara, California, USA) was performed by Maria Monette Lopez in a CPA-accredited clinical cellular pathology department using antibodies optimised to national diagnostic standards (NEQAS). Antibody expression was scored in triplicate by a single consultant breast histopathologist using a Zeiss AxioCam MRc5 microscope (Zeiss, Cambridge, UK) and mean calculated. Ki-67 and p53 were scored as a percentage in 5% increments. ALDH-1 (stromal) and EGFR were scored as negative/low-staining (0), moderate staining (1) or high (2) staining. CK5/6 was scored as either positive or negative, following a semi-quantitative appraisal. For SMA, CD3, CD4, CD8, FOXP3, MHC, GLUT1 and ERAP1,

expression was scored as negative/low-staining (1), moderate staining (2) or high staining (3). PD-L1 was marked separately for tumoural and lymphocyte expression.

### 2.2.7 Assessment of TILs

Full-face H&E sections were assessed for stromal TIL infiltration by a single consultant breast histopathologist using a Zeiss AxioCam MRc5 microscope (Zeiss, Cambridge, UK). The scoring method followed guidelines published by the International TILs Working Group(31)**.** In brief, all mononuclear cells in the stromal compartment and within the borders of the invasive tumour were evaluated. The edges of the tumour and areas of necrosis were avoided and polymorphonuclear leukocytes were excluded. To allow for intratumoural heterogeneity, 10 high power fields (x40) were scored in 5% increments and then an average count taken. The pathologist was blinded to clinical information.

### 2.2.8 Categorisation of immune and molecular markers

For the purposes of survival analysis, biomarker expression was categorised into low and high. Thresholds were determined from literature review. For p53, high was determined as ≥10% of tumour cells staining positive. In a study of 7739 patients with all subtypes of breast cancer, a ROC analysis identified 10% as the optimal cut-off for IHC staining of p53 predictive for survival (205). This cut-off has also been used in TNBC(206). The association of p53 with survival was also evaluated in two categories based on mutant pattern as previously defined(207): mutant (0% or >50% positive) or wild-type (1-49% positive). For Ki67, a cut-off of ≥20% was used in line with other studies(208-210). EGFR and CK5/6 are considered positive if ≥1% cytoplasmic or membranous tumour cell staining is positive(211, 212). For ALDH1, negligible staining (≤0.333 out of 2) was considered absent/low, as per previous work in breast cancer(213, 214). For TILs, survival was

evaluated according to whether the case was TIL high (lymphocyte predominant breast cancer) (>60%), moderate (11-59%) or low (≤10%) as previously defined(39).

For the immune markers used in Chapter 5, there were no clearly defined thresholds in the literature. Therefore, the median score for each marker was determined and the median was then used as the threshold between high/low.

In order to assess the impact of being classified as basal, cases were categorised by their expression of the markers EGFR and CK5/6. Both markers were assessed as above: positive if ≥1% cytoplasmic or membranous tumour cell staining is positive(211, 212). Basal cancers were defined as those that were positive for either marker. In order to determine whether cases were "high basal", they were also categorised based on a mean intensity score of ≥1.5/3 for EGFR and ≥0.5/1 for CK5/6. These thresholds were based on work showing ≥50% EGFR or CK5/6 staining to be prognostic in TNBC(215, 216). "High basal" cases were defined as those that were high-positive for either marker.

### 2.2.9 Digital analysis

#### 2.2.9.1 Image acquisition

For TMA analysis, image acquisition of whole slide images of TMA cores of TNBC tumours from the POSH cohort was performed by Hayley McKenzie using a Zeiss AxioScan Z1 (Zeiss, Cambridge, UK) at resolution equivalent to x20 magnification.

For multiplex analysis, image acquisition of whole slide images of TNBC tumours from the POSH cohort was performed by Monette Maria Lopez using a Zeiss AxioScan Z1 (Zeiss, Cambridge, UK) at resolution equivalent to x20 magnification.

For mouse tumours, image capture of select fields was performed at x20 magnification by Gareth Thomas using an Olympus CKX41 microscope (Olympus, Tokyo, Japan) using cell^ B imaging (Olympus, Tokyo, Japan) software.

#### 2.2.9.2 Image processing

Digital image analysis of TMAs, multiplex slides and mouse tumours was performed using Definiens Developer XD version 2.7 (Munich, Germany). Analysis is composed of three distinct stages: Tissue Identification – separation of tissue and glass regions at resolution equivalent to x5 magnification; Exclusion of Artefacts – identification of necrotic and blood regions, and any folds or staining anomalies; and Cellular Analysis – identification of nuclear staining and segmentation of cellular structures.

Two methods of stain identification were employed for each batch of slides (with each individual stain of all tissues treated as a single batch): classic thresholding based on the source stain; and background exclusion using sliding window filtering. Classic thresholding is based on identification of significant increases and peaks in a histogram derived from centiles of the stain positive region (each centile gives a threshold that identifies the given percentage of pixels with the highest values). The background exclusion method uses two modified representations of the source stain; a *coarse* grained representation generated using a 251x251 pixel window, which removes the local peaks in staining giving the overall stain character, and a *fine* grained representation (7x7 pixel window), which shows local peaks. Subtraction of *coarse* from *fine* gives a normalised representation of local peaks.

Following identification of staining, conjoined nuclei were separated using size and morphology based constraints.

For mouse tissues, there was a high degree of non-specific staining so an enhanced background reduction method was employed to identify true positive staining. A coarsely filtered (moving average in a >151 pixel square) image is generated to represent the background stain level, smoothing out peaks and troughs. This is then subtracted from each original pixel value, leaving only local maxima with a value greater than background(217).

#### 2.2.9.3 TMA analysis

The whole slide images containing the TMAs were segmented automatically, followed by manual checks to ensure all cores were accurately separated, and then assigned to a grid system for reporting.

For survival analysis using digital automated scores, the median automated score was calculated. The median was then used as the threshold between high/low.

#### 2.2.9.4 Multiplex analysis

The multiplex image sets comprised a consistent counterstain and tumour marker stain (haematoxylin and DAB stained pancytokeratin (CK), respectively) plus temporary AEC stained secondary markers. Images were registered using the DAB stain in ImageJ (National Institute of Health, Bethesda, Maryland, USA). This produced a pseudo fluorescent representation of the multiplex AEC staining. Following identification of

nuclear objects using the original brightfield image, cellular regions were described by expanding each nucleus to a distance of ~40µm from the nucleus border. These cellular objects were then classified as positive or negative for the various multiplex stains. Distance from tumour (DAB staining) was also measured and exported.

For the spatial analysis, nine cases were randomly selected from the following two categories: TIL-high (>15%) and TIL-low (≤15%).

The proportion of positive cells was determined by dividing the cell count for a particular stain combination (e.g. CD8+/SMA-ve) by the total cell count. For the inside tumour to outside tumour comparison, the proportion of positive cells within the tumour border (as determined by cytokeratin) was divided by the number of positive cells outwith the tumour border. For comparison of inside:outside ratios between two groups, an unpaired t-test with Welch's correction was performed.

## 2.3 Bioinformatics

For the bioinformatics analysis of TCGA, weighted-gene co-expression network analysis (WGCNA) was applied to TCGA RNA-seq data from triple-negative breast cancer, as described previously(218). Briefly, after data pre-processing, a measure of topological overlap was calculated based on pairwise correlations between 14,958 expressed genes across 66 samples. Within this dissimilarity matrix, 89 modules were identified by unsupervised clustering. To assign cell type to modules, each gene signature was statistically evaluated by Fisher's exact test for overlap with gene signatures from Immgen and xCell databases(219, 220).

## 2.4 In Vitro Fibroblasts and Tumour Cells

### 2.4.1 Prevention of contamination

In order to reduce the risk of contamination with other DNA/RNA samples, a number of precautions were taken. All PCR work took place within a dedicated laminar flow cabinet. The cabinet was cleaned with RNaseZap™ (Thermo Fisher Scientific, Swindon, UK) prior to use. All pipettes, pipette tips and other consumables were exposed to UV radiation within the PCR hood prior to commencement of work. All pipette tips had aerosol filters and were PCR-grade. Nuclease-free, PCR-grade water was used. A clean lab coat was worn and fresh gloves were used prior to DNA purification and then another fresh set prior to the PCR reaction. All cell line work took place in a class II laminar flow hood using aseptic technique.

### 2.4.2 Origin of cells

Breast fibroblasts were outgrown from mammary tissue of wild type Balb/c female weaning mice. The isolation technique is described in 2.4.4.

Balb-neu-T murine cancer-associated fibroblasts were kindly donated by Kirsty Ford. They had been outgrown from spontaneous mammary tumour tissue of Balb/c mice transgenic for the rat Her-2/neu gene (Balb-neu-T). The mice were kindly donated by Natalya Savelyeva.

4T1 cells are a 6-thioguanine resistant transplantable epithelial breast cancer cell line derived from Balb/c mice. This cell line is known to be highly oncogenic and invasive; it is capable of spontaneous metastasis(221). Frozen vials were kindly donated by Kirsty Ford,

previously donated by Professor John F Marshal from Bart's cancer institute, Queen Mary University of London.

### 2.4.3 Cell culture

D-10 medium was prepared by adding 5 mls L-glutamine, 5mls penicillin/streptomycin and 50mls FCS to a 500ml bottle of High-glucose DMEM.

D-20 medium was prepared by adding 5 mls L-glutamine, 5mls penicillin/streptomycin and 100mls FCS to a 500ml bottle of High-glucose DMEM.

All medium was sterile filtered when first prepared using a 0.22uM filter (Millipore). All media was stored at 4°C for a maximum of one month.

25mls of media was added to a T-75 culture flask (Corning, New York, USA). Cells were cultured in a 37°C incubator.

Fibroblasts and CAFs were routinely cultured in D-20 and 4T1 cells in D-10.

Cells were split when confluence reached 70-80% and the medium was replaced every 2-4 days depending on the cell type. To split the cells, 15mls of media was transferred into a universal and the remaining 10mls disposed. The flask was washed with 5mls of PBS. In order to lift the cells off from the flask, 5mls of EDTA-Trypsin was added. The flask was placed in the 37°C incubator for between 2 and 10 minutes, until a single cell suspension was formed. The suspension was then pipetted into the universal containing 15mls of media. The universal was placed in the centrifuge at 200 x g for 5 minutes. The supernatant was poured off then the cells resuspended. 10mls of PBS was added. If required, a cell count was performed by adding 15µl of cells to 15µl of Trypan-Blue, then 10µl loaded onto a haemocytometer and viewed under the microscope. 1mls of cell suspension was added to 24mls of D10 medium in a T-75 flask and placed back in the 37°C incubator.

### 2.4.4 Isolation of cells

Mammary tissue from Balb/c female weaning mice was dissected and cut into small pieces in sterile PBS with 4% penicillin/streptomycin and 0.25ug/ml amphotericin B. These pieces were washed three times with this PBS in a class II laminar flow hood. A 12 well tissue culture plate was scored with a scalpel to create a groove in the centre of each well in the shape of a cross. A small piece of tissue was placed on the centre of this cross using tweezers. 750µl of D-20 was added to each well around the periphery to keep the tissue in place. The entire plate was then left in a hypoxic (3% $O_2$ /5% $CO_2$) (37°C) incubator for 5 days. After this time the media was changed every 72 hours. After approximately 14 days, cells were growing in the media and the antibiotic concentration was lowered to 2%. When confluence reached 75-90%, the wells were washed with PBS and then detached with 0.05% trypsin/EDTA solution. These wells were aspirated then added to growth medium to neutralize the trypsin solution. They were centrifuged at 200 x g for 5 minutes. They were then added to a T75 tissue flask (Corning, New York, USA) in 15mls of D-20.

Once a sufficient number of cells were grown in culture, they were cryopreserved and stored in liquid nitrogen.

### 2.4.5 Cell defrosting

After removing vials from liquid nitrogen, they were transferred on dry ice to a 37°C water bath. They were thawed rapidly by agitation in the bath then aspirated and added to a T-75 flask containing pre-warmed medium. The flask was left overnight in a 37°C

incubator. After 18-24 hours, the medium was changed to remove all remaining DMSO, which can be toxic to cells when defrosted.

### 2.4.6 Maintenance of cells in culture

Medium was changed every 72 hours while cells were in culture. Once cells reached 75-90% confluence, medium was removed and the cells were washed with PBS. 0.05% trypsin-EDTA solution was added, then the cells left at 37°C for approximately 2 to 5 minutes until detached from the flask. They were then aspirated and added to warmed medium to neutralize the trypsin solution. They were then centrifuged at 200 x g for 5 minutes, then resuspended in medium. They were then diluted in further medium to allow splitting into new sterile flasks. For fibroblasts/CAFs, this was usually 1 in 3; for cancer cells this was usually 1 in 5.

### 2.4.7 Freezing of cells

In order to maintain cell stocks, or if cells were reaching too high a passage number, cells would be frozen. They were washed, trypsinised and centrifuged in the same method as the previous section. However, they would be resuspended in freezing medium (cooled to 4°C) containing 90% FCS and 10% DMSO. DMSO was used as it is a cryoprotectant, in order to reduce ice crystal formation, which can result in cell membrane rupture with freezing and thawing. $1x10^6$ cells were added to 1ml of freezing medium in a croyovial. They were frozen in a Mr. Frosty™ freezing container (Thermo Fisher Scientific, Swindon, UK) filled with isopranolol at -80°C overnight. The following day they were transferred to liquid nitrogen and stored indefinitely.

### 2.4.8 Senescing cells

To induce senescence confluent fibroblasts were detached and irradiated in suspension with 10Gy ϒ-rays and plated at sub-confluent densities. Cells acquired a fully senescent phenotype 4-6 days post-treatment.

### 2.4.9 Metabolism Assay

The MTT assay is used to measure cellular metabolic activity as a correlate of cellular viability, proliferation and cytotoxicity. This colorimetric assay is based on the reduction of a yellow tetrazolium salt (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide or MTT) to purple formazan crystals by metabolically active cells. First, cells were plated on a 96-well plate. Four wells were reserved per condition for MTT assay. After culture, at the time of the assay the supernatant was aspirated off. 20µl of MTT reagent (Promega, Southampton, UK) and 100µl D20 medium were added to each well. The plate was left to incubate in the 37°C incubator for three hours. The plate absorbance was then analysed on a Varioskan™ Flash (Thermo Fisher Scientific, Swindon, UK) microplate reader at 490nm.

### 2.4.10 Cell Viability Assay

To further assess proliferation, cells were counted using a CASY® counter (Roche-Innovatis, Bielefeld, Germany). This works by exposing the cells to a low voltage; if the membrane is intact the voltage does not pass through the cell. Thereby the machine is able to measure the number of viable cells by detecting the amplitude of electrical signal.

Cells were plated on a 96-well plate. Two wells were reserved per condition for viability assay. After culture, at the time of the assay the supernatant was aspirated off. The wells were washed with PBS then 100µl of trypsin added to each well until effectively trypsinised. For each well, all 100µl of cells were added to 9.9mls of CASYton (Roche-Innovatis, Bielefeld, Germany) (an electrolyte for cell suspension). The cell count was then performed on the CASY counter, using two replicates per well.

### 2.4.11 Measurement of nucleic acid concentrations

The concentration of all DNA and RNA samples was measured by analysing a 1µl sample on a NanoDrop™ 1000 spectrophotometer (NanoDrop Technologies, Inc. Wilmington, USA), according to the manufacturer's instructions. As nucleic acids have an absorbance maximum at 260nm, the ratio of this maximum to the absorbance at 280nm is used as a measure of DNA/RNA purity. A ratio of 1.8 is accepted as pure for DNA and 2.0 as pure for RNA.

### 2.4.12 RT-quantitative PCR from 6-well culture plate

#### 2.4.12.1 RNA extraction

6-well plates requiring extraction of RNA were washed with PBS, then placed on dry ice, then transferred to the -80°C freezer. RNA extraction was then performed using the RNeasy mini kit (Qiagen, Manchester, UK).

30µl of beta-mercaptoethanol was added to 3mls of Buffer RLT. 350µl of the RLT mixture was added to each well and the cell scraper used to lift the cells from the plate surface.

350µl of 70% Ethanol was then added to each well and pipetted vigorously to lyse cells. The mixture was then pipetted into a spin column. This was placed in the centrifuge for 15 seconds at 8,000 x g and the flow-through discarded.

For the DNAse step, 350µl of Buffer RW1 was added to the spin column and placed in the centrifuge for 15 seconds at 8,000 x g and the flow-through discarded. 10µl of DNAse was added to 70µl of Buffer RDD, mixed and then all 80µl added to each spin column. This was incubated at room temperature for 15 minutes. 350µl of Buffer RW1 was then added then the column was placed in the centrifuge for 15 seconds at 8,000 x g and the flow-through discarded.

500µl of RPE was added and then the column was placed in the centrifuge for 15 seconds at 8,000 x g and the flow-through discarded. 500µl of RPE was added and then the column was placed in the centrifuge for two minutes at 8,000 x g and the flow-through discarded. The column was placed in a new Eppendorf and then placed in the centrifuge for one minute at 8,000 x g. The column was placed in a new Eppendorf again and 20-25µl of RNAase-free water added. This was then placed in the centrifuge for one minute at 8,000 x g and placed on ice. The concentration was checked using a NanoDrop™ 1000 spectrophotometer (NanoDrop Technologies, Inc. Wilmington, USA). If not being used immediately, the RNA was stored at -80°C.

#### 2.4.12.2 cDNA synthesis

Using a template, the volume of RNA and Millipore water required per sample was calculated. These volumes were then added to a small PCR tube, along with 4µl of 5xrt buffer, 2µl of 20x dNTP, 1µl 20x random primer and 1µl of multiscribe enzyme from the RevertAid First Strand cDNA synthesis kit (Thermo Fisher Scientific, Swindon, UK). The cDNA was placed in a PTC-200 thermal cycler (MJ Research) at 25°C for five minutes 42°C

for 60 minutes and 70°C for five minutes then cooled to 4°C. The cDNA was then stored at -80°C or used immediately for rtPCR.

### 2.4.12.3 Quantitative PCR

Primers were kindly donated by Kirsty Ford; she had previously designed them using Primer Blast online software. The oligonucleotide sequences were manufactured by Sigma Aldrich (Gillingham, UK), resuspended to 100µM in $H_20$ (autoclaved) and stored at -20°C (Table 2.1). cDNA was diluted to 12.5x by adding 5.6µl of cDNA to 64.4µl PCR-grade water. 3 wells were made per target per sample. In a 96-well PCR plate, in each well, 5µl of diluted PCR was added to 12.5µl SybrGreen (10x stock) (Life Technologies, Thermo Fisher Scientific, Swindon, UK) and the desired working concentration of each primer. The remaining volume was made up of PCR-grade water to a total volume of 25µl per well. The plate was then run on the 7599 Real Time PCR System (Applied Biosystems, Thermo Fisher Scientific, Swindon, UK) at 50°C for two minutes, 95°C for 10 minutes, 95°C for 15 minutes, then 60°C for one hour, for 40 cycles. Relative gene expression was calculated by comparing target gene cycle threshold to housekeeping gene cycle threshold (delta-delta cycle threshold used for analysis). A fold-change value was then calculated by comparing these normalized gene values.

**Table 2-1. PCR Primers Used**

| Gene target | Forward sequence | Reverse sequence | Working [] (uM) |
|---|---|---|---|
| ACTA2 | CCTATGCCATCATGCGTCT | AATCTCACGCTCGGCAGTAG | 0.1 |
| Col1A1 | GTGTTCCCTACTCAGCCGTC | ACTCGAACGGGAATCCATCG | 0.4 |
| Col1A2 | GCTGGTGTAATGGGTCCTCC | CGACCGGCATCTCCATTAGG | 0.2 |
| ATM | CGATATGCCAGTCTTTTCAGGG | CTCAAGGCTGCCCTTACTCA | 0.5 |

| | | | |
|---|---|---|---|
| Fibronectin1 | GAAGACAGATGAGCTTCCCCA | GGTTGGTGATGAAGGGGGTC | 0.1 |
| Beta2M | TGCTATCCAGAAAACCCCTCAA | GGATTTCAATGTGAGGCGGG | 0.05 |

### 2.4.13 Western Blot

To analyse protein from frozen cells in monolayer in wells, cells were first lysed. 100µl of chilled RIPA buffer, containing 1% protease and phosphatase inhibitors (Calbiochem, Merck Millipore, Watford, UK) was added to each well of a 6-well culture plate. A cell scraper was used to remove the cells from the plate surface and then aspirated into an eppendorf and pipetted vigorously to lyse. The eppendorf was left on ice for 30 minutes. The eppendorf was placed in the centrifuge at 16,000 x g at 4°C for 20 minutes. The supernatant (lysate) was then placed in a fresh eppendorf and kept on ice.

In order to calculate the quantity of protein in the lysate, the DC Bio Rad protein assay kit (Bio-Rad, Watford, UK) was used according to manufacturer instructions. 24.5µl of Reagent A, 3µl of autoclaved water and 0.5µl of Reagent S was pipetted into each well of a 96-well plate. Two wells were used per lysate and two wells were used for negative control. 2µl of lysate was added to each well and 2ul of RIPA mix to the negative control wells. 200µl of reagent B was then added to each well and left at room temperature for 15 minutes. The plate was then run on the Varioskan™ plate reader at 650nm (Thermo Fisher Scientific, Swindon, UK).

The Varioskan absorbances (subtracting the negative control wells) were used to calculate a standard quantity of protein from each sample (up to 50µg). The desired volume of each lysate was then made up to 37.5µl with Millipore water in an eppendorf. 12.5µl of NuPage LDS Sample buffer (4X) (Invitrogen, Thermo Fisher Scientific, Swindon, UK) and 5ul of Reducing Agent (Invitrogen, Thermo Fisher Scientific, Swindon, UK) were then added to the eppendorf. The eppendorf was placed on the heating block at 95°C for

five minutes. The eppendorfs were then placed in the centrifuge at 16,000 x g for 30 seconds. Using a 4-12% 10-well polyacrylamide BIS NuPAGE precast gel (Invitrogen, Thermo Fisher Scientific, Swindon, UK) loaded into a gel tank (Bio-Rad, Watford, UK), 10ul of a Page Ruler pre-stained protein ladder (Thermo Fisher Scientific, Swindon, UK) was loaded and then 35µl of the boiled lysate loaded. The gel was run at 60V for 20 minutes, MES SDS NuPage running buffer (Invitrogen, Thermo Fisher Scientific, Swindon, UK) redistributed if required and then increased to 130V for 60 to 90 minutes.

Transfer of the polyvinylidene membrane (Merck Millipore, Watford, UK) was then performed. Methanol was poured onto the membrane to activate it then the membrane was soaked in NuPage transfer buffer (Invitrogen, Thermo Fisher Scientific, Swindon, UK). Sponges and blotting paper were soaked in transfer buffer. The transfer unit was assembled then run overnight at 27V at 4°C.

The membrane was then washed well in TBS-T then soaked in 5% milk in TBS-T at room temperature on the rocker for one hour, then washed in TBS-T. The primary antibody (Table 2.2) was diluted in BSA/tween then the membrane incubated in this for two hours. The membrane was washed three times for five minutes in TBS-T. The secondary antibody was prepared in 2.5% milk in TBS-T at a concentration of 1:5000. The membrane was soaked in the appropriate horseradish peroxidase (HRP) conjugated secondary antibody (Dako, Agilent, Santa Clara, California, USA) (Table 2.2) at a dilution of 1:5000 for one hour at room temperature on the rocker. The membrane was washed three times for five minutes in TBS-T. The membrane was then developed using 400µl each vial from a chemo luminescent substrate, ECL Super Signal West Pico (Thermo Fisher Scientific, Swindon, UK). A signal booster of 100µl of each vial from ECL Super Signal West Femto (Thermo Fisher Scientific, Swindon, UK) was used for weaker antibodies. Photographs were taken using a ChemiDoc-It Imagine System (UVP, Upland, California, USA) and VisionWorks (UVP, Upland, California, USA) software. Quantification of Western Blot bands was performed using the area under the curve function in ImageJ (National Institute of Health, Bethesda, Maryland, USA).

**Table 2-2. Western Blot antibodies**

| Antibody | Manufacturer | Dilution | Secondary (HRP) | Molecular weight (kDa) |
|---|---|---|---|---|
| Collagen1 | Novus | 1 in 1000 | Anti-rabbit | 139 |
| Hsc70 | Santa Cruz | 1 in 1000 | Anti-mouse | 73 |
| SMA | Sigma Aldrich | 1 in 1000 | Anti-mouse | 42 |

### 2.4.14 Immunofluorescence

On a 96-well opaque black plate, cells were cultured. Before each step, the solution from the previous step was aspirated.

Each well was washed with 150µl of PBS three times. They were then fixed in 100µl of 4% paraformaldehyde at room temperature for 20 minutes. Each well was washed with 150µl of 0.1% Triton-X diluted in PBS three times. 0.5% Triton-X was then left on the cells for ten minutes. Ammonium chloride was then left on the cells for 10 minutes. Each well was washed with 150µl of 0.1% Triton-X diluted in PBS three times. 100ul of 2% BSA in 0.1% Triton-X (block) was added to each well for at least 90 minutes. The primary antibody (Table 2.3) was diluted in blocking solution and 100µl added to each well, incubation occurred overnight at 4°C. Each well was washed with 150µl of 0.1% Triton X diluted in PBS three times for five minutes on the rocker. The appropriate fluorescent secondary antibody (Thermo Fisher Scientific, Swindon, UK) (Table 2.3) was diluted to 1:250 in blocking solution and 100µl added to each well in the dark, incubation occurred over 45 minutes at room temperature. Each well was washed with 150µl of 0.1% Triton X diluted in PBS three times for five minutes on the rocker (covered). DAPI was diluted to 1:1000 in PBS then 100µl added to each well in the dark. Each well was washed quickly with PBS twice. 200µl of PBS was then added to each well and the plate stored covered at

4°C or immediately viewed on an Olympus IX81 (Olympus, Tokyo, Japan) fluorescent microscope using Xcellence imaging software (Olympus, Tokyo, Japan).

DAPI nuclear staining allowed the selection of fields of view with similar numbers of cells. Quantification of images was performed using ImageJ software(222) (National Institute of Health, Bethesda, Maryland, USA). Mean intensity was calculated as a mean of four fields of view per variable. In addition, a control condition with cells exposed to secondary antibody only was prepared to ensure specificity of secondary antibody binding.

**Table 2-3 Immunofluorescence antibodies**

| Antibody | Manufacturer | Dilution | Secondary |
|---|---|---|---|
| SMA | Sigma Aldrich | 1 in 750 | Goat anti-mouse 546 (red) |
| Fibronectin | Sigma Aldrich | 1 in 800 | Goat anti-rabbit 488 (green) |
| Collagen1 | Abcam | 1 in 500 | Goat anti-rabbit 546 (red) |

### 2.4.15 siRNA transfection

RNA interference can be used to silence or knock down the expression of a target gene, by inactivating its corresponding mRNA by double-stranded RNA. Here the target mRNA (for ATM) is degraded by short interfering RNA (siRNA).

$5x10^4$ CAFs were plated in each well of a 6-well plate. The plate was left overnight in a 37°C incubator. On day 2, siRNA was prepared according to manufacturer's instructions: 50µl of siRNA Buffer (Thermo Fisher Scientific, Swindon, UK) and 250µl of RNA-se free H20 was added to the siRNA tube (on-target mATM) (Dharmacon, Lafayette, Colorado, USA), then vortexed for 2 minutes. The negative control silencer (Ambion AM4635;

Invitrogen, Thermo Fisher Scientific, Swindon, UK) was prepared in the same way. 200µl of optimem was added to 2.75µl of siRNA or negative control solution for each well and both eppendorfs pipetted vigorously. 3.3µl of oligofectamine (Invitrogen, Thermo Fisher Scientific, Swindon, UK) and 13.2µl of optimem for each well was then added together in an eppendorf. This solution was left at room temperature for 5 minutes. Then 16.5µl per well of the oligofectamine/optimem solution was added to each eppendorf and pipetted. This was then placed in a 37°C incubator for 20 minutes. The medium was then aspirated off the 6-well plates, then 800µl of optimem added to each well. 210µl from the siRNA or negative silencer eppendorfs was then added to relevant wells in a droplet fashion, with swirling, to ensure adequate coverage of all cells with the siRNA.

On day 3, 1ml of 40% DMEM was added to each well in order to prevent prolonged serum starvation.

On day 4, day 2 transfection was repeated.

On day 5, 1ml 1ml of 40% DMEM was again added to each well.

The plates were frozen at -80°C on day 8.

## 2.5 In Vivo Fibroblasts and Tumour Cells

All persons conducting *in vivo* work held a valid Procedure Individual Licence (PIL) License. Accordingly, all attended a mandatory Home Office accredited PIL course. Experiments were conducted under a project license at a licensed establishment, according to Home Office guidelines.

### 2.5.1 The 4T1 model

4T1 breast cancer cells were cultured *in vitro* and harvested by trypsinisation as per Section 2.4.6. 4T1 cells were either injected alone, or in combination with Balb/c breast fibroblasts, with or without *ex vivo* treatment with TGF-β 2ng/ml for one week prior, Balb-neu-T CAFs or Balb/c breast fibroblasts irradiated to make senescent (see 2.4.8). The cells were resuspended in 100µl PBS for injection.

Cells were injected subcutaneously into the top right mammary fat pads of Balb/c mice aged 6-8 weeks. Tumours typically developed 4-6 days after injection and were then measured every 48-72 hours by calipers. Tumour volume was calculated using the formula $4/3\pi xr^3$, described previously(223), where r is the radius and is calculated as an average of length and width. Mice were culled when the tumour size exceeded 15x15mm in diameter or tumours were ulcerated for more than five days (considered humane end points). Upon culling, the tumours were excised and divided in half with a scalpel. One half was placed in 4% PFA (Sigma Aldrich, Gillingham, UK) and the other half was snap frozen in OCT (CellPath, Powys, UK) in foil packets on dry ice. The latter were then stored at -80°C.

### 2.5.2 AZD0156

Captisol (Glentham, Corsham, UK) solution was prepared by adding 12g to sterile $H_20$ to a volume of 40mls. This was vortexed and then placed on the roller for 5 minutes or until it became transparent. It was then filtered using a 0.22uM filter (Merck Millipore, Watford, UK). This solution was kept at 4°C for up to one month. For the control solution (without the drug), 1mls of DMSO was added to 9mls of 30% captisol and stored at 4°C. Prior to administration, the solution was stirred at room temperature for one hour.

The ATM inhibitor AZD0156 (Astrazeneca, Cambridge, UK) was prepared by adding 28.9mg to 656.82µl of DMSO, vortexing for one minute then sonicating for 30 mins at

level 7, 5 seconds on, 2 seconds off. DMSO solvent was added to increase the solubility of the drug in captisol solution. The DMSO-AZD0156 mixture was then diluted in captisol to the desired concentration, to administer 100µl by oral gavage to each mouse, e.g. for 20ml/kg, the mixture was added to 5911.36mls of captisol, with each 100µl dose containing 0.44mg of AZD0156. This solution was stored while being stirred at 4°C for one week. Prior to administration, the mixture was stirred at room temperature for one hour.

### 2.5.3 CTLA4

Intravenous anti-CTLA4, InVivoPlus BP0164 (Bio X Cell, Lebanon, New Hampshire, USA) was given at a dose of 200µg (20mg/kg) in 200µl PBS, every 48 hours for 3 doses from the day tumours became palpable. The dilution was stored at 4°C.

### 2.5.4 Immunohistochemistry of mouse tissues

#### 2.5.4.1 Frozen OCT tissue

The frozen OCT tumours in foil were placed on the cryostat (Microcom HM500) and 5-10µm sections cut through the centre of the tumour onto Super Frost slides (Thermo Fisher Scientific, Swindon, UK). These were left overnight to dry. All subsequent steps took place at room temperature. The slides were then fixed in 100% acetone for 10 minutes. They were then air dried for 2 minutes then an Immedge hydrophobic barrier pen (Thermo Fisher Scientific, Swindon, UK) used to demarcate the tissue. The samples were then washed three times with PBS containing 0.05% Tween 20 (PBS/Tw). 100µl of peroxidase suppressor (Thermo Fisher Scientific, Swindon, UK) was applied for 15 minutes. They were again washed, twice with PBS/Tw. They were then blocked with 2.5% normal goat serum diluted to 2.5% in PBS/Tw for 30 minutes. The blocking buffer was

then tapped off and 100µl of primary antibody (Table 2.4), diluted in PBS/Tw, applied. This was incubated for two hours. The slides were then washed three times with PBS/Tw. One drop of the appropriate HRP:Polymer (ImmunoPress, Vector Laboratories, Peterborough, UK) was then applied to sections for 30 minutes. The tissue was again washed three times with PBS/Tw. Vector NovaRed chromagen (Vector Laboratories, Peterborough, UK) was formulated according to the manufacturer's instructions and one drop applied to each section for up to 10 minutes. The slides were washed with PBS twice and then placed into a slide pot, then washed in tape water twice. The pot was then immersed in haematoxylin (Vector Laboratories, Peterborough, UK) for 30 seconds, then again washed in tap water twice. The slides were then dehydrated in alcohol sequentially: 90%, then 100%. They were then cleared in Histoclear (National Diagnostics, Atlanta, Georgia, USA). A small drop of Vectamount mounting solution (Vector Laboratories, Peterborough, UK) was then applied to a coverslip, which was then lowered down onto each slide. The slides were left to dry overnight and thereafter stored indefinitely at room temperature.

Sections were viewed on an Olympus CDKX41 (Olympus, Tokyo, Japan) microscope using cell^B imaging (Olympus, Tokyo, Japan) software.

**Table 2-4. Immunohistochemistry antibodies for mouse tissues**

| Antibody | Manufacturer | Dilution |
|---|---|---|
| CD8 | In House | 1 in 3000 |
| NK p46 | Biolegend | 1 in 50 |
| F4/80 | Biorad | 1 in 100 |

#### 2.5.4.2 PFA-fixed tissue

Tissue fixed in 4% PFA was sent to Research Histology, Pathology, Southampton General Hospital, for sectioning and staining for SMA. This was performed by Monette Maria Lopez.

#### 2.5.4.3 Images and quantification

Image capture of select fields was performed at x20 magnification by Gareth Thomas using an Olympus CKX41 microscope (Olympus, Tokyo, Japan) using cell^B imaging (Olympus, Tokyo, Japan) software. Quantification was performed using ImageJ software (National Institute of Health, Bethesda, Maryland, USA), with appropriate thresholding and calculated from a control stained tissue.

### 2.5.5 Pre-clinical Statistics

For analysis within Section 5.2.4.2 and Chapters 6 and 7, GraphPad Prism version 8 was used to perform statistical analysis and produce graphs. The number of repeats performed is given in each individual set of data. For the analysis of multiple groups together, a one-way ANOVA was performed. For the comparison between two groups, an unpaired two-tailed t-test was performed. Where control values were fixed across experiments (e.g. 1 for qPCR fold change), then unpaired t-tests were performed against control. When comparing tumour volumes, an AUC comparison was performed for each group, in addition to a two-way repeated measure ANOVA over time.

Unless otherwise specified, figures portray means with error bars displaying the standard error of the mean (SEM). Significance p-values were denoted by: ns=≥0.05, *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001. Specific p values are written when values fall just below the 5% significant level.

## 2.6 i-c-TNBC

During the course of my work here, I designed a trial, “i-c-TNBC”, to facilitate the recruitment of newly diagnosed TNBC patients, in order to collect blood and tumour samples to enable genomic work. A secondary purpose was to provide a validation cohort for the analyses performed here. In order to set the study up, I wrote the protocol (Appendix A) and obtained ethical approval. The i-c-TNBC study is a non-interventional, prospective, observational cohort study at a single site. Patients were recruited from the breast cancer chemotherapy clinic at University Hospital Southampton NHS Foundation Trust between 2016 and 2018. It should be noted that recruitment was suspended for a period of 12 months during maternity leave. Only patients with stage I-III disease were included. All patients were recruited prior to the start of systemic anti-cancer therapy. Data regarding patients’ diagnosis, treatment and clinical outcome was recorded and anonymised. Access to diagnostic histopathological H&E slides for TIL review and use of excess tissue block specimens was granted. Ethical approval was granted by HRA (16/YH/0385).

# Chapter 3 Clinical Prognostic Factors in Young Breast Cancer

## 3.1 Introduction

As discussed in the introduction, TNBC has a higher likelihood of disease recurrence and death compared to other subtypes of breast cancer. There are currently no NICE-approved targeted therapies. The incidence of breast cancer in young women (aged less than 40) is low, but increasing(17). Young women are more likely to have breast cancer with adverse biological features, including higher grade, absence of hormone receptors, lymph node involvement and vascular invasion(196). Young age has been consistently shown to be an adverse prognostic factor, with a higher risk of distant recurrence, independent of other factors (17, 224). There is limited published data on the prognostic features of young women with TNBC (a particularly adverse combination) from prospective studies in the literature.

Although young women present more frequently with stage III breast cancer than older patients, de novo metastatic breast cancer (dnMBC) is found infrequently (1% of those aged under 40 in one retrospective study) (225). The POSH study, a prospective observational study of almost 3000 patients aged 40 years or younger with a first diagnosis of invasive breast cancer (196), provides a unique opportunity to study the natural history of dnMBC in young women. Initially I aimed to characterise the clinical features, pattern of disease progression and survival of young breast cancer patients who present with metastatic disease (dnMBC), compared to those who later develop distant metastases after initially localised disease (recurrent MBC; rMBC), in this large prospective cohort genotyped for germline *BRCA1/2*. With particular relevance to the work presented here, I investigated the outcomes stratified by whether they had TNBC or a non-TNBC subtype.

Given the poor outcomes for TNBC, I also undertook a detailed clinicopathological assessment of early stage TNBC in this cohort of young women. Parameters included were ethnicity, BMI and *BRCA* status, in addition to traditional pathological prognostic factors such as grade, tumour size and nodal staging. I have correlated these clinicopathological factors with survival, in order to ascertain prognostic features in this cohort.

This chapter therefore comprises a comprehensive analysis of clinicopathological features and their association with survival in young women with metastatic breast cancer and with early-stage TNBC.

## 3.2 Results

### 3.2.1 Metastatic cohort

#### 3.2.1.1 Characteristics of the metastatic dataset

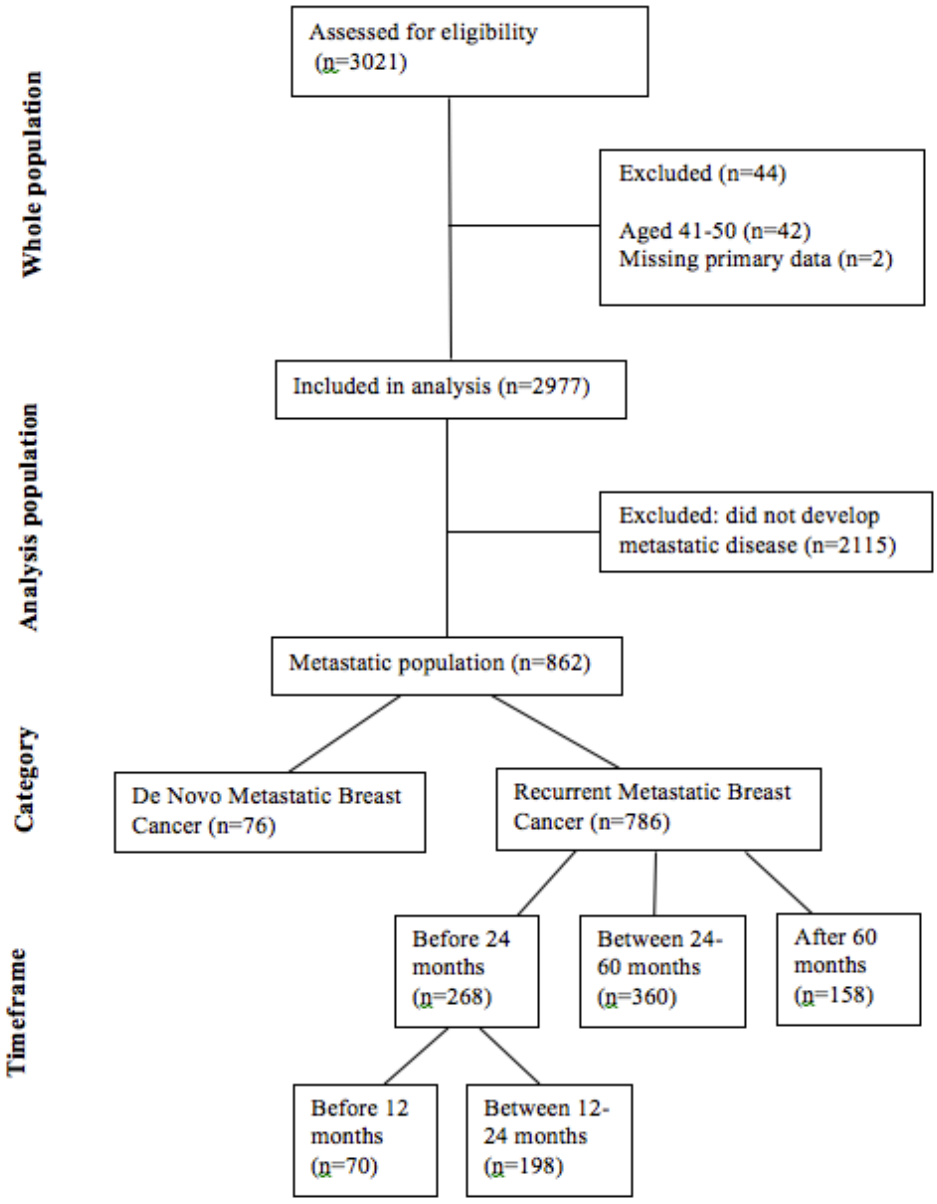



**Figure 3-1. CONSORT diagram for the metastatic cohort.**
**Includes analyses conducted on follow-up data received until 26 June 2016.**

A total of 3021 eligible women were recruited to the POSH study. For this study, 44 women were excluded (42 were aged 41-50 years and 2 had missing primary tumour data). Those aged 41-50 were included in the original POSH criteria if they had a known BRCA mutation, in order to enrich for mutation carriers. They were not however included in outcome analyses to date(197). Of the 2977 women included, 862 (29.0%) were diagnosed with metastatic disease and comprise the analysis population. There were 76 women (2.6%) who presented with dnMBC. To date, the distant recurrence rate amongst the 2901 women with localised disease at presentation is 27.1% (n=786) (CONSORT diagram in Figure 3.1).

Of patients with rMBC, there were 70 women (8.9%) who developed metastases within 12 months of diagnosis (early12). There were 268 women (34.1%) who developed metastatic disease within 24 months of their first diagnosis (early24), 360 (45.8%) within 24 to 60 months (early24-60) and 158 (20.1%) after 60 months (late). Median follow-up of the analysis population was 11.00 years (95% CI 10.79 to 11.59; n=862).

#### 3.2.1.2 Demographics and clinicopathological features of the metastatic cohort

For the 862 women diagnosed with metastatic disease at diagnosis or during follow-up, clinicopathological data is presented in Table 3.1. Only one patient with dnMBC had a *BRCA1* mutation (1.3%; 1/76), whereas there was a larger proportion with a *BRCA2* mutation (11.8%; 9/76). However, amongst the early24 patients, 8.6% (23/268) had a pathogenic *BRCA1* mutation whereas only 3.0% (8/268) of them had a *BRCA2* mutation.

On review of tumour characteristics, TNBC was least common in the late relapse group (9.5%) and most common in the early24 group (37.3%); the proportion of TNBCs

decreased with increased time to relapse. TNBC was uncommon in the dnMBC group (10.5%).

**Table 3-1. Demographic table for patients in the POSH cohort diagnosed with metastatic disease at any time.**

**N=862.**

| | **M1 (dnMBC)** (n=76) | **RMBC < 12 months** (n=70) | **RMBC < 24 months** (n=268) | **RMBC 24 to 60 months** (n=360) | **RMBC 60+ months** (n=158) | **Total** (n=862) |
|---|---|---|---|---|---|---|
| **Age (years)** | | | | | | |
| 18 to 25 | 4 (5.3%) | 0 | 4 (1.5%) | 6 (1.7%) | 2 (1.3%) | 16 (1.9%) |
| 26 to 30 | 6 (7.9%) | 13 (18.6%) | 37 (13.8%) | 37 (10.3%) | 8 (5.1%) | 88 (10.2%) |
| 31 to 35 | 27 (35.5%) | 26 (37.1%) | 95 (35.4%) | 117 (32.5%) | 53 (33.5%) | 292 (33.9%) |
| 36 to 40 | 39 (51.3%) | 31 (44.3%) | 132 (49.3%) | 200 (55.6%) | 95 (60.1%) | 466 (54.1%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| **Ethnicity** | | | | | | |
| Caucasian | 67 (88.2%) | 60 (85.7%) | 236 (89.7%) | 328 (91.6%) | 136 (86.6%) | 767 (89.8%) |
| Black | 6 (7.9%) | 7 (10.0%) | 17 (6.5%) | 18 (5.0%) | 13 (8.3%) | 54 (6.3%) |
| Asian | 3 (3.9%) | 3 (4.3%) | 8 (3.0%) | 10 (2.8%) | 7 (4.5%) | 28 (3.3%) |
| Other | 0 | 0 | 2 (0.8%) | 2 (0.6%) | 1 (0.6%) | 5 (0.6%) |
| Total | 76 (100%) | 70 (100%) | 263 (100%) | 358 (100%) | 157 (100%) | 854 (100%) |
| Missing | 0 | 0 | 5 (1.9%) | 2 (0.6%) | 1 (0.6%) | 8 (0.9%) |
| **Body mass index** | | | | | | |
| Underweight | 32 (45.1%) | 30 (44.8%) | 119 (45.4%) | 167 (47.2%) | 80 (51.3%) | 398 (47.2%) |
| Overweight | 21 (29.6%) | 21 (31.3%) | 85 (32.4%) | 105 (29.7%) | 38 (24.4%) | 249 (29.5%) |
| Obese | 18 (25.4%) | 16 (23.9%) | 58 (22.1%) | 82 (23.2%) | 38 (24.4%) | 196 (23.3%) |
| Total | 71 (100%) | 67 (100%) | 262 (100%) | 354 (100%) | 156 (100%) | 843 (100%) |
| Missing | 5 (6.6%) | 3 (4.3%) | 6 (2.2%) | 6 (1.7%) | 2 (1.3%) | 19 (2.2%) |
| **BRCA (BRCA1 or 2) status** | | | | | | |

| | M1 (dnMBC) (n=76) | RMBC < 12 months (n=70) | RMBC < 24 months (n=268) | RMBC 24 to 60 months (n=360) | RMBC 60+ months (n=158) | Total (n=862) |
|---|---|---|---|---|---|---|
| BRCA- | 66 (86.8%) | 62 (88.6%) | 237 (88.4%) | 325 (90.3%) | 138 (87.3%) | 766 (88.9%) |
| BRCA+ | 10 (13.2%) | 8 (11.4%) | 31 (11.6%) | 35 (9.7%) | 20 (12.7%) | 96 (11.1%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| **BRCA1 status** | | | | | | |
| BRCA1- | 75 (98.7%) | 66 (94.3%) | 245 (91.4%) | 346 (96.1%) | 148 (93.7%) | 814 (94.4%) |
| BRCA1+ | 1 (1.3%) | 4 (5.7%) | 23 (8.6%) | 14 (3.9%) | 10 (6.3%) | 48 (5.6%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| **BRCA2 status** | | | | | | |
| BRCA2- | 67 (88.2%) | 66 (94.3%) | 260 (97.0%) | 339 (94.2%) | 148 (93.7%) | 814 (94.4%) |
| BRCA2+ | 9 (11.8%) | 4 (5.7%) | 8 (3.0%) | 21 (5.8%) | 10 (6.3%) | 48 (5.6%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| **Histological Grade** | | | | | | |
| 1 | 2 (2.9%) | 0 | 1 (0.4%) | 10 (2.8%) | 6 (3.9%) | 19 (2.3%) |
| 2 | 23 (33.8%) | 9 (13.0%) | 42 (16.0%) | 123 (35.0%) | 74 (48.7%) | 262 (31.4%) |
| 3 | 43 (63.2%) | 60 (87.0%) | 220 (83.7%) | 218 (62.1%) | 72 (47.4%) | 553 (66.3%) |
| Total | 68 (100%) | 69 (100%) | 263 (100%) | 351 (100%) | 152 (100%) | 834 (100%) |
| Missing | 8 (10.5%) | 1 (1.4%) | 5 (1.9%) | 9 (2.5%) | 6 (3.8%) | 28 (3.2%) |
| **Histological Type** | | | | | | |
| Ductal | 65 (86.7%) | 68 (97.1%) | 239 (89.8%) | 308 (86.8%) | 136 (87.2%) | 748 (87.8%) |
| Ductal and Lobular | 4 (5.3%) | 2 (2.9%) | 7 (2.6%) | 12 (3.4%) | 4 (2.6%) | 27 (3.2%) |
| Lobular | 3 (4.0%) | 0 | 6 (2.3%) | 21 (5.9%) | 14 (9.0%) | 44 (5.2%) |
| Medullary | 0 | 0 | 2 (0.8%) | 2 (0.6%) | 0 | 4 (0.5%) |
| Metaplastic | 0 | 0 | 4 (1.5%) | 0 | 0 | 4 (0.5%) |
| Mixed | 2 (2.7%) | 0 | 4 (1.5%) | 4 (1.1%) | 0 | 10 (1.2%) |
| Other | 1 (1.3%) | 0 | 1 (0.4%) | 8 (2.3%) | 2 (1.3%) | 12 (1.4%) |

| | M1 (dnMBC) (n=76) | RMBC < 12 months (n=70) | RMBC < 24 months (n=268) | RMBC 24 to 60 months (n=360) | RMBC 60+ months (n=158) | Total (n=862) |
|---|---|---|---|---|---|---|
| Unclassified | 0 | 0 | 3 (1.1%) | 0 | 0 | 3 (0.4%) |
| Total | 75 (100%) | 70 (100%) | 266 (100%) | 355 (100%) | 156 (100%) | 852 (100%) |
| Missing | 1 (1.3%) | 0 | 2 (0.7%) | 5 (1.4%) | 2 (1.3%) | 10 (1.2%) |
| **Surgical Margins (mm)** | | | | | | |
| 0 | 5 (12.2%) | 6 (10.0%) | 27 (11.9%) | 33 (12.0%) | 15 (11.7%) | 80 (11.9%) |
| >0 to <1 | 0 | 3 (5.0%) | 7 (3.1%) | 3 (1.1%) | 3 (2.3%) | 13 (1.9%) |
| >=1 to <=5 | 19 (46.3%) | 36 (60.0%) | 115 (50.9%) | 155 (56.4%) | 66 (51.6%) | 355 (53.0%) |
| >5 | 17 (41.5%) | 15 (25.0%) | 77 (34.1%) | 84 (30.5%) | 44 (34.4%) | 222 (33.1%) |
| Total | 41 (100%) | 60 (100%) | 226 (100%) | 275 (100%) | 128 (100%) | 670 (100%) |
| Missing | 35 (46.1%) | 10 (14.3%) | 42 (15.7%) | 85 (23.6%) | 30 (19.0%) | 192 (22.3%) |
| **LVI**[1] | | | | | | |
| Absent | 18 (31.0%) | 15 (23.1%) | 68 (27.4%) | 121 (35.6%) | 60 (41.1%) | 267 (33.7%) |
| Present | 40 (69.0%) | 50 (76.9%) | 180 (72.6%) | 219 (64.4%) | 86 (58.9%) | 525 (66.3%) |
| Total | 58 (100%) | 65 (100%) | 248 (100%) | 340 (100%) | 146 (100%) | 792 (100%) |
| Missing | 18 (23.7%) | 5 (7.1%) | 20 (7.5%) | 20 (5.6%) | 12 (7.6%) | 70 (8.1%) |
| **No. Positive Lymph Nodes** | | | | | | |
| 0 | 11 (23.4%) | 11 (16.2%) | 66 (25.5%) | 96 (26.8%) | 49 (31.4%) | 222 (27.1%) |
| 1 to 3 | 12 (25.5%) | 21 (30.9%) | 89 (34.4%) | 128 (35.8%) | 70 (44.9%) | 299 (36.5%) |
| 4 to 9 | 11 (23.4%) | 19 (27.9%) | 52 (20.1%) | 86 (24.0%) | 30 (19.2%) | 179 (21.8%) |
| 10+ | 13 (27.7%) | 17 (25.0%) | 52 (20.1%) | 48 (13.4%) | 7 (4.5%) | 120 (14.6%) |
| Total | 47 (100%) | 68 (100%) | 259 (100%) | 358 (100%) | 156 (100%) | 820 (100%) |
| Missing | 29 (38.2%) | 2 (2.9%) | 9 (3.4%) | 2 (0.6%) | 2 (1.3%) | 42 (4.9%) |
| **ER status** | | | | | | |
| Negative | 23 (30.7%) | 41 (58.6%) | 146 (54.7%) | 99 (27.6%) | 25 (15.8%) | 293 (34.1%) |
| Positive | 52 (69.3%) | 29 (41.4%) | 121 (45.3%) | 260 (72.4%) | 133 (84.2%) | 566 (65.9%) |
| Total | 75 (100%) | 70 (100%) | 267 (100%) | 359 (100%) | 158 (100%) | 859 (100%) |

| | M1 (dnMBC) (n=76) | RMBC < 12 months (n=70) | RMBC < 24 months (n=268) | RMBC 24 to 60 months (n=360) | RMBC 60+ months (n=158) | Total (n=862) |
|---|---|---|---|---|---|---|
| Missing | 1 (1.3%) | 0 | 1 (0.4%) | 1 (0.3%) | 0 | 3 (0.3%) |
| **PR status** | | | | | | |
| Negative | 24 (40.7%) | 41 (66.1%) | 161 (70.6%) | 122 (41.9%) | 27 (21.8%) | 334 (47.6%) |
| Positive | 35 (59.3%) | 21 (33.9%) | 67 (29.4%) | 169 (58.1%) | 97 (78.2%) | 368 (52.4%) |
| Total | 59 (100%) | 62 (100%) | 228 (100%) | 291 (100%) | 124 (100%) | 702 (100%) |
| Missing | 17 (22.4%) | 8 (11.4%) | 40 (14.9%) | 69 (19.2%) | 34 (21.5%) | 160 (18.6%) |
| **HER2 status** | | | | | | |
| Negative | 38 (52.1%) | 44 (62.9%) | 177 (67.6%) | 224 (66.3%) | 106 (75.7%) | 545 (67.0%) |
| Positive | 35 (47.9%) | 26 (37.1%) | 85 (32.4%) | 114 (33.7%) | 34 (24.3%) | 268 (33.0%) |
| Total | 73 (100%) | 70 (100%) | 262 (100%) | 338 (100%) | 140 (100%) | 813 (100%) |
| Missing | 3 (3.9%) | 0 | 6 (2.2%) | 22 (6.1%) | 18 (11.4%) | 49 (5.7%) |
| **TNBC status** | | | | | | |
| Not TNBC | 68 (89.5%) | 46 (65.7%) | 168 (62.7%) | 305 (84.7%) | 143 (90.5%) | 684 (79.8%) |
| TNBC | 8 (10.5%) | 24 (34.3%) | 100 (37.3%) | 55 (15.3%) | 15 (9.5%) | 178 (20.2%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| Missing | 0 | 0 | 0 | 0 | 0 | 0 |
| **Focality** | | | | | | |
| Localised | 30 (61.2%) | 40 (63.5%) | 161 (66.5%) | 202 (61.6%) | 89 (64.5%) | 482 (63.7%) |
| Multifocal | 19 (38.8%) | 23 (36.5%) | 81 (33.5%) | 126 (38.4%) | 49 (35.5%) | 275 (36.3%) |
| Total | 49 (100%) | 63 (100%) | 242 (100%) | 328 (100%) | 138 (100%) | 757 (100%) |
| Missing | 27 (35.5%) | 7 (10.0%) | 26 (9.7%) | 32 (8.9%) | 20 (12.7%) | 105 (12.2%) |
| **Path T stage**[2] | | | | | | |
| T0 | 1 (2.1%) | 1 (1.5%) | 7 (2.7%) | 7 (2.0%) | 3 (1.9%) | 18 (2.2%) |
| T1 | 16 (33.3%) | 16 (23.9%) | 79 (30.6%) | 136 (38.7%) | 68 (43.9%) | 299 (36.8%) |
| T2 | 22 (45.8%) | 34 (50.7%) | 128 (49.6%) | 168 (47.9%) | 70 (45.2%) | 388 (47.8%) |
| T3 | 9 (18.8%) | 14 (20.9%) | 41 (15.9%) | 38 (10.8%) | 14 (9.0%) | 102 (12.6%) |
| T4 | 0 | 2 (3.0%) | 3 (1.2%) | 2 (0.6%) | 0 | 5 (0.6%) |

| | M1 (dnMBC) (n=76) | RMBC < 12 months (n=70) | RMBC < 24 months (n=268) | RMBC 24 to 60 months (n=360) | RMBC 60+ months (n=158) | Total (n=862) |
|---|---|---|---|---|---|---|
| Total | 48 (100%) | 67 (100%) | 258 (100%) | 351 (100%) | 155 (100%) | 812 (100%) |
| Missing | 28 (36.8%) | 3 (4.3%) | 10 (3.7%) | 9 (2.5%) | 3 (1.9%) | 50 (5.8%) |
| **Path N stage**[3] | | | | | | |
| N0 | 11 (23.4%) | 11 (16.2%) | 66 (25.5%) | 96 (26.8%) | 49 (31.4%) | 222 (27.1%) |
| N1 | 36 (76.6%) | 57 (83.8%) | 193 (74.5%) | 262 (73.2%) | 107 (68.6%) | 598 (72.9%) |
| Total | 47 (100%) | 68 (100%) | 259 (100%) | 358 (100%) | 156 (100%) | 820 (100%) |
| Missing | 29 (38.2%) | 2 (2.9%) | 9 (3.4%) | 2 (0.6%) | 2 (1.3%) | 42 (4.9%) |
| **Max tumour size (invasive) (mm)** | | | | | | |
| Median | 35 | 32 | 30 | 27 | 25 | 28 |
| Range | 2 to 80, | 3 to 160, | 2 to 160, | 0 to 199, | .5 to 102, | 0 to 199, |
| IQR[4] | 18 to 49 | 25 to 60 | 20 to 47 | 18 to 41 | 18 to 40 | 19 to 43 |
| Missing | 29 (38.2%) | 5 (7.1%) | 18 (6.7%) | 21 (5.8%) | 6 (3.8%) | 74 (8.6%) |
| **Surgery** | | | | | | |
| No surgery | 26 (34.2%) | 2 (2.9%) | 5 (1.9%) | 1 (0.3%) | 1 (0.6%) | 33 (3.8%) |
| Surgery | 50 (65.8%) | 68 (97.1%) | 263 (98.1%) | 359 (99.7%) | 157 (99.4%) | 829 (96.2%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| **Surgical Type** | | | | | | |
| BCS[5] | 16 (21.1%) | 18 (25.7%) | 97 (36.2%) | 120 (33.3%) | 59 (37.3%) | 292 (33.9%) |
| Mastectomy | 33 (43.4%) | 50 (71.4%) | 166 (61.9%) | 238 (66.1%) | 98 (62.0%) | 535 (62.1%) |
| Nodal surgery | 1 (1.3%) | 0 | 0 | 1 (0.3%) | 0 | 2 (0.2%) |
| None | 26 (34.2%) | 2 (2.9%) | 5 (1.9%) | 1 (0.3%) | 1 (0.6%) | 33 (3.8%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| **Site of metastases at any time** | | | | | | |
| Bone | 8 (10.5%) | 7 (10.0%) | 31 (12.0%) | 31 (8.9%) | 14 (10.0%) | 84 (10.2%) |
| Bone-Brain | 4 (5.3%) | 4 (5.7%) | 7 (2.7%) | 13 (3.7%) | 2 (1.4%) | 26 (3.2%) |

| | M1 (dnMBC) (n=76) | RMBC < 12 months (n=70) | RMBC < 24 months (n=268) | RMBC 24 to 60 months (n=360) | RMBC 60+ months (n=158) | Total (n=862) |
|---|---|---|---|---|---|---|
| Bone-Visc[6] | 22 (28.9%) | 12 (17.1%) | 84 (32.4%) | 128 (36.8%) | 65 (46.4%) | 299 (36.3%) |
| Bone-Visc-Brain | 20 (26.3%) | 15 (21.4%) | 41 (15.8%) | 68 (19.5%) | 19 (13.6%) | 148 (18.0%) |
| Brain | 3 (3.9%) | 4 (5.7%) | 18 (6.9%) | 11 (3.2%) | 5 (3.6%) | 37 (4.5%) |
| Nodal | 4 (5.3%) | 3 (4.3%) | 8 (3.1%) | 15 (4.3%) | 10 (7.1%) | 37 (4.5%) |
| Visc | 12 (15.8%) | 21 (30.0%) | 53 (20.5%) | 55 (15.8%) | 17 (12.1%) | 137 (16.6%) |
| Visc-Brain | 3 (3.9%) | 4 (5.7%) | 17 (6.6%) | 27 (7.8%) | 8 (5.7%) | 55 (6.7%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 348 (100%) | 140 (100%) | 823 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 12 (3.3%) | 18 (11.4%) | 39 (4.5%) |
| **Site of 1st metastases** | | | | | | |
| Bone | 23 (30.3%) | 16 (22.9%) | 52 (20.1%) | 73 (21.1%) | 30 (21.7%) | 178 (21.7%) |
| Bone-Brain | 0 | 2 (2.9%) | 4 (1.5%) | 6 (1.7%) | 0 | 10 (1.2%) |
| Bone-Visc | 20 (26.3%) | 12 (17.1%) | 73 (28.2%) | 108 (31.2%) | 58 (42.0%) | 259 (31.6%) |
| Bone-Visc-Brain | 0 | 5 (7.1%) | 13 (5.0%) | 22 (6.4%) | 2 (1.4%) | 37 (4.5%) |
| Brain | 1 (1.3%) | 4 (5.7%) | 18 (6.9%) | 13 (3.8%) | 5 (3.6%) | 37 (4.5%) |
| Nodal | 12 (15.8%) | 4 (5.7%) | 17 (6.6%) | 23 (6.6%) | 15 (10.9%) | 67 (8.2%) |
| Visc | 20 (26.3%) | 25 (35.7%) | 73 (28.2%) | 88 (25.4%) | 23 (16.7%) | 204 (24.9%) |
| Visc-Brain | 0 | 2 (2.9%) | 9 (3.5%) | 13 (3.8%) | 5 (3.6%) | 27 (3.3%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 346 (100%) | 138 (100%) | 819 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 14 (3.9%) | 20 (12.7%) | 43 (5.0%) |
| **Brain Metastases** | | | | | | |
| Yes | 30 (39.5%) | 27 (38.6%) | 83 (32.0%) | 119 (34.2%) | 34 (24.3%) | 266 (32.3%) |
| No | 46 (60.5%) | 43 (61.4%) | 176 (68.0%) | 229 (65.8%) | 106 (75.7%) | 557 (67.7%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 348 (100%) | 140 (100%) | 823 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 12 (3.3%) | 18 (11.4%) | 39 (4.5%) |

| | M1 (dnMBC) (n=76) | RMBC < 12 months (n=70) | RMBC < 24 months (n=268) | RMBC 24 to 60 months (n=360) | RMBC 60+ months (n=158) | Total (n=862) |
|---|---|---|---|---|---|---|
| **Bone Metastases** | | | | | | |
| Yes | 54 (71.1%) | 38 (54.3%) | 163 (62.9%) | 240 (69.0%) | 100 (71.4%) | 557 (67.7%) |
| No | 22 (28.9%) | 32 (45.7%) | 96 (37.1%) | 108 (31.0%) | 40 (28.6%) | 266 (32.3%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 348 (100%) | 140 (100%) | 823 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 12 (3.3%) | 18 (11.4%) | 39 (4.5%) |
| **Visceral Metastases** | | | | | | |
| Yes | 57 (75.0%) | 52 (74.3%) | 195 (75.3%) | 278 (79.9%) | 109 (77.9%) | 639 (77.6%) |
| No | 19 (25.0%) | 18 (25.7%) | 64 (24.7%) | 70 (20.1%) | 31 (22.1%) | 184 (22.4%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 348 (100%) | 140 (100%) | 823 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 12 (3.3%) | 18 (11.4%) | 39 (4.5%) |

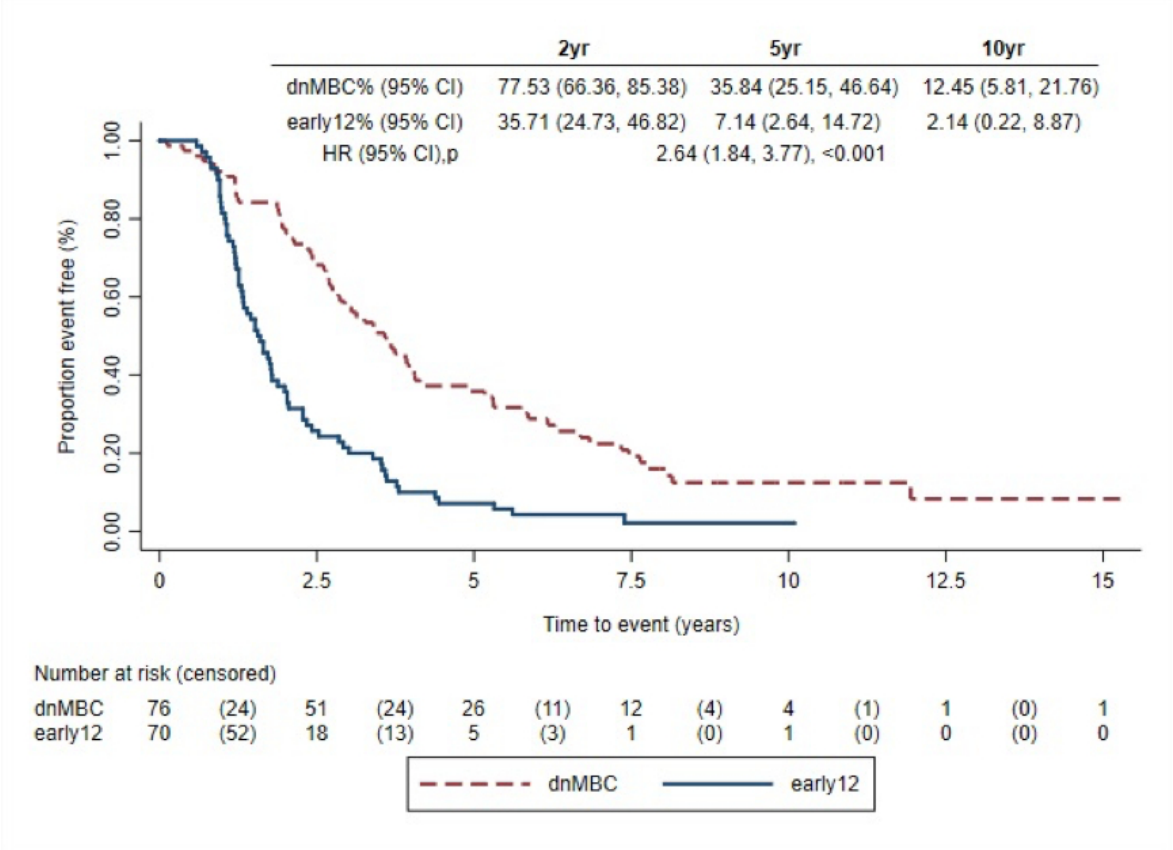
[1]LVI = lymphovascular invasion
[2]Path T stage = Pathological T-stage; T1=≤2cm, T2=2-5cm, T3=>5cm, T4=involvement chest wall, skin or inflammatory
[3]Path N stage = Pathological nodal-stage; N0= no cancer cells in regional lymph nodes, N1= regional nodal
[4]IQR = interquartile range
[5]BCS = breast-conserving surgery
[6]Visc = visceral metastatic disease

### 3.2.1.3 Survival of the metastatic cohort

Those who relapsed within 12 months (n=70) had a significantly worse OS than the dnMBC group (Figure 3.2A), with a HR for death of 2.64 (1.84-3.77; p=<0.001). For those who relapsed after 24 months (n=518), the OS varied over time, consistent with the delay from first diagnosis of breast cancer to onset of metastatic disease (clearly the HR for death at 2 years was very small). However, by 5 years the risk of death for those who relapsed between 24 and 60 months (n=360) was increased compared to the dnMBC

group with a 5yr HR of 1.55 (1.10 – 2.18; p=0.013) and 10yr HR 2.21 (1.02 – 4.77; p=0.044) (Figure 3.2B). Similarly, for those who relapsed after 60 months (n=158), compared to the dnMBC group, the 10yr HR was 1.74 (0.80 – 3.78; p=0.160). There was longer survival following distant metastases for dnMBC compared to all other groups who developed metastases, including those with late relapse after 60 months (HR 2.46; 1.79-3.38, p<0.001) (Figure 3.3).

**A.**



**B.**

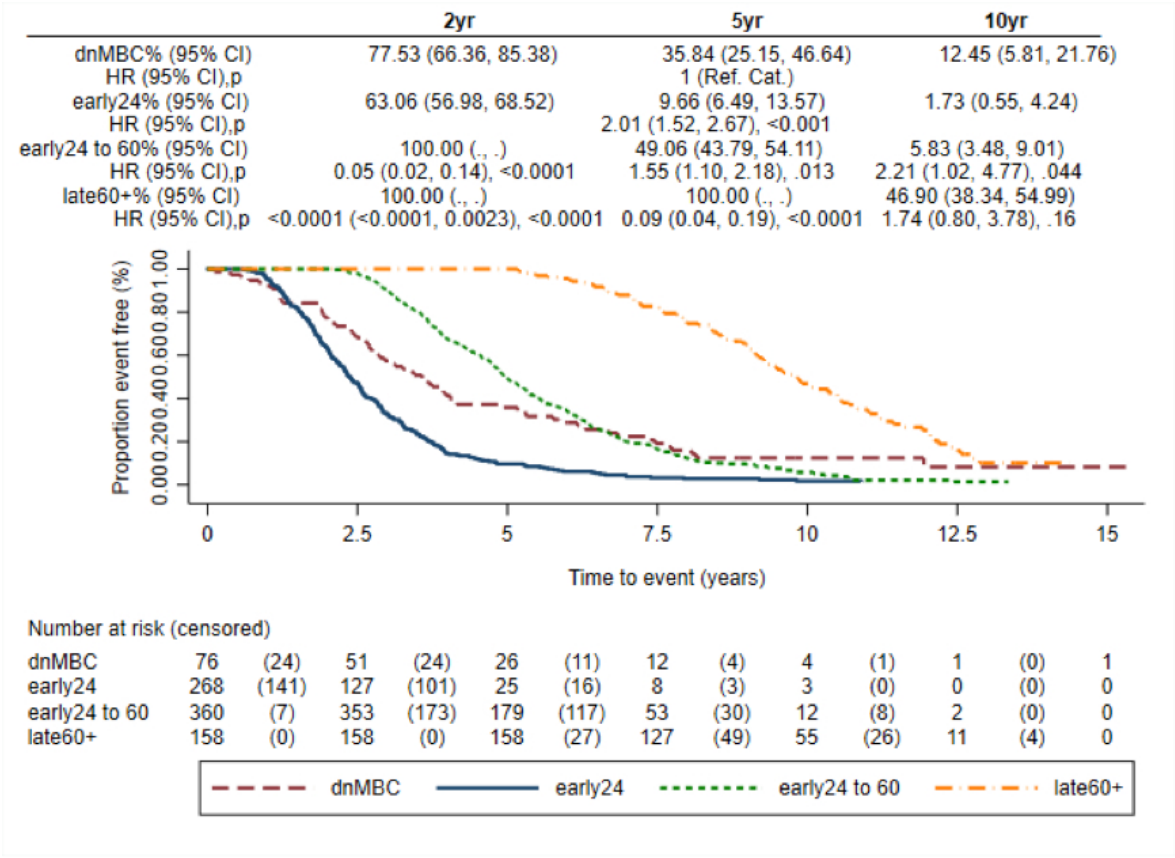
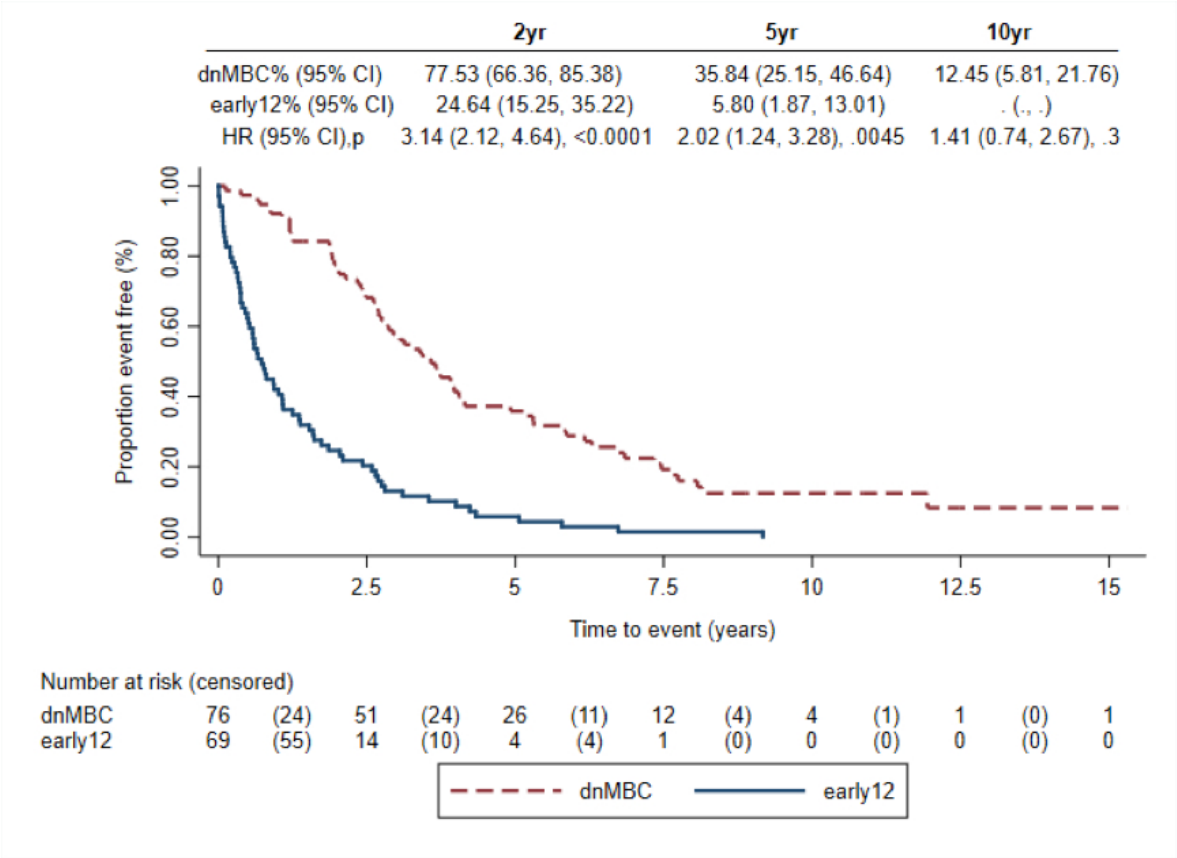



**Figure 3-2. Kaplan Meier analyses of OS.**

**A. OS for dnMBC vs. early12 and B. OS for dnMBC vs. early24, early24 to 60, and early 60+; reference category: dnMBC. Time-varying hazard ratio (and 95% CI) for death following initial diagnosis shown for early24 to 60 and late60+, relative to dnMBC, is shown. For early24, an overall HR, relative to dnMBC, is shown. The proportion of patients alive, at 2, 5 and 10 years is also shown for each category.**

**A.**


2yr 5yr 10yr
dnMBC% (95% CI) 77.53 (66.36, 85.38) 35.84 (25.15, 46.64) 12.45 (5.81, 21.76)
early12% (95% CI) 24.64 (15.25, 35.22) 5.80 (1.87, 13.01) . (., .)
HR (95% CI),p 3.14 (2.12, 4.64), <0.0001 2.02 (1.24, 3.28), .0045 1.41 (0.74, 2.67), .3
Proportion event free (%)
0.00 0.20 0.40 0.60 0.80 1.00
0 2.5 5 7.5 10 12.5 15
Time to event (years)
Number at risk (censored)
dnMBC 76 (24) 51 (24) 26 (11) 12 (4) 4 (1) 1 (0) 1
early12 69 (55) 14 (10) 4 (4) 1 (0) 0 (0) 0 (0) 0
dnMBC
early12

**B.**

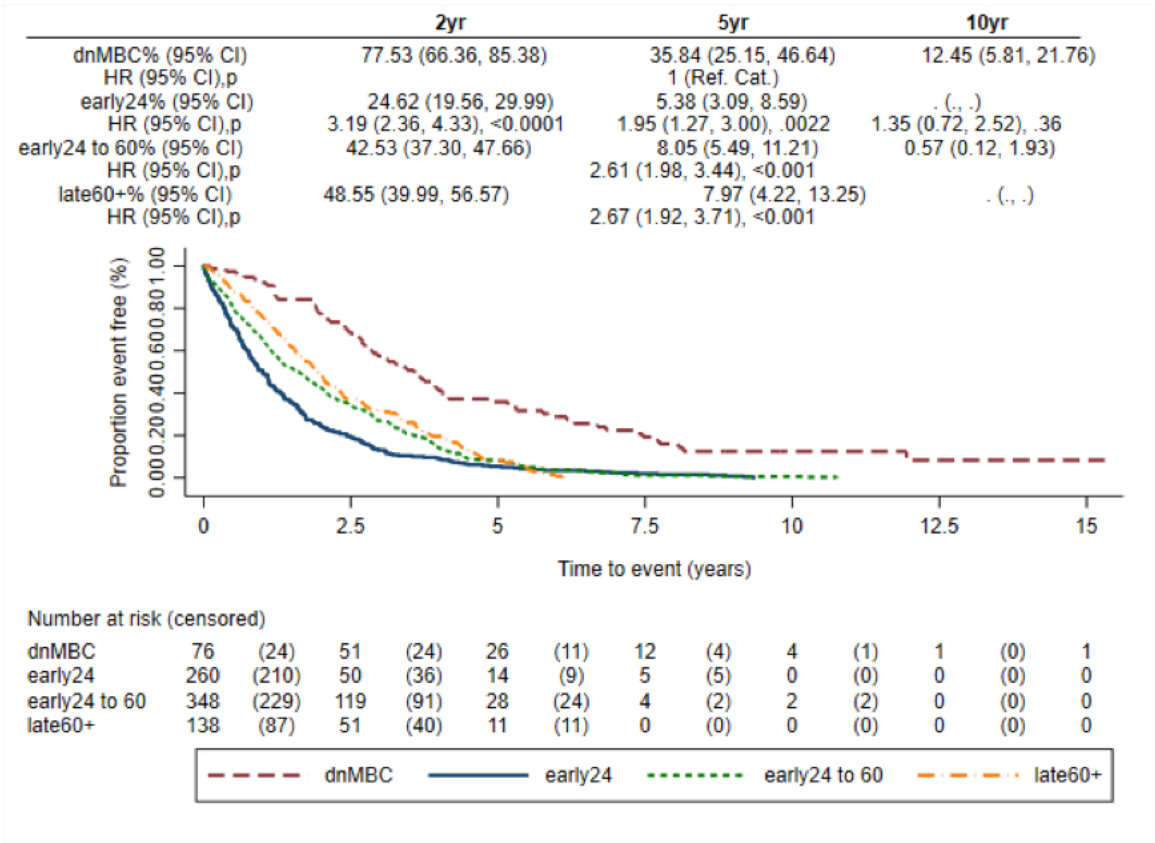



**Figure 3-3. Kaplan Meier analyses of PDRS.**

**A. PDRS for dnMBC vs. early12 and B. PDRS for dnMBC vs. early24, early24 to 60, and late 60+; reference category: dnMBC. Time-varying hazard ratio (and 95% CI) for death following diagnosis of metastatic disease shown for early12, relative to dnMBC, For early24 to 60 and late60+, an overall HR is shown, relative to dnMBC. The proportion of patients alive, at 2, 5 and 10 years is also shown for each category.**

### 3.2.1.4 Parameters affecting survival of the metastatic cohort

A multivariable analysis (MVA) to assess for factors related to duration of survival in those with dnMBC vs. early12 was performed. For OS, early12 patients maintained a significantly worse OS compared to dnMBC after adjustment for other factors (HR 3.76; 2.22-6.38; p<0.001). Positive axillary nodes were found to be associated with significantly higher risk of death (HR 2.29; 1.17-4.47; p=0.015) whilst patients with HER2-negative

tumours were at reduced risk of death (HR 0.500; 0.311-0.802; p=0.004). Similar results were also found in the MVA for PDRS (results not shown).

For early24 patients, PDRS was worse compared to dnMBC patients after adjustment for other factors at 2 and 5 years (HR 2.84; 1.72-4.69; p<0.001 and HR 1.68; 1.03-2.74; p=0.036). Again, positive axillary nodal status was found to be an independent risk factor for earlier distant relapse (HR 1.41; 1.04-1.90; p=0.026) whilst patients with HER2 positive tumours had longer survival (HR 0.690; 0.53-0.89; p=0.004). ER-positive status was protective for survival at 2 years compared to ER-negative (HR 0.50; 0.37-0.66; p<0.001), but not at 5 or 10 years.

#### 3.2.1.5 Survival of the metastatic cohort stratified by TNBC status

In order to assess whether TNBC status is a negative prognostic factor in the setting of metastatic disease in young women, Kaplan Meier survival analyses were performed for rMBC and dnMBC separately, when stratified by TNBC vs. non-TNBC. For patients who developed rMBC, the outcomes (OS, DFS and PDRS) were worse for those with TNBC, compared to those without TNBC (p<0.0001) (Figure 3.4).

For patients with *de novo* metastatic disease, those with TNBC had a significantly worse OS than those without (Figure 3.5).

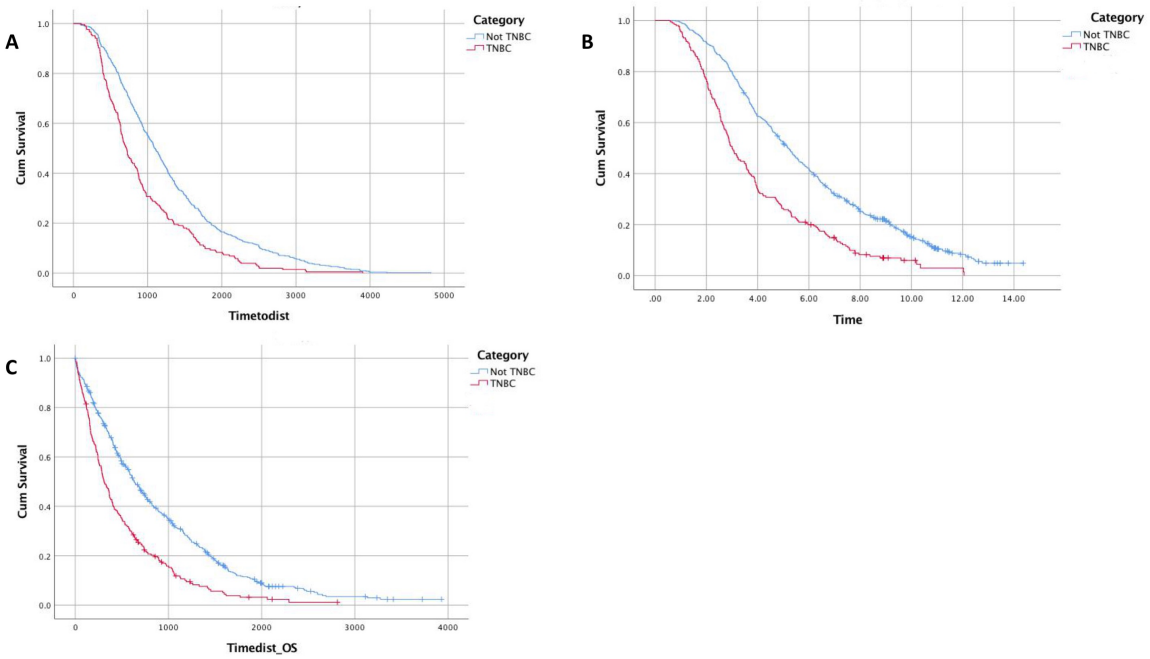



**Figure 3-4. Kaplan Meier survival analyses, stratified by TNBC status.**

**X-axis = Time from the development of metastases to outcome (death or date of last follow-up) in days.**

**A. DFS, B. OS and C. PDRS, for rMBC patients, TNBC (n=170) vs. non-TNBC (n=662), from the POSH cohort. Log-rank mantel cox P<0.0001 (all). A. Median DFS 716 (633-799) days vs. 1091 (1012-1170) days, B. Median OS 3.010 (2.588-3.432) years vs. 5.180 (4.825-5.535) years, C. Median PDRS 312 (257-367) days vs. 632 (558-706) days.**

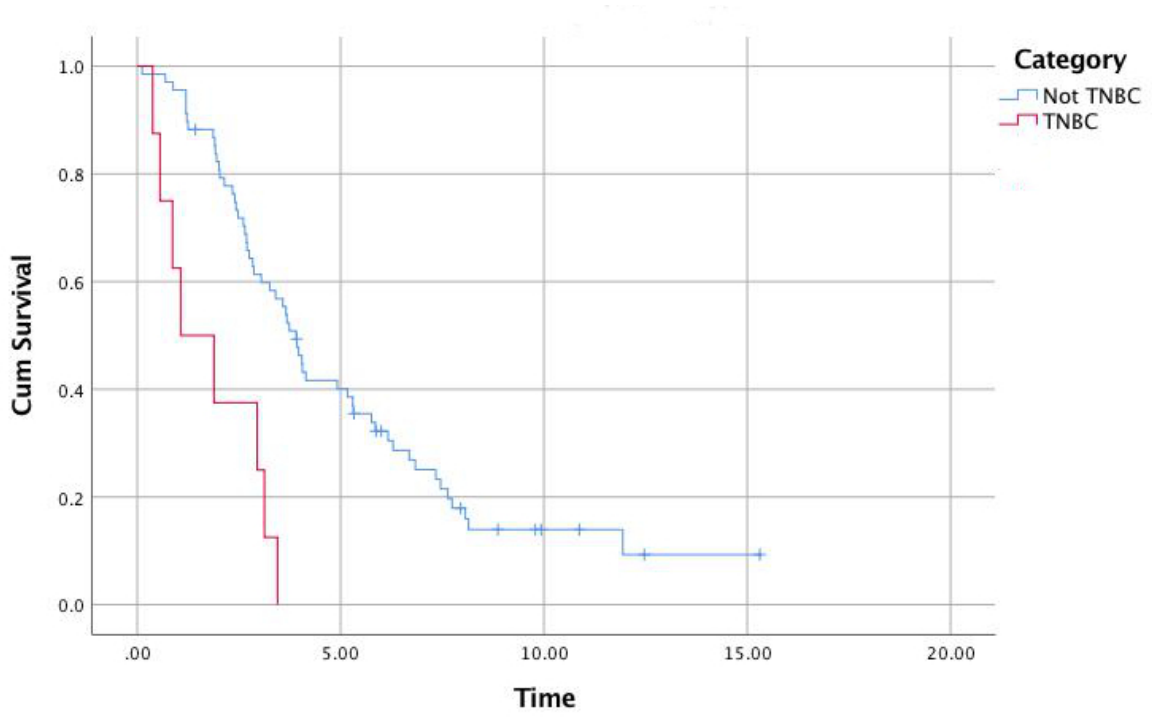



**Figure 3-5. Kaplan Meier survival analysis for dnMBC stratified by TNBC status.**

**X-axis = Time from diagnosis to outcome (death or date of last follow-up) in years. Includes dnMBC patients in the POSH cohort, OS TNBC (n=8) vs. non-TNBC (n=68). Log-rank mantel cox P<0.0001. Median OS 1.070 (0.000-2.484) years vs. 3.910 (3.380-4.440) years.**

#### 3.2.1.6 Sites of metastases

Regarding sites of metastases (at any time during disease course), I defined widespread disease as those who developed bone, visceral and brain metastases. The dnMBC and early12 groups had a higher frequency of widespread disease (26.3% and 21.4% respectively) than the late relapse group (13.6%). Early12 patients were the most likely to develop only visceral disease (30.0%). Of those who presented with a late relapse, 46.4% developed bone and visceral disease (compared to 17.1% in the early12 group). Patients with dnMBC were most likely to develop brain metastases (39.5%), compared to 24.3% in the late relapse group. The presence of bone metastases at any time during the disease course was least common in the early12 group (54.3%, compared to 71.1% of the dnMBC group). Visceral metastases at any time were equally prevalent throughout all groups.

In the rMBC group with TNBC, patients who were diagnosed with brain metastases at any time during the disease course had a reduced HR for death, compared to those with TNBC who did not develop brain metastases (HR 0.68, 95% CI 0.48-0.97, p=0.030) (Figure 3.6A). When PDRS was compared, there was no significant difference (Figure 3.6B).

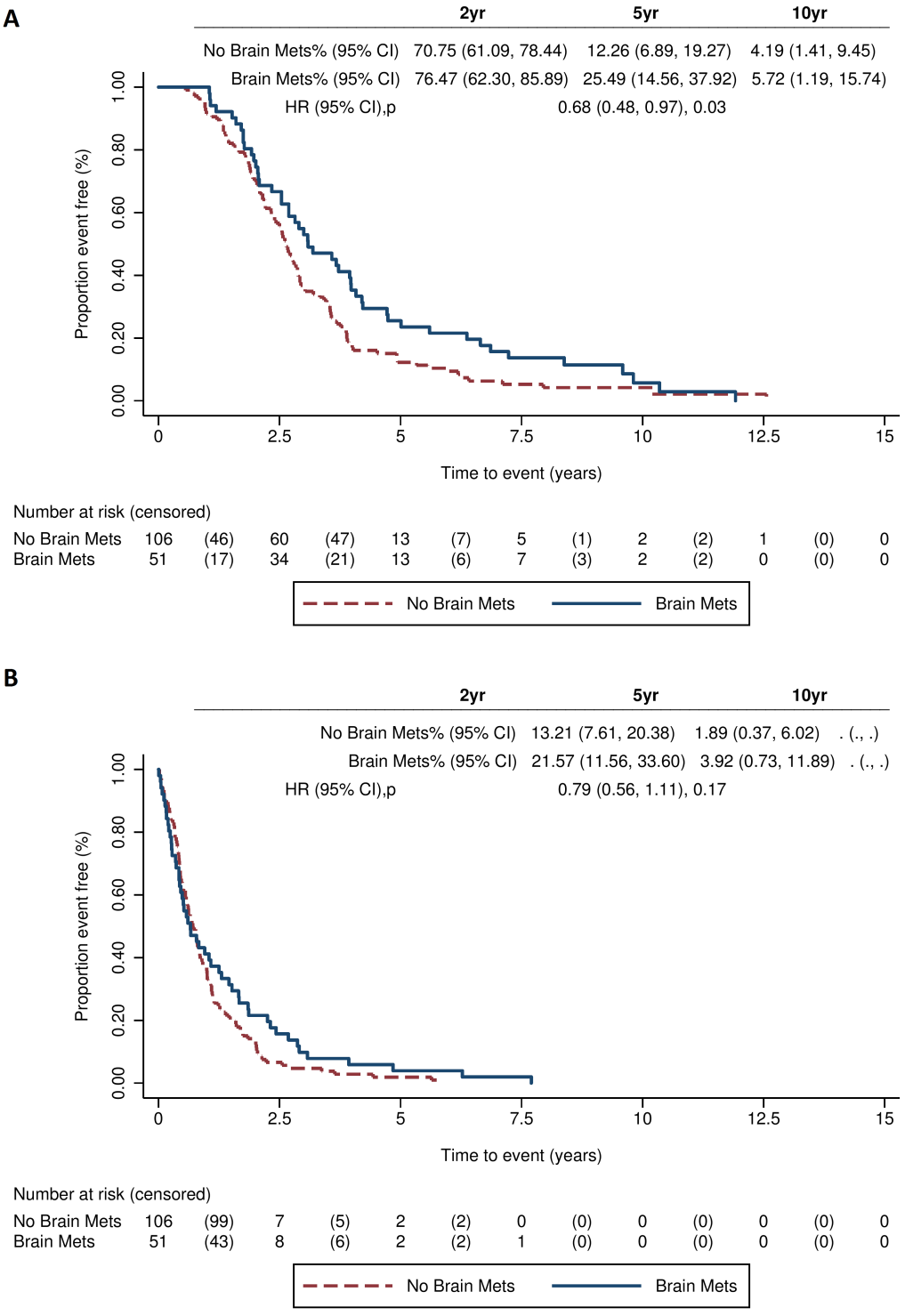


**Figure 3-6. Kaplan Meier survival analyses for relapsed metastatic TNBC patients, stratified by brain metastases status.**

**A. OS and B.DFS for patients with TNBC treated with adjuvant/neoadjuvant intent but later experienced recurrent metastatic disease, stratified by 1) developed (n=51) and 2) never developed (n=106) brain metastases during course of disease. P=0.030 (A) and P=0.170 (B); reference category: no brain metastases.**

### 3.2.2 The TNBC dataset

I have shown above that the outcomes for those with metastatic TNBC are worse than those with metastatic non-TNBC. I then proceeded to evaluate the characteristics and outcomes for those with early-stage TNBC amongst the POSH cohort. There were 609 patients from the POSH cohort with early-stage TNBC at diagnosis. Those treated with neoadjuvant chemotherapy (n=104) were excluded, leaving 505 patients treated with either adjuvant (n=491) or no chemotherapy (n=14).

#### 3.2.2.1 Demographics and clinicopathological features of the TNBC cohort

Baseline characteristics for the primary cohort of 505 and the pathology cohort (within the primary cohort) of 350 (discussed in the next chapter) can be seen in Table 3.2. The majority (93.4%) had grade 3 tumours (466/499). With regards to T-stage, only 6.0% were T3 (30/502), none were T4. The majority (62.4%) were node-negative (314/503). Almost all patients (97.2%) received adjuvant chemotherapy (491/505). One quarter (25.9%) had a *BRCA* mutation (131/505), 122 in *BRCA1* and 9 in *BRCA2.* The clinicopathological features of the pathology cohort (n=350) were compared to the entire cohort (n=505) and were not found to be significantly different (non-significant p-values for each variable when comparing the two groups). The disease pattern for those who developed metastatic disease amongst the early-stage cohort (n=120) is also shown (Table 3.3).

**Table 3-2. Baseline characteristics of early-stage TNBC patients in POSH cohort.**

**N=505. Excluding those with metastatic disease at diagnosis or those treated with neoadjuvant intent.**

| | No. POSH cohort (n=505) | %[1] | No. pathology cohort (n=350) | % |
|---|---|---|---|---|
| **Age at diagnosis (years)** | | | | |
| 18-25 | 7 | 1.4 | 4 | 1.1 |
| 26-30 | 52 | 10.3 | 39 | 11.1 |
| 31-35 | 178 | 35.2 | 134 | 38.3 |
| ≥36 | 268 | 53.1 | 173 | 49.4 |
| **BMI[2]** | | | | |
| ≤25.0 | 249 | 50.2 | 170 | 49.3 |
| >25.0 to ≤30.0 | 142 | 28.6 | 102 | 29.6 |
| >30.0 to ≤40.0 | 90 | 18.1 | 65 | 18.8 |
| >40.0 | 15 | 3.0 | 8 | 2.3 |
| Not known | 9 | 1.8 | 5 | 1.4 |
| **Ethnicity** | | | | |
| White/Caucasian | 461 | 92.4 | 324 | 93.6 |
| Black | 18 | 3.6 | 12 | 3.5 |
| Asian | 16 | 3.2 | 9 | 2.6 |
| Other | 4 | 0.8 | 1 | 0.3 |
| Not known | 6 | 1.2 | 4 | 1.1 |
| **T-stage[3]** | | | | |
| 1 | 233 | 46.4 | 156 | 44.6 |
| 2 | 239 | 47.6 | 172 | 49.1 |
| 3 | 30 | 6.0 | 22 | 6.3 |
| Not known | 3 | 0.6 | 0 | 0.0 |
| **N-stage[4]** | | | | |
| 0 | 314 | 62.4 | 210 | 60.2 |

| | No. POSH cohort (n=505) | %[1] | No. pathology cohort (n=350) | % |
|---|---|---|---|---|
| 1 | 134 | 26.6 | 100 | 28.7 |
| 2 | 55 | 10.9 | 39 | 11.2 |
| Not known | 2 | 0.4 | 1 | 0.3 |
| **Grade** | | | | |
| 1 | 2 | 0.4 | 2 | 0.6 |
| 2 | 31 | 6.2 | 16 | 4.7 |
| 3 | 466 | 93.4 | 326 | 94.8 |
| Not known | 6 | 1.2 | 6 | 1.7 |
| **Lymphovascular invasion** | | | | |
| Present | 188 | 38.4 | 133 | 39.2 |
| Absent | 302 | 61.6 | 206 | 60.8 |
| Not known | 15 | 3.0 | 11 | 3.1 |
| **Focality** | | | | |
| Localised | 412 | 86.4 | 284 | 85.5 |
| Multifocal | 65 | 13.6 | 48 | 14.5 |
| Unknown | 28 | 5.5 | 18 | 5.1 |
| **NPI[5] category** | | | | |
| Excellent | 1 | 0.2 | 1 | 0.3 |
| Good | 13 | 2.6 | 7 | 2.0 |
| Moderate | 342 | 68.5 | 231 | 67.2 |
| Poor | 143 | 28.7 | 105 | 30.5 |
| Not calculable | 6 | 1.2 | 6 | 1.7 |
| **Adjuvant chemotherapy** | | | | |
| FEC[6] +/- AC[6] +/- docetaxel | 232 | 45.9 | 164 | 46.9 |
| ECMF[6] | 121 | 24.0 | 82 | 23.4 |

| | No. POSH cohort (n=505) | %[1] | No. pathology cohort (n=350) | % |
|---|---|---|---|---|
| EC[6] +/- paclitaxel +/- gemcitabine | 93 | 18.4 | 68 | 19.4 |
| AC +/- paclitaxel +/- docetaxel | 31 | 6.1 | 17 | 4.9 |
| ACMF[6], CAF[6], CMF[6] | 11 | 2.2 | 9 | 2.6 |
| Other | 3 | 0.6 | 0 | 0.0 |
| No chemotherapy | 14 | 2.8 | 10 | 2.9 |
| **Chemotherapy category** | | | | |
| Anthracycline and taxane | 114 | 22.6 | 86 | 24.6 |
| Anthracycline | 373 | 73.9 | 252 | 72.0 |
| Taxane | 1 | 0.2 | 0 | 0.0 |
| Other (CMF) | 3 | 0.6 | 2 | 0.6 |
| No chemotherapy | 14 | 2.8 | 10 | 2.9 |
| **Adjuvant radiotherapy** | | | | |
| Yes | 426 | 84.4 | 307 | 87.7 |
| No/unknown | 79 | 15.6 | 43 | 12.3 |
| **Surgery** | | | | |
| Breast-conserving surgery | 331 | 65.5 | 235 | 67.1 |
| Mastectomy | 173 | 34.3 | 114 | 32.6 |
| Nodal-only surgery | 1 | 0.2 | 1 | 0.3 |
| **BRCA status** | | | | |
| BRCA-1 gene mutation | 122 | 24.2 | 82 | 23.4 |
| BRCA-2 gene mutation | 9 | 1.8 | 7 | 2.0 |
| No gene mutation | 374 | 74.1 | 261 | 74.6 |
| **Distant recurrence** | | | | |
| Yes | 120 | 23.8 | 85 | 24.3 |
| Died of breast cancer[7] | 6 | 1.2 | 4 | 1.1 |

| | No. POSH cohort (n=505) | %[1] | No. pathology cohort (n=350) | % |
|---|---|---|---|---|
| No | 379 | 75.0 | 261 | 74.6 |

[1] = Percentages calculated from total, excluding cases where factor is "not known". Not known percentage is calculated from number in total cohort.

[2] = BMI = body mass index

[3] = pathological staging; T1=≤2cm, T2=2-5cm, T3=>5cm in greatest dimension

[4] = pathological staging; N0= no lymph node metastases, N1=metastases in 1-3 lymph nodes, N2= Metastases in >4 lymph nodes

[5] = Nottingham Prognostic Index(226) = (0.2xS) + N + G; S is the size of the index lesion in centimetres, N is the node status: 0 nodes = 1, 1-3 nodes = 2, >3 nodes = 3, G is the grade of tumour: Grade I =1, Grade II =2, Grade III =3. Scoring classification: ≤2.4 = Excellent, >2.4 to ≤3.4=Good, >3.4 to ≤5.4 = Moderate, >5.4 = Poor

[6] = FEC=5-fluorouracil, epirubicin and cyclophosphamide, EC= epirubicin and cyclophosphamide, AC=doxorubicin and cyclophosphamide, ACMF= Doxorubicin Plus Cyclophosphamide, Methotrexate, and 5-fluorouracil, CAF= Cyclophosphamide, Doxorubicin, 5-Fluorouracil, CMF= Cyclophosphamide, Methotrexate, and 5-fluorouracil

[7] = Not documented as having a relapse, but breast cancer listed as cause of death

**Table 3-3. Characteristics of metastatic disease for early-stage TNBC patients in POSH cohort.**
**Excluding those with metastatic disease at diagnosis or those treated with neoadjuvant intent. Includes only patients who have developed recurrent metastatic disease (n=120).**

| Site of 1st metastasis | N=120 | | N=85 | |
|---|---|---|---|---|
| Bone | 13 | 11.1 | 8 | 9.4 |
| Bone and visceral | 23 | 19.7 | 17 | 20.0 |
| Bone and CNS[1] | 3 | 2.6 | 3 | 3.5 |
| Bone, visceral and CNS | 8 | 6.8 | 4 | 4.7 |
| Brain | 8 | 6.8 | 6 | 7.1 |
| Nodal only | 15 | 12.8 | 11 | 12.9 |
| Visceral | 39 | 33.3 | 30 | 35.3 |
| Visceral and CNS | 8 | 6.8 | 6 | 7.1 |
| Not determined | 3 | 2.5 | 0 | 0.0 |

| **Site of all metastases at any time** | | | | |
|---|---|---|---|---|
| Bone | 9 | 7.7 | 6 | 7.1 |
| Bone and visceral | 25 | 21.4 | 20 | 23.5 |
| Bone and CNS | 5 | 4.3 | 4 | 4.7 |
| Bone, visceral and CNS | 13 | 11.1 | 10 | 11.8 |
| Brain | 10 | 8.5 | 8 | 9.4 |
| Nodal only | 9 | 7.7 | 6 | 7.1 |
| Visceral | 28 | 23.9 | 22 | 25.9 |
| Visceral and CNS | 12 | 10.3 | 9 | 10.6 |
| Not determined | 3 | 2.5 | 0 | 0.0 |
| **Brain metastases at any time** | | | | |
| Yes | 43 | 36.8 | 31 | 36.5 |
| No | 74 | 63.2 | 54 | 63.5 |
| Not determined | 3 | 2.5 | 0 | 0.0 |
| **Bone metastases at any time** | | | | |
| Yes | 58 | 49.6 | 40 | 47.1 |
| No | 59 | 50.4 | 45 | 52.9 |
| Not determined | 3 | 2.5 | 0 | 0.0 |
| **Visceral metastases at any time** | | | | |
| Yes | 84 | 71.8 | 61 | 71.8 |
| No | 33 | 28.2 | 24 | 28.2 |
| Not determined | 3 | 2.5 | 0 | 0.0 |

[1]= Central Nervous System (i.e. cerebral metastases or leptomeningeal disease)

#### 3.2.2.2 Survival of the TNBC cohort

Of the primary cohort of 505 patients with TNBC treated with adjuvant intent (or surgery alone), there have been 124 deaths (90 of these were amongst the scored cohort). In two cases the cause was another cancer (oesophageal and unknown) in patients with *BRCA1* and *TP53* mutations respectively. In one case the cause was infection in a patient with a recent distant relapse of their breast cancer and in another case, the cause was unknown (in a patient who died less than a year after their breast cancer diagnosis). In the remaining 120 cases, the cause of death was breast cancer. Estimated median OS was 12.87 years at a median follow-up of 8.27 years.

#### 3.2.2.3 Clinical factors affecting survival of the TNBC cohort

The impact of traditional clinical and pathological features on survival was assessed using both univariable and multivariable analyses.

In order to determine which clinical parameters to include in MVA, backwards elimination in a Cox regression model for OS was performed (see Methods 2.2.5). It was found that lymphovascular invasion (LVI) and nodal stage were significant prognostic indicators for OS. However, a proportional hazards test identified that LVI is not proportional. Therefore for MVA in this section I have adjusted for N-stage and stratified by LVI.

**Table 3-4. Hazard ratios for mortality for clinicopathogical parameters for all stage I-III TNBC cases from the POSH cohort.**

**Excludes those treated with neoadjuvant chemotherapy (n=505). Univariable and multivariable hazard ratios obtained from Cox regression analysis.**

| | Univariable HR (95% CI) | p-value | Multivariable HR[1] | p-value |
|---|---|---|---|---|
| **Age** | | | | |
| 18-30 | 1 | | 1 | |
| 31-35 | 1.497 (0.753-2.975) | 0.250 | 1.364 (0.683-2.724) | 0.380 |
| 36-40 | 1.621 (0.835-3.148) | 0.153 | 1.565 (0.802-3.055) | 0.189 |
| 41-50 | 0.513 (0.065-4.028) | 0.525 | 0.809 (0.101-6.457) | 0.842 |
| **BMI** | | | | |
| Normal (≤25.0) | 1 | | 1 | |
| Overweight (>25.0 to ≤30.0) | 1.780 (1.194-2.653) | **0.005** | 1.578 (1.045-2.380) | **0.030** |
| Obese (>30.0) | 1.228 (0.761-1.983) | 0.401 | 1.329 (0.818-2.160) | 0.251 |
| **Ethnicity** | | | | |
| Asian/Other | 1 | | 1 | |
| White | 1.753 (0.556-5.531) | 0.338 | 4.375 (0.607-31.514) | 0.143 |
| Black | 1.723 (0.384-7.726) | 0.477 | 3.520 (0.390-31.758) | 0.262 |
| **T-stage** | | | | |
| 1 | 1 | | | |
| 2 | 1.663 (1.134-2.439) | **0.009** | 1.296 (0.874-1.923) | 0.197 |
| 3 | 2.370 (1.222-4.598) | **0.011** | 1.171 (0.560-2.449) | 0.675 |
| **N-stage** | | | | |
| 0 | 1 | | 1 | |
| 1 | 2.773 (1.862-4.130) | **<0.0001** | 2.202 (1.426-3.399) | **0.000368** |
| 2 | 4.226 (2.616-6.825) | **<0.0001** | 2.970 (1.748-5.044) | **<0.0001** |

| | Univariable HR (95% CI) | p-value | Multivariable HR[1] | p-value |
|---|---|---|---|---|
| **NPI[2]** | | | | |
| Poor | 1 | | 1 | |
| Moderate | 0.338 (0.236-0.485) | **<0.0001** | 0.703 (0.386-1.283) | 0.251 |
| Good/Excellent | 0.312 (0.096-1.011) | 0.052 | 0.849 (0.211-3.408) | 0.817 |
| **LVI** | | | | |
| Absent | 1 | | 1 | |
| Present | 2.741 (1.908-3.938) | **<0.0001** | 1.836 (1.226-2.751) | **0.003** |
| **Focality** | | | | |
| Localised | 1 | | 1 | |
| Multifocal | 1.450 (0.909-2.313) | 0.119 | 0.954 (0.580-1.569) | 0.852 |
| **BRCA mutation status** | | | | |
| None | 1 | | 1 | |
| Pathogenic BRCA1 mutation | 0.667 (0.426-1.045) | 0.077 | 0.703 (0.446-1.106) | 0.127 |
| Pathogenic BRCA2 mutation | 0.807 (0.199-3.273) | 0.764 | 0.757 (0.185-3.095) | 0.698 |
| **Chemotherapy** | | | | |
| Anthracycline and taxane | 1 | | 1 | |
| Anthracycline, no taxane | 0.649 (0.440-0.959) | **0.030** | 1.003 (0.656-1.534) | 0.987 |
| Other/None | 0.220 (0.051-0.943) | **0.042** | 0.552 (0.122-2.495) | 0.440 |
| **Radiotherapy** | | | | |
| No | 1 | | 1 | |
| Yes | 0.880 (0.555-1.396) | 0.588 | 0.587 (0.359-0.961) | **0.034** |
| **Breast surgery** | | | | |
| BCS | 1 | | 1 | |

| | Univariable HR (95% CI) | p-value | Multivariable HR[1] | p-value |
|---|---|---|---|---|
| Mastectomy | 1.654 (1.161-2.356) | **0.005** | 1.258 (0.861-1.838) | 0.235 |
| **Brain metastases at any time** | **N=117** | | | |
| No (74) | 1 | | 1 | |
| Yes (43) | 0.706 (0.477-1.044) | 0.081 | 0.707 (0.461-1.086) | 0.113 |
| **Bone metastases at any time** | **N=117** | | | |
| No (59) | 1 | | 1 | |
| Yes (58) | 0.954 (0.654-1.392) | 0.808 | 0.965 (0.648-1.437) | 0.862 |
| **Visceral metastases at any time** | **N=117** | | | |
| No (33) | 1 | | 1 | |
| Yes (84) | 1.205 (0.787-1.845) | 0.385 | 1.226 (0.782-1.923) | 0.374 |

[1]Adjusted for N-stage and stratified by LVI

[2]Nottingham Prognostic Index(226) = (0.2xS) + N + G; S is the size of the index lesion in centimetres, N is the node status: 0 nodes = 1, 1-3 nodes = 2, >3 nodes = 3, G is the grade of tumour: Grade I =1, Grade II =2, Grade III =3. Scoring classification: ≤2.4 = Excellent, >2.4 to ≤3.4=Good, >3.4 to ≤5.4 = Moderate, >5.4 = Poor

The univariable and multivariable hazard ratios (HRs) for survival for each clinicopathological factor separately can be seen in Table 3.4 (Forest plot for univariable ratios shown in Figure 3.7). For univariable analysis, improved survival was significantly associated with a Nottingham Prognostic Index (NPI)(226) score of moderate or better. Worse survival was significantly associated with a body mass index (BMI) of 25-30 (Figure 3.8), nodal stage of 2/3, presence of LVI and multifocal tumours.

In addition to nodal involvement/LVI, being categorised as overweight remained a significant negative prognostic factor with a HR of 1.578 (1.045-2.380; p=0.030) after adjustment in MVA (Table 3.4). On univariable analysis, treatment with adjuvant radiotherapy was not prognostic, but after MVA it was found to be significantly prognostic with an HR of 0.587.

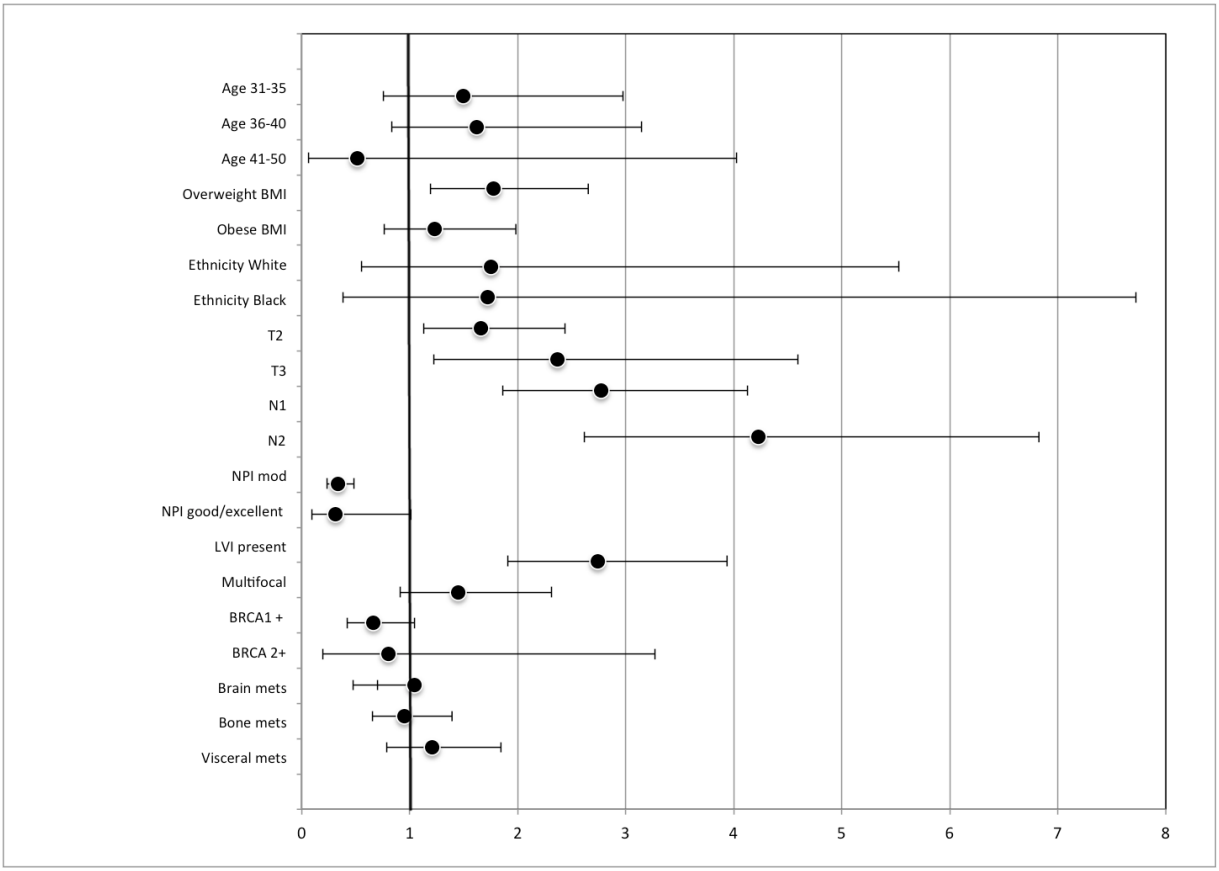



**Figure 3-7. Forest plot of univariable hazard ratios of clinical factors for overall survival for all patients with TNBC.**

**Includes patients in POSH study treated with adjuvant or neoadjuvant intent (n=505). Reference line shown (1.00) - for reference categories, please refer to Table 3.3.**

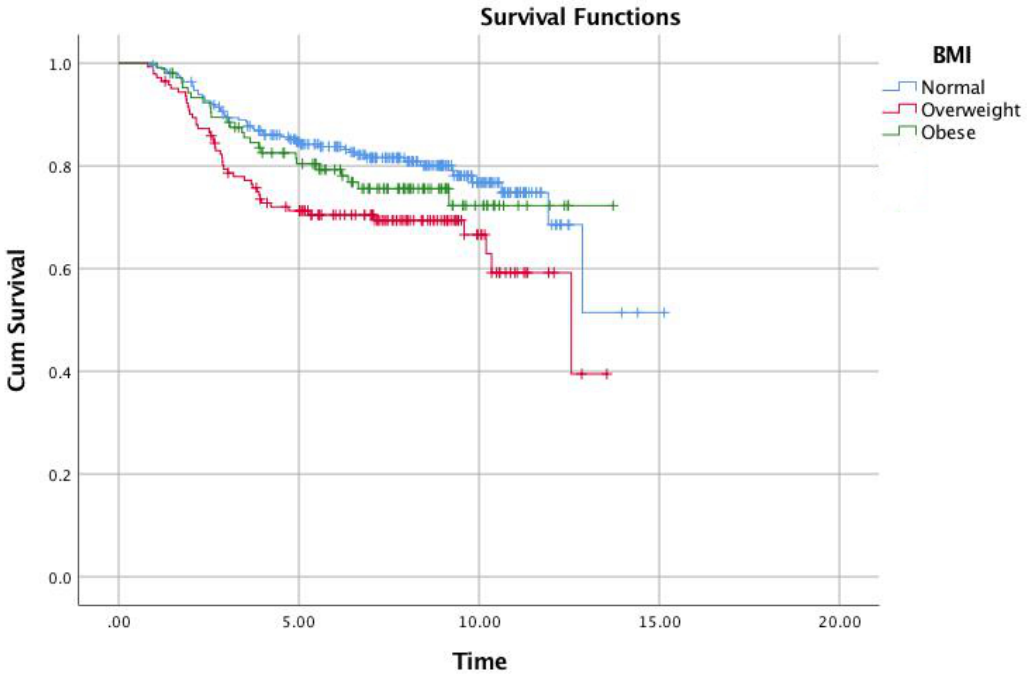



**Figure 3-8. Kaplan Meier survival analysis for stage I-III TNBC cases from the POSH cohort, stratified by BMI.**

**Normal/underweight (n=249), overweight (n=142) and obese (n=105).**

**Log rank (mantel-cox) p= 0.016.**

## 3.3 Discussion

Here I have presented a comprehensive review of the outcomes for young patients with breast cancer who present with *de novo* vs. recurrent metastatic breast cancer. I have then analysed the TNBC cohort, including factors such as clinicopathological parameters at diagnosis of the primary cancer, and evaluated associations with survival.

### 3.3.1 Survival, incidence and features of dnMBC and rMBC amongst young women

This is the largest published prospective study to evaluate the clinicopathological features, patterns of disease and outcomes in young patients with metastatic breast cancer (Appendix D)(227). These data have shown that young women who develop secondary metastatic disease, even if greater than five years after diagnosis, have shorter survival time following diagnosis of metastases compared to those who present with *de novo* metastatic disease. When survival from initial diagnosis (OS) was compared, this was superior for those who relapsed after 24 months, compared to the patients with dnMBC.

Only 2.6% of this cohort had metastatic disease at presentation, lower than the national prevalence of 5.2% from the National Cancer Registry from 2012-2017(228). In the USA this figure is 6% according to SEER data(229). Late stage at diagnosis is reported to be more common in women aged greater than 80 and so this likely contributes to the higher figures in the unselected age population (1, 225). Previous retrospective studies have identified *de novo* rates in the same age group as 1-3.9% (225, 230-232).

It is notable that 62.1% of patients had a mastectomy, a comparatively high rate. It is likely that this is related to adverse tumour features at original diagnosis in those that developed recurrent metastatic breast cancer (e.g. nodal involvement in 72.9%).

### 3.3.2 Metastatic outcome stratified by tumour subtype

Amongst women with rMBC, those with TNBC had significantly worse outcomes for OS, DFS and PDRS compared to other subtypes as a whole. Women with dnMBC and TNBC also had worse outcomes than those without.

Of note, amongst patients with rMBC of the triple negative subtype, the development of brain metastases did not negatively impact PDRS, with a slight improvement in OS. It is possible that for those who survived longer from initial diagnosis of breast cancer, the longer interval time to develop metastases resulted in a higher incidence of dissemination to the central nervous system. It is also likely that many of these women had uncontrolled and rapidly progressive visceral metastatic disease, so the brain metastases did not have a significantly additional adverse impact on their limited prognosis.

Of note, only a small proportion of dnMBC cases were of the triple negative subtype (10.5%), in common with previous studies showing dnMBC to have a favourable biological profile(8). The reason for this is unclear but could perhaps be related to larger, more detectable primary lesions, with less time to develop distant metastases before diagnosis.

### 3.3.3 BRCA prevalence in the metastatic cohort

The POSH cohort is unique not only for its age but for completeness of germline *BRCA* gene mutation testing. A notable proportion of patients with *de novo* disease (11.8%) had a *BRCA2* mutation, whereas just one (1.3%) had a *BRCA1* mutation. The overall *BRCA2* mutation rate in the metastatic population was 5.6% (48/862). Although the 69.3% ER-positivity rate in this group may explain this to some extent, the ER-positivity was higher in the early24to60 and late relapse groups with a lower *BRCA2* prevalence (5.8% and 6.3% respectively). Across the cohort of 862 patients with metastatic disease, the BRCA2 mutation rate was 5.6%. We have previously reported a *BRCA2* mutation rate of 5.0% across the POSH cohort as a whole (excluding dnMBC patients); with a lower proportion of Grade 3 tumours and greater oestrogen receptor expression (and relatively few cases of triple negative tumours) amongst them (201). Therefore it is surprising that the *BRCA2* mutation prevalence amongst the dnMBC cases was relatively high.

It is unknown whether a *BRCA1/2* mutation confers independent prognostic risk for *de novo* metastatic breast cancer(118). Tumours associated with a *BRCA1* mutation are associated with high grade, triple-negative, basal-like tumours, while *BRCA2*-mutated tumours are usually ER-positive/HER2-negative, high grade tumours(201). Compared to ER-positive sporadic tumours, *BRCA2* tumours do not have easily identifiable features at an immunohistochemical level(233). There is limited published data on *de novo* vs. recurrent metastatic disease in patients with a *BRCA2* gene mutation. In one Scandinavian study, 54 breast cancer patients from 22 families with *BRCA2* germline mutations found that *BRCA2*-associated cases were more often clinical stage IV at presentation, compared to controls (OR=4.6; 95% CI 1.3-17, p=0.021)(234). We know that overall a *BRCA2* mutation does not confer a worse outcome in the POSH cohort(201), but perhaps in a subset the combination of homologous recombination repair deficiency with other somatic mutations leads to early invasive behaviour. However, it is also possible that those with a *BRCA2* mutation had a strong family history of breast and ovarian cancer and were subsequently more vigilant regarding symptoms suggestive of metastatic disease. Another possibility is that bone pain, resulting from bone metastases (more common in ER-positive disease, the phenotype associated with a *BRCA2* mutation) in a young person is a red flag symptom that led to early imaging. The proportion of *BRCA1* mutation

carriers was highest in the early24 group; this is likely to reflect the high frequency of ER-negative cases in this group.

### 3.3.4 Clinicopathological features affecting early-stage TNBC survival

In the second half of this chapter, I evaluated the clinicopathological features of the early-stage TNBC cohort (n=505). The majority of this cohort received adjuvant chemotherapy. Approximately one quarter had a pathogenic *BRCA* mutation. The majority was grade 3 but node-negative.

I then evaluated the prognostic impact of the baseline clinicopathological features. As would be expected, N-stage, T-stage, LVI and NPI score all had a significant impact on survival.

A BMI of 25-29.9kg/$m^2$ (categorised as overweight by the World Health Organisation) was also significantly associated with an increased risk of death and this remained significant after adjusting for nodal stage and stratifying by LVI. The negative impact of BMI on prognosis on the POSH cohort as a whole has been published(200), but has not been studied in the TNBC subtype within this cohort specifically. A large meta-analysis found an association between obesity and poorer breast cancer specific survival in pre- and post-menopausal women, but did not analyse by subtype(235). A previous study of just 118 non-metastatic TNBC cases found an association between obesity and mortality(236). Of note, they did not find this association for the other breast cancer subtypes. The authors of this study noted previous work highlighting a possible interaction between obesity, elevated adipokines (e.g. leptin) and the prognosis of triple negative tumours. There is increasing evidence that the chronic inflammatory environment of adiposity contributes to enhanced proliferation and metastatic potential of breast cancer cells(237).

The positive prognostic effect of anthracycline chemotherapy relative to anthracycline and taxane combination is likely to be related to treatment selection bias; patients treated with anthracyclines only were likely to have tumours with comparatively favourable baseline features; combination chemotherapy would have been given to those with more high-risk disease. In the period during which POSH recruited, taxanes were only given to those with node-positive tumours. In keeping with this, the effect disappeared after MVA. The survival results were similar for mastectomy but with a negative prognostic effect that was not present after MVA.

Adjuvant radiotherapy was not prognostic on univariable analysis, but it was found to be significantly prognostic after MVA (HR 0.587). This implies a suppression effect of N-stage and LVI in univariable analysis. The reason for this may be that radiotherapy was given to higher risk patients (e.g. higher T-stage), meaning that the adverse pathological features were confounding factors, resulting in radiotherapy not being found to have a survival benefit in univariable analysis. However when N-stage is controlled for in MVA, a beneficial effect of radiotherapy is found. A large meta-analysis has shown the beneficial effect of radiotherapy in reducing breast cancer recurrence and death, but this only included patients treated with BCS (238). The analysis here also includes patients who have undergone a mastectomy and is therefore likely to have more confounding factors due to the heterogeneity of the population. When this is controlled for, to some degree, in my MVA, the beneficial effect of radiotherapy is revealed.

### 3.3.5 Limitations

This study is limited by some elements of the POSH trial. Patients were not recruited consecutively. There is a possibility there was a degree of selection bias; instead of enrolling in an observational trial, young patients with metastatic disease may have preferred to opt into an interventional trial. As invitation to patients to participate in the

trial was at the discretion of the treating physician/surgeon, there was potentially bias against recruiting patients with metastatic disease.

Additionally, recruitment for this study took place between 2002 and 2008. This has the benefit of long follow-up but means that some treatment modalities may be outdated. There have been changes in systemic options available since recruitment took place, although most patients in this study were treated with anthracycline +/- taxane chemotherapy and approximately half received trastuzumab if HER2+(196). Trastuzumab was only approved as adjuvant therapy in 2005, a number of years after the trial commenced recruitment. Therefore patients recruited early in the POSH timeline may have received it only in the metastatic setting. Additionally, recent developments in the treatment of TNBC, particularly with regards to platinum therapy and PARP inhibitors, means that outcomes will likely be improved for patients now diagnosed with TNBC. In terms of the metastatic work, there is an increasingly proactive approach to staging investigations, including the use of advanced imaging, such as positron emission tomography. This may mean that more patients are now diagnosed with occult metastases at presentation, affecting the characteristics of the *de novo* group. Finally, the youth of the cohort means that our findings may not be translatable to the TNBC population as a whole. However, as detailed previously, POSH participants recruited from England represented 23% of the available population during the recruitment period and comparison with cancer registry data confirmed that the cohort is representative of the wider population of those with breast cancer diagnosed at 40 years or younger(201). The strengths of this study include the large cohort size and complete germline *BRCA* testing. There are few missing data (with HER2 status missing in only 5.7% of cases) and long follow-up, with only a small number of patients lost to follow-up.

For the early-stage TNBC dataset, cases treated with neoadjuvant chemotherapy (n=104) were excluded from the dataset for reasons explained in Section 2.2.4. This does leave the data susceptible to a certain amount of selection bias. For instance, some of the larger tumours (few were T4 in this dataset) with more adverse biological features are likely to have been excluded. To some extent, this limits the applicability of the data

presented, as there is increasing use of neoadjuvant chemotherapy in the modern era. Future analyses could incorporate the 104 cases excluded here.

## 3.4 Summary

In this chapter, I have shown that young women who develop secondary metastatic disease, even if greater than five years after diagnosis, have shorter survival time following diagnosis of metastases compared to those who present with *de novo* metastatic disease. For young women who develop metastatic disease, outcomes are worse for those with TNBC compared to the other subtypes as a whole. Of note, brain metastases did not confer a worse outcome for those with metastatic TNBC. For the early-stage TNBC cohort, an increased BMI was associated with a poor prognosis. Other prognostic features were as anticipated (N-stage, T-stage and LVI).

Advances in genomic and transcriptonomic analysis have revealed a host of potential targets in TNBC, including immune signatures and the tumour microenvironment. There is a drive to identify patients who will benefit from immunotherapy and means of enhancing immunotherapy efficacy. I therefore proceeded to assess the prognostic associations of a number of potential molecular biomarkers in the early-stage TNBC cohort.

# Chapter 4 Molecular Prognostic Factors in Young Breast Cancer

## 4.1 Introduction

A biomarker is an objective characteristic that can be measured in a reproducible way. A prognostic biomarker influences or predicts the outcome from a disease, whereas a predictive biomarker indicates the likely benefit derived from a particular treatment. Biomarkers can be useful to guide clinical decision-making and are increasingly being used to guide therapeutic strategies and patient selection(239). Some possible biomarkers identified in basal breast cancer/TNBC include p53, Ki67, ALDH1, EGFR and CK5/6 (213, 240-245). In the introduction, I outlined the evidence underlying TILs as a prognostic biomarker (Section 1.4). These biomarkers have not been evaluated specifically in a young age group. Breast cancer arising in young women has been shown to have different gene expression patterns compared to the older age group and biology in this cohort is more aggressive(246).

As markers of the basal phenotype, EGFR and cytokeratin 5/6 (CK 5/6) have long been established as potential markers of poor prognosis in TNBC(211, 242). However, discordance between the PAM50 intrinsic subtype and the IHC-based surrogate markers has been reported and basal TNBCs have even been suggested to be more chemosensitive (with an associated improved prognosis) compared to their EGFR-ve/CK 5/6-ve counterparts(247, 248). Therefore, further evaluation of the prognostic value of these markers in a large cohort of TNBC cases is required.

Ki-67 is a proliferation marker and its level has been associated with poor prognosis in a variety of cancers(241). It has been established as a useful prognostic marker in ER-

positive breast cancer(208). Its significance in TNBC is unclear as it has been associated with a poor prognosis in some studies but not in others(244, 249, 250). Similarly, p53 has variably been associated with poor prognosis in TNBC(240, 244). ALDH-1 is a marker of normal and malignant breast stem cells(251). Cancer stem cells (CSCs) are believed to be highly tumorigenic; effective markers to define CSCs have not been fully delineated and may differ between tumour subtypes(252). ALDH-1 has been correlated with a poor prognosis in breast cancer in general(253). Studies evaluating its prognostic relevance in TNBC have been small and contradictory(252, 254-257). Further evaluation is therefore required in a large cohort.

The hypothesis was that EGFR, p53, Ki67, ALDH1 and CK 5/6 would be independent adverse prognostic features. It was anticipated that increasing stromal TILs would be protective, given this has previously been found in the all-ages population with TNBC. Using tissue microarrays (TMAs), I performed a pre-planned molecular analysis of the TNBC cohort using a panel of these potential biomarkers. I looked for a correlation between these markers and survival, in order to ascertain prognostic features in this cohort. I adjusted for established prognostic clinicopathological parameters to determine whether these markers were independently prognostic.

## 4.2 Results

### 4.2.1 TNBC Cohort from POSH

#### 4.2.1.1 The pathology cohort

There were 505 patients (see Section 3.2.2.1) who were defined as TNBC and were treated with either adjuvant (n=491) or no chemotherapy (n=14). There was no archival pathology material available for 134 patients, either because the material was being used as part of another study or the tissue was never received from the local centre. In a further 21 cases, there was insufficient or absent tumour tissue available, for instance in some cases there were already tissue cores removed for other studies and only DCIS remaining on the block. The pathology cohort was therefore the 350 patients from whom there was suitable pathology material available (69.3% of the total cohort of 505) (Figure 4.1).

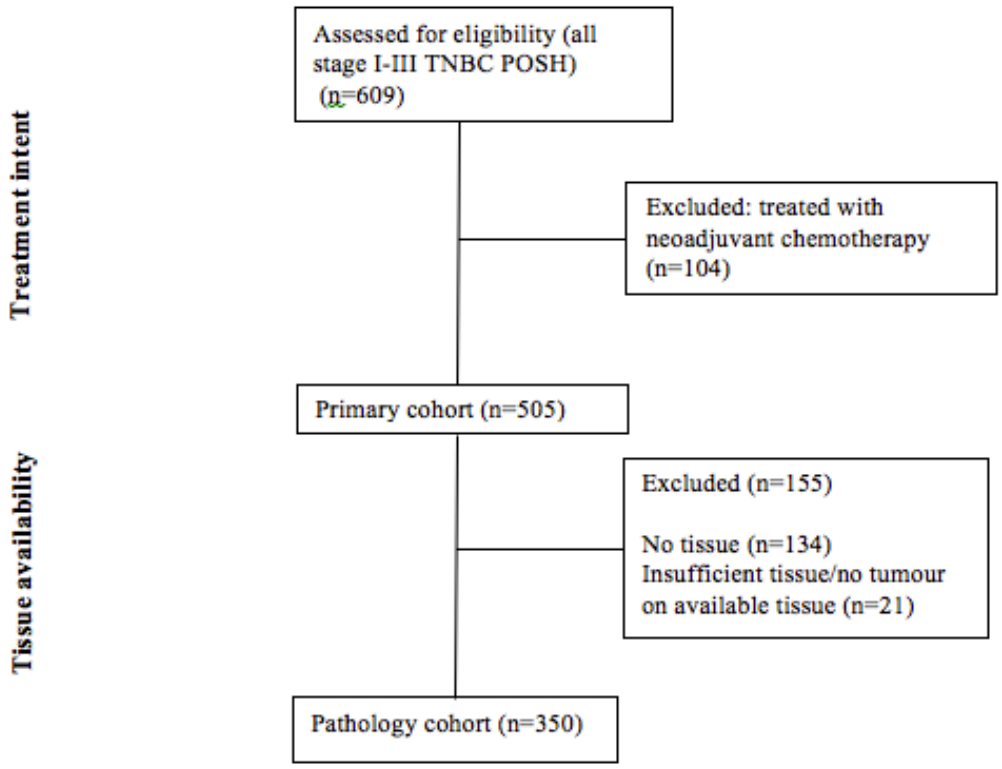



**Figure 4-1. CONSORT diagram for all stage I-III TNBC patients in POSH study.**

**N=609, with 505 patients included in clinicopathological analysis and 350 patients with tissue available for further testing.**

#### 4.2.1.2 Molecular analysis and correlation with survival

The frequencies of expression of the biomarkers, in addition to the univariable and multivariable hazard ratios for survival for each marker, are shown in Table 4.1. The univariable analyses can be viewed in a forest plot in Figure 4.2. The thresholds for each marker were chosen according to literature review and are described in Methods (Section 2.2.8). Stromal TIL scoring revealed that 7.1% had ≥60% infiltration or "lymphocyte-predominant breast cancer" (LPBC) (25/350). The remainder were categorised as moderate (45.4%) or low (47.4%). The majority were CK5/6 and/or EGFR-positive, consistent with a basal phenotype. Most cases were low in ALDH1. Approximately half of cases were p53-high and half were Ki67-high.

**Table 4-1. Hazard ratios for mortality for molecular panel performed on tissue microarray samples from stage I-III TNBC patients from the POSH cohort. Excluding those treated with neoadjuvant chemotherapy (n=350). Univariable and multivariable hazard ratios obtained from Cox regression analysis.**

| | No. (%[1]) | Univariable HR (95% CI) | p-value | Multivariable HR[2] | p-value |
|---|---|---|---|---|---|
| **P53** | | | | | |
| Low (<10%) | 130 (40.1) | 1 | | 1 | |
| High (>=10%) | 194 (59.9) | 1.131 (0.725-1.764) | 0.589 | 0.966 (0.610-1.531) | 0.884 |
| Unknown[3] | 26 (7.4) | | | | |
| **Ki67** | | | | | |
| Low (<20%) | 144 (44.0) | 1 | | | |
| High (>=20%) | 183 (56.0) | 0.830 (0.539-1.277) | 0.396 | 0.781 (0.492-1.241) | 0.296 |
| Unknown | 23 (7.0) | | | | |
| **CK5/6** | | | | | |
| Negative | 111 (35.0) | 1 | | 1 | |
| Positive (≥1%) | 206 (65.0) | 0.778 (0.499-1.214) | 0.269 | 0.813 (0.514-1.287) | 0.378 |
| Unknown | 33 (9.4) | | | | |
| **Stromal TILs** | | | | | |
| Low (≤10%) | 166 (47.4) | 1 | | 1 | |
| Moderate (11-59%) | 159 (45.4) | 0.492 (0.317-0.765) | **0.002** | 0.528 (0.336-0.829) | **0.006** |
| High (≥60%) | 25 (7.1) | 0.095 (0.013-0.684) | **0.019** | 0.139 (0.019-1.024) | 0.053 |
| **ALDH1** | | | | | |
| Low (absent/weak[4]) | 243 (76.2) | 1 | | 1 | |

| | No. (%[1]) | Univariable HR (95% CI) | p-value | Multivariable HR[2] | p-value |
|---|---|---|---|---|---|
| High (moderate/strong) | 76 (23.8) | 0.723 (0.412-1.267) | 0.257 | 0.822 (0.467-1.447) | 0.497 |
| Unknown | 31 (8.9%) | | | | |
| **EGFR** | | | | | |
| Negative | 32 (9.9%) | 1 | | 1 | |
| Positive (≥1%) | 292 (90.1%) | 1.869 (0.756-4.622) | 0.176 | 2.273 (0.918-5.632) | 0.076 |
| Unknown | 26 (7.4%) | | | | |

[1]Percentages are calculated as a total of the available data for that category. For unknown data, the percentage is calculated for the whole cohort (n=350)

[2]Multivariable analysis, adjusted for N-stage and stratified by LVI

[3]Unknown data is due to the absence of tissue blocks (n=3) or due to failure of the stain/TMA, e.g. due to absence of tumour tissue on all three cores or significant artefact/folding

[4] Mean expression calculated, range 0-2. For moderate/strong expression, the mean was greater than 0.333, out of a maximum of 2. For more details see Methods.

Of all the prognostic biomarkers studied, the only marker with significant impact on OS was stromal TILs. A high TIL score (≥60%) was markedly associated with a reduced risk of death from breast cancer, with a hazard ratio of 0.095 (0.013-0.684; p=0.002) relative to low (p=0.040). The hazard ratio for those with moderate TIL infiltration was 0.492 (0.317-0.765; p=0.002), relative to low. Therefore, there has been a failure to reject the null hypotheses for all the other markers.

When classified by p53 mutation pattern, with a mutant pattern being absent or >50% staining positive (according to previously defined criteria(207)), there was no difference in survival. There were 273 cases labelled as mutant and 51 as wild-type.

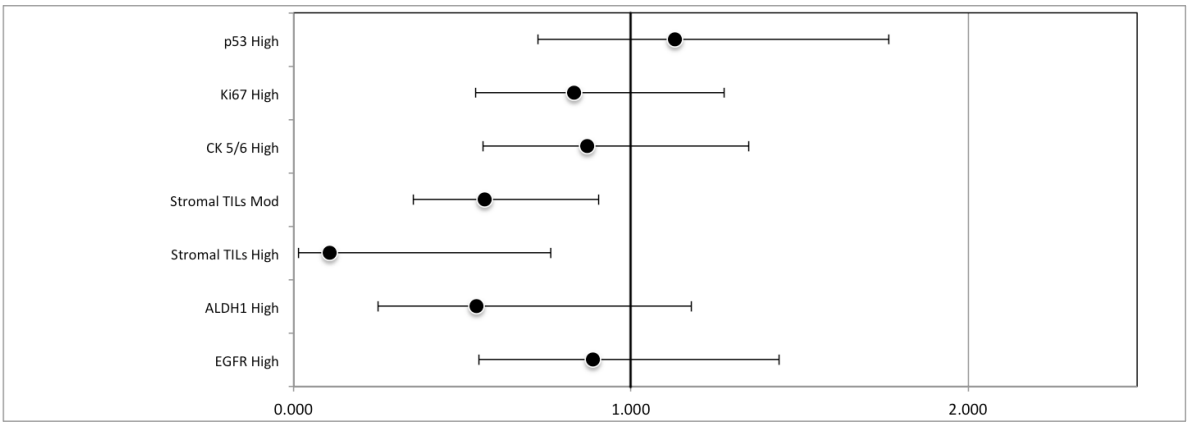



**Figure 4-2. Forest plot of univariable hazard ratios of histological factors for overall survival for stage I-III TNBC cases from the POSH cohort.**

**N=350. Reference line shown (1.00) – set as low for all variables (see Table 4.1).**

#### 4.2.1.3 Basal markers

The univariable and multivariable HRs for basal categories (dependent on the expression of EGFR and CK5/6) are shown in Table 4.2. EGFR is shown at two cut-offs: positive (≥1%) and negative, in addition to high (mean intensity score ≥1.5/3) and low. CK5/6 is shown at two cut-offs: positive (≥1%) and negative, in addition to high (mean intensity score ≥0.5/1) and low. The frequency of basal (either CK5/6 or EGFR positive) and "high basal" (either CK5/6 or EGFR scored as high) are also shown. The rationale behind these cut-offs is described in Section 2.2.8.

There was no relationship between g*BRCA* mutation status and p53, EGFR, CK5/6 or Ki67. Its relationship with TIL score will be discussed in the next Chapter.

**Table 4-2. Additional frequencies and hazard ratios for death for basal criteria in stage I-III TNBC patients from the POSH cohort.**
**Excluding those treated with neoadjuvant chemotherapy (n=350). Univariable and multivariable hazard ratios obtained from Cox regression analysis.**

| | No. (%[1]) | Univariable HR (95% CI) | p-value | Multivariable HR[2] | p-value |
|---|---|---|---|---|---|
| **CK5/6** | | | | | |
| Negative | 111 (35.0%) | 1 | | 1 | |
| Positive | 206 (65.0%) | 0.778 (0.499-1.214) | 0.269 | 0.813 (0.514-1.287) | 0.378 |
| Unknown[3] | 33 (9.4%) | | | | |
| **CK5/6** | | | | | |
| Low | 133 (42.0%) | 1 | | | |
| High | 184 (58.0%) | 0.871 (0.562-1.350) | 0.538 | 0.968 (0.614-1.526) | 0.889 |
| Unknown | 33 (9.4%) | | | | |
| **EGFR** | | | | | |
| Negative | 32 (9.9%) | 1 | | 1 | |
| Positive | 292 (90.1%) | 1.869 (0.756-4.622) | 0.176 | 2.273 (0.918-5.632) | 0.076 |
| Unknown | 26 (7.4%) | | | | |
| **EGFR** | | | | | |
| Low | 183 (56.5%) | 1 | | 1 | |
| High | 141 (43.5%) | 0.841 (0.542-1.303) | 0.438 | 0.849 (0.545-1.325) | 0.472 |
| Unknown | 26 (7.4%) | | | | |
| **Basal**[4] | | | | | |
| Non-basal | 13 (4.0%) | 1 | | 1 | |
| Basal | 308 (96.0%) | 1.263 (0.399- | 0.691 | 1.237 (0.389- | 0.719 |

| | No. (%[1]) | Univariable HR (95% CI) | p-value | Multivariable HR[2] | p-value |
|---|---|---|---|---|---|
| | | 4.003) | | 3.935) | |
| Unknown | 29 (8.3%) | | | | |
| **High basal**[5] | | | | | |
| Non/low | 78 (24.7%) | 1 | | 1 | |
| High | 238 (75.3%) | 0.806 (0.497-1.307) | 0.382 | 0.835 (0.507-1.374) | 0.478 |
| Unknown | 34 (9.7%) | | | | |
| **High basal category**[6] | | | | | |
| Positive for EGFR or CK5/6 | 151 (63.4%) | 1 | | | |
| Positive for both | 87 (36.6%) | 0.852 (0.496-1.463) | 0.561 | 0.910 (0.521-1.590) | 0.742 |

[1]Percentages are calculated as a total of the available data for that category. For unknown data, the percentage is calculated for the whole cohort (n=350)

[2]Multivariable analysis, adjusted for N-stage and stratified by LVI

[3]Unknown data is due to the absence of tissue blocks (n=3) or due to failure of the stain/microarray, e.g. due to absence of tumour tissue on all three cores or significant artifact/folding

[4]Basal is defined as CK5/6 and/or EGFR expression as ≥1%

[5]High basal is defined as mean expression CK5/6 ≥0.5/1 and/or EGFR mean expression ≥1.5/3

[6]High basal – positive for one high basal marker (CK5/6 or EGFR) is used as a reference for Cox regression, negative for both/unknown excluded from survival analysis for this category

#### 4.2.1.4 Further survival analysis using TIL score

Given the significance of the univariable Cox regression for survival according to TIL-score (Table 4.1), a Kaplan Meier survival analysis was performed. A TIL score of moderate or high were both associated with improved OS (p=0.0002 log-rank Mantel Cox) (Figure 4.3).

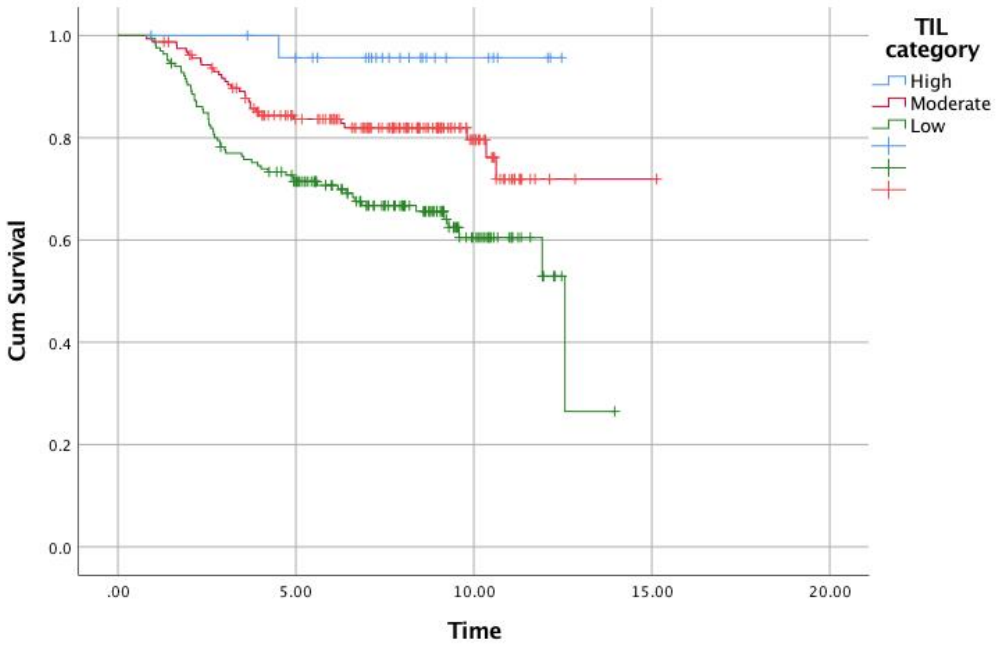



**Figure 4-3. Kaplan Meier survival analysis for stage I-III TNBC patients in POSH cohort with stromal TIL score evaluated.**

**TIL-score: High (lymphocyte-predominant breast cancer) = ≥60% TILs (n=25), Moderate = 11-59% TILs (n=159), Low = ≤10% (n=166). Log-rank Mantel Cox p=0.0002.**

A multivariable Cox regression analysis using TILs as a continuous numerical variable, adjusting for N-stage and stratified by LVI, is shown in Table 4.3. It shows that TIL percentage is associated with a reduced risk of death for every unit increase and is an independent prognostic factor for survival.

**Table 4-3. Multivariable Cox regression analysis for overall survival. N=337.**

| Variable | Hazard Ratio | Standard Error | P-value | 95% Confidence Interval |
|---|---|---|---|---|
| TIL percentage[2] | 0.983 | 0.007 | 0.019 | (0.968, 0.997) |
| N-stage 1[1] | 2.390 | 0.261 | 0.001 | (1.433, 3.986) |
| N-stage 2[1] | 3.367 | 0.315 | <0.001 | (1.818, 6.236) |

[1] N-stage 0=reference category. A HR>1 represents an increased risk in death for patients compared to N-stage 0 patients.

[2] A HR<1 represents a reduced risk of death for every unit increase in TIL%

## 4.3 Discussion

In this chapter, I performed TMA staining to evaluate the prognostic effect of a panel of biomarkers that have been variably associated with prognosis in TNBC in the literature, but require further evaluation in a large cohort of this kind. There were 350 tumours from stage I-III TNBC cases from the POSH cohort included.

### 4.3.1 Ki67

In terms of the biomarkers assessed, Ki67 is a well-established marker of cell proliferation, although not in routine use clinically(241). I used a cut-off of ≥20% as the threshold between high and low proliferation categories, in line with other studies(208-210). In a study of 1577 ER+ breast cancer cases, 20% was defined as optimal to stratify high-risk patients (reduced disease-free and overall survival)(208). Using TMAs from 707 breast cancers (all subtypes), this cut-off again was most useful for prognosis, with a high proliferation index being associated with large tumours and nodal metastasis(209). However, this cut-off was developed in studies using ER+ cases and may not be appropriate in TNBC specifically.

Ki67 is a nuclear protein, the expression of which varies throughout the cell cycle. It has been established as a prognostic biomarker in luminal cancer, but its importance in TNBC is unknown. It is a difficult marker to standardise. A high Ki67 has been associated with adverse outcomes in one study of 156 cases of TNBC(244). Concordant results have been found in another smaller study(243). In the latter study using the same cut-off, Ki67 positivity was 74.7%. The rate was higher (83.3%) in the larger study but a cut-off of 14% was used(244). The positivity rate in this present study was lower at 56.0%. This is comparatively low compared to other studies. Given that both the quoted studies appeared to use full-face sections, it may be that the nature of TMAs has resulted in some

areas of high-Ki67 staining being missed by random selection. Additionally this marker is traditionally used in ER+ disease and may not be prognostic in the younger TNBC population. This marker could be used as a continuous variable instead of categorical in future survival analyses.

### 4.3.2 p53

P53 is a tumour suppressor gene that regulates multiple processes, including the cell cycle and apoptosis, giving it the title of “guardian of the genome”(258). As discussed in the introduction, p53 is one of the most commonly mutated genes in breast cancer, particularly in the TNBC subtype.

Mutant p53 by missense mutations often result in increased stability of the protein and accumulation/detection by IHC(206). A truncating or nonsense p53 mutation can result in undetectable p53. Additionally, normal p53 can accumulate at detectable levels due to DNA damage(259). Therefore, p53 expression by IHC appears to be a poor surrogate for the underlying mutation status of the gene. This may explain why previous IHC studies attempting to correlate p53 expression with outcome in breast cancer have been contradictory (244).

In one study of 156 TNBC cases, p53 expression (positive in 71.3%) correlated with poor outcomes(244). However in two larger studies, p53 was positive in 49.9% and 56.4% and associated with better OS(258, 259). In one of these studies, the authors suggested that wild-type p53 accumulating as a mechanism to repair DNA damage accounts for its expression being associated with an improved prognosis(258). In this study we found “any” p53 positivity to occur at a rate of 68.2%. Using a stricter criteria of >=10% of tumour cells staining positive, the rate was 59.9%. In a study of 7739 patients with all subtypes of breast cancer, a ROC analysis was performed to define the optimal cut-off for

IHC staining of p53 predictive for survival. The lowest p-value to predict OS across all cases (and luminal A specifically) was at 10% (205). There was no association between p53 expression and OS found in this study and so unfortunately the relationship remains ambiguous.

### 4.3.3 ALDH-1

ALDH-1 is a marker of normal and malignant breast stem cells(251). Cancer stem cells (CSCs) are believed to be highly tumorigenic; effective markers to define CSCs have not been fully delineated and may differ between tumour subtypes(252). In a small study that included 56 TNBCs, ALDH1 correlated with the basal subtype and with more extensive lymphocytic infiltration(252). In another small study of 180 breast cancers (unclear number of TNBCs), higher ALDH-1 stromal expression was correlated with longer survival(245). However, this study was limited by short survival times and the relevance of ALDH-1 within subtypes was not clear. ALDH-1 has been correlated with ER/PR-negativity and the basal subtype, but its prognostic effect within those subtypes is not clear(252, 253, 260). There have been no previous studies specifically evaluating the prognostic impact of this marker in TNBC. Here, we have shown no significant correlation between ALDH-1 and survival. Its relationship with lymphocytic infiltrate will be explored in the next chapter.

### 4.3.4 Basal markers

EGFR and CK5/6 can be considered surrogate IHC markers of basal-like breast cancer(211, 261). They are considered positive if ≥1% cytoplasmic or membranous tumour cell staining is positive(211, 212). It has previously been suggested that tumours classified as basal confer the poor prognosis amongst the TNBC group (with the addition of these two markers being significantly more prognostic than TNBC status alone) (211). However,

currently, classification into basal/non-basal subgroups is overall not felt to be clinically useful as a prognostic marker as the vast majority of patients with TNBC still require neoadjuvant or adjuvant chemotherapy (and 56-95% classify as basal, so not particularly discriminatory)(262). However, given the evolution of potentially effective and specific treatments for TNBC, such as immunotherapy, PARP inhibitors and platinums, further molecular classification may evolve to have a role in precision therapeutics. In a study of 258 consecutive TNBC cases in Italy, EGFR positivity was found to be associated with worse OS for thresholds of ≥30% and ≥50% (more pronounced for the latter) but not for ≥10%(215). In another study of 192 consecutive TNBC cases in the USA, combined expression of EGFR and CK5/6 (but only at thresholds of 15% and 50% respectively; defined by ROC analysis) was predictive of poor outcome, but neither marker was prognostic alone(216). In another study of 352 TNBCs, EGFR expression (threshold ≥1%; positive in 53.5%) did not correlate with DFS or OS.(206). Using fluorescent in-situ hybridization (FISH), Sanger and next-generation sequencing and qPCR of mRNA (all from FFPE tissue), the authors of this study stated that EGFR gene amplification/mutations are unusual in TNBC. Of note, when EGFR and p53 IHC were evaluated together, they found patients with EGFR-ve/p53-ve or EGFR+ve/p53+ve to have the best prognosis.

The proportion of cases that were classified as basal was high (96.0%). It must be noted that EGFR and CK5/6 are only considered surrogate IHC markers of this molecular classification. The unusually high proportion classified as basal here may be related to the staining thresholds used. Typically, EGFR positivity in TNBC has been reported as 42-76%(206). Here 90.1% of cases were EGFR positive if the ≥1% criteria were used. Using a threshold of mean intensity score ≥1.5/3 (“EGFR high”), 43.5% were positive. CK5 was positive in 71-76% of TNBC cases in the studies discussed above(206, 216); 65.0% here. Using stricter criteria of mean intensity score ≥0.5/1 (“CK5/6 high”), 56.3% were positive. There was no impact on prognosis for any of these categories, even when combined into basal/non-basal or high-basal/low-basal categories. When the high basal group was evaluated alone, there was no impact on prognosis when double-positive (both CK5/6 and EGFR “high”) tumours were compared to single-positive (only one marker “high”)

tumours, unlike in the study above where there was a significantly worse prognosis for double-positives(216).

### 4.3.5 Stromal TILs

Of this panel of biomarkers, only stromal TILs were found to be prognostic for overall survival. The Kaplan Meier survival analysis using TILs as a categorical variable was striking (Figure 4.3). Using TILs as a continuous variable, after MVA adjustment for nodal stage and stratification by LVI, TILs were independently prognostic with a hazard ratio of 0.983 (0.968-0.997; p=0.019) for every 1% increase in TIL percentage.

To my knowledge, TIL infiltration has never been specifically evaluated in the young breast cancer population before. In one of the largest studies evaluating the prognostic impact of TILs in TNBC managed with adjuvant treatment, only 88 patients (of 647) were aged less than 40. In this study, younger patients had a higher median TIL score compared to other age groups (26%; IQR 10-52%)(45). Given that biology is more adverse in this younger age group with worse outcomes and that there may be a difference in TIL infiltration, it was prudent to confirm that TILs are prognostic in this cohort.

### 4.3.6 Limitations

This study is limited by our methodology of scoring EGFR and CK5/6. Instead of the percentage of cells staining positive, we scored in staining intensities of 0-3 and 0-1 respectively. This was done for efficiency reasons; each antibody was evaluated by a single pathologist on greater than 1000 TMA cores. The use of a single pathologist did mean that marking was more likely to be consistent, given the absence of inter-observer variability. In order to determine which markers are expressed above the threshold of

≥1% this marking system is acceptable, as a score of zero was only given for absence of staining. However, we are unable to compare our results to studies using a threshold of, e.g. 15% or 50%, as we do not have the equivalent information. For high expression, instead of 50% threshold, we have used mean intensity ≥0.5/1 or ≥1.5/3 as thresholds but this is not expected to be equivalent. In the future I would utilise a semi-quantitative system to evaluate these markers to increase the complexity and discrimination of the scoring. In order to optimise cut-off thresholds for use in TNBC specifically, further statistical work, e.g. using ROC analysis, could be employed.

Immunohistochemistry (IHC)-based classification is a somewhat crude surrogate for the complex genomic alterations resulting in altered proteins. However, its use is accessible and affordable, so when clinically relevant can be utilised widely in clinical practice. Gene expression profiles could be performed on a select number of cases and compared to the IHC classification in order to confirm the accuracy of these markers as a surrogate of the underlying genetic changes.

Finally, there may be a degree of selection bias in terms of the blocks that were available for analysis. As work has been done on some of the tissue previously, it may be that some of the smaller tumours have insufficient tissue left and are thus not included in this analysis. In order to investigate this, the characteristics of the cases excluded based on tissue availability could be compared to the dataset. On the other hand some of the larger tumours would have received neoadjuvant chemotherapy; neoadjuvant cases were not included in the stage I-III cohort here (n=104/609).

## 4.4 Summary

I have shown that, notwithstanding nodal stage, TIL is the most prognostic variable predictive of outcome in young women with TNBC. In the next chapter, I explore this

further and attempt to delineate some correlates of lymphocytic infiltration in order to further understand possible mechanisms affecting this in TNBC.

# Chapter 5 Immunomodulation in Triple Negative Breast Cancer

## 5.1 Introduction

In the previous chapter, I used a panel of possible molecular markers to assess their correlation with prognosis in a cohort of young women with stage I-III TNBC. Of these markers, the most prognostic for overall survival was TILs. In this chapter, I explore TIL infiltration in this cohort in more detail, including correlation with clinicopathological parameters, association with immune markers and a more detailed prognostic analysis. The mechanisms by which some tumours are rendered immune "hot" and others immune "cold" in TNBC are unclear (see Section 1.4). There has been no association found between immunogenicity and the number of somatic mutations or neoantigens in breast cancer or melanoma(24, 148).

Here, I evaluated for associations between TIL score and baseline clinicopathological features. The POSH cohort is relatively unique for its complete germline *BRCA* testing. Tumours associated with a *BRCA1* mutation have long been suggested to have higher lymphocytic infiltration, although this has been contested(124, 263). This cohort offers an ideal resource for further evaluation of lymphocytic infiltration in *BRCA*-mutated cases, given the relatively large number of cases involved.

It is not clear whether a specific immune cell marker could be more predictive of outcome than TIL score alone and the impact of lymphocytic subtypes warrants further investigation. Comparing mRNA gene signatures amongst the TNBC molecular subtypes, a group have recently found differing proportions of adaptive and innate cells in the different subtypes(169). The immunomodulatory subtype was most enriched in adaptive

cells, whereas the mesenchymal stem cell like subtype was mainly composed of innate cells. Therefore, the immune cell composition was heterogenous between subtypes. The prognostic impact of FOXP3 expression in TNBC is unclear as few studies are powered to evaluate their prognostic impact within molecular subtypes(264). As discussed in the Introduction (Section 1.7.1), the impact of PD-L1 expression on prognosis in TNBC, particularly when evaluated in the context of the TIL infiltrate, is not clear.

In order to determine factors that may affect immune infiltration in TNBC more generally, I have utilised bio-informatics to interrogate a publicly available dataset, the Cancer Genome Atlas (TCGA)(73). This enables the inclusion of genomic and transcriptomic data within this exploration of immune modulation in TNBC.

I sought to utilize digital pathology in order to validate pathologist scoring and determine whether automated analysis could be used for prognostic evaluation in the future. In addition, this technology offers an opportunity to explore the spatial relationship between immune cells, by using multiplex immunohistochemistry (mIHC) of a selected number of full-face cases. mIHC enables detection of multiple markers on a single tissue section(265). Primarily I have sought evidence of the fully inflamed, stroma restricted and margin restricted phenotypes previously described(169).

Ultimately, by utilising tissue microarrays (TMAs), digital technology and bioinformatics, I have explored immunomodulation in TNBC. Aside from the TCGA dataset, the cases are from the POSH cohort. Although a limitation of the POSH data set is its limited age range, the size of the study, its prospective nature and completeness of data (including BRCA status) mean it is a unique and valuable resource. Additionally, the study was set up locally - this means I was well placed to access data and tissue from the trial.

## 5.2 Results

### 5.2.1 TILs & association with clinicopathological variables

In the previous chapter, it was shown that stromal TILs have been assessed in the 350 patients from whom there was suitable pathology material available (69.3% of the total cohort of 505). Of these, 25 had lymphocyte-predominant breast cancer (LPBC; (≥60% stromal TILs) (7.1%) (H&E examples of TIL scores are shown in Figure 5.1). The median TIL score was 15% (range 0-80%; IQR 5-35%) (Figure 5.2).

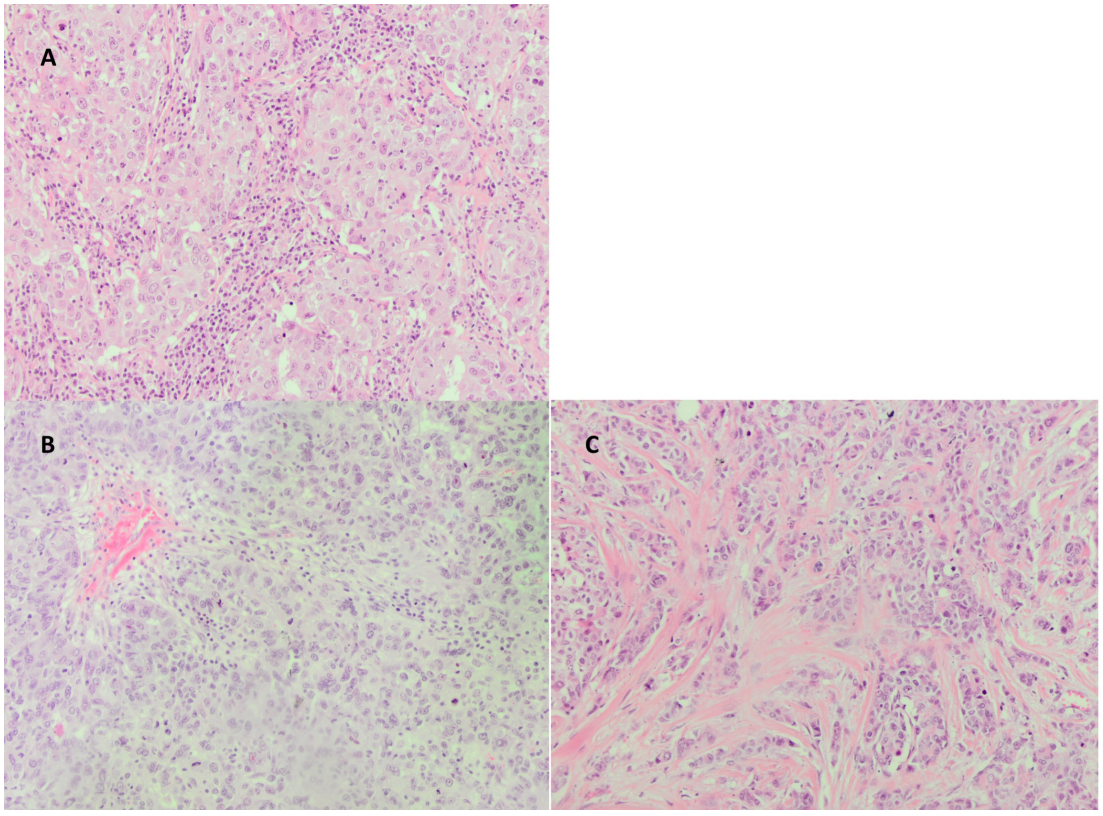



**Figure 5-1. H&E example images of varying TIL concentration.**

**A. Lymphocyte-predominant breast cancer (≥60%), B. TIL count 11-59%, C. TIL count ≤20%, from stage I-III TNBC cases from the POSH cohort. Stromal TILs measured as per International TILs Working Group guidelines(31). Resolution X20.**

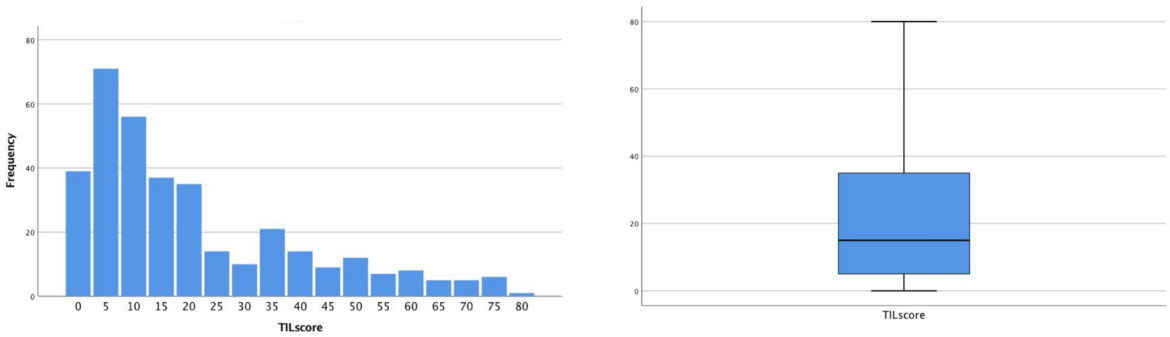



**Figure 5-2. Distribution of TIL score frequency.**
**Bar chart and box plot with median (15%) and IQR (5-35%) shown, for all patients with stage I-III TNBC included from the POSH cohort (n=350).**

There was a clear association with survival, with improved outcomes for those with moderate or high TIL infiltration (p=0.0002 log-rank Mantel Cox; see previous chapter for Kaplan Meier survival analysis).

The cumulative hazard ratio for survival over time, per TIL increments of 10%, shows the highest risk of recurrence or death in those with a TIL score of 0-10% and the lowest in those with LPBC (≥60%) (Figure 5.3). In the previous chapter I performed multivariable Cox regression, which showed that TILs were an independent prognostic factor for survival for every 1% increase.

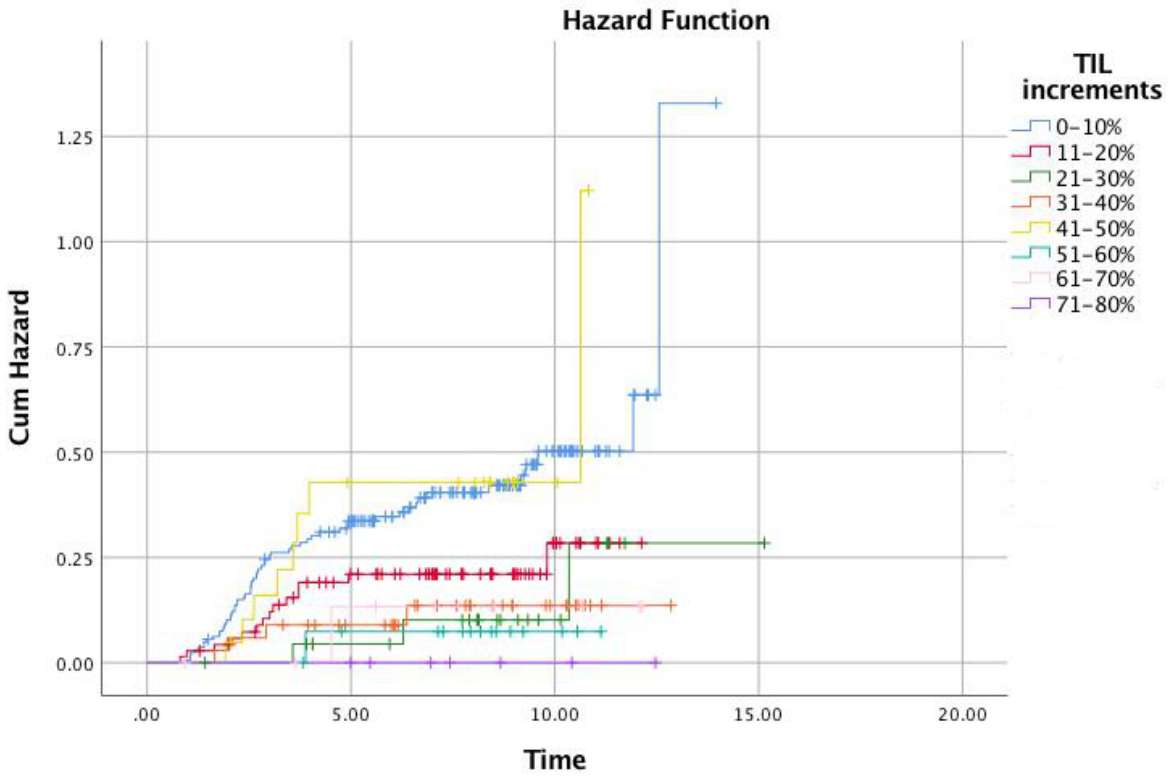



**Figure 5-3. Cumulative hazard estimates analysis for recurrence or death for TNBC patients in POSH cohort with TIL score performed.**

**Shown in increments of 10% (n=350). Log-rank (Mantel Cox) p=0.002**

I assessed for associations between TIL score and clinicopathological parameters. Using Pearson correlation coefficient, an association between TIL score and numerical variables was sought (Table 5.1). Of these, only age was found to have a significant correlation with TILs, with an r of -0.144 (p=0.007).

Using the non-parametric median test in SPSS, it was identified that the median TIL score was significantly lower in those with multifocal, LVI-positive and node-positive tumours (Table 5.2). The TIL score was higher in tumours from patients with a germline *BRCA1* or *2* gene mutation (20 vs. 10, p-value 0.043) (Figure 5.4).

**Table 5-1. Correlation of TIL score with ordinal variables.**
**N=350**

| Factor | Pearson's correlation coefficient | p-value |
|---|---|---|
| Age | **-0.144** | **0.007** |
| BMI | -0.081 | 0.133 |
| Tumour size | -0.078 | 0.146 |

**Table 5-2. Correlation of TIL score with categorical variables.**
**N=350**

| Factor | Category 1 (number) Median | Category 2 (number) Median | p-value |
|---|---|---|---|
| Focality | Multifocal (48) 10 | Localised (284) 15 | **0.023** |
| LVI | Present (133) 10 | Absent (206) 15 | **0.002** |
| Nodal involvement | Present (139) 10 | Absent (210) 15 | **0.018** |
| BRCA | Mutation (89) 20 | No mutation (261) 10 | **0.043** |
| Brain metastases | Yes (31) 10 | No (54) 10 | 0.635 |
| Bone metastases | Yes (40) 10 | No (45) 10 | 0.964 |
| Visceral metastases | Yes (61) 10 | No (24) 7.5 | 0.676 |
| 1$^{st}$ site of metastases | Non-visceral (19) 10 | Visceral (66) 10 | 0.423 |

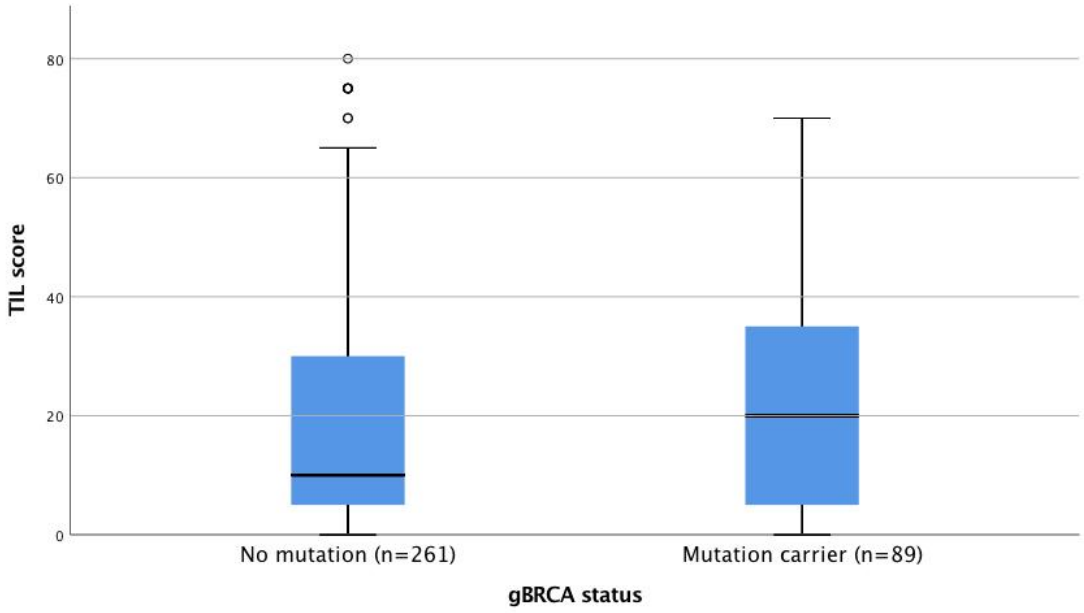



**Figure 5-4. Box plot showing median TIL score with IQR in cases, categorised by BRCA status.**

**Median TIL score 10 for those without a germline BRCA mutation (n=261) vs. 20 for those with a germline BRCA mutation (n=89) (p=0.043, SPSS non-parametric median test p-value). Stage I-III cases from the POSH cohort.**

### 5.2.2 TCGA analysis

Weighted gene co-expression analysis utilizes an algorithm to identify correlations between gene pathways in the context of a system network and has been used to good effect before in head and neck cancer(218). TNBC cases from the Cancer Genome Atlas (TCGA) database (a publically available dataset) were subdivided into Immunogenomic categories using the Thorsson criteria(266). The lymphocyte-deplete (C4) subtype was lowest and the IFN-γ _Dominant (C2) subtype highest in CD8 expression (Figure 5.5). Macrophage gene module expression was highest in C2 and C4 (Figure 5.6). Expression of the extracellular matrix (ECM) modules was consistently highest in the lymphocyte-deplete subtype (Figure 5.7).

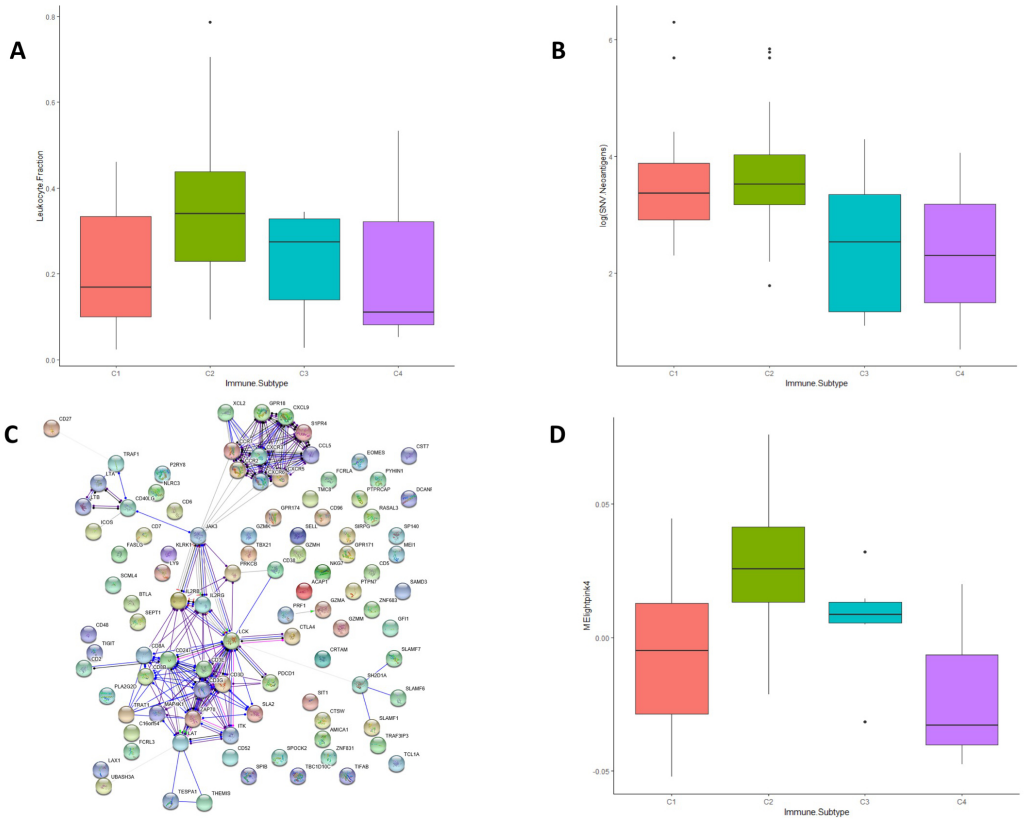



**Figure 5-5. TNBC cases from the Cancer Genome Atlas (TCGA) database categorised into Immunogenomic subtypes and analysed by lymphocyte markers.**

**Subtyped using the Thorsson criteria(249).**

**C1= wound healing, C2= IFN-gamma dominant, C3= inflammatory, C4= lymphocyte-depleted.**

**A. Percentage leukocyte fraction**

**B. SNV-neoantigens**

**C. Weighted gene coexpression network analysis (WCGNA) for CD8 T-cell module (lightpink4); image depicts the gene modules included in the analysis in the graph in 5.5D. The algorithm (WCGNA package in R) for WCGNA provides information regarding both the correlation between two genes and the degree of similarity between the correlation between these two genes and their correlation structure within the network as a whole. This contributes to the understanding of the interaction between the tumour, stroma and the immune system. This is described in more detail in Ottensmeier *et al.***

**D. CD8 T-cell module p<0.0001.**

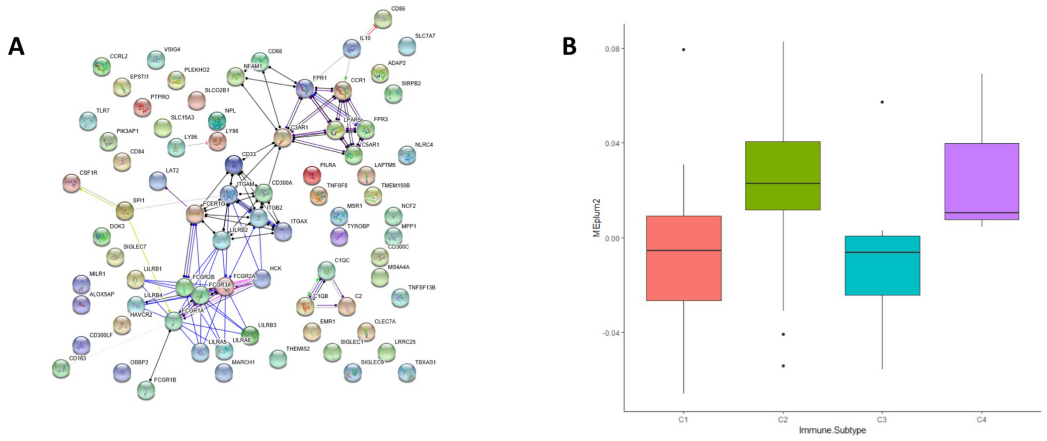



**Figure 5-6. TNBC cases from the Cancer Genome Atlas (TCGA) database categorised into Immunogenomic subtypes and analysed by macrophage gene module expression.**

**Subtyped using the Thorsson criteria(249).**

**C1= wound healing, C2= IFN-gamma dominant, C3= inflammatory, C4= lymphocyte-depleted. A. network analysis for CD68 macrophage (plum2) and B. Macrophage module p<0.0001.**

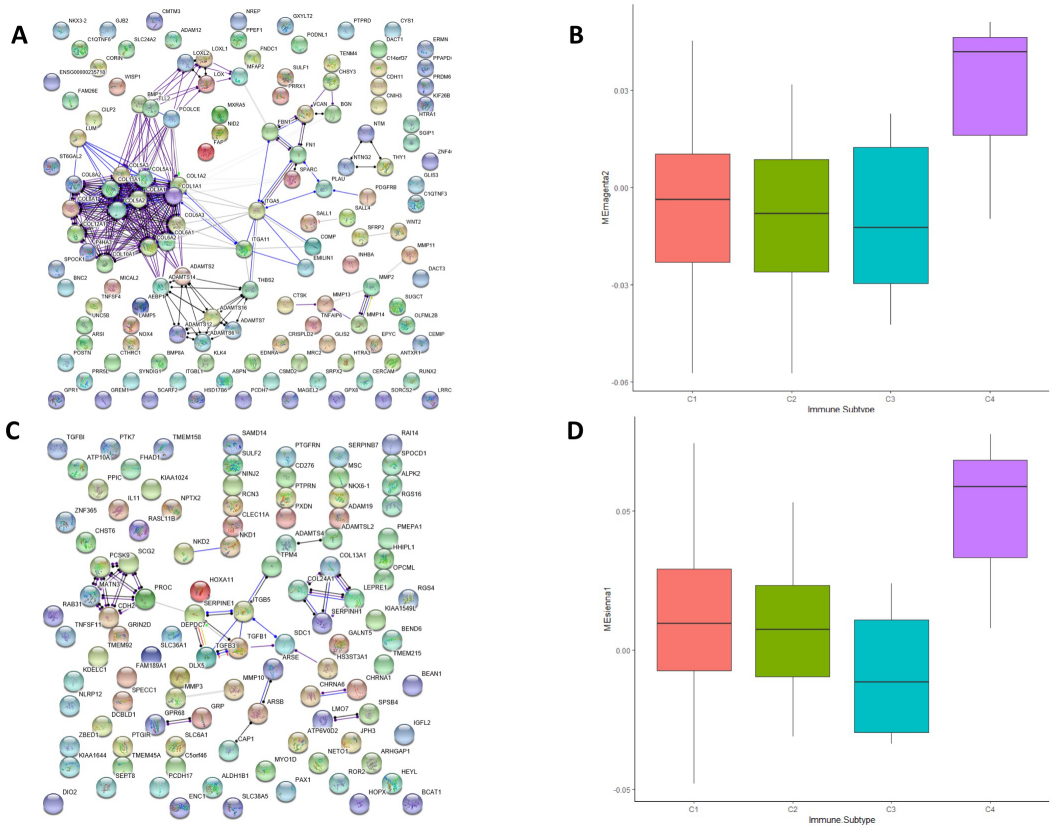



**Figure 5-7. TNBC cases from the Cancer Genome Atlas (TCGA) database categorised into Immunogenomic subtypes and analysed by ECM expression.**

**Subtyped using the Thorsson criteria(249).**
**C1= wound healing, C2= IFN-gamma dominant, C3= inflammatory, C4= lymphocyte-depleted. A. network analysis for ECM (magenta2 FN1), B. ECM magenta module p<0.0001, C. network analysis for ECM (sienna1 TGFB1/3), D. ECM sienna module p<0.0001.**

This raised the hypothesis of the extracellular matrix inhibiting CD8 expression. As fibroblasts are the primary cell type involved in production of the ECM, I decided to investigate smooth muscle actin (SMA; a marker of cancer-associated fibroblasts) expression in this cohort.

Using the same dataset, an inverse correlation between glycolysis and CD8 expression was found. I therefore included a marker of glycolysis, GLUT1, in the following analyses.

### 5.2.3 Immune cells

In order to undertake a more detailed analysis of the immune cells present in early-stage TNBC, I performed an immune panel using the TMAs, consisting of: CD3, CD4, CD8, FOXP3, smooth muscle actin (SMA), MHC 1, GLUT1, ERAP-1 and PD-L1. I then went on to assess for a correlation between these immune parameters and survival, using both univariable and multivariable (adjusting for nodal stage and stratifying by LVI) Cox regression analysis (Table 5.3). A Forest plot is shown in Figure 5.8. Thresholds for these categories were chosen as per Section 2.2.8.

**Table 5-3. Hazard ratios for mortality for histological parameters of stage I-III TNBC patients in POSH cohort.**
**Excluding those treated with neoadjuvant intent (n=350). Univariable and multivariable hazard ratios obtained from Cox regression analysis.**

| | No. (%[1]) | Univariable HR (95% CI) | p-value | Multivariable HR[2] | p-value |
|---|---|---|---|---|---|
| **Stromal TILs** | | | | | |
| Low (≤10%) | 166 (47.4) | 1 | | 1 | |
| Moderate (11-59%) | 159 (45.4) | 0.492 (0.317-0.765) | **0.002** | 0.528 (0.336-0.829) | **0.006** |
| High (≥60%) | 25 (7.1) | 0.095 (0.013- | **0.019** | 0.139 (0.019- | 0.053 |
| **SMA** | | | | | |
| Low (<2)[3] | 192 (58.9) | 1 | | 1 | |
| High (≥2) | 134 (41.1) | 1.207 (0.782- | 0.395 | 1.107 (0.711- | 0.652 |
| Unknown[4] | 24 (6.9) | | | | |
| **CD3** | | | | | |
| Low (<2) | 134 (41.0) | 1 | | 1 | |
| High (≥2) | 193 (59.0) | 0.598 (0.388- | **0.020** | 0.684 (0.439- | 0.094 |

| | No. (%[1]) | Univariable HR (95% CI) | p-value | Multivariable HR[2] | p-value |
|---|---|---|---|---|---|
| Unknown | 23 (6.6) | | | | |
| **CD4** | | | | | |
| Low (<1.333) | 171 (51.8) | 1 | | 1 | |
| High | 159 (48.2) | 0.514 (0.328- | **0.004** | 0.464 (0.265- | **0.007** |
| Unknown | 20 (5.7) | | | | |
| **CD8** | | | | | |
| Low (<1.667) | 165 (50.3) | 1 | | 1 | |
| High | 163 (49.7) | 0.508 (0.325- | **0.003** | 0.574 (0.365- | **0.016** |
| Unknown | 22 (6.3) | | | | |
| **FOXP3** | | | | | |
| Low (≤1) | 166 (50.5) | 1 | | 1 | |
| High (>1) | 163 (49.5) | 0.562 (0.360- | **0.011** | 0.658 (0.418- | 0.072 |
| Unknown | 21 (6.0) | | | | |
| **MHC 1** | | | | | |
| Low (<2) | 119 (36.3) | 1 | | 1 | |
| High (≥2) | 209 (63.7) | 0.530 (0.344- | **0.004** | 0.748 (0.471- | 0.219 |
| Unknown | 22 (6.3) | | | | |
| **GLUT1** | | | | | |
| Low (<2.333) | 161 (49.2) | 1 | | 1 | |
| High | 166 (50.8) | 0.819 (0.534- | 0.358 | 0.997 (0.600- | 0.992 |
| Unknown | 23 (6.6) | | | | |
| **ERAP1** | | | | | |
| Low (<2) | 49 (31.6) | 1 | | 1 | |

| | No. (%[1]) | Univariable HR (95% CI) | p-value | Multivariable HR[2] | p-value |
|---|---|---|---|---|---|
| High (≥2) | 106 (68.4) | 0.765 (0.412- | 0.397 | 0.737 (0.393- | 0.341 |
| Unknown | 195 (55.7) | | | | |
| **PD-L1 (tumour)** | | | | | |
| Absent[5] | 204 (63.4) | 1 | | 1 | |
| Present | 118 (36.6) | 0.653 (0.407- | 0.077 | 0.798 (0.488- | 0.369 |
| Unknown | 28 (8.0) | | | | |
| **PD-L1 (lymphocytes)** | | | | | |
| Absent | 169 (52.6) | 1 | | 1 | |
| Present | 152 (47.4) | 0.424 (0.267- | **<0.001** | 0.516 (0.321- | **0.006** |
| Unknown | 29 (8.3) | | | | |

[1]Percentages are calculated as a total of the available data for that category. For unknown data, the percentage is calculated for the whole cohort (n=350)

[2]Multivariable analysis, adjusted for N-stage and stratified by LVI

[3] = For SMA, CD3, CD4, CD8, FOXP3, MHC, GLUT1 and ERAP1, mean expression calculated, range 1-3. For high expression, the mean was greater than or equal to the median; for FOXP3 high expression was calculated as greater than the median to ensure the groups were balanced

[4]Unknown data is due to the absence of tissue blocks (n=3) or due to failure of the stain/microarray, e.g. due to absence of tumour tissue on all three cores or significant artefact/folding

[5] = For PD-L1, expression was categorised as ≥1% or absent

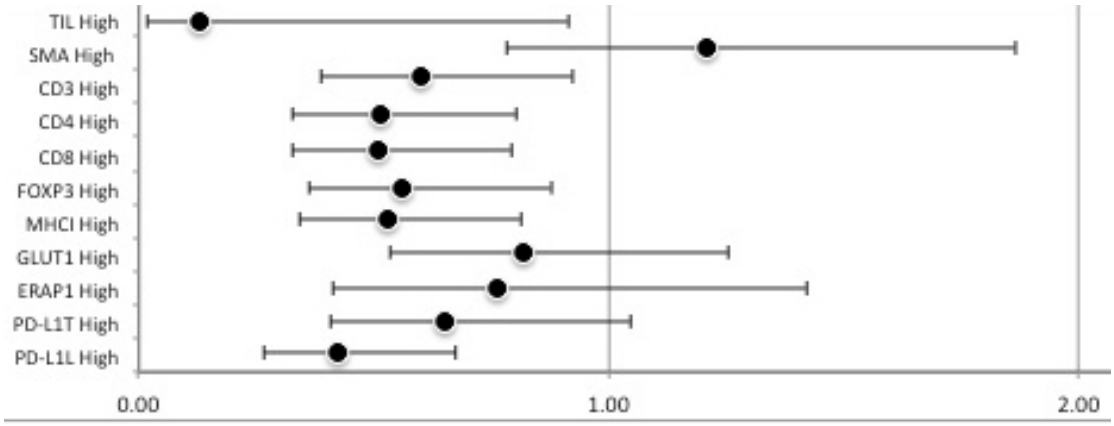



**Figure 5-8. Forest plot of univariable hazard ratios for immune markers and overall survival for stage I-III TNBC cases from the POSH cohort.**

**N=350. Immune markers determined by TMA scoring. Reference line shown (1.00) – set as low for all variables (see Table 5.3).**

Survival was significantly longer for those with CD3 high vs. low, CD4 high vs. low, FOXP3 high vs. low, CD8 high vs. low, MHC I high vs. low and PD-L1 (lymphocyte) high vs. low (Table 5.3; Figure 5.8). This remained significant for CD4, CD8 and PD-L1 following MVA. Selected Kaplan Meier survival analyses are shown in Figure 5.9.

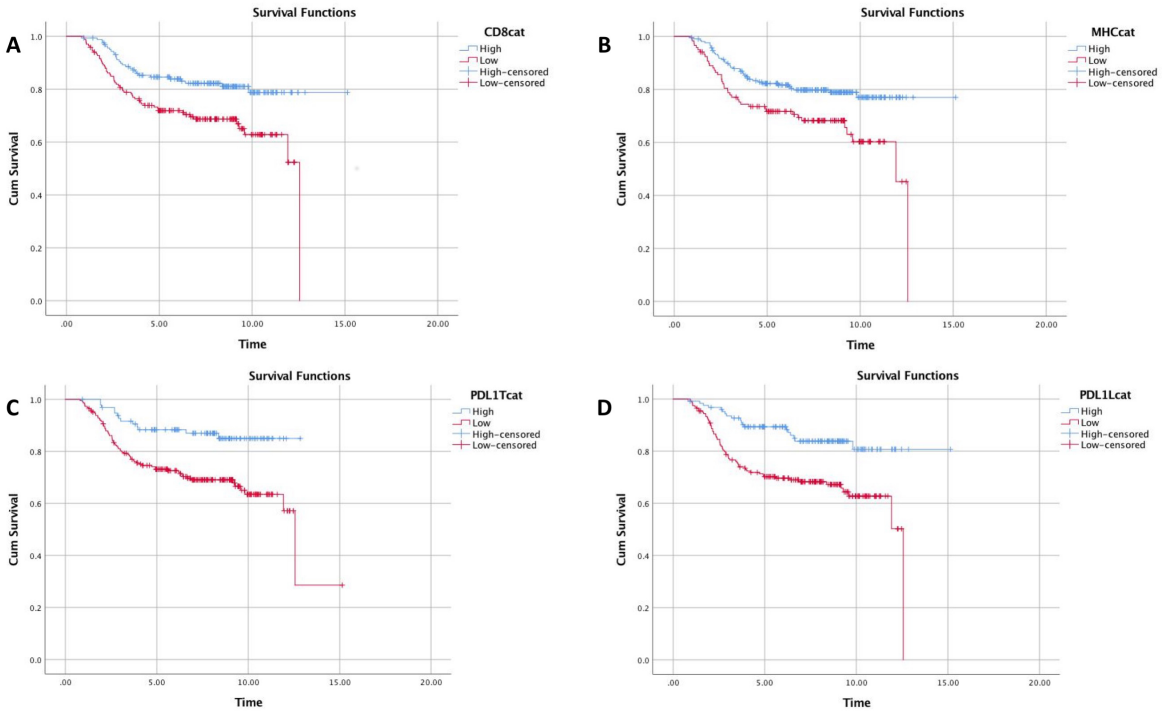



**Figure 5-9. Kaplan Meier survival analysis for stage I-III TNBC cases from the POSH cohort, stratified by immune markers.**

**A. CD8 high (n=157) vs. low (n=161), p=0.002, B. MHC I high (n=209) vs. low (n=119), p=0.003, C. PD-L1 (tumoural) high (n=97) vs. low (n=225), p=0.001, D. PD-L1 (lymphocytes) high (n=125) vs. low (n=196), p=0.000170.**

**p-values from log-rank Mantel Cox test.**

In order to explore whether or not PD-L1 was prognostic within TIL subgroups, I then divided cases into 4 categories based on TIL high/low (cut-off >15%) and PD-L1 (lymphocytes) present/absent and performed a Kaplan Meier survival analysis (Figure 5.10). A threshold of 15% (the median TIL score) was chosen as the threshold to divide TIL infiltration into two evenly split groups. TIL-positive, PDL1-positive cases had the best survival and TIL-negative, PDL1-negative the worst. The remaining two survival curves overlapped.

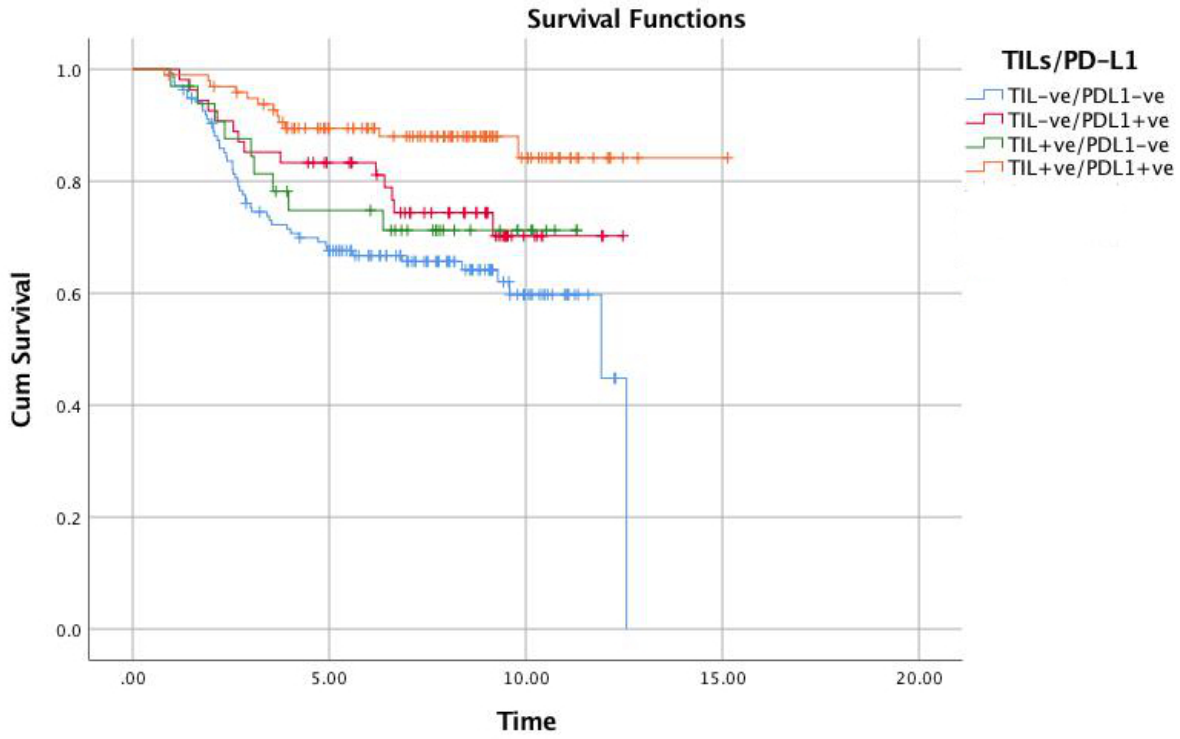



**Figure 5-10. Kaplan Meier survival analysis for overall survival for patients with stage I-III TNBC from the POSH cohort.**

**N=320**

**TIL+ve/PDL-1+ve: n=98.**
**TIL-ve/PDL-1+ve: n= 54.**
**TIL+ve/PDL-1-ve: n=32.**
**TIL-ve/PDL-1-ve: n=136.**
**Reference line shown (1.00) – set as low for all variables.**
**P<0.0001 log-rank mantel cox**

A ROC analysis using 3-year overall survival as the primary outcome was performed using backwards elimination with 24 initial variables, including clinical variables in addition to CD8 and PD-L1 score. The initial positive predictive value (PPV) was 0.8324 (95% CI 0.7728-0.8919) (Table 5.4). After elimination, the most prognostic factors were TIL percentage (PPV 0.6823) and the number of pathological lymph nodes involved (PPV 0.7049). Combined, the PPV for these two factors was 0.7603 (Figure 5.11).

**Table 5-4. ROC association statistics, using 3-year survival as outcome, with TNBC cases from the POSH cohort.**

**N=350**

| ROC Model | Mann-Whitney | | | |
|---|---|---|---|---|
| | Area | Standard Error | 95% Wald Confidence Limits | |
| **TIL Percentage** | 0.6868 | 0.0476 | 0.5935 | 0.7801 |
| **TIL Category** | 0.6335 | 0.0368 | 0.5613 | 0.7057 |
| **CD3** | 0.5653 | 0.0522 | 0.4631 | 0.6675 |
| **CD4** | 0.6176 | 0.0482 | 0.5233 | 0.7120 |
| **CD8** | 0.6562 | 0.0414 | 0.5751 | 0.7372 |
| **FOXP3** | 0.6328 | 0.0441 | 0.5463 | 0.7192 |
| **CD4:CD8 Ratio** | 0.4577 | 0.0528 | 0.3543 | 0.5612 |
| **FOXP3:CD8 Ratio** | 0.5090 | 0.0535 | 0.4042 | 0.6138 |
| **GLUT1** | 0.5472 | 0.0502 | 0.4488 | 0.6457 |
| **SMA** | 0.4891 | 0.0517 | 0.3878 | 0.5904 |
| **P53** | 0.5117 | 0.0495 | 0.4147 | 0.6088 |
| **KI67** | 0.4959 | 0.0536 | 0.3909 | 0.6010 |
| **ALDH1** | 0.5732 | 0.0402 | 0.4944 | 0.6520 |
| **CK56** | 0.5377 | 0.0496 | 0.4404 | 0.6350 |
| **EGFR** | 0.4965 | 0.0477 | 0.4031 | 0.5900 |
| **MHC** | 0.6297 | 0.0483 | 0.5350 | 0.7244 |
| **PDL1 (lymphocytes)** | 0.6609 | 0.0377 | 0.5870 | 0.7349 |

| ROC Model | Mann-Whitney | | | |
|---|---|---|---|---|
| | Area | Standard Error | 95% Wald Confidence Limits | |
| **Age at Diagnosis** | 0.5579 | 0.0506 | 0.4587 | 0.6571 |
| **BMI** | 0.5304 | 0.0507 | 0.4310 | 0.6298 |
| **Clinical T Stage** | 0.6234 | 0.0459 | 0.5333 | 0.7134 |
| **Path T Stage** | 0.6403 | 0.0411 | 0.5597 | 0.7209 |
| **Max invasive tumour size** | 0.6353 | 0.0453 | 0.5465 | 0.7240 |
| **Max overall tumour size** | 0.6188 | 0.0456 | 0.5294 | 0.7082 |
| **Heterozygote Risk** | 0.5497 | 0.0550 | 0.4419 | 0.6575 |
| **Combination** | 0.8324 | 0.0304 | 0.7728 | 0.8919 |

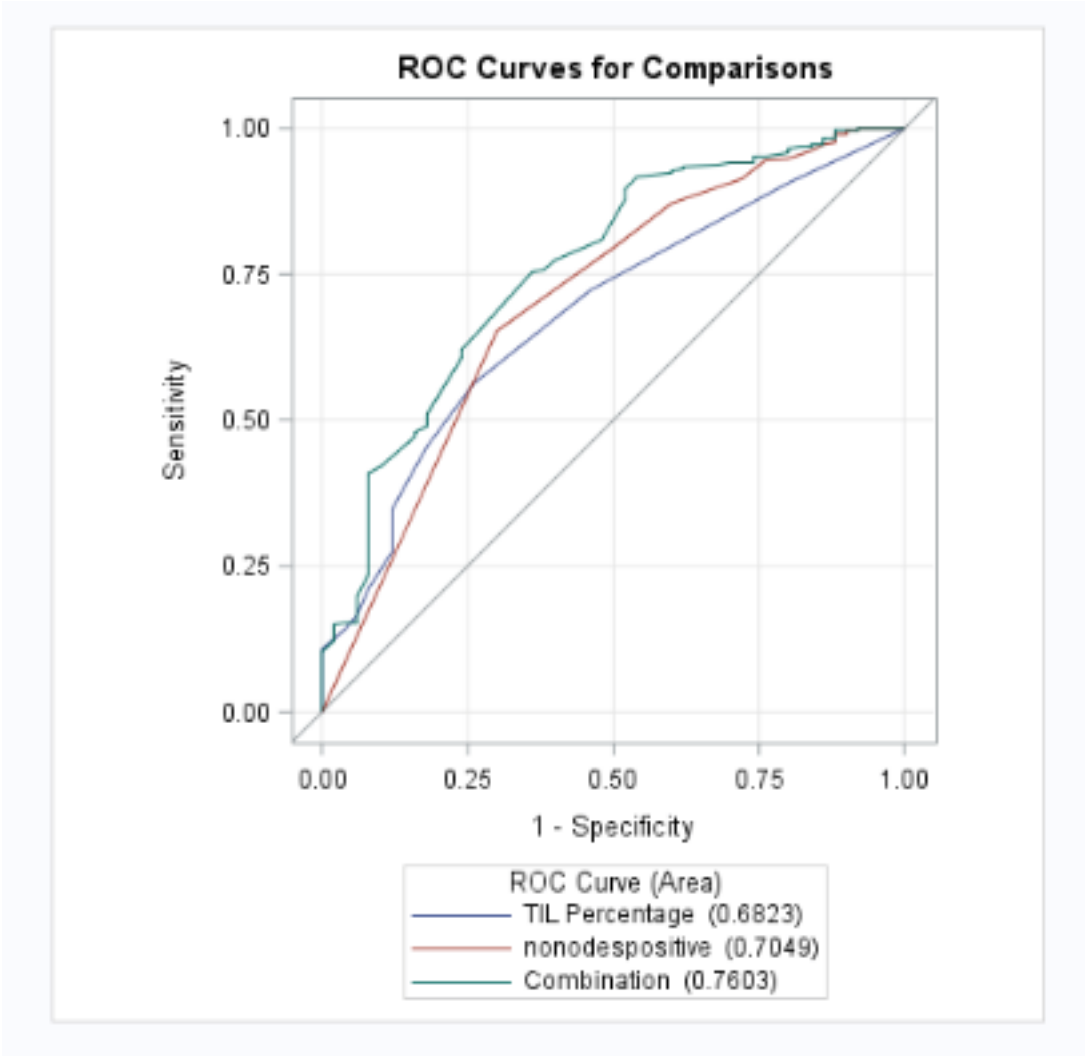



**Figure 5-11. ROC analysis, using 3-year OS as outcome, of TNBC cases from the POSH cohort.**

**N=350. Backwards elimination of 24 variables resulted in two highly prognostic variables: TIL percentage (PPV 0.6823) and pathological lymph nodes ("nonodespositive") (PPV 0.7049). The combined PPV was +0.7603.**

I assessed for correlations between TIL score and the expression of all the markers studied, including those from the previous chapter (Table 5.5). There was a very significant positive correlation between TIL score and CD3, CD4 and CD8. FOXP3 also positively correlated. There was also a significant positive correlation between TIL score and ERAP1, MHC I, Ki67, ALDH1 and PD-L1 (both tumoural and on lymphocytes). There was no correlation between TILs and GLUT1, p53, CK 5/6 or EGFR. The only marker that inversely correlated with TILs was SMA (r=-0.211, p=0.000126).

**Table 5-5. Correlation of TIL score with expression markers. N=350**

| Factor | Pearson's correlation | p-value |
|---|---|---|
| CD3 | **+0.494** | **1.5825E-21** |
| CD4 | **+0.416** | **3.1829E-15** |
| CD8 | **+0.638** | **4.9466E-39** |
| FOXP3 | **+0.398** | **5.7689E-14** |
| SMA | **-0.211** | **0.000126** |
| GLUT1 | -0.050 | 0.370 |
| ERAP1 | **+0.329** | **0.000029** |
| MHC1 | **+0.436** | **1.2702E-16** |
| P53 | +0.062 | 0.267 |
| Ki67 | **+0.231** | **0.000025** |
| ALDH1 | **+0.300** | **4.6976E-8** |
| CK 5/6 | +0.016 | 0.782 |
| EGFR | +0.059 | 0.291 |
| PDL1T | **+0.433** | **3.6849E-16** |
| PDL1L | **+0.598** | **1.792E-32** |

### 5.2.4 Digital imaging

#### 5.2.4.1 Automated TMA analysis

The TMAs were scanned in and then Definiens software was used to analyse the images of TMAs stained with CD8 and PD-L1 (see Methods 2.2.9). I then looked for a correlation between automated and pathologist (manual) scoring (Figure 5.12). There was a close correlation for both CD8 and PD-L1 (R=+0.822 and +0.790 respectively). The distribution of automated scoring is shown in Figure 5.13; there was a negative skew for both markers, but this more extreme for PD-L1.

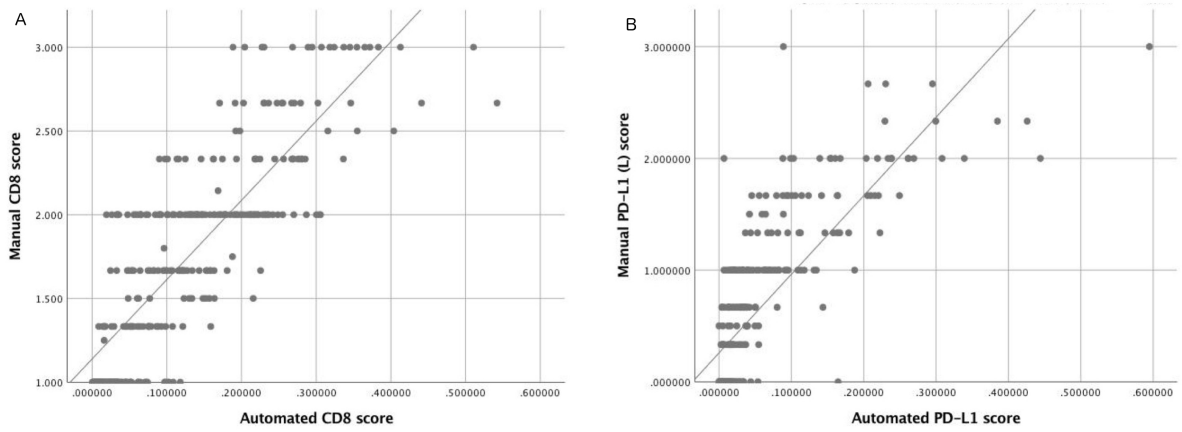



**Figure 5-12. Pearson's two-tailed correlation coefficient between automated and manual CD8/PD-L1 scoring.**

**A. CD8 (R=+0.822, n=325, p=<0.001) and B. PD-L1 (R=+0.790,n=320, p=<0.001).**

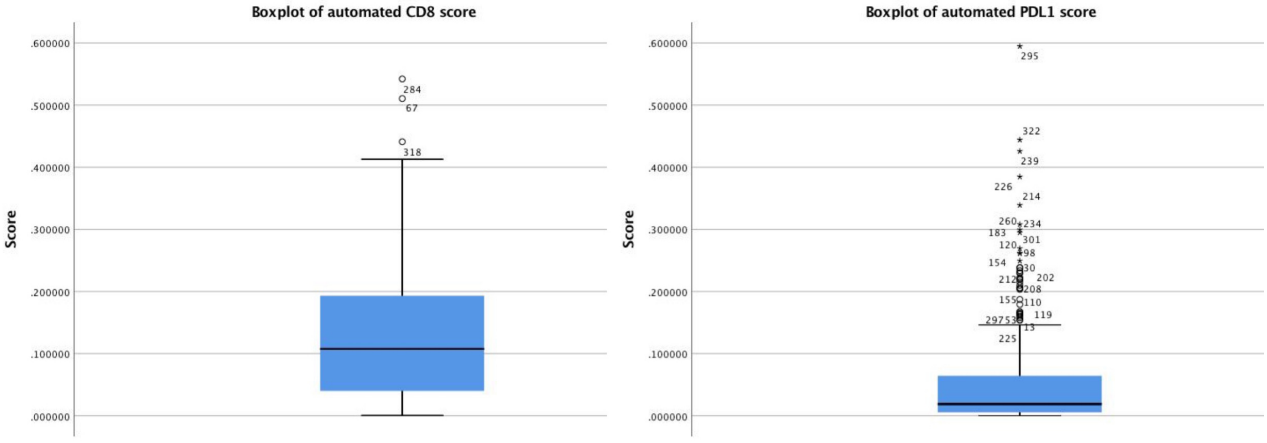



**Figure 5-13. Boxplot for automated CD8 and PD-L1 scoring of stage I-III TNBC cases from the POSH cohort, using Definiens software.**

**Median (CD8=0.107618, PD-L1=0.018681) and interquartile range shown.**

I used Kaplan Meier survival analysis to assess the prognostic impact of automated CD8/PD-L1 scoring (Figure 5.14). They were categorised into high and low by using the median as a threshold. Like manual scoring, automated scoring of CD8 high and PD-L1 high were prognostic for better outcome, compared to low scoring. On Cox regression univariable analysis, the HR for CD8 (automated) high vs. low was 0.470 (0.300-0.738; p=0.001) and for PD-L1 (automated) high vs. low was 0.350 (0.220-0.559; p<0.0001). On MVA (adjusted for nodal stage and stratified by LVI), the HRs were 0.496 (0.315-0.780; p=0.002) and 0.425 (0.263-0.685; p=0.0004) so they both remained independently statistically significant. When PD-L1 and CD8 positive/negative were combined into four categories, CD8+PDL1+ had the best survival (Figure 5.14C).

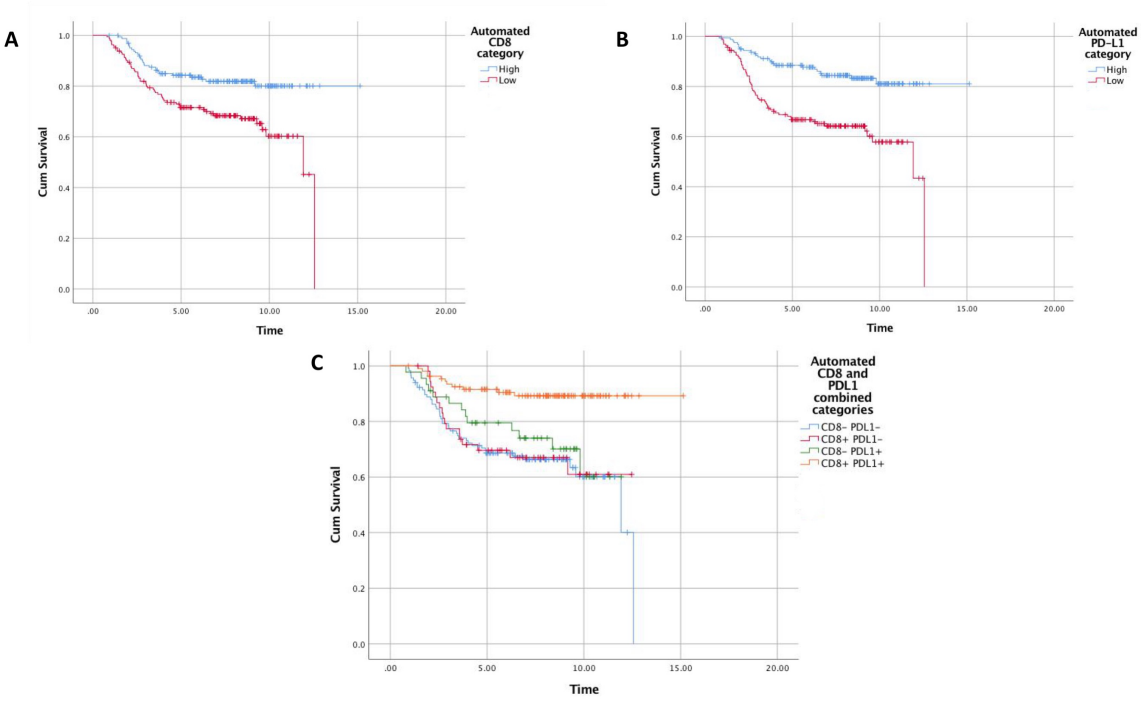



**Figure 5-14. Kaplan Meier survival analysis for automated CD8/PDL-1 scores.**

**A. Automated CD8, high (n=163) and low (n=162), p=0.001, B. Automated PD-L1, high (n=160) and low (n=160), p<0.0001 and C. Automated CD8+/PDL1+ve (n=109), CD8+/PDL1-ve (n=54), CD8-ve/PDL1+ve (n=45) and CD8-ve/PDL1-ve (n=117), p<0.0001. p-values calculated with log-rank Mantel cox.**

Finally, I assessed for a correlation between TIL score and automated CD8 and PD-L1 score. Both correlations were positive at +0.613 (p=<0.0001) and +0.475 (p=<0.0001).

#### 5.2.4.2 Multiplex full-face analysis

In order to undertake spatial analysis in detail using a small number of cases, I selected 9 TIL-high and 9 TIL-low cases from the POSH cohort at random. The hypothesis was that tumours with pathologist-defined high TILs would have a lower number of SMA-positive cells within the tumour border and that the pathologist-defined SMA-high cases would have a lower number of CD8-positive cells within the tumour border, with a higher number outside the tumour border (indicating an "exclusion" effect). Multiplex staining was performed using cytokeratin (tumour cell marker), CD8 (CD8+ T-cell marker) and SMA (CAF marker). The number of cells which stained positively for either, or both, within the tumour border and outside the tumour border were determined, then divided by the total number of cells in this area, in order to calculate the proportion of cells that were positive in each compartment. The results, stratified by TIL status moderate/high (>15%) vs. low (≤15%), are shown (Figure 5.15). I then stratified the cases by whether they were SMA-high or –low (according to previous criteria) (Figure 5.16). Finally, I evaluated the ratio of the positive cells inside the tumour divided by the positive cells outside the tumour (Figure 5.17).

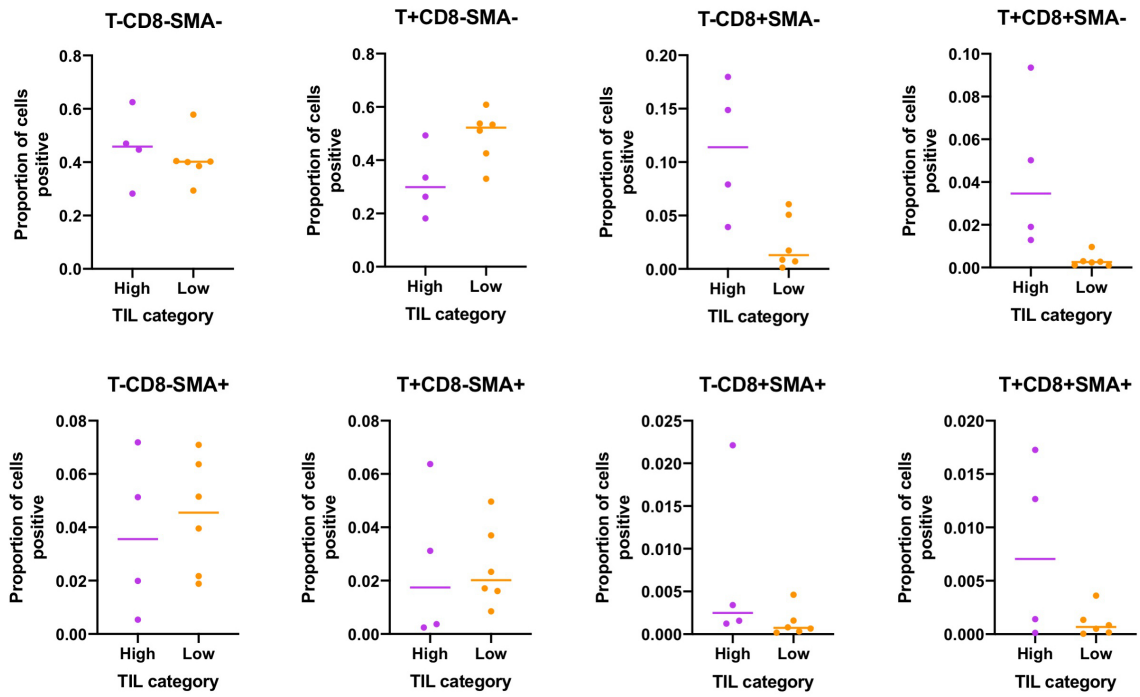



**Figure 5-15. Scatter dot plots shown for proportion of cells positive for CD8+/- and SMA+/- stratified by TIL category for full-face images from stage I-III TNBC cases from the POSH cohort.**

**Multiplex staining with CD8, SMA and cytokeratin. Median value shown by line in each graph. TIL-high, n=9; TIL-low, n=9. Unpaired t-test with Welch's correction performed, no significant difference found.**

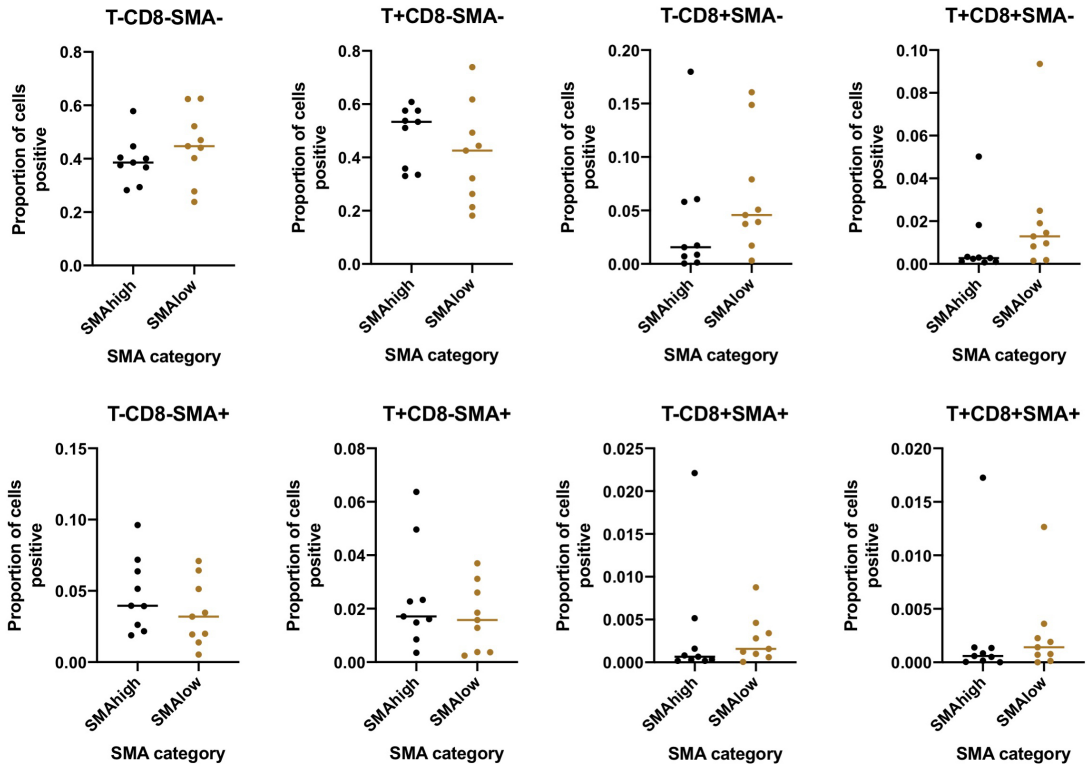



**Figure 5-16. Scatter dot plots shown for proportion of cells positive for CD8+/- and SMA+/- stratified by SMA category for full-face images from stage I-III TNBC cases from the POSH cohort.**

**Multiplex staining with CD8, SMA and cytokeratin. Median value shown by line in each graph. SMA-high, n=9; SMA-low, n=9. Unpaired t-test with Welch's correction performed, no significant difference found.**

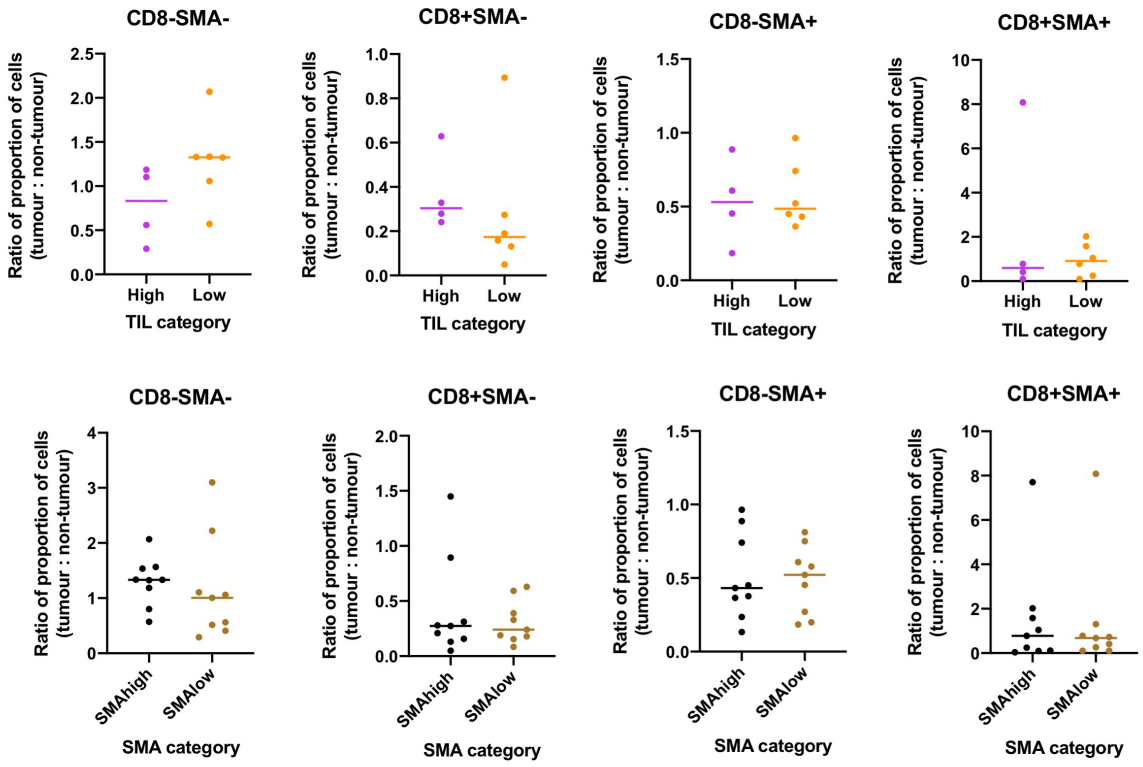



**Figure 5-17. Scatter dot plots shown for ratio of positive cells inside the tumour border to outside the border for full-face images from stage I-III TNBC cases from the POSH cohort.**

**Multiplex staining with CD8, SMA and cytokeratin. Median value shown by line in each graph. TIL-high, n=8; TIL-low, n=10. SMA-high, n=9; SMA-low, n=9. Unpaired t-test with Welch's correction performed, no significant difference found.**

There were no significant differences in the median proportion of cells or the ratios studied. There were more CD8+ T-cells (CD8+/SMA-ve) inside and outside the tumour for TIL high tumours, vs. low, but the difference was not significant. When evaluated according to SMA status, the CD8+ T-cells (CD8+/SMA-ve) were lower both inside and outside the tumour for SMA-high tumours. The ratio of CD8+SMA-ve cells was lower for TIL-low tumours. Overall, this shows that the number of CD8 cells was generally high for the TIL-high tumours and generally low for the SMA-high tumours. Compared to TIL-high tumours, the TIL-low tumours had more cells inside the tumour border, relative to outside, i.e. appeared to be more cellular. In contrast, TIL-low tumours had less CD8 cells inside the tumour relative to outside.

## 5.3 Discussion

### 5.3.1 Prognostic value of TILs

In this cohort of 350 patients aged less than 40 with stage I-III TNBC treated with adjuvant or no chemotherapy, we have shown that stromal TILs are prognostic, with a hazard ratio for those with high (≥60%) infiltration (LPBC) of 0.095 (0.013-0.684; p=0.019) compared to those with low (≤10%) infiltration. For those with moderate (11-59%) infiltration the hazard ratio was 0.492 (0.317-0.765; p=0.002) compared to low. Although consistent with previous studies(36, 43-45), this has not previously been reported specifically in patients aged less than 40 years. Generally, TNBC cohorts have included only small numbers of women aged less than 40 (less than a quarter of the total number) (36, 44). The median stromal TIL score was 15%, consistent with previous reports of 10-18% in TNBC(36, 44, 45). TIL score was an independent prognostic factor when evaluated as a continuous variable and adjusted for nodal status and stratified by LVI. When evaluated as part of a backwards elimination model, using 3 year survival as the primary outcome, the TIL percentage was second only to the number of pathological lymph nodes in terms of prognostic significance. Traditional risk factors, such as clinical T stage and grade, were less prognostic. The combined PPV for TILs and number of lymph nodes was +0.7603.

### 5.3.2 TILs and clinicopathological associations

I evaluated the association of TILs with clinicopathological features. TILs and MHC class I (but not ERAP1) negatively correlated with age in this cohort, an interesting finding given this is an already comparatively youthful cohort. Immunosenescence is the phenomenon of impaired immunity as we age, usually used in reference to the elderly. The adaptive

immune system appears to be particularly affected. Older people have a lower number of circulating naïve CD8+ T-cells (with potentially impaired immune responses to neoantigens) and have reduced T-cell receptor diversity (267-269). This is likely to contribute to the increasing incidence of infections, autoimmune diseases and neoplasms with age. Following infection with an intracellular organism, old animals have been shown to have a reduction in expansion of antigen-specific CD8+ T-cells, with an impaired effector response(270). Elderly people respond less effectively to vaccines. In a murine model of metastatic breast cancer, old mice did not develop a CD8 response to Mage-b vaccination(271). Although these changes are reasonably well characterised, the impact on immunosurveillance and prognosis in cancer and responses to immunotherapy are unclear. It is interesting that I have found a correlation between TIL scores and age in this cohort, given they are all comparatively young. It would be worth exploring this association further in TNBC in a cohort with indiscriminate age in future work.

Patients with a *BRCA* mutation had a higher TIL score than those without (20 vs. 10, p=0.043). It has recently been established that the 136 patients with a *BRCA* mutation and TNBC in the POSH cohort (TNBC cohort n=558) might have a survival advantage in the first few years after diagnosis; the OS difference was significant at two years(201). Although this could be due to increased chemotherapy sensitivity, the results here provide evidence for increased immunogenicity in *BRCA*-mutation carriers and the strategy of pursuing a role for immunotherapy in those patients. Combined immunotherapy has been used successfully *in vivo* for mammary tumours in *BRCA1*-deficient mice(122). In general, inactivated BRCA is felt to be a predictive marker for response to immunotherapy (157). Further results in clinical trials are however awaited; a phase II trial of anti-PD-L1 therapy for advanced breast cancer in the context of a germline *BRCA* mutation is underway (NCT03025035). However, the degree of immunogenicity found in BRCA-mutated patients is wide and many *BRCA*-deficient tumours still have low lymphocytic infiltrate. The degree of homologous recombination and the presence of allele-specific loss of heterozygosity have been implicated to influence the immunogenicity of BRCA-deficient tumours(126). Therefore, it is likely that

further molecular stratification is required to develop precision treatments even within the cohort with a *BRCA* mutation.

Across all cases, localised (vs. multifocal) and LVI-negative (vs. positive) cases had a statistically higher median TIL count. I hypothesise that cases with a high TIL infiltrate are in a state of "near-equilibrium" in the process of immunoediting, with the immune system keeping the tumour in check with potentially lower clonal diversity and consequently a more localised, less aggressive tumour burden.

### 5.3.3 Bioinformatics analysis

To identify gene signatures whose expression differed amongst TNBC subsets, we integrated immunogenomic classification with weighted gene coexpression network analysis of TNBC tumours from the TCGA dataset and found that ECM modules FN1 and TGF-β1/3 were highly expressed in the lymphocyte-deplete subset. This is consistent with work demonstrating promotion of an immunosuppressive TME by TGF-β signalling (which promotes the differentiation of fibroblasts into myofibroblasts) and possible exclusion of T-cells by CAFs (which upregulate fibronectin)(173, 187). This would support further investigation of the role of CAFs in TNBC tumour progression.

A strong correlation between 13 immune metagenes and histologic TIL counts has been observed in breast cancer cases of all subtypes(148). In this bioinformatics analysis, an analysis of an association with histologic TIL counts in the TCGA dataset was not performed. In Safonov *et al*, they also attempted to identify DNA level alterations that were associated with immune gene expression and were unable to identify any(148). Here there has not been an attempt to identify DNA level alterations or neoantigens, instead the focus has been on correlations between more broad gene modules.

### 5.3.4 Immune panel

Using multivariable analysis, we found that CD4 and CD8 expression were all associated with improved prognosis (FOXP3 near significance). FOXP3 expression has previously been shown to correlate with poor outcomes in breast cancer(56, 272), however when ER-negative cases were evaluated alone results have been in agreement with what is shown here(55, 273). In general, regulatory T-cells are thought to suppress anti-tumour immunity. However in some cancers, such as colorectal cancer, they have been associated with an improved prognosis (274). Mechanisms might include inhibition of pro-inflammatory cells (e.g. Th17 cells) and suppression of microbial-driven inflammation(275). In colorectal cancer, there have been shown to be two distinct subtypes of FOXP3+ cells exerting different effects, with the “inflammatory” subtype being protective(276). Further work needs to be performed to understand the phenotype and function of FOXP3+ cells in TNBC.

Of note, pathologist assessment of TIL score outperformed CD8 expression, in terms of prognostication. This is likely to be related to the larger variability within the TIL score (0 to 100%) relative to CD8 expression (low, moderate, high), in addition to the use of full-face images for TILs and the inclusion of all lymphocyte phenotypes, meaning a more complete assessment of the tumour.

In this cohort, PD-L1 expression on both tumour cells and lymphocytes were both correlated with TIL score and improved outcome. A number of recent reports in smaller TNBC cohorts have been consistent with this finding(149, 154) and high PD-L1 expression has been associated with the triple negative subtype(153), as well as response to neoadjuvant chemotherapy(151). In a spatial analysis study that demarcated TNBC into microenvironment subtypes, the good prognosis, Fully Inflamed subtype had the highest gene expression of several immune checkpoints, including CTLA-4 and PD-L1(170).

It is now well established that PD-L1 plays a key role in limiting the cytotoxic anti-tumour response via its interaction with PD-1 receptor on immune cells. In previous studies it has been suggested that the finding of PD-L1 being associated with improved prognosis in TNBC is solely a reflection of its co-localisation with TILs(153) and thus an anti-tumour inflammatory response (as part of a negative feedback loop). However, in this cohort we have found the prognostic impact of PD-L1 to be independent of TILs. Although PD-1 has been established as a marker of T-cell exhaustion, human breast tumour CD8+ TILs expressing PD-1 have recently been found to retain cytokine production and degranulation capability(155). This was found to be in contrast to melanoma. This study highlighted the need to be cautious about labelling all PD-1+/CD8+ TILs as being of the exhausted phenotype across all cancers. Further work to determine the usefulness of PD-1/PD-L1 as a predictive biomarker when used in isolation for response to checkpoint blockade in TNBC is required. An understanding of the complex and dynamic nature of the interaction between the immune system and the tumour is required to define appropriate predictive biomarkers and a number of markers may need to be incorporated into a "predictive score", e.g. TILs, tumour mutation burden, in addition to PD-L1.

In addition, there is a need to validate and standardise PD-L1 scoring. The assays used to score expression of PD-L1 have been developed in conjunction with the individual drugs, using different antibody clones, scoring algorithms and cut-offs for determining whether a patient is PD-L1 "negative" or "positive"(277). There is controversy with regards to optimal cut-off "threshold" for PD-L1 scoring, in addition to the type of cells on which PD-L1 expression is most relevant(278). In addition, there are a number of different assays used to assess PD-L1 expression clinically and the use of 1% as a cut-off has been shown to decrease concordance between assays(279). As this is commonly the cut-off used to determine whether or not patients are eligible for checkpoint inhibition clinically, it is clearly important that the most robust assay is being used. In the future, centralisation of PD-L1 scoring and multi-parameter algorithms (with PD-L1 expression being just one input) may be necessary.

Here I have also shown that MHC class I and ERAP1 expression are correlated with TIL infiltration in TNBC. MHC I was also associated with improved survival, on univariable analysis only. MHC class I proteins are membrane-bound proteins expressed on almost all nucleated cells; they are encoded by *HLA-A, -B* and *–C* genes. The ERAP1 protein is involved in N-terminal trimming during antigen processing and appears to be critical in the generation of the immunopeptidome(139). Downregulation of MHC class I has been reported in different tumour types and correlated with poor prognosis(134-136). Clearly the absence of antigen presentation is likely to assist in immune evasion via reduced antigenicity, although this has not been reported extensively in TNBC. In a small study utilising proteomic profiling of TNBC, downregulation of proteins involved in the MHC class I pathway (such as ERAP1 and TAP 1/2) were associated with shorter OS(133).

In this cohort, Ki67 and ALDH1 positively correlated with TILs, while SMA negatively correlated, but none of these markers were individually correlated with prognosis. Ki67 is a well-established marker of cell proliferation, although not in routine use clinically(241). ALDH-1 is a marker of normal and malignant breast stem cells(251). To our knowledge, neither Ki67 nor ALDH-1 has been correlated with TILs before. However, in one study of colorectal cancer, cancer-initiating cells (characterised by ALDH-1 expression) have been shown to have low immunogenicity(280). SMA is a marker of cancer-associated fibroblasts (CAFs), which are increasingly thought to mediate progression of breast cancer through mechanisms such as extracellular matrix remodelling and secretion of growth factors and cytokines(171). Increasingly, fibroblasts are being implicated in the exclusion of T-cells, mitigating resistance to immunotherapy(187). Much like FOXP3+ cells, there are likely to be multiple subtypes of CAFs, with differing phenotypes(173).

In one study using 116 consecutive pre-menopausal cases of TNBC, the lymphocyte-predominant subtype was associated with a lower frequency of basal cytokeratins and EGFR(281). Here we found no association between TIL score and either of the basal markers evaluated. P53 has been positively correlated with TILs in TNBC, but I did not find this here(282). As discussed in the previous chapter, the associations between p53 and outcomes in TNBC have been variable in previous studies; p53 was not found to be

prognostic in this cohort. P53 IHC is a poor surrogate marker of the underlying gene status regardless and it may not be associated with immunogenicity in TNBC.

### 5.3.5 Digital pathology validation

I utilised digital pathology analysis to compare with pathologist scoring and performed spatial analysis. Digital pathology and automated image analysis offer many benefits including reduced subjectivity and human error (such as transcribing errors)(283). Additionally, there are opportunities to produce data that are not visible with the naked eye, but have potential significant prognostic or predictive implication. Although digital imaging is not used clinically widely at present, it is likely to be adopted with increasing frequency, due to the potential for improved efficiency and quality of pathology reporting. The evaluation of expression staining is certainly subject to inter-observer variability.

The correlation between pathologist and automated scoring of CD8 and PD-L1 was strong, with an R of +0.822 ($p<0.0001$;n=325) and +0.790 ($p<0.0001$; n=320). Using Cox regression multivariable analysis to assess the automated scores in a categorical manner, the HRs were 0.496 ($p=0.002$) and 0.425 ($p=0.0004$) for CD8 and PD-L1 high vs. low, indicating they are independently prognostic. When combined into one category by automated scoring, CD8+/PD-L1+ had the best prognosis compared to the other three categories, e.g. CD8+/PD-L1-. The combination of PD-L1-positivity and high TILs/CD8s conferring a positive prognosis has been found in previous work(152) and it is hypothesised that TIL-rich tumours that also express PD-L1 have better immune surveillance. With some further validation, future analyses could utilize this automated method.

### 5.3.6 Spatial analysis

Finally, I performed a spatial analysis of CD8- and SMA- positive cells in the context of the location of the tumour (cytokeratin-positive cells) using full-face TNBC cases from the POSH cohort. The ratio of CD8-ve cells inside the tumour relative to outside was highest for TIL-low tumours. It is hypothesised that these are tumour cells or immunosuppressive cells. It was expected that the SMA-high tumours would have a reduced ratio of CD8+ cells inside relative to outside (due to CAFs physically excluding CD8 cells outside the tumour border), but this was not the case. It therefore may be that CAFs exert an immunosuppressive effect in TNBC through cytokines and other secreted factors in the TME, which affect stromal regions out with the tumour border, rather than simply forming a physical barrier. In the microenvironment classification paper discussed above, the immune desert and margin restricted phenotypes (both with low CD8 expression) had the poorest prognosis and showed elevated signatures of activated stroma, fibrosis (including TGF-β-dependent signatures) and fibrotic foci(170). Here, the TIL-low tumours had less CD8+ cells inside the tumour relative to outside, indicating some evidence of a possible exclusion/margin-restricted phenotype.

### 5.3.7 Limitations

This study is limited by the use of TMAs, which may under or overestimate expression of markers due to intratumoural heterogeneity. In order to validate TMA scoring, a number of cases could be selected for full-face scoring and these results could be compared to the TMA scores. We also acknowledge that scoring is not standardised for all markers (such as MHC class I), which limits reproducibility to a degree. The method of scoring was relatively simplified, in the interests of practicality. It is much quicker to score a core as “low, medium or high” than to score in a semi-quantitative way, which requires the estimation of the percentage of cells positive in addition to intensity. Further work could utilise a more in-detail, semi-quantitative system, e.g. including both percentage of cells

staining positive in addition to staining intensity. A classic example of this is the Allred score for ER expression, which combines the number of cells staining positive with the intensity of staining to produce a combined score(284). There are a number of quantitative methods being developed, such as dot-counting systems(285). In one study evaluating a quantitative antigen/antibody complex IHC method, the novel method was found to be as reliable at detection of HER2 as standard methods, with a broader range and higher sensitivity(285). Given the results here and the promise of reproducibility and efficiency, I would incorporate digital technology in future IHC analyses. The automated analysis generated a continuous variable, which could be used in survival analyses instead of dividing the continuous variable into categories.

## 5.4 Summary

In this study we have shown that a higher TIL score is significantly associated with improved survival in young women with TNBC. H&E TIL scoring outperformed CD8 expression and was second only to number of involved lymph nodes using a backwards elimination ROC analysis for 3-year survival. High PD-L1 expression is also associated with improved survival in this cohort. This indicates a need to further understand the function of the PD-L1/PD-1 pathway and the phenotype of TILs in TNBC, given this divergence from other tumour types. This may better inform combination approaches with immunotherapy in this disease. Given the importance of lymphocytic infiltrate in prognosis, it is vital to understand mechanisms of immune exclusion in this disease. Using a systems biologic approach, we identified upregulation of TGF-B/fibronectin gene modules in lymphocyte-deplete TNBC. In this cohort, we also found the presence of CAF in the tumour stroma to inversely correlate with TILs. Together these data suggest that these cells may be immunosuppressive. Inhibition of CAFs may therefore be a strategy to improve clinical response to immunotherapy. I therefore subsequently developed a murine CAF-rich TNBC model in which to explore the efficacy of CAF targeting.

# Chapter 6 Developing a CAF-rich Model of Triple Negative Breast Cancer

## 6.1 Introduction

There is increasing interest in the role of the stroma in neoplastic progression and metastatic spread. Fibroblasts/cancer-associated fibroblasts (CAFs) are the most abundant cell type in the tumour microenvironment (TME) and are likely to be central to stromal disruption in malignancy. The signalling pathways that mediate the transformation from normal fibroblasts into CAFs have not been fully elucidated, but transforming growth-factor-beta (TGF-β) is clearly a key growth factor/cytokine. *In vitro*, TGF-β can induce fibroblasts to differentiate into the CAF phenotype(172). Myofibroblasts can also release TGF-β, providing a positive feedback loop(174). Our group has previously shown that fibroblast to myofibroblast transdifferentiation was dependent on Nox4 (via the generation of a delayed phase of intracellular ROS)(183).

Although CAFs are generally accepted to have a myofibroblastic phenotype, this is not universally true. CAFs can also be senescent; this phenotype can be induced by DNA damage or repeated cell division. Although the impact of senescent CAFs within the tumour microenvironment is not fully elucidated, our group has shown that senescent HFFF2 fibroblasts developed a myofibroblast-like phenotype i*n vitro* and activated canonical TGF-β signaling(195). However, gene expression was different between myofibroblastic and senescent CAFs and the senescent cells appeared to be less able to remodel the ECM. The relative abundance and importance of myofibroblastic vs. senescent CAFs in human cancer are unclear and warrant further study.

In breast cancer around 80% of stromal fibroblasts acquire the activated CAF phenotype(172); it is now well established that patients with an increased stroma:tumour ratio tend to have worse survival(286). Using microdissection of 53 primary breast tumours, a "stroma-derived prognostic predictor" was derived from gene expression profiles, reflecting the importance of stromal factors in malignant progression(287). Increased expression of α-SMA has been correlated with a higher grade, lymph node spread and a worse prognosis, in addition to drug resistance and reduction of anti-tumour immunity(171, 173). In chapter 5 I have demonstrated that SMA is negatively correlated with CD8 cells in triple negative breast cancer.

There was therefore a rationale for further studying the role of CAFs and the potential for stromal targeting in breast cancer. Here I have explored the effect of TGF-β stimulation on breast fibroblasts and CAFs *in vitro* and the impact of different fibroblast phenotypes on breast tumour growth *in vivo*. The hypothesis was that normal breast fibroblasts would be transformed into the myofibroblast phenotype when stimulated by TGF-β. In order to evaluate the phenotype with and without TGF-β, cell count, metabolism, quantitative PCR, western blot and immunofluorescence were performed. The *in vivo* hypothesis was that myofibroblasts (CAFs or TGF-β-treated fibroblasts) would augment 4T1 tumour growth and inhibit CD8 infiltration compared to control (no fibroblast) tumours.

## 6.2 Results

### 6.2.1 TGF-β stimulation in vitro

In order to proceed with *in vitro* experiments, normal breast fibroblasts were isolated from tumours from Balb/c mice as described in Section 2.4 .4. Breast "CAFs" originated from Balb-neu-T mice. It is acknowledged that the "CAF" cells isolated would be accurately described as tumour-derived fibroblasts, likely to be heavily enriched with CAFS. From here onwards they are referred to as CAFs.

In order for fibroblasts to differentiate, transcription of myofibroblast genes must take place and this is primarily regulated by the cytokine TGF-β(288). To assess the proliferative effect of TGF-β, breast fibroblasts and CAFs were treated with TGF-β in culture and the viability count and metabolism assessed after 3 and 5 days.

Figure 6.1 demonstrates that, for fibroblasts, there was a small, non-significant increase in cell number and metabolism with TGF-β treatment on Day 3, but a significant increase on Day 5. CAFs similarly showed a small non-significant increase in cell number on Day 3, but a significant increase on Day 5. The metabolism decreased with TGF-β treatment (significant at Day 3), despite the small increase in cell count. Therefore the proliferative effect of TGF-β was more pronounced on fibroblasts.

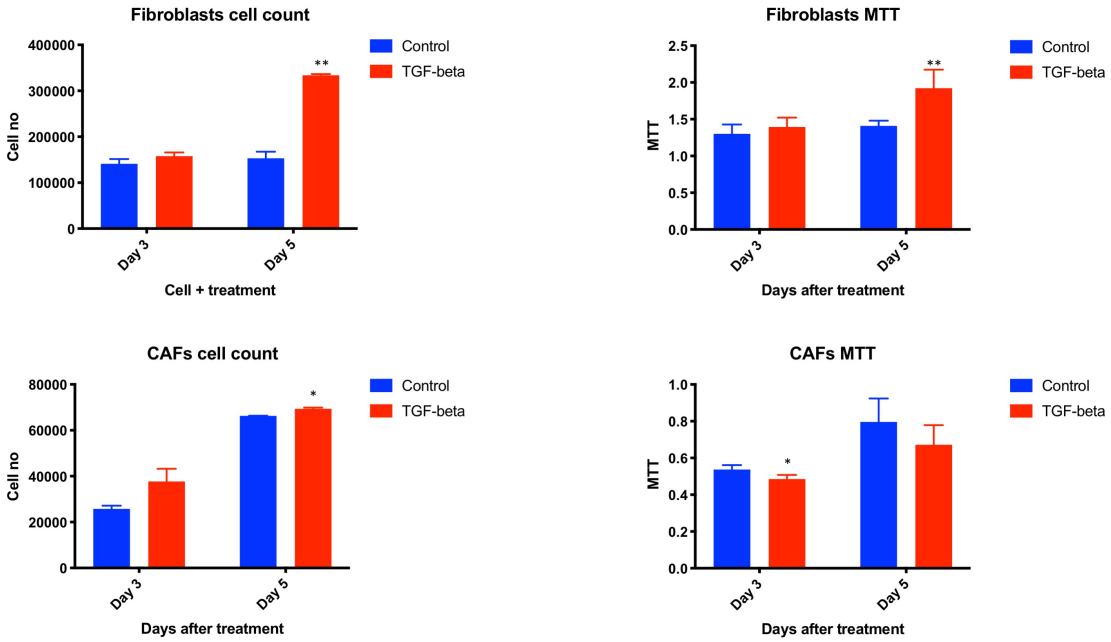



**Figure 6-1. Balb/c murine fibroblasts and Balb-NeuT murine CAFs, cell count and MTT 3 and 5 days after TGF-β treatment.**

**Standard deviation shown for n=2 (cell count) and n=4 (MTT).**

**p-values shown for TGF-β vs. control, unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

In order to assess whether TGF-β was effective at differentiating cells into the myofibroblastic phenotype, fibroblasts and CAFs were treated with TGF-β in culture. They were harvested after 3 or 6 days and analysed by RT-PCR and Western blotting.

qtPCR was performed, analysing ACTA2 (SMA), Fibronectin, Collagen1A1 (Col1A1) and Collagen1A2 (Col1A2); beta-2-microglobulin (β2M) was used as a control. The fold change in mean expression was calculated relative to β2M. Figure 6.2 demonstrates that, for fibroblasts, TGF-β treatment augmented transcription of three of the four myofibroblastic genes studied (not significant for ACTA). This indicates that TGF-β induced a myofibroblastic phenotype in the breast fibroblasts isolated.

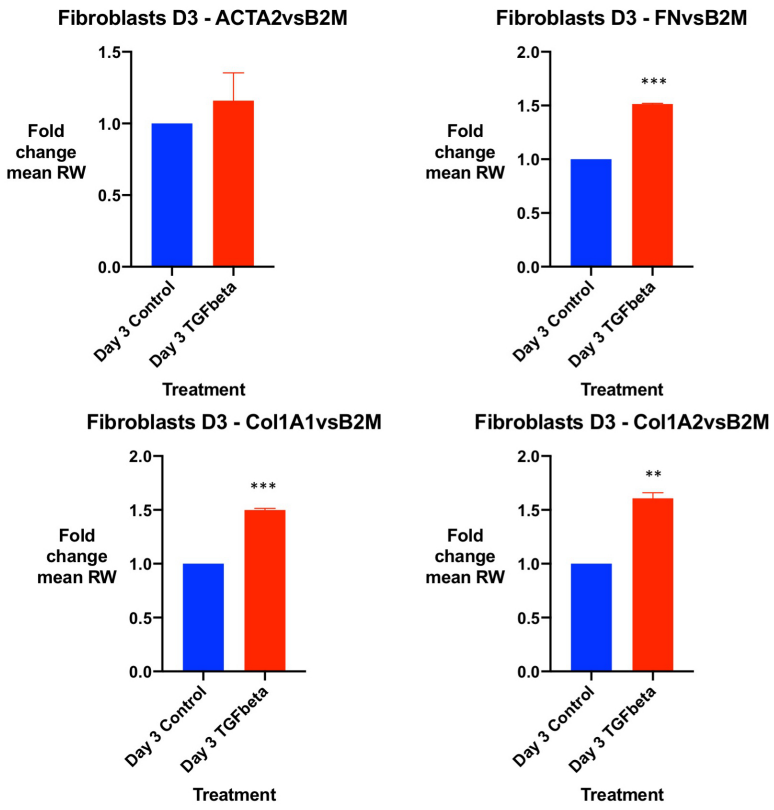



**Figure 6-2. qPCR normal Balb/c fibroblasts, after 3 days of treatment with TGF-β, vs. control.**

**In each graph, control was set at 1.00 fold change mean RW and standard error of the mean is shown (experiment n=2).**

**p-values shown for TGF-β vs. control, unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

Figure 6.3 demonstrates that TGF-β augmented ACTA-2 and Col1A1 expression in CAFs. TGF-β did not change expression of fibronectin or Col1A2. There are no p-values performed because the experiment was only performed once (albeit with PCR performed in triplicates).

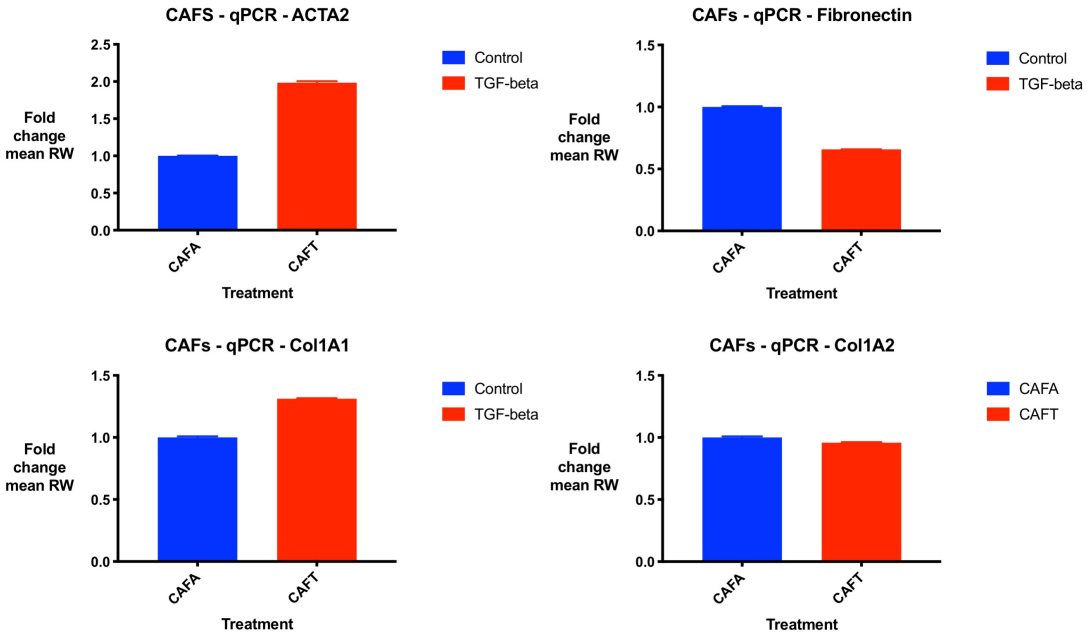



**Figure 6-3. qPCR Balb-neu-T CAFs, after 3 days treatment with TGF-β, vs. control.**

**In each graph, control was set at 1.00 fold change mean RW and standard error of the mean is shown.**

**CAFA = control, CAFD=ATMi, CAFT= TGF-β, CAFTD=TGF-β and ATMi**

To assess protein expression following stimulation by TGF-β, Western blot of fibroblast plates following 3 days of TGF-β treatment was performed. SMA and Col1A1 antibodies were used, with Hsc70 as a positive control. Figure 6.4 demonstrates that TGF-β appeared to increase protein expression of Col1A1 in fibroblasts (despite under loading of the housekeeper extraction on the gel). There was minimal effect on CAFs. SMA was not enhanced by TGF-β in fibroblasts. If anything, there was a marginal reduction in SMA in CAFs treated with TGF-β. Therefore, TGF-β positively effected myofibroblastic protein expression in fibroblasts, but not in CAFs.

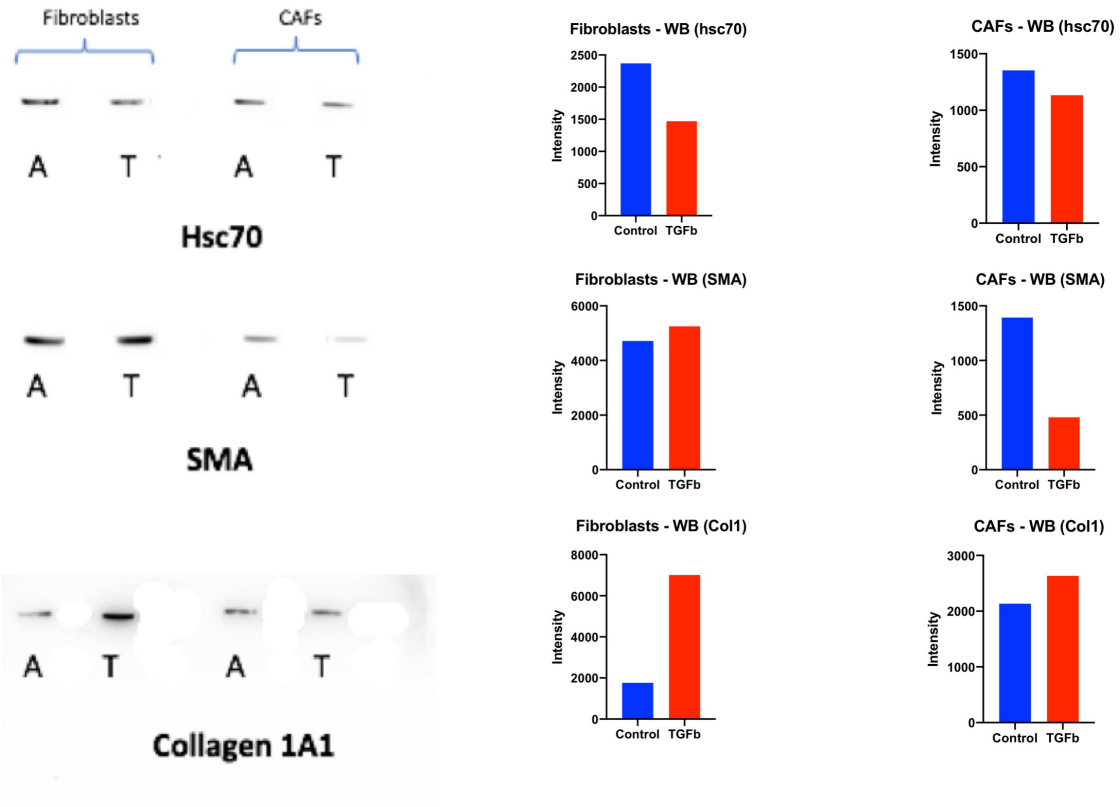



**Figure 6-4. Western blot, using normal Balb/c breast fibroblasts or Balb-Neu-T CAFs, treated +/- TGF-β.**

**A= control, T/TGFb= TGF-β treatment. Graphs show intensity calculated using area under curve in ImageJ.**

In order to further assess the effect of TGF- β on the phenotype of fibroblasts *in vitro*, immunofluorescence was performed after seven days of treatment of fibroblasts/CAFs (Figure 6.5 and 6.6). TGF-β significantly augmented immunofluorescence intensity of SMA, fibronectin and collagen on fibroblasts. There was no change of intensity of fibronectin or SMA using CAFs treated with TGF-β, but collagen was significantly increased.

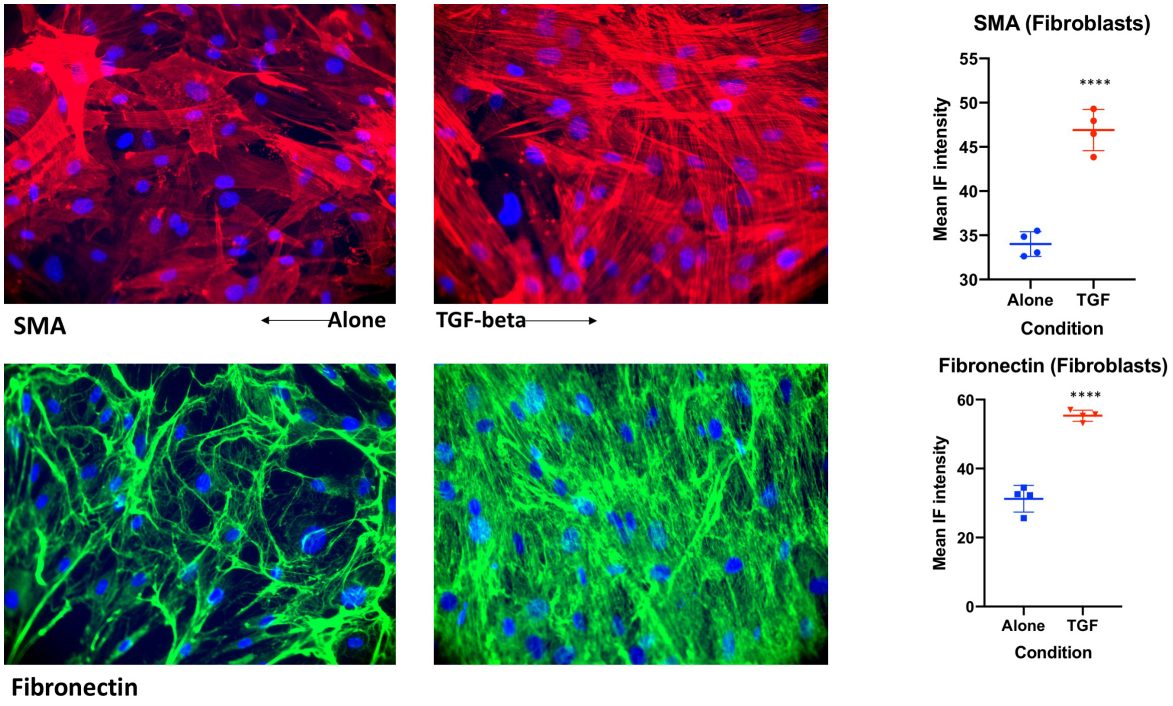



**Figure 6-5. Immunofluorescence with Normal Balb/c breast fibroblasts, control vs. 7 days treatment with TGF-β.**

**Graphs show mean colour intensity of immunofluorescence with fibroblasts, using Image-J (n=4 per condition).**

**p-values shown for TGF-β vs. control, unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

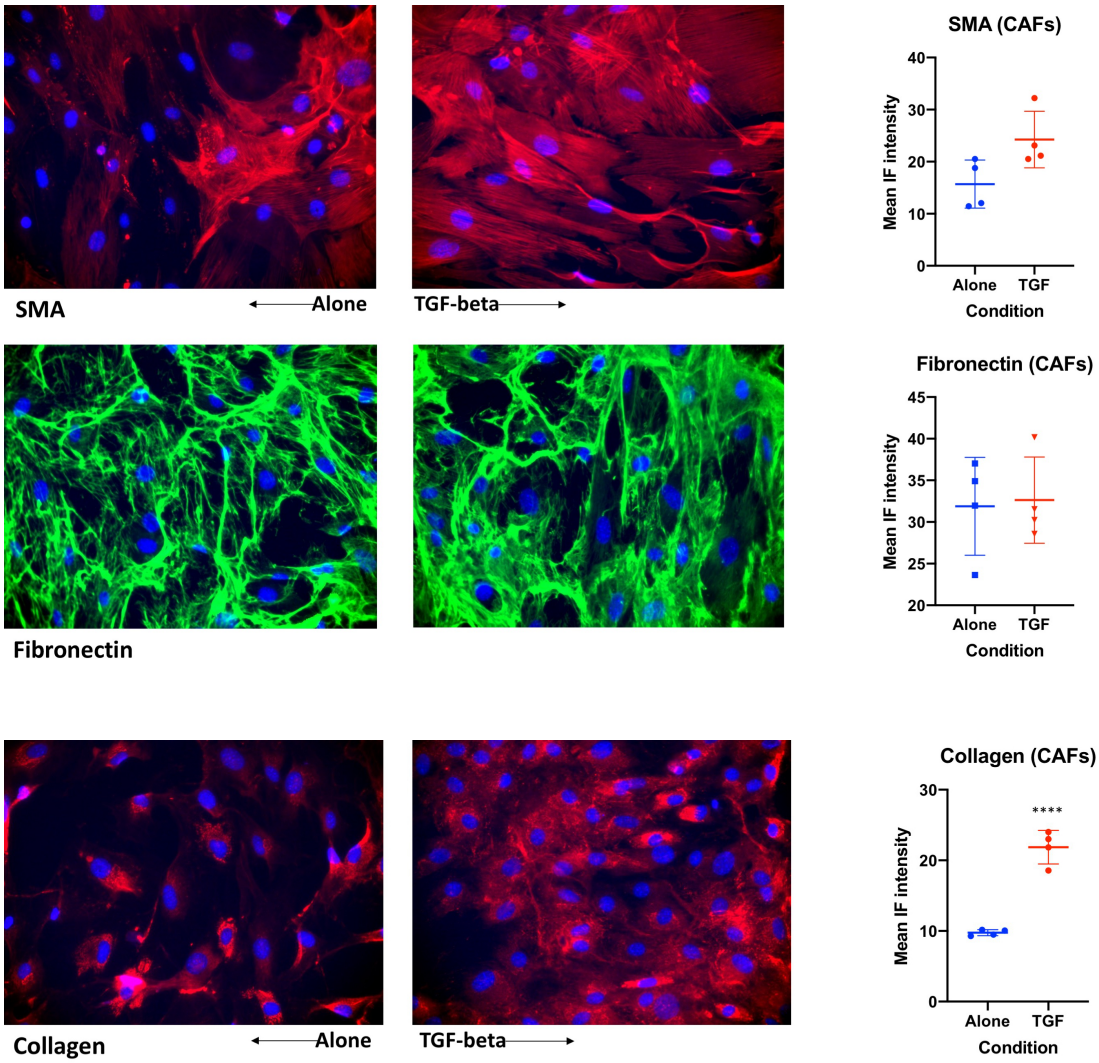



**Figure 6-6. Immunofluorescence with Balb-neu-T CAFs, control vs. 7 days treatment with TGF-β.**

**Graphs show mean colour intensity of immunofluorescence with fibroblasts, using Image-J (n=4 per condition).**

**p-values shown for TGF-β vs. control, unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

### 6.2.2 Effect of fibroblasts on 4T1 tumours in vivo

It has previously been suggested that CAFs within the tumour microenvironment facilitate an immune suppressed environment, meaning that cancer cell proliferation continues in the absence of immune recognition(180). In order to assess the effect of different

fibroblast phenotypes on 4T1 tumours *in vivo,* 30 Balb/c mice were treated with subcutaneous mammary co-injection. Each mouse was injected on Day 0 with 25,000 4T1 cells, in addition to 150,000:

A: Control - nil

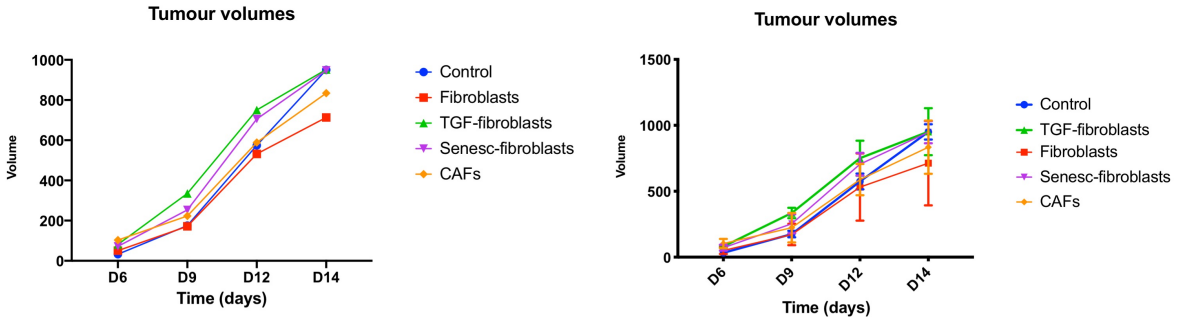
B: Normal Balb/c breast fibroblasts

C: TGF-β treated normal Balb/c breast fibroblasts – treated 6 days prior to injection

D: Senescent fibroblasts – irradiated 5 days prior to injection

E: Balb-Neu-T CAFs

Tumours were measured from Day 6 when palpable and the mice were culled on Day 14 at humane endpoint. Tumour volumes are shown in Figure 6.7.




**Area under curve**

| | Control | Fibroblasts | TGF-beta fibroblasts | Senescent fibroblasts | CAFs |
|---|---|---|---|---|---|
| Total Area | 1241 | 1086 | 1602 | 1469 | 1281 |
| Std. Error | 133.3 | 248.2 | 324.5 | 221.6 | 154.2 |
| 95% CI | 980.2 to 1503 | 599.2 to 1572 | 966.1 to 2238 | 1034 to 1903 | 978.6 to 1583 |

**Figure 6-7. Tumour volumes for Balb/c mice injected with 4T1 cells and other fibroblast types, ratio 6:1.**

**Mean tumour size and mean tumour size with SD shown in graphs. Area under curves shown in Table. No significant difference using two-way repeated measure ANOVA over time.**

Given the overlapping confidence intervals, there is no statistically significant difference in the areas under the curve.

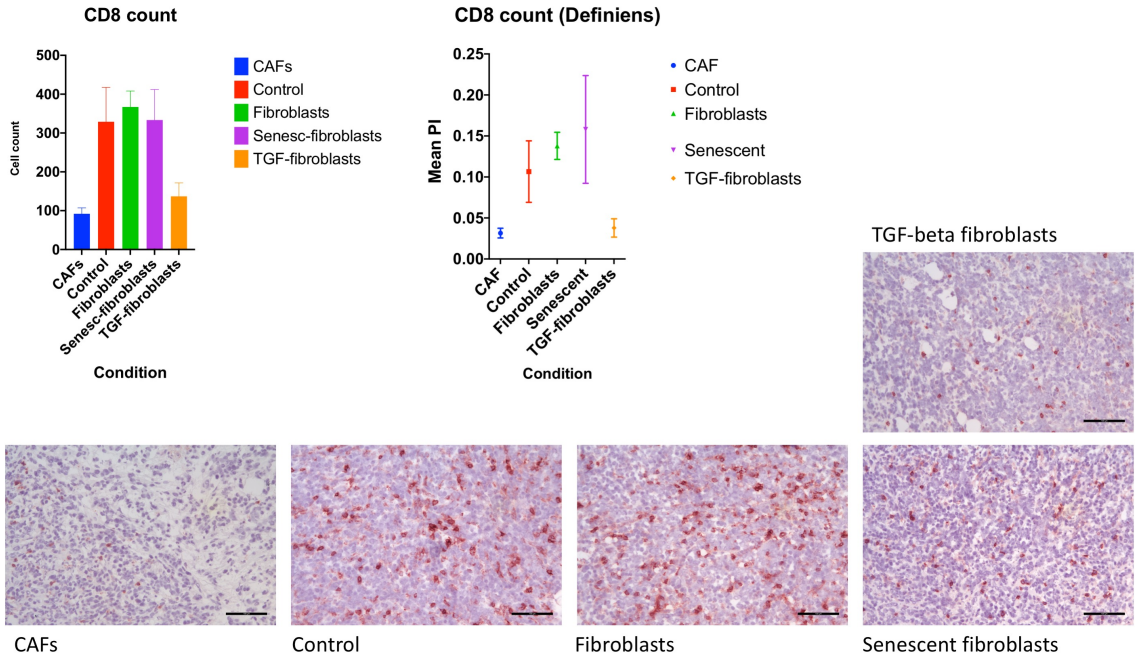


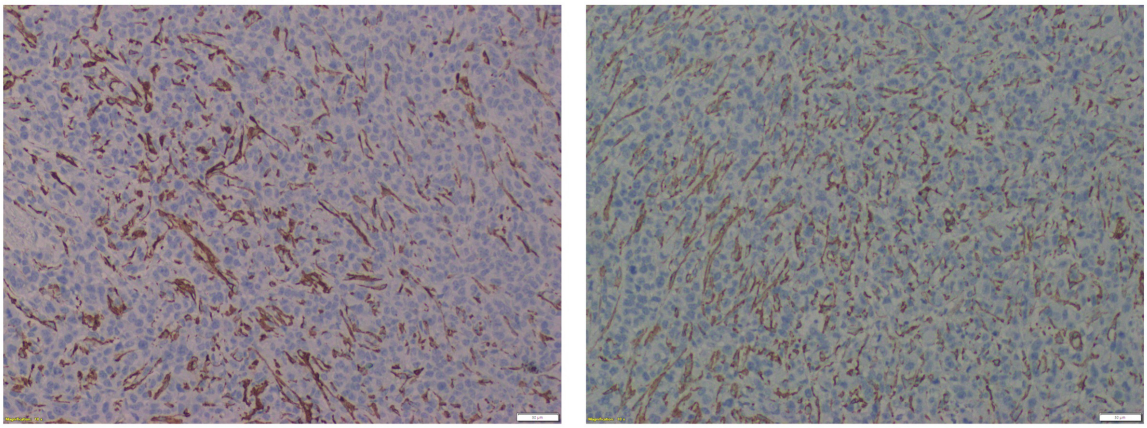

**Figure 6-8. Balb/c mice injected with 4T1 cells and other fibroblast types, ratio 6:1, immunohistochemistry performed with CD8 antibody.**

**For both graphs, N= 4 mice per group, 3 slides per case. Mean with SEM shown. Two-tailed unpaired t-test performed (each test condition compared to control). p-values *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance)**

**For first graph, CD8 count calculated per slide in ImageJ after thresholding, CAFs vs. control: p=0.0728.**

**For second graph, CD8 count calculated in Definiens after thresholding, producing a positivity index, CAFs v. control: p=0.0955.**

**SMA**



Control CAFs

**Figure 6-9. Balb/c mice injected with 4T1 cells and Balb-Neu-T CAFs, ratio 6:1, immunohistochemistry performed with SMA antibody.**

**A single H&E example shown.**
**Magnification x20.**

In order to determine whether the fibroblasts effected any change in effector T-cell or CAF infiltration, CD8 and SMA immunohistochemistry staining were performed (Figures 6.8 and 6.9). There was no significant difference in CD8 expression for any of the cell types used. For CAFs vs. Control, the cell count was lower with a p-value near significance at 0.0728 and 0.0955 for Image-J and Definiens calculations respectively. SMA was universally high throughout all fibroblast conditions studied, therefore not calculated in ImageJ (National Institute of Health, Bethesda, Maryland, USA).

### 6.2.3 Effect of change of fibroblast ratio in vivo

In the previous section, I showed that there was an abundance of fibroblasts present in the 4T1-control tumours and no impact on tumour growth with the addition of TGF-β-activated fibroblasts or CAFs at a 6:1 ratio. Therefore, I repeated the experiment with an increase in fibroblast to tumour ratio from 6:1 to 10:1, in order to evaluate whether this changed the tumour growth and histological fibroblast density between conditions. Thirty

Balb/c mice (n=6 per group) were injected on Day 0 with 15,000 4T1 cells, in addition to 150,000:

A: Control - nil

B: Normal Balb/c breast fibroblasts

C: TGF-β treated normal Balb/c breast fibroblasts – treated 6 days prior to injection

D: Senescent fibroblasts – irradiated 5 days prior to injection

E: Balb-Neu-T CAFs

Tumours were measured from Day 5 when palpable and the mice were culled on Day 16 at humane endpoint.

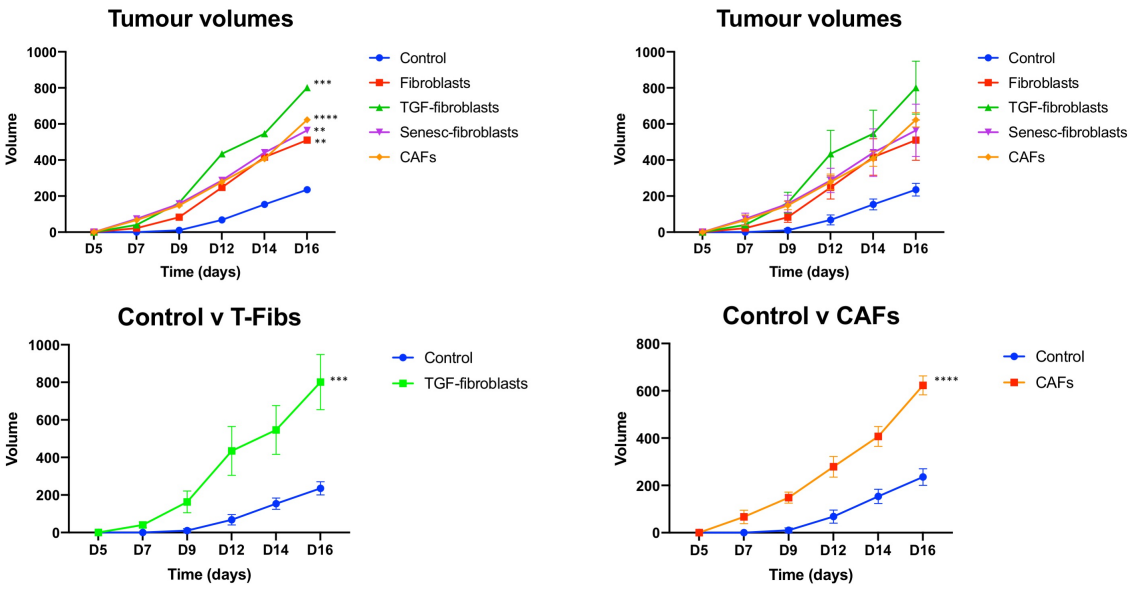
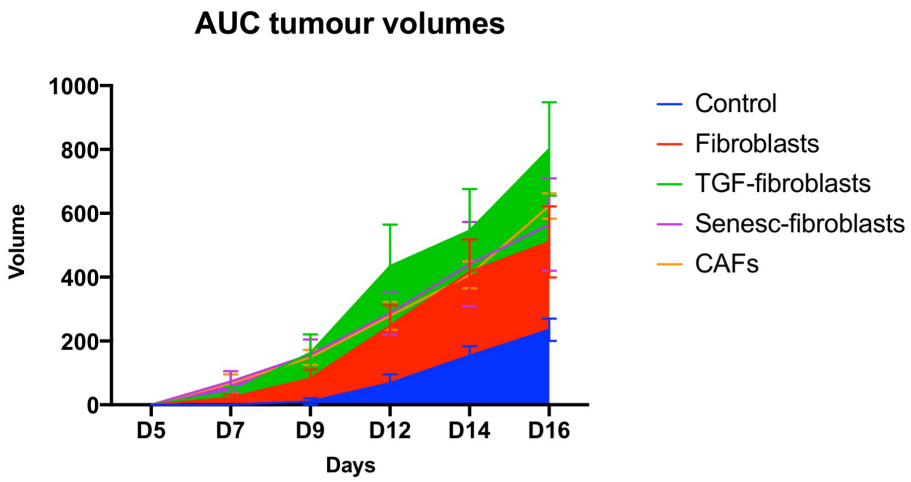



**Figure 6-10. Tumour volumes for Balb/c mice injected with 4T1 cells and varying fibroblast types at a 10:1 ratio.**

**The top left graph shows the mean tumour volume; the top right graph shows the mean with SEM, for each variable. The mean with SEM is also shown separately in the bottom two graphs for TGF-β-activated fibroblasts and CAFs, compared to control.**
**Difference was assessed using two-way repeated measure ANOVA over time, compared to control. p-values *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**




**Area under curve**

| | Control | Fibroblasts | TGF-fibroblasts | Senescent-fibroblasts | CAFs |
|---|---|---|---|---|---|
| **Area** | 349.6 | 1025 | 1585 | 1241 | 1212 |
| **Std. Error** | 84.61 | 253.8 | 379.8 | 327.8 | 132.1 |
| **95% Confidence Interval** | 183.8 to 515.5 | 527.3 to 1522 | 841.0 to 2330 | 598.2 to 1883 | 953.5 to 1471 |

**Figure 6-11. AUC for tumour volumes for Balb/c mice injected with 4T1 cells and varying fibroblast types at a 10:1 ratio.**

**Area under the curves, in addition to 95% CI, for tumour size shown.**

The tumours were significantly larger for all conditions relative to control (Figure 6.10).

The confidence intervals for the area under the curve were not overlapping (Figure 6.11).

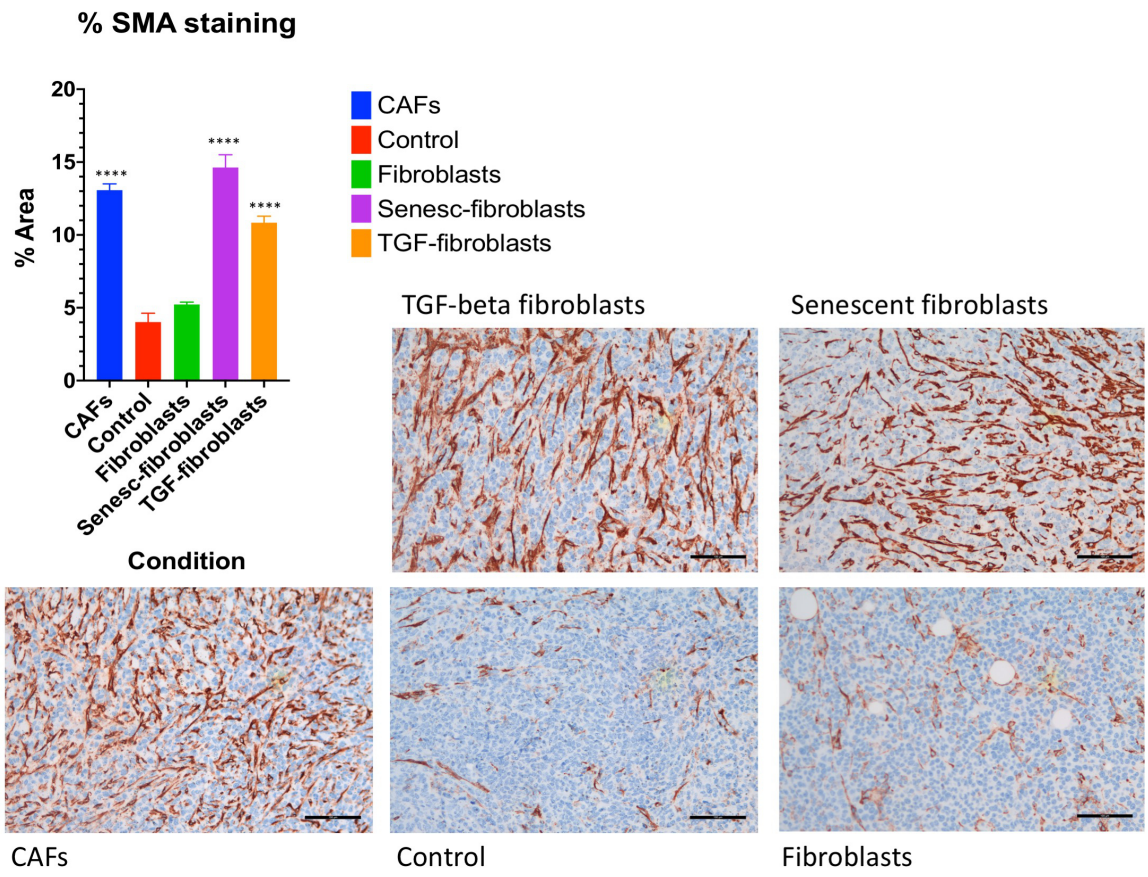



**Figure 6-12. SMA immunohistochemistry using tumours from Balb/c mice injected with 4T1 cells and varying fibroblast types at a 10:1 ratio.**

**N= 4 mice per group, 3 slides per case. Mean with SEM shown. SMA percentage of total area calculated per slide in Image-J after thresholding.**

**Two-tailed unpaired t-test performed (test condition compared to control). p-values *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

In order to determine whether there was a change in abundance of myofibroblasts between the different conditions, SMA immunohistochemistry staining was performed (Figure 6.12). SMA expression was significantly higher, relative to control, for senescent fibroblasts, TGF-β-activated fibroblasts and CAFs. The normal fibroblast arm had no change in SMA expression relative to control.

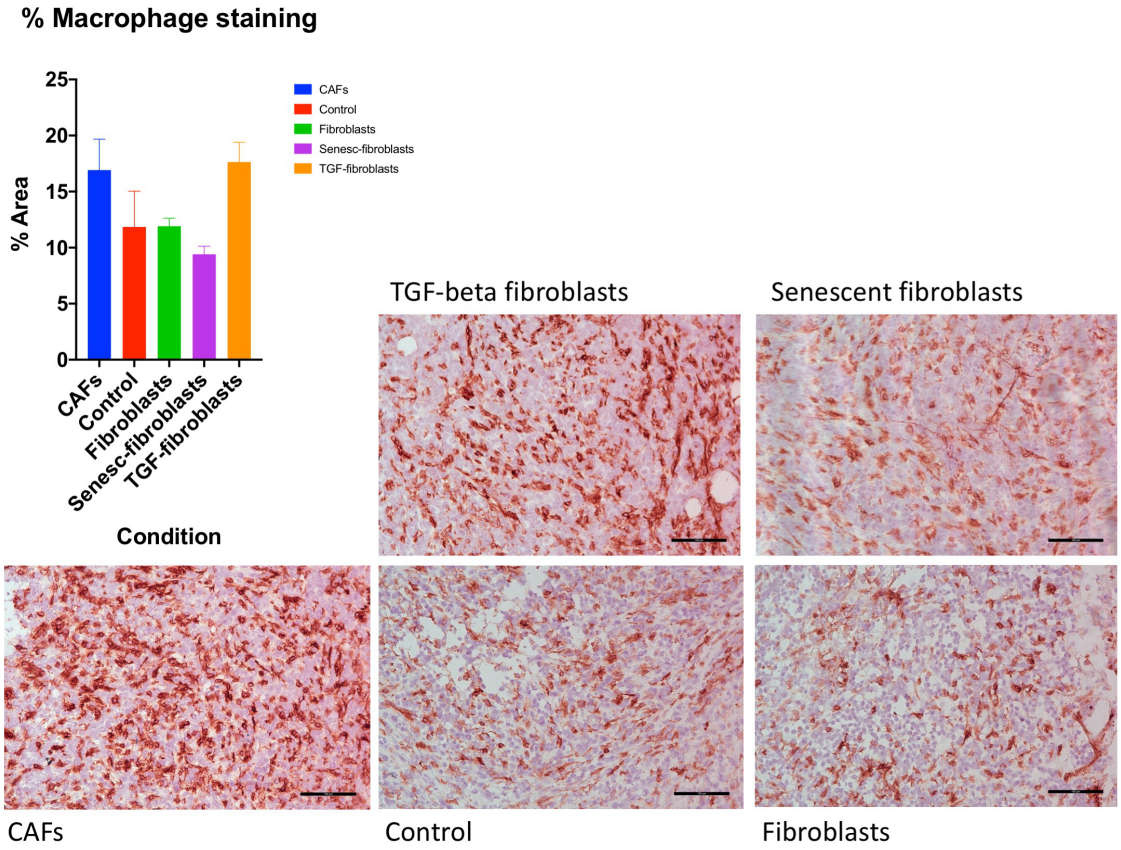



**Figure 6-13. F4/80 immunohistochemistry using tumours from Balb/c mice injected with 4T1 cells and varying fibroblast types at a 10:1 ratio.**

**N= 4 mice per group, 3 slides per case. Mean with SEM shown. F4/80-positive percentage of total area calculated per slide in Image-J after thresholding.**

**Two-tailed unpaired t-test performed (test condition compared to control). p-values *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

Macrophages are an important component of the leukocyte infiltrate present in tumours and can attract fibroblasts via the secretion of chemokines(289, 290). Tumour-associated macrophages (TAMs) have previously been correlated with poor prognosis in breast cancer(291). In order to determine whether the different fibroblast phenotypes had an effect on macrophage infiltration, F4/80 staining was performed (Figure 6.13). There was no significant difference in expression between each of the experiment arms, compared to control.

## 6.3 Discussion

### 6.3.1 *In vitro* – Stimulation of fibroblasts with TGF-β

In this chapter, I have demonstrated successful isolation of breast CAFs from murine tumours. In vitro, TGF-β stimulates the development of myofibroblastic features in normal breast fibroblasts, such as increased collagen and fibronectin expression. I have shown this using qPCR, western blot, immunofluorescence, cell count and metabolism assays.

### 6.3.2 *In vivo* – Effect on 4T1 tumours by costimulation with fibroblasts

I then proceeded to explore the effect of fibroblast co-injection *in vivo.* Initially, using a 6:1 ratio of CAFs/fibroblasts to tumour cells, co-injection did not alter the tumour growth of the 4T1 mouse model. The model used is a 4T1 tumour in immunocompetent Balb/c mice; this is an accepted spontaneous, metastatic murine model of TNBC(292). However, on review of the immunohistochemistry, I noted that the control tumours were already significantly fibrotic. It was therefore hypothesized that a significant number of fibroblasts would need to be co-injected to alter the fibrosis within the tumour sufficiently for a change in growth. Using a 10:1 ratio, the growth in co-injected mice significantly increased. This was most significant for the CAF, senescent fibroblast and TGF-β fibroblasts arms. SMA expression was significantly increased in these groups, compared to the control tumours.

Given that normal breast fibroblasts also increased the growth of 4T1 tumours *in vivo*, it is hypothesised that after co-injection, some of these cells were activated within the tumour, resulting in differentiation to CAFs during the course of the experiment. A

number of different cell types can differentiate into CAFs, including resident tissue fibroblasts(184). Fibroblast-activating factors, tumour-derived exosomes and epigenetic modification could all contribute to this process(184). TGF-β is certainly involved in promoting the differentiation into activated fibroblasts. CAFs secrete large amounts of this cytokine resulting in the generation of an autocrine loop(178). Tumour cells can also secrete enough TGF-β to activate primary fibroblasts; indeed breast cancer cells have been shown to induce matrix-metalloproteinase 9 (via TGF-β) in fibroblasts(293). Even a transient interaction between breast cancer cells and normal fibroblasts appears sufficient to modify the tumour microenvironment, increase neoplastic behaviour and enhance metastatic potential of cancer cells *in vivo* (modified by TGF-β) (294).

However, it should be noted that in this experiment, SMA expression was not significantly higher for mouse tumours resulting from co-injection with normal fibroblasts relative to control. Therefore, it is presumed that not all of the cells have transformed. One of the most unique features of CAFs is their ability to remodel and synthesise the extracellular matrix, resulting in mechanical stiffening of tissue. Increasing stromal stiffness has been correlated with poor prognosis in breast cancer (295). The activated normal fibroblasts in the 4T1 model may not have had sufficient time to evolve to this pro-tumorigenic behaviour prior to the end of the experiment, given the relatively short length of murine experiments.

Fibroblasts activated by TGF-β prior to co-injection increased tumour growth and SMA expression, similar to CAFs. Given the key role of TGF-β in the differentiation of precursor cells into CAFs, this is perhaps not surprising(296, 297). Once nearby each other, tumour cells and CAFs participate in cross-talk which facilitates ECM remodelling and deposition, immune cell polarization, cellular proliferation metabolic reprogramming and promotion of stem-like behaviour(296). Breast stromal fibroblasts remaining in “normal” tissue following surgical resection have even been shown to have gene expression comparable to that of CAFs; this was hypothesised to occur secondary to TGF-β or metabolic programming and to provide an environment for local recurrence(298).

### 6.3.3 Senescent fibroblasts

In these experiments, senescent fibroblasts have performed similarly to CAFs and TGF-β-activated fibroblasts *in vivo*. Our group has previously shown that induction of fibroblast senescence appears to generate a myofibroblastic-type cell(195). In those experiments, senescent fibroblasts did not appear to deposit collagenous, pro-tumorigenic ECM at the same rate as TGF-β-activated fibroblasts. However, senescent human fibroblasts are known to secrete factors that stimulate malignant breast epithelial cells (such as matrix metalloproteinases) and deposit matrix that is pro-tumorigenic(299). Here, it has been shown that senescent fibroblasts, along with normal breast fibroblasts, promote growth of the tumour and are associated with increased SMA expression. However, it is unclear to what degree the senescent nature of these cells contributes to tumour-promoting behaviour and, if so, by which mechanisms. The relative importance of the secretory and matrix-depositing functions of myofibroblastic cells in tumour promotion needs further clarification.

### 6.3.4 Macrophages

Although not significant, there was a trend to a larger number of macrophages in the TGF-β and CAF arms compared to the control arm. Tumour-associated macrophages (TAMs) are known to be involved in neoangiogenesis, tumour cell invasion and immunosuppressive effects(300). TAMs can be polarized towards the M1 or M2 subtype; the latter is pro-tumorigenic while the former exerts opposing effects(175). Tumour cells can promote polarization of TAMs towards the M2 subtype; this effect is mediated by several cytokines including TGF-β(300). Indeed CAFs have previously been implicated in the recruitment and proliferation of M2 macrophages in the TME(175). CAFs have been correlated with TAMs in TNBC(301). In future experiments, it would be important to

understand what proportion of M1 vs. M2 TAMs are present in CAF targeting experiments. Expansion of human TMA staining to incorporate macrophage staining would also be useful.

### 6.3.5 Limitations

As discussed in 6.2.1, the "CAFs" isolated are acknowledged to be tumour-derived fibroblasts, likely to be enriched in CAFs. In order to identify the percentage of these isolated cells that are CAFs, further evaluation of the cells present in this population could be performed (e.g. with flow cytometry) at baseline. In addition, the CAFs have been isolated from Balb-neu-T mice, which is a murine HER2+ breast cancer model. There is potential that the phenotype of CAFs from a HER2+ breast tumour is phenotypically different to those from a TNBC tumour. It has been acknowledged that CAF subsets (at least four) accumulate differently in different subtypes of breast cancer, with each subset having different properties and effects on the tumour microenvironment(173).

The *in vivo* experiment performed here would need to be repeated to validate the important results found. When the optimised ratio experiment is repeated, I would perform CD8 staining, as despite three attempts the CD8 staining for this experiment was unsuccessful. To better assess the functional status of CD8 cells, other markers could be assessed, e.g. by flow cytometry, such as LAG3 and TIM3. Some of the *in vitro* experiments have only been performed once and also warrant repeating. Functional assays, such as gel contraction studies, could also be performed *in vitro* to measure changes in the myofibroblastic phenotype.

## 6.4 Summary

Here I have shown that TGF-β differentiates normal murine breast fibroblasts into myofibroblasts *in vitro*. In the 4T1 "triple negative" mouse model, myofibroblasts and senescent fibroblasts augment tumour growth, associated with increase SMA deposition. This result is in keeping with the relationship between SMA and outcome/CD8 expression in human TNBC described in Chapter 5, indicating that CAFs may contribute to an immune exclusion phenotype. Given this, I have decided to explore the possibility of targeting CAFs in further *in vivo* work(184).

# Chapter 7 Targeting Cancer-associated Fibroblasts in Breast Cancer

## 7.1 Introduction

As outlined in Section 1.8.3, CAF inhibition could be useful in combination with other therapy; its adjunctive role with immunotherapy is of particular interest. The NOX/ATM/p53 axis has been identified as being important in myofibroblast differentiation (see Section 1.8.4) and is therefore a possible therapeutic target in fibrotic disorders and cancer. Our group has explored NOX inhibition in previous work(183). ATM has been shown to mediate TGF-β fibrotic responses in a redox-sensitive manner(192). The extent to which ATM activity is required for myofibroblast differentiation is unclear.

The ATM inhibitor (ATMi), AZD0156, is a potent, first-in-class, bioavailable drug undergoing clinical testing in a phase I trial (NCT02588105)(302). It is being tested as a monotherapy but also in combination with targeted therapy including PARP inhibitors, in the treatment of solid tumours(191). *In vitro*, it has been shown to be a radio-sensitiser. *In vivo*, it has been shown to enhance lung tumour control using radiotherapy. It has also been shown to improve the efficacy of olaparib, using TNBC xenograft models(302). In general, DDR inhibitors are being used in combination with other therapies, with the aim of exploiting synthetic lethality, or in patients who have specific genomic alterations (e.g. DNA repair deficiencies) in tumour cells(191). However, these inhibitors may also have a role to play in targeting the tumour microenvironment; their application may therefore be useful in a wider patient group.

The potential for ATM targeting to inhibit the fibrotic, tumour-promoting microenvironment in cancer has not been explored. I have therefore assessed the effect

of ATM inhibition on breast cancer CAFs *in vitro*, initially using the dose advised by AstraZeneca (Cambridge, UK). I then evaluated for a change in effect at different doses. Finally, I investigated whether ATM inhibition influences tumour growth and immune cell influx *in vivo*.

## 7.2 Results

### 7.2.1 Initial ATM-inhibitor *in vitro* testing

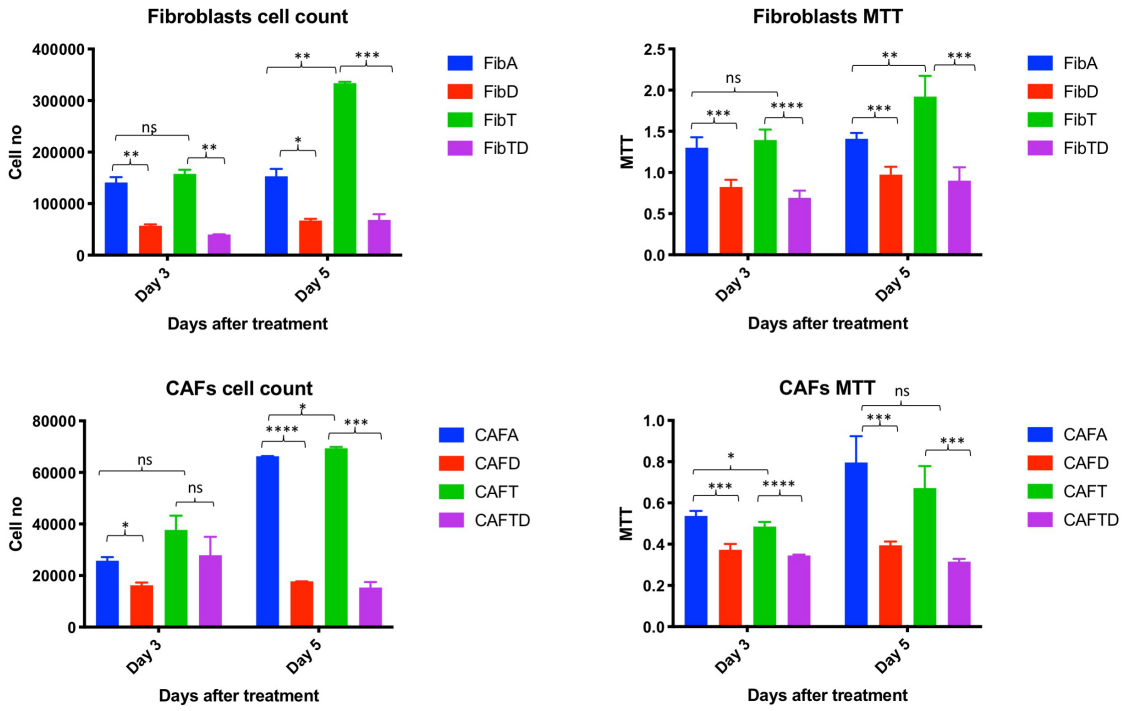



**Figure 7-1. Balb/c murine fibroblasts and Balb-NeuT murine CAFs, cell count and MTT 3 and 5 days after TGF-β and/or ATM inhibitor (AZD0156) [1.25uM] treatment.**

**Mean with standard deviation shown for n=2 (cell count) and n=4 (MTT).**

**p-values shown for treatment arm vs. control, unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

**FibA/CAFA = control, FibD/CAFD=ATMi, FibT/CAFT= TGF-β, FibTD/CAFTD=TGF-β and ATMi**

In order to assess for the effect of ATM inhibition on the proliferation and metabolism of breast fibroblasts and CAFs, cell count and MTT assays were performed after 3 and 5 days of treatment with AZD0156 (Figure 7.1). The ATM inhibitor was effective at inhibiting metabolism for both fibroblasts and CAFs, with or without TGF-β treatment. AZD0156

reduced the number of viable cells in every condition, raising concerns about cellular toxicity.

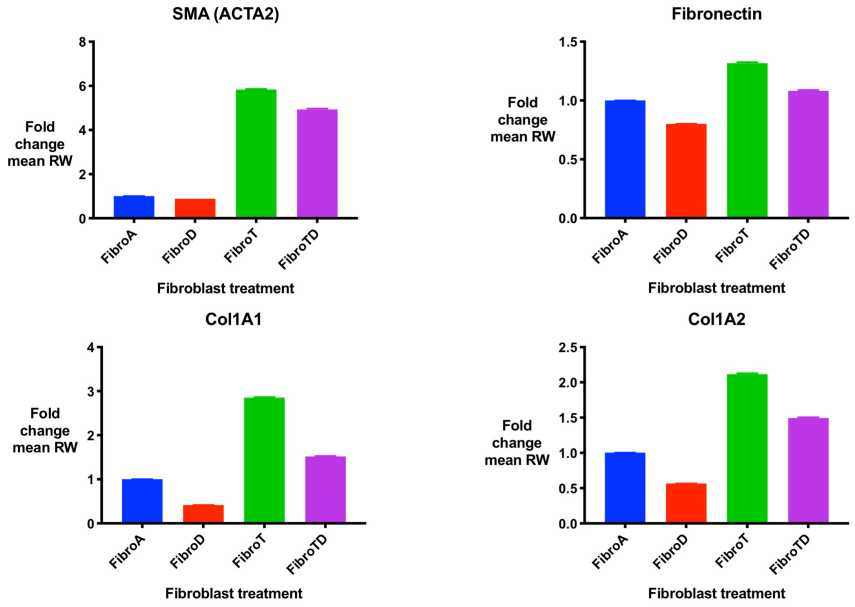



**Figure 7-2. qPCR normal Balb/c fibroblasts, after 3 days of treatment with TGF-β and/or ATM inhibitor (AZD0156) [1.25uM], vs. control.**

**In each graph, control was set at 1.00 fold change mean RW and mean with standard error of the mean is shown.**

**FibA= control, FibD=ATMi, FibT=TGF-β, FibTD=TGF-β and ATMi**

In order to test whether ATM inhibition was capable of inhibiting the myofibroblastic phenotype of breast fibroblasts and CAFs, the cells were treated with 3 days of AZD0156 and/or TGF-β and thereafter qtPCR and Western Blot performed. QtPCR compares relative gene expression for the chosen target genes relative to the housekeeper gene (beta-2 microglobulin). A fold-change value was then calculated by comparing these to the normalised gene values. The cells alone (without the addition of TGF-β or ATM inhibitor) were used as the control arm. The qtPCR experiments showed a moderate reduction in myofibroblastic gene expression when fibroblasts were treated with the ATM-inhibitor, slightly more pronounced after pretreatment with TGF-β (Figure 7.2).

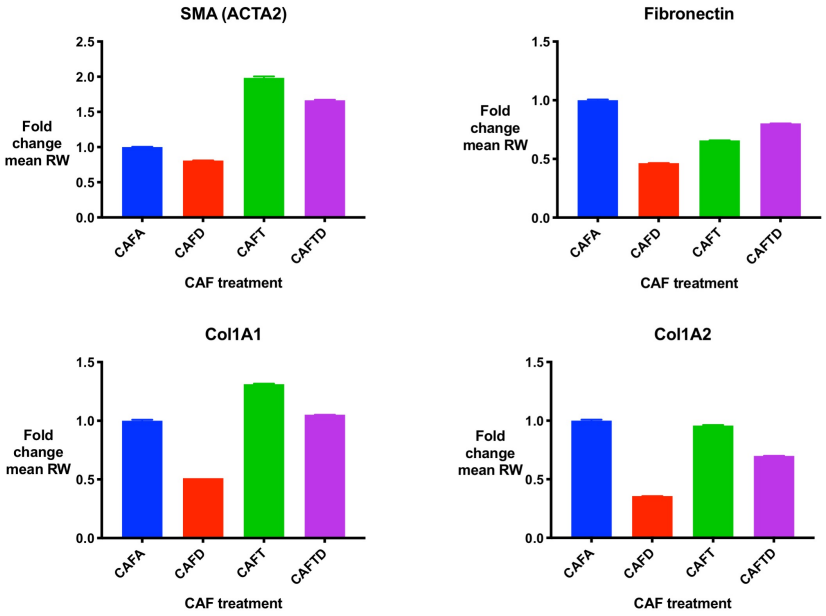



**Figure 7-3. qtPCR Balb-neu-T CAFs, after 3 days of treatment with TGF-β and/or ATMi inhibitor (AZD0156) [1.25uM], vs. control.**

**In each graph, control was set at 1.00 fold change mean RW and mean with standard error of the mean is shown.**

**CAFA= control, CAFD=ATMi, CAFT=TGF-β, CAFTD=TGF-β and ATMi.**

For CAFs, ATM inhibition again reduced myofibroblastic gene expression - this was more pronounced without pretreatment with TGF-β (and fibronectin expression actually increased after treatment with the drug in combination with TGF-β) (Figure 7.3). There are no p-values performed because the experiment was only performed once (albeit with PCR performed in triplicates).

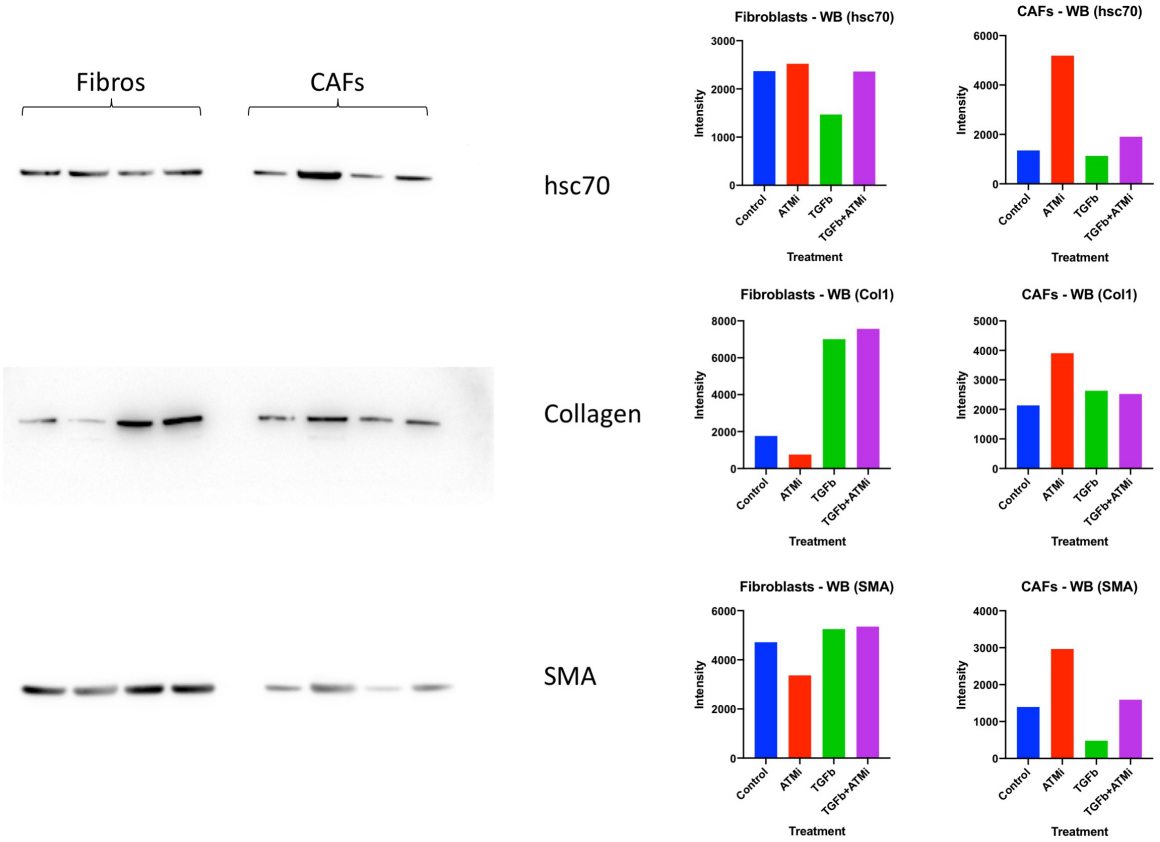
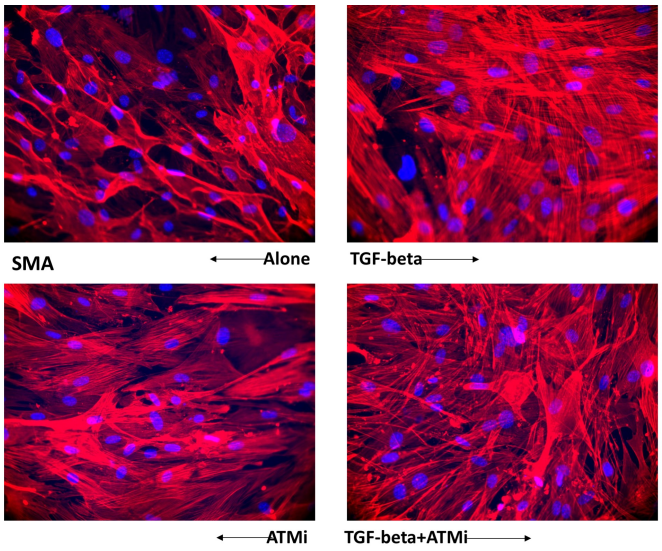
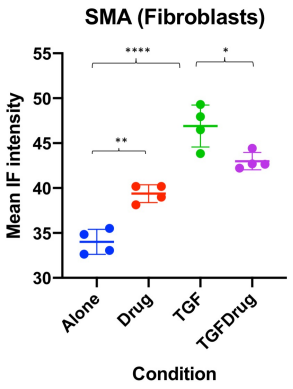


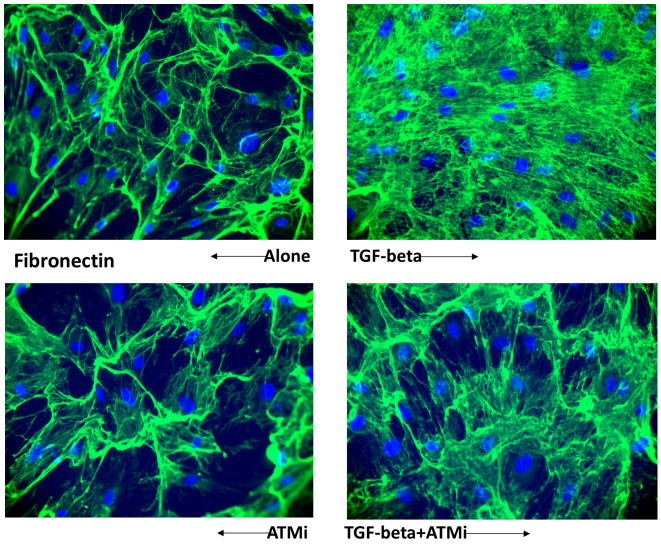

**Figure 7-4. Western blot, using normal Balb/c breast fibroblasts or Balb-Neu-T CAFs, treated with 3 days of TGF-β and/or ATM inhibitor (AZD0156) [1.25uM], vs. control.**

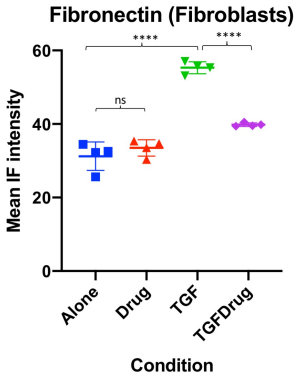
**Graphs show intensity calculated using area under curve in ImageJ.**

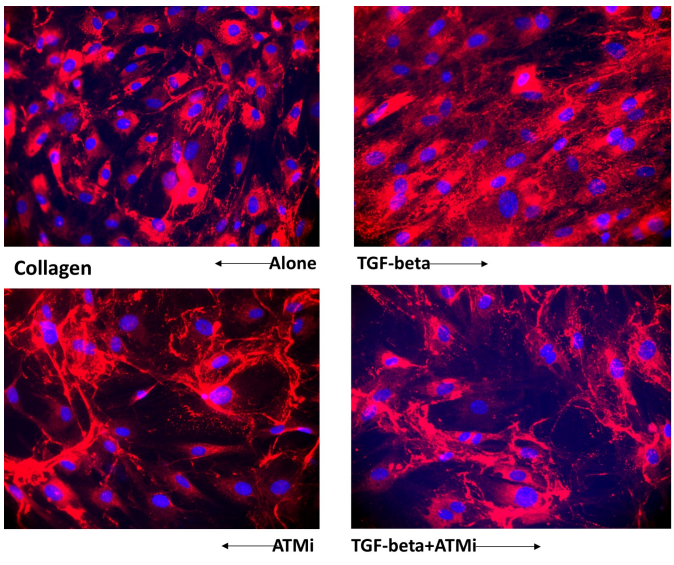
To assess for changes in protein expression following ATM inhibition, western blot was performed (Figure 7.4). For fibroblasts, ATM inhibition slightly reduced the expression of SMA and collagen. Once cells were treated with TGF-β however, ATM inhibition did not reduce their expression. For CAFs treated with TGF-β, ATM inhibition did not reduce the expression of these antibodies.


SMA
Alone
TGF-beta
ATMi
TGF-beta+ATMi


SMA (Fibroblasts)
Mean IF intensity
55
50
45
40
35
30
****
*
**
Alone
Drug
TGF
TGFDrug
Condition


Fibronectin
Alone
TGF-beta
ATMi
TGF-beta+ATMi


Fibronectin (Fibroblasts)
Mean IF intensity
60
40
20
0
****
****
ns
Alone
Drug
TGF
TGFDrug
Condition


Collagen
Alone
TGF-beta
ATMi
TGF-beta+ATMi

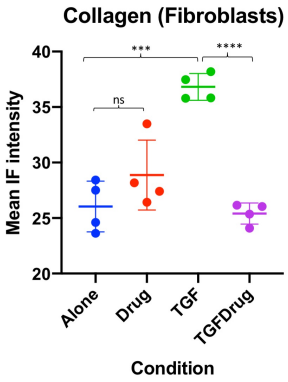
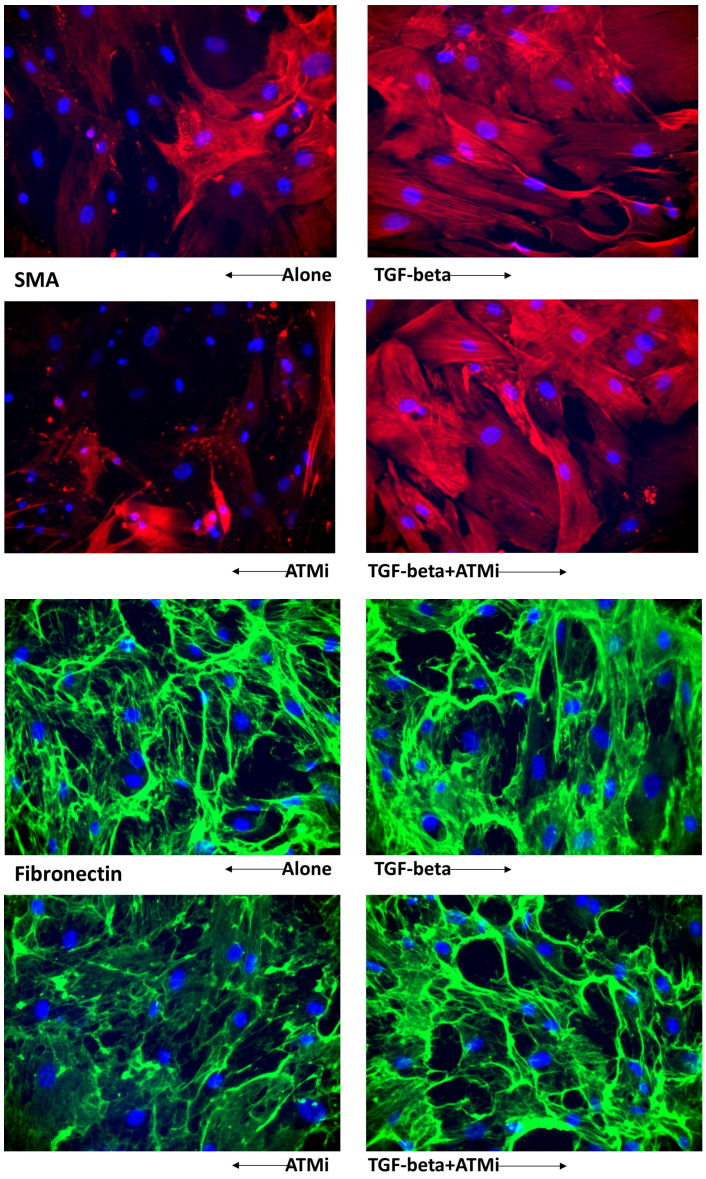
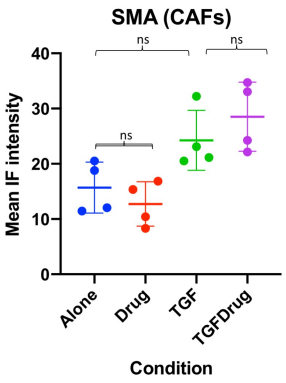

Collagen (Fibroblasts)
Mean IF intensity
40
35
30
25
20
***
****
ns
Alone
Drug
TGF
TGFDrug
Condition

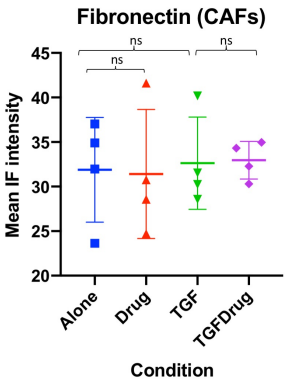
**Figure 7-5. Immunofluorescence with Normal Balb/c breast fibroblasts, control vs. 7 days of treatment with TGF-β and/or ATM-inhibitor (AZD0156) [1.25uM].**

**Graphs show mean and standard deviation of colour intensity of immunofluorescence with fibroblasts, using ImageJ (n=4 per condition).**
**p-values for unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

To further assess for a change in phenotype with ATM inhibition, immunofluorescence was performed after the cells were treated for 7 days with ATM-inhibitor and/or TGF-β (Figures 7.5 and 7.6). For fibroblasts alone, the drug increased expression of SMA and made no difference to expression of collagen or fibronectin. For fibroblasts activated with TGF-β, the drug inhibited expression of all three markers. For CAFs, the drug only changed the immunofluorescence of collagen and only after pretreatment with TGF-β.


SMA
Alone
TGF-beta
ATMi
TGF-beta+ATMi
Fibronectin
Alone
TGF-beta
ATMi
TGF-beta+ATMi


SMA (CAFs)
ns
ns
ns
Mean IF intensity
40
30
20
10
0
Alone
Drug
TGF
TGFDrug
Condition


Fibronectin (CAFs)
ns
ns
ns
Mean IF intensity
45
40
35
30
25
20
Alone
Drug
TGF
TGFDrug
Condition

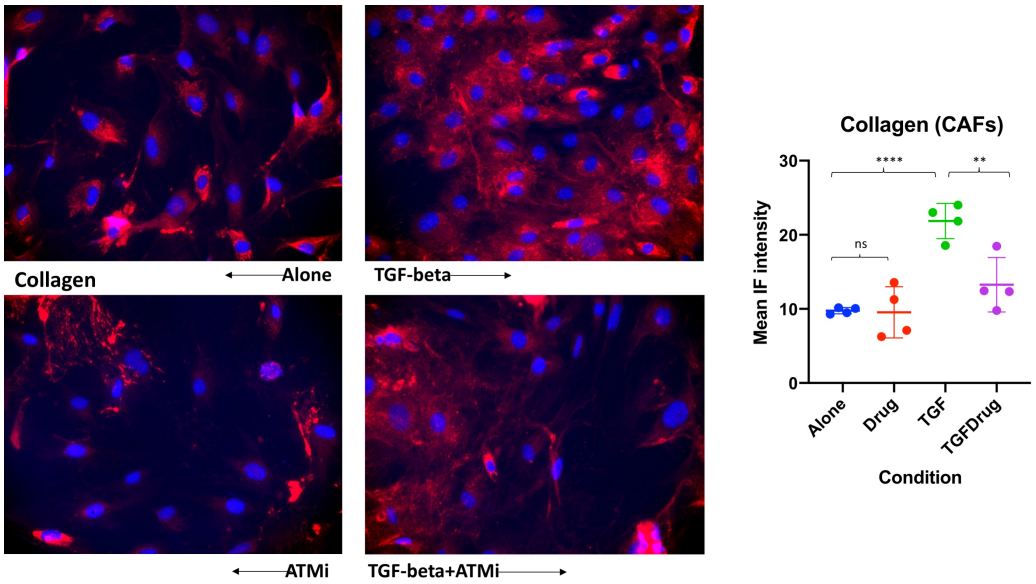



**Figure 7-6. Immunofluorescence with Balb-neu-T breast CAFs, control vs. 7 days treatment with TGF-β and/or ATM-inhibitor (AZD0156) [1.25uM].**

**Graphs show mean and standard deviation of colour intensity of immunofluorescence with fibroblasts, using ImageJ (n=4 per condition).**

**p-values for unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

Given the significant reduction in viable cells (for both CAFs and fibroblasts) found after treatment with the ATM-inhibitor, I was concerned about the toxicity of the recommended *in vitro* dose of 1.25uM. I therefore decided to repeat the above experiments with varying doses of the drug to explore whether the myofibroblastic phenotype could be effectively inhibited with a lower, less toxic dose.

### 7.2.2 ATM dosing *in vitro*

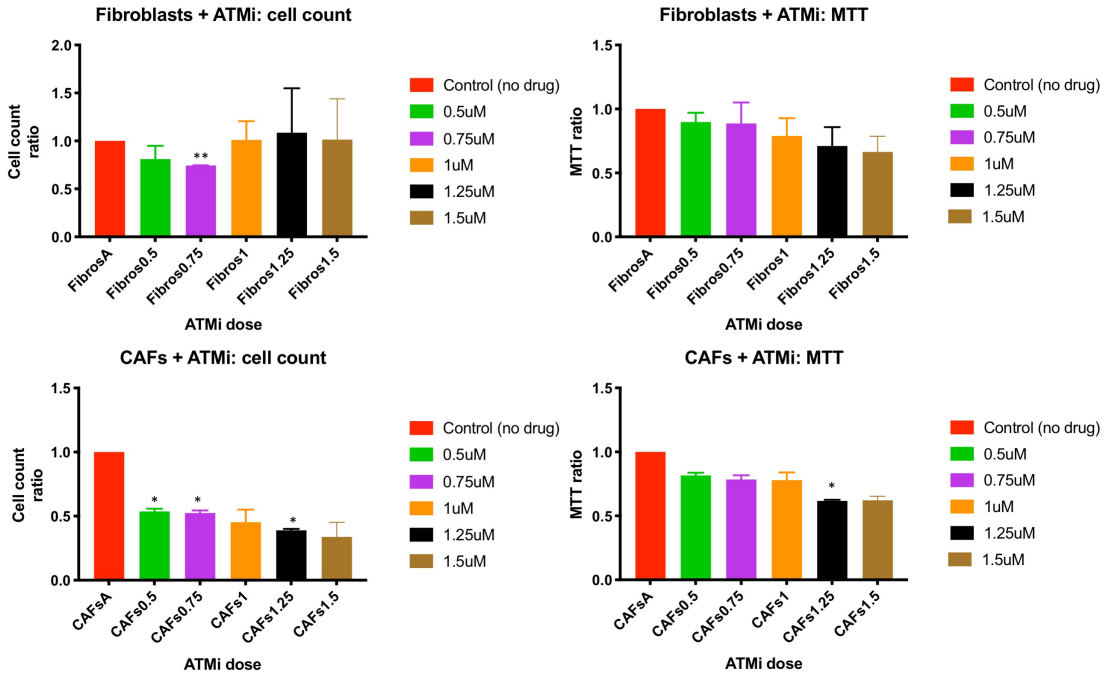



**Figure 7-7. Balb/c murine fibroblasts and Balb-NeuT murine CAFs, cell count and MTT 3 days after ATM inhibitor (AZD0156) treatment, control then escalating doses from 0.5 to 1.5uM.**

**Mean with standard deviation shown (experiment n=2). Samples per experiment: n=2 (cell count) and n=4 (MTT).**

**p-values shown for treatment arm vs. control, unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

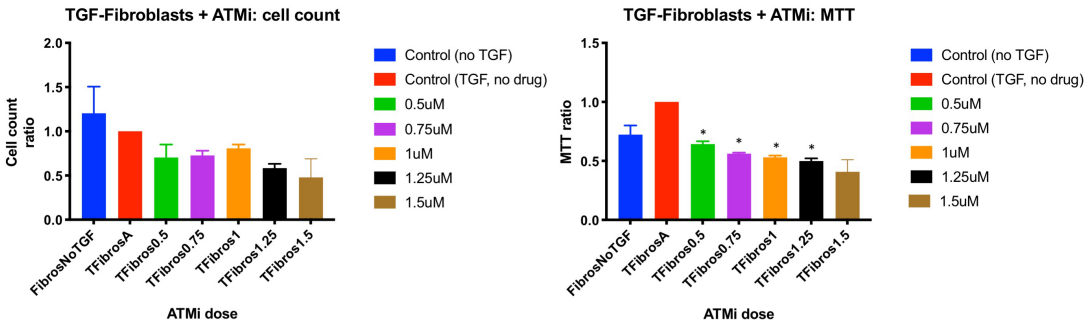


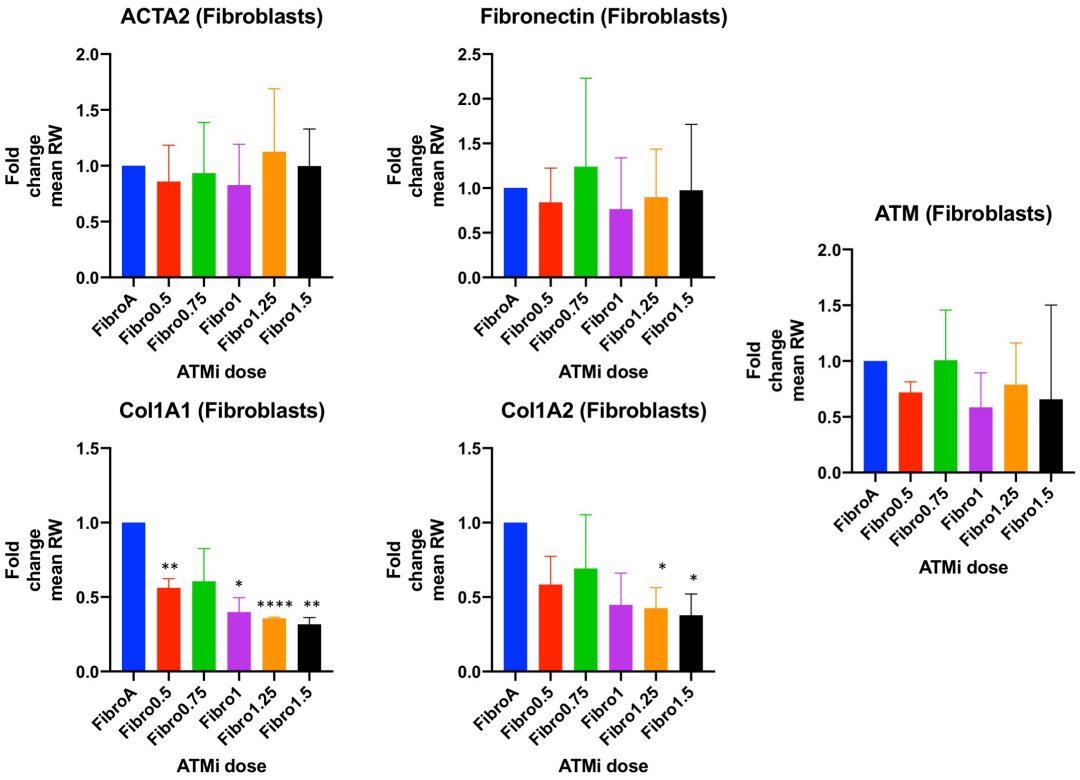

**Figure 7-8. Balb/c murine normal breast fibroblasts activated with TGF-β, cell count and MTT 3 days after ATM inhibitor (AZD0156) treatment, control then escalating doses from 0.5 to 1.5uM.**

**Mean with standard deviation shown (experiment n=2). Samples per experiment: n=2 (cell count) and n=4 (MTT).**

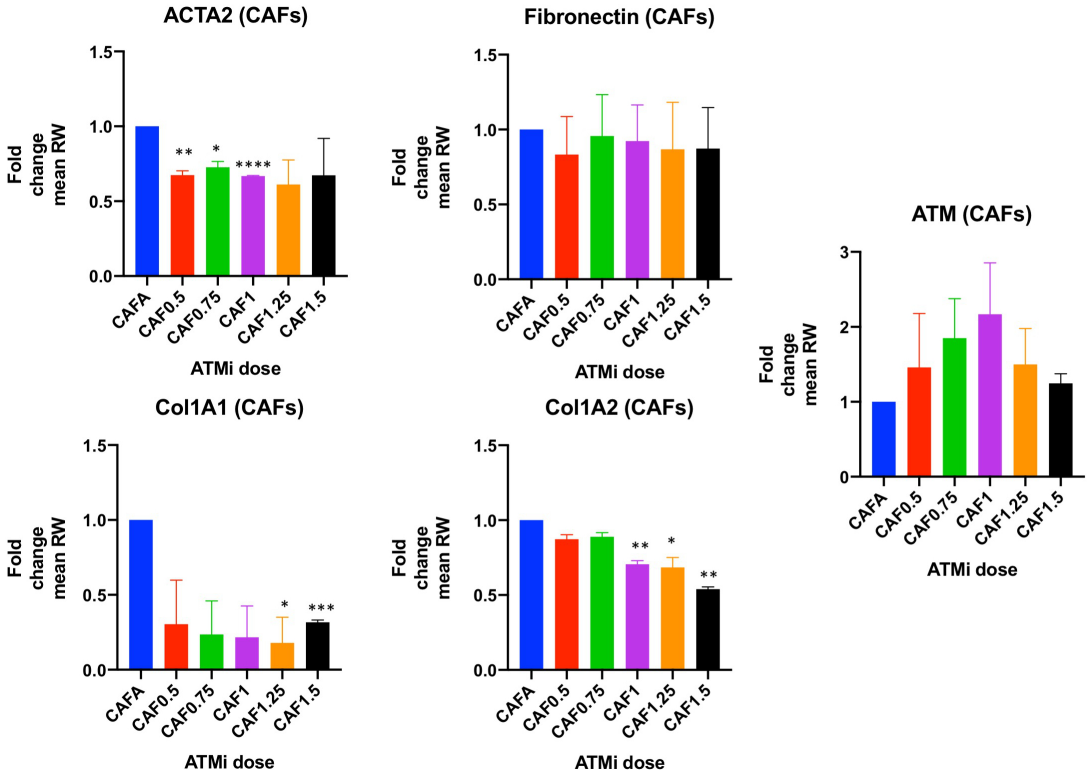
**p-values shown for treatment arm vs. control, unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

To determine whether changing the dose of the ATM-inhibitor had an effect on cellular toxicity, cell count and metabolism assays were performed for normal breast fibroblasts, CAFs and fibroblasts activated with TGF-β with escalating doses of ATM-inhibitor relative to control (Figures 7.7 and 7.8). For fibroblasts alone, ATM inhibition did not significantly affect the cell count or metabolism. For fibroblasts pretreated with TGF-β, the ATM-inhibitor significantly reduced the viable cell count and metabolism for all doses studied. There was no significant difference in toxicity between successive doses.

**A.**


ACTA2 (Fibroblasts)
Fibronectin (Fibroblasts)
ATM (Fibroblasts)
Col1A1 (Fibroblasts)
Col1A2 (Fibroblasts)
Fold change mean RW
FibroA
Fibro0.5
Fibro0.75
Fibro1
Fibro1.25
Fibro1.5
ATMi dose

**B.**


ACTA2 (CAFs)
Fibronectin (CAFs)
ATM (CAFs)
Col1A1 (CAFs)
Col1A2 (CAFs)
Fold change mean RW
CAFA
CAF0.5
CAF0.75
CAF1
CAF1.25
CAF1.5
ATMi dose

**C.**

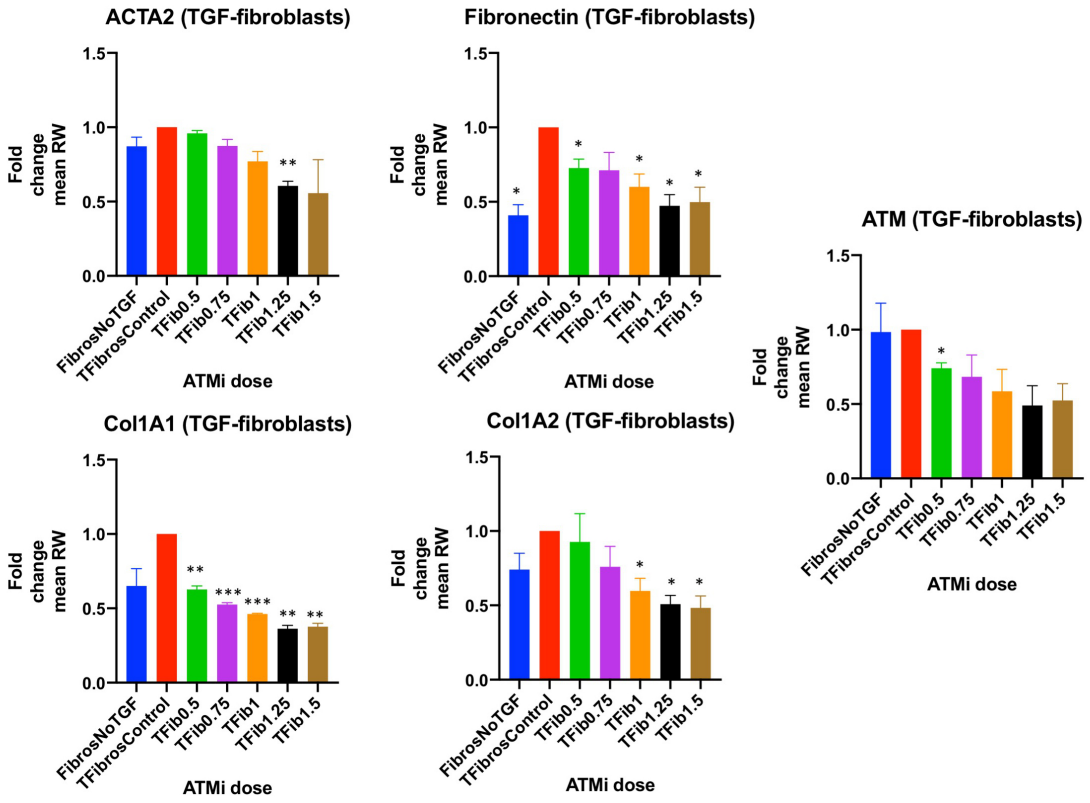



**Figure 7-9. qPCR of Balb/c breast fibroblasts of Balb-neu-T CAFs treated with TGF-β and carrying doses of ATM-inhibitor (AZD0156).**

**A. normal Balb/c breast fibroblasts, B. Balb-neu-T CAFs and C. normal Balb/c breast fibroblasts pretreated with TGF-β, after 3 days of treatment with varying doses of ATM-inhibitor (AZD0156), in addition to control and a no-TGF-β control for C. In each graph, control was set at 1.00 fold change mean RW and mean with standard deviation is shown (experiment n=2). p-values shown for treatment arms vs. control, unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

To determine whether the phenotypic change in cells was found at different doses of ATM inhibitor, qtPCR and Western Blot experiments were repeated at varying doses. qtPCR of normal fibroblasts showed that the ATM-inhibitor reduced expression of collagen1A1 and collagen1A2, more significantly for the higher doses of 1.25uM and 1.5uM (Figure 7.9A). The results were similar for CAFs but ACTA2 was also reduced (non-significant for higher doses)(Figure 7.9B). For TGF-β activated fibroblasts, the expression of all myofibroblastic genes was reduced by the ATM-inhibitor (not significant at some

doses)(Figure 7.9C). For ATM, expression was only reduced by the drug when the TGF-β activated fibroblasts were treated, and only significantly at one dose.

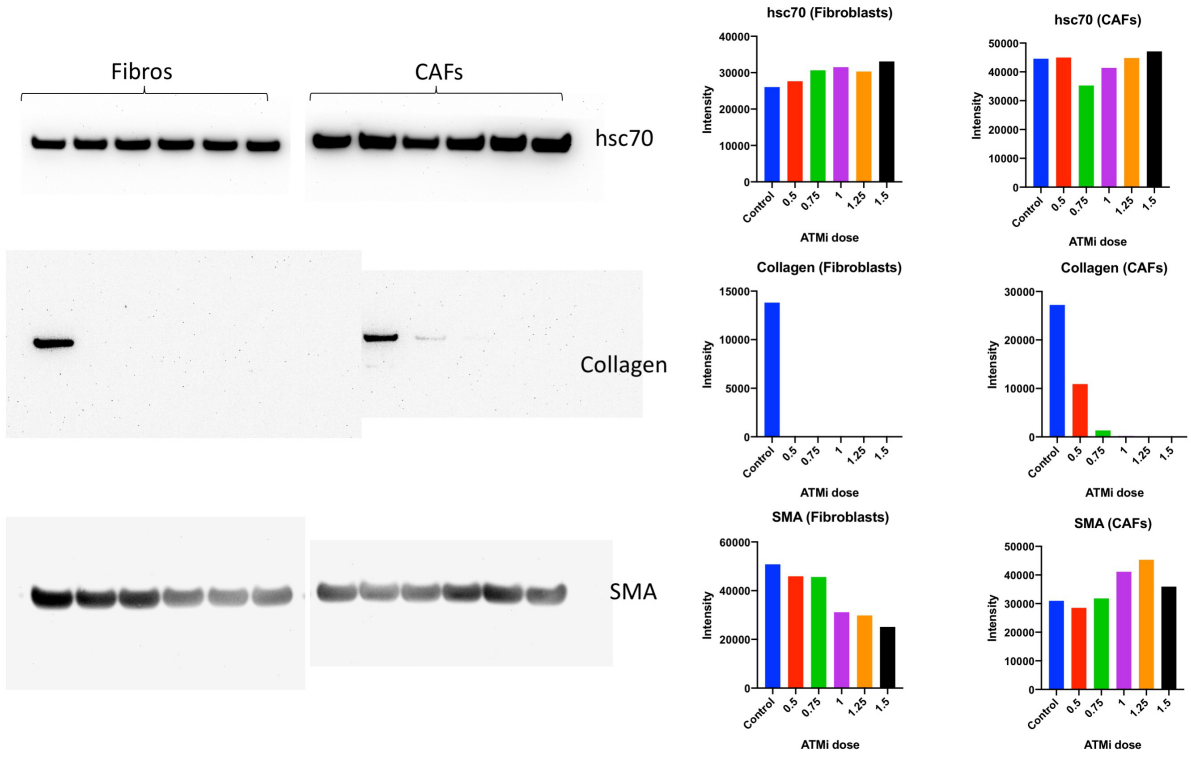



**Figure 7-10. Western blot, using normal Balb/c breast fibroblasts or Balb-Neu-T CAFs, treated with 3 days of ATM-inhibitor (AZD0156) at varying doses, 0.5 to 1.5uM.**

**Graphs show intensity calculated using area under curve in ImageJ.**

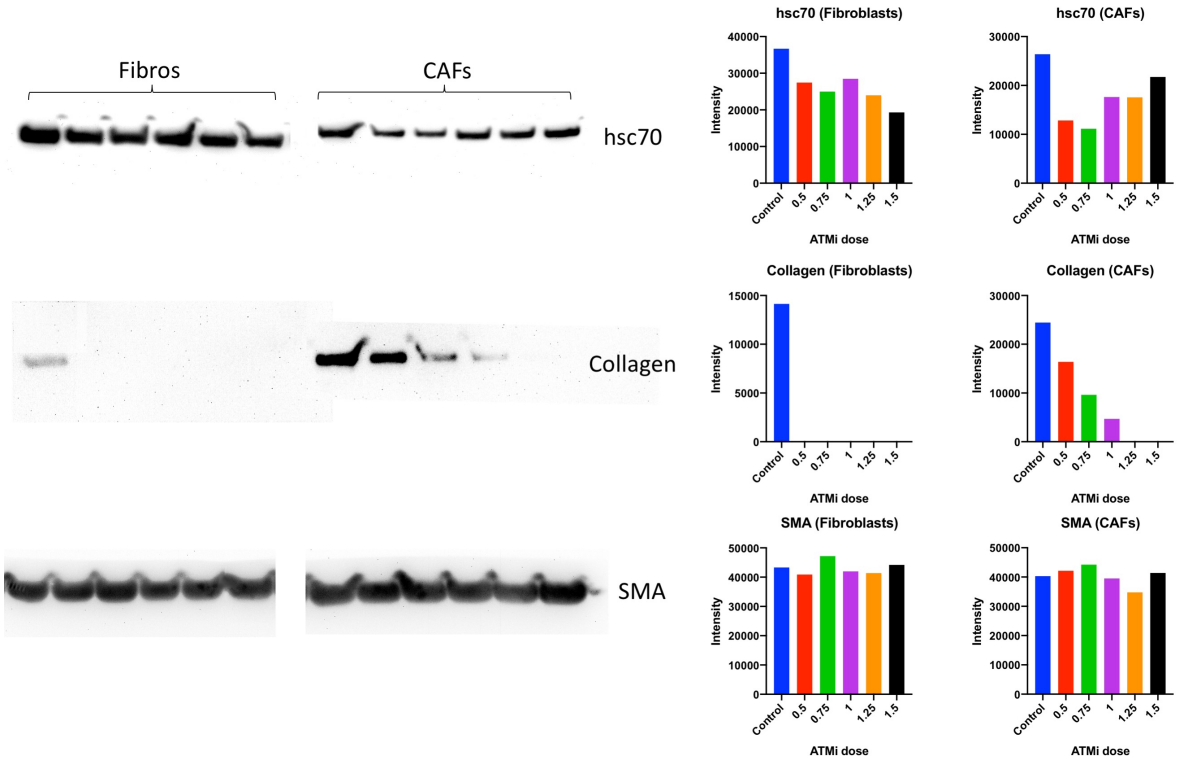



**Figure 7-11. Western blot, using normal Balb/c breast fibroblasts or Balb-Neu-T CAFs, treated with 3 days of ATM-inhibitor (AZD0156) at varying doses, 0.5 to 1.5uM.**

**Graphs show intensity calculated using area under curve in ImageJ.**

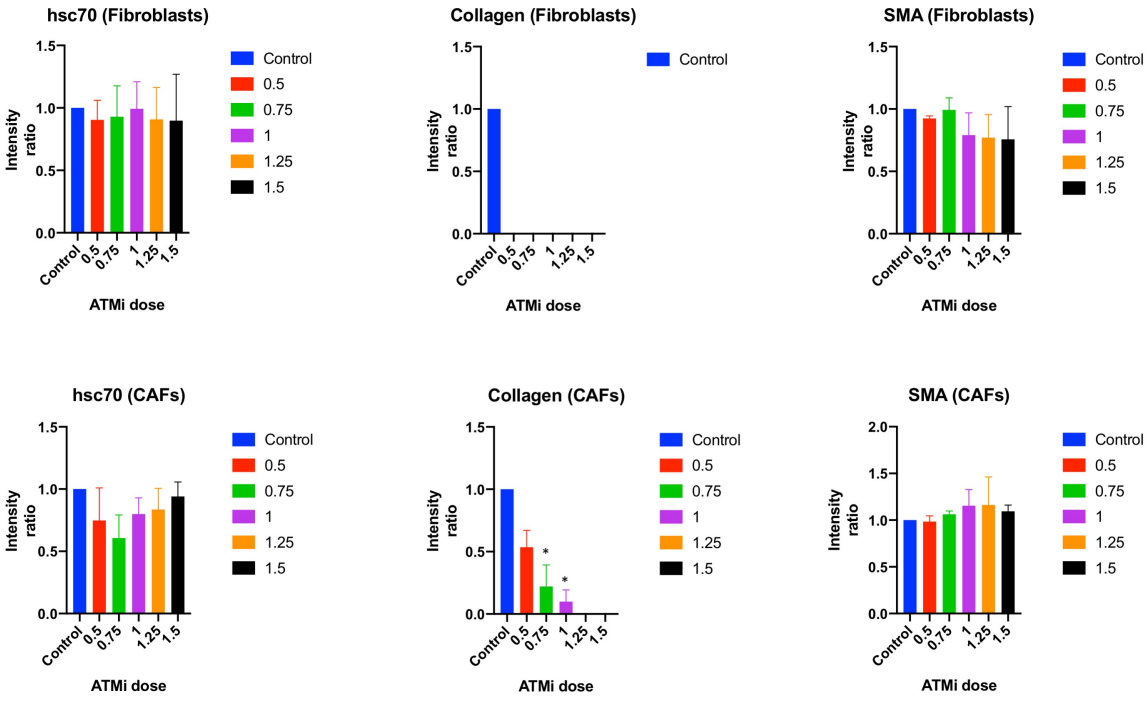


**Figure 7-12. Combined results of 2 western blot experiments, using normal Balb/c breast fibroblasts or Balb-Neu-T CAFs, treated with 3 days of ATM-inhibitor (AZD0156) at varying doses, 0.5 to 1.5uM.**

**Graphs show intensity calculated using area under curve in ImageJ. Mean with standard deviation shown.**

**p-values shown for treatment arms vs. control, unpaired two-tailed t test: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

Western blot using normal fibroblasts and varying doses of ATM-inhibitor showed collagen was not detected in any of the treated arms (Figure 7.10, 7.11 and 7.12). For CAFs, the ATM-inhibitor successfully inhibited collagen, with an increasing degree correlated to dose. SMA was not inhibited using either cell type.

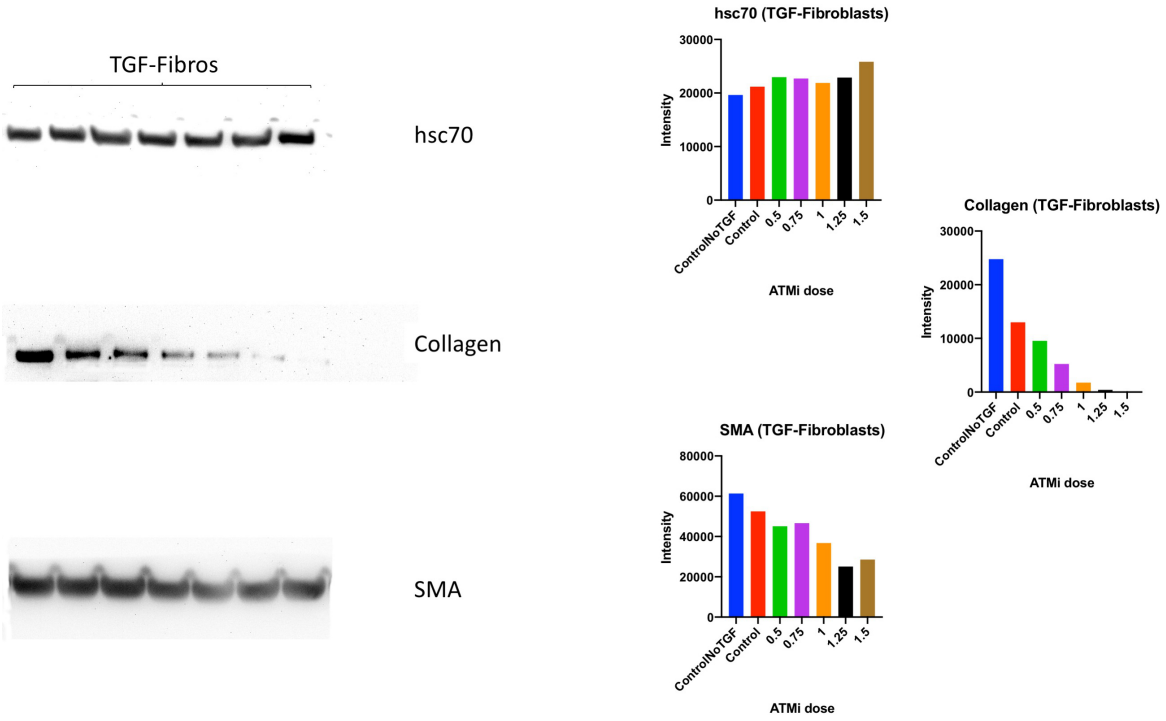



**Figure 7-13. Western blot, using normal Balb/c breast fibroblasts activated with TGF-β, then treated with 3 days of ATM-inhibitor (AZD0156) at varying doses, 0.5 to 1.5uM.**

**Graphs show intensity calculated using area under curve in ImageJ.**

**No p-values shown as experiment n=1.**

For TGF-β activated fibroblasts treated with varying doses of the ATM-inhibitor, the drug successfully inhibited collagen to an increasing degree with higher doses (Figure 7.13). SMA was similarly inhibited under the same conditions.

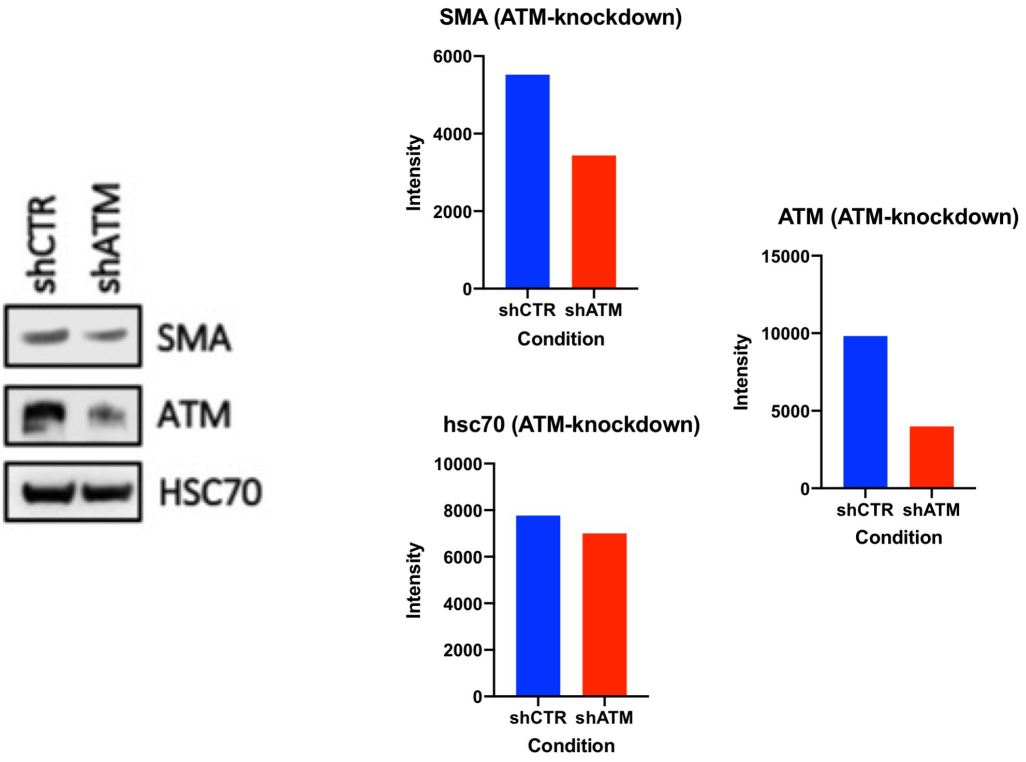



**Figure 7-14. Western blot, using Balb-neu-T CAFS +/- ATM-knockdown (siATM).**

**No p-values shown as experiment n=1.**

To confirm that ATM targeting suppresses the myofibroblast phenotype, gene knockdown of ATM by siRNA (see Methods 2.4.15), with a negative silencer used for the control cells, in breast CAFs was performed. Protein expression by Western Blot was then compared to assess for change in the myofibroblastic phenotype. Cells treated with ATM-knockdown showed a reduction in both SMA and ATM expression compared to control (Figure 7.14), suggesting that ATM is an effective target for reversing the myofibroblastic phenotype.

### 7.2.3 Pilot study of ATM inhibitor *in vivo*

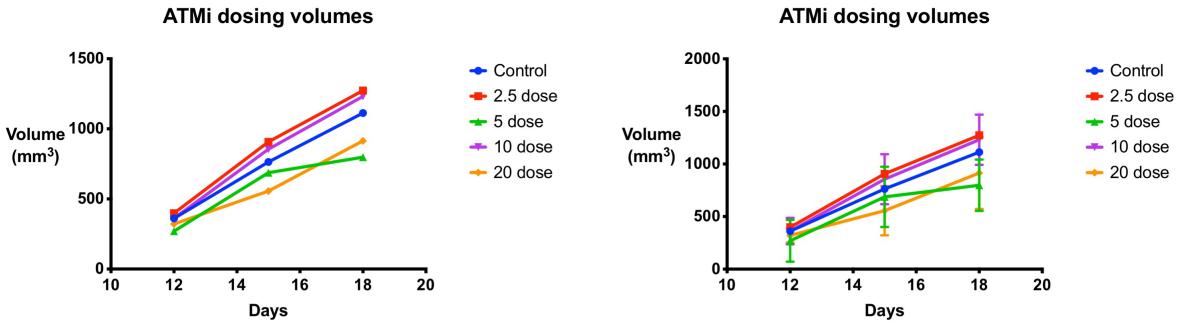
In the previous sections, I have shown that pharmacological inhibition of ATM can inhibit TGF-β induced myofibroblastic trans-differentiation *in vitro.* In order to proceed with an assessment of the drug's efficacy *in vivo*, I needed to assess the toxicity of AZD0156 in mice. I therefore performed a pilot study using a small number of mice with 4T1/CAF tumours given different doses of the drug.

Each mouse was injected on Day 0 with 25,000 4T1 cells and 150,000 Balb-neu-T breast CAFs. On day 6 treatment commenced with daily oral gavage with the ATM inhibitor AZD0156 (Astrazeneca, Cambridge, UK), resuspended in 10% DMSO and 90% captisol (Glentham, Corsham, UK). Four mice were treated with vehicle, two with 20mg/kg, two with 10mg/kg, two with 5mg/kg and two with 2.5mg/kg (all given at a volume of 100µl

Tumour sizes were measured, volumes calculated and the tumours were resected at the end of the experiment for immunohistochemistry. The lungs were also dissected to assess for metastases. All mice were culled on day 18.

The lungs did not show any metastases. There was no observed toxicity from the drug, although it did not appear to affect tumour size (Figure 7.15). However, the drug was commenced on Day 6 (two days after the tumours became palpable) and some doses were omitted for pragmatic reasons. In addition the 4T1:CAF ratio was 6:1 rather than the optimal ratio of 10:1. Immunotherapy with CTLA4 has been used successfully in the 4T1 murine model in the adjuvant setting, with concomitant infiltration of CD8+ T-cells into the tumours(303). However, the 4T1 model is not considered to be particularly immunogenic. It is hypothesised that anti-fibrotic therapies could be used in combination with immunotherapy in tumours where immunotherapy is not effective alone. Therefore it was decided to repeat the experiment using the largest (recommended) dose of 20mg/kg and compare the results to treatment with immunotherapy with CTLA-4

antibody. The intention was to compare anti-tumour efficacy and CD8 infiltration between these arms and the control arm, with a view to a combination therapeutic experiment in the future.




**Area under curve**

| | Control | 2.5 dose | 5 dose | 10 dose | 20 dose |
|---|---|---|---|---|---|
| Total Area | 4564 | 5228 | 3664 | 4956 | 3521 |
| Std. Error | 1143 | 478.3 | 770.6 | 915.1 | 721.7 |
| 95% CI | 2325 to 6804 | 4291 to 6166 | 2154 to 5175 | 3162 to 6749 | 2106 to 4935 |

**Figure 7-15. Tumour volumes for Balb/c mice injected with 4T1 cells and Balb-neu-T CAFs, ratio 6:1, then treated with varying doses of ATM-inhibitor (AZD0156).**

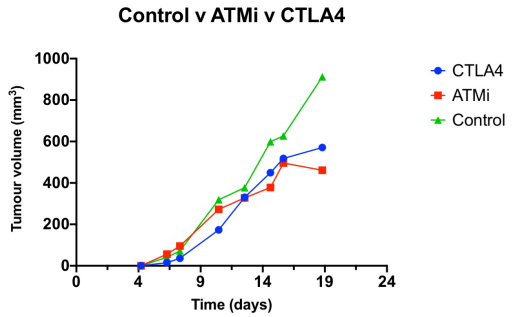
**Mean tumour volume and mean tumour volume with SD shown in graphs (n=4 in control group, n=2 in test groups). Area under curves shown in Table. No significant difference using two-way repeated measure ANOVA over time.**

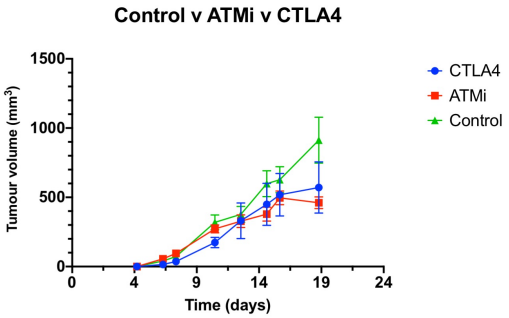
### 7.2.4 Therapeutic experiment with ATM inhibitor *in vivo*

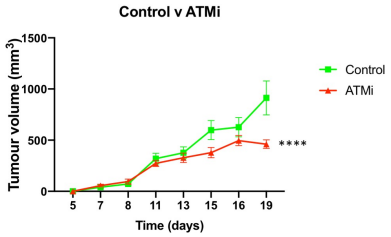
For this experiment, each mouse was injected on Day 0 with 15,000 4T1 cells and 150,000 Balb-neu-T breast CAFs. On day 4 treatment commenced with oral gavage of the ATM inhibitor AZD0156 (Astrazeneca, Cambridge, UK), resuspended in 10% DMSO and 90% captisol (Glentham, Corsham, UK). Eight mice were treated with oral vehicle, eight with 20mg/kg AZD0156 and eight with oral vehicle and intravenous anti-CTLA4 (Bio X Cell,

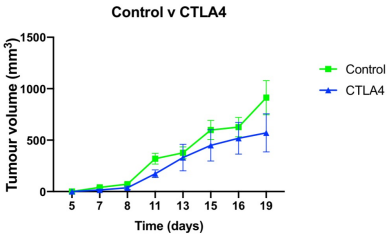
Lebanon, New Hampshire, USA) at a dose of 200ug (20mg/kg) given every 48 hours for 3 doses, commencing on day 4.

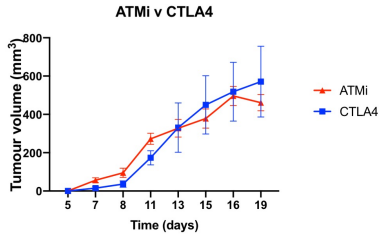
Tumour sizes were measured, volumes calculated and the tumours were resected at the end of the experiment for immunohistochemistry. All mice were culled on day 19.

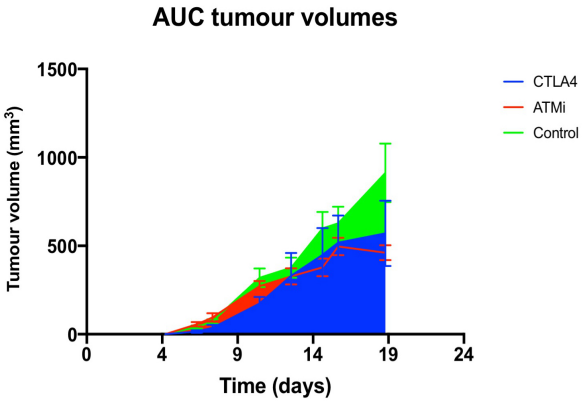



**Area under curve**

| | CTLA4 | ATMi | Control |
|---|---|---|---|
| Total Area | 156.6 | 160.9 | 220.0 |
| Std. Error | 53.11 | 17.40 | 39.20 |
| 95% CI | 52.51 to 260.7 | 126.8 to 195.1 | 143.1 to 296.8 |

**Figure 7-16. Balb/c mice injected with 4T1 cells and Balb-neu-T CAFs, ratio 10:1, then treated with vehicle only (Control), CTLA-4 antibody or ATM-inhibitor (AZD0156).**

**Mean tumour size with standard deviation shown in graphs (n=8 per group). Area under curves shown in Table.**

**p-values shown for treatment arms vs. control, difference using two-way repeated measure ANOVA over time.: *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance).**

With a 10:1 ratio and ATM inhibitor treatment starting from Day 4, the drug was effective at reducing tumour growth (Figure 7.16). The CTLA-4 antibody did not significantly inhibit tumour growth.

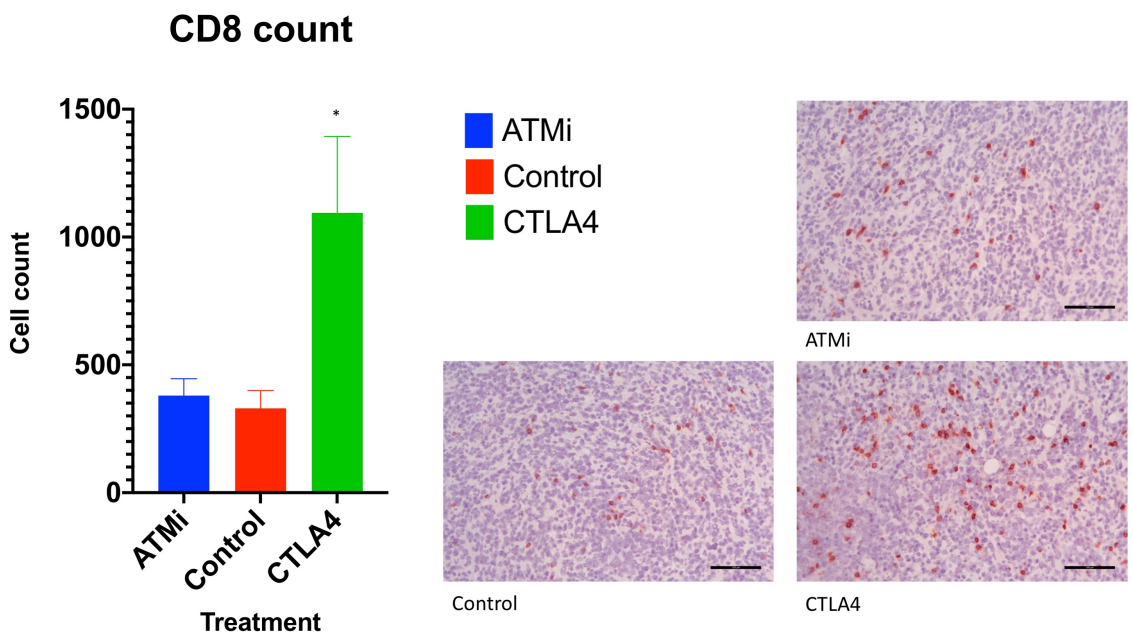



**Figure 7-17. Balb/c mice injected with 4T1 cells and Balb-neu-T CAFs, ratio 10:1, immunohistochemistry performed with CD8 antibody.**

**N= 4 mice per group, 3 slides per case. Mean with standard error of the mean shown. CD8 count calculated per slide in ImageJ after thresholding.**

**Two-tailed unpaired t-test performed. p-values *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance)**

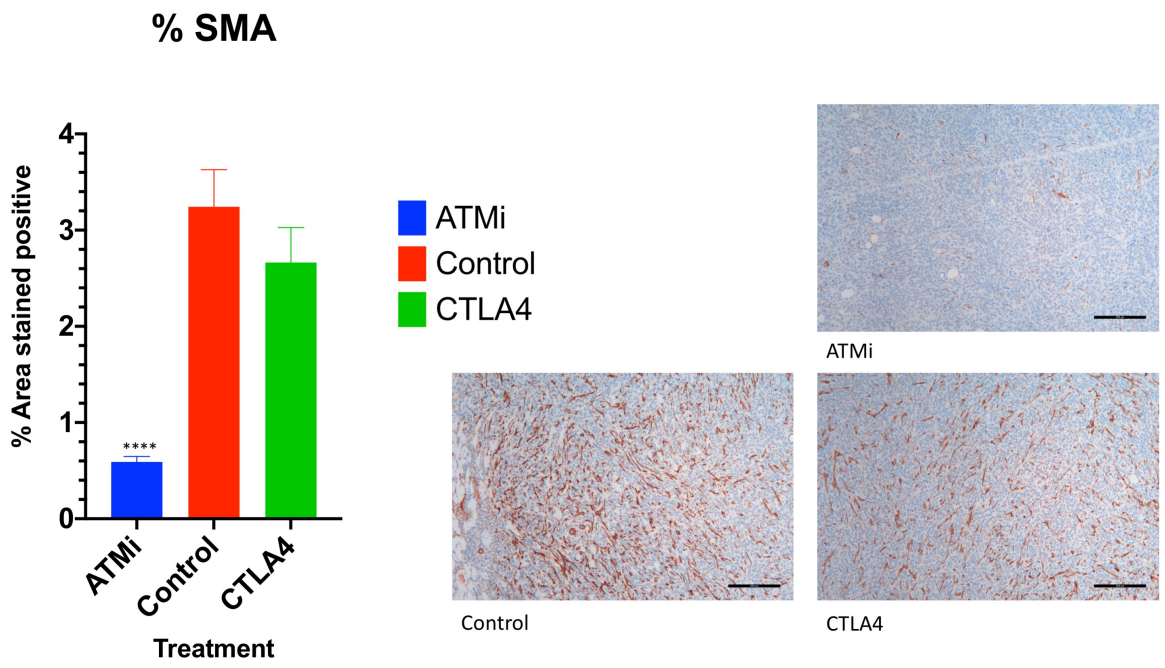



**Figure 7-18. Balb/c mice injected with 4T1 cells and Balb-neu-T CAFs, ratio 10:1, immunohistochemistry performed with SMA antibody.**

**N= 4 mice per group, 3 slides per case. Mean with standard error of the mean shown. SMA percentage of total area calculated per slide in Image-J after thresholding.**

**Two-tailed unpaired t-test performed. p-values *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance)**

To assess whether ATM inhibition or immunotherapy altered CAF infiltration, SMA staining was performed (Figure 7.18). The ATM-inhibitor significantly reduced SMA expression within the tumour, indicating a reduction in CAF density. The CTLA-4 antibody did not affect SMA expression.

Given the role tumour-associated macrophages (TAMs) play in orchestration of the immune microenvironment, I wanted to assess whether there was any change in macrophage infiltration found with the treatments administered in this experiment by staining for F4/80.

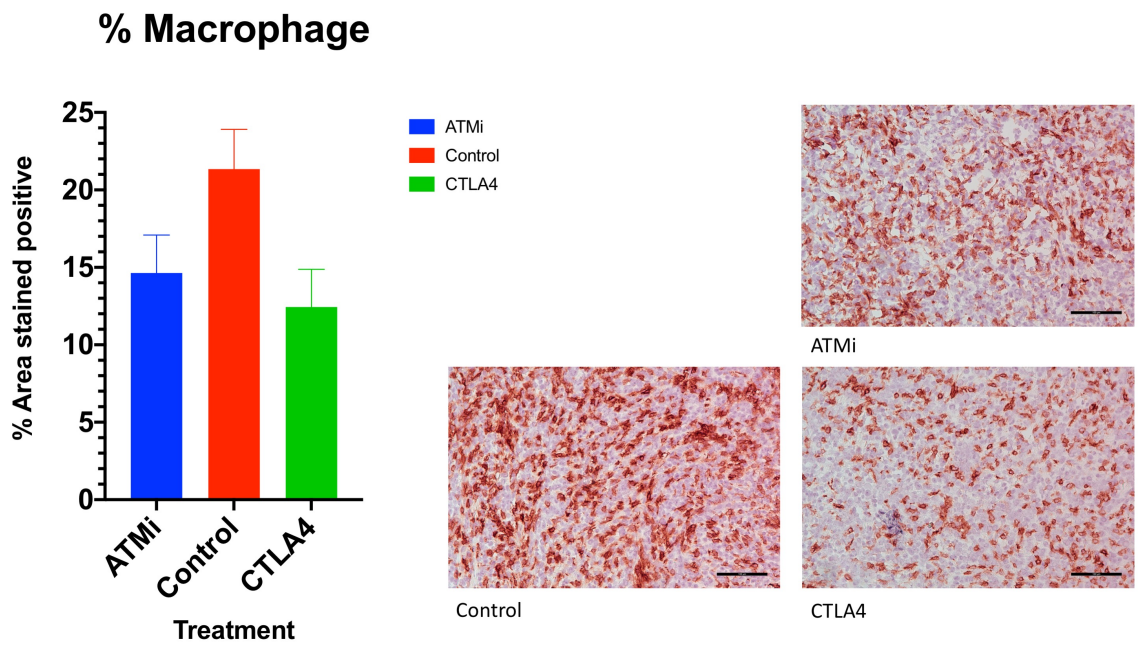



**Figure 7-19. Balb/c mice injected with 4T1 cells and Balb-neu-T CAFs, ratio 10:1, immunohistochemistry performed with F4/80 antibody.**

**N= 4 mice per group, 3 slides per case. Mean with standard error of the mean shown. F4/80 percentage of total area calculated per slide in Image-J after thresholding.**

**Two-tailed unpaired t-test performed. p-values *=<0.05, **=<0.01 ***=<0.001, ****=<0.0001 (no asterisk indicates no significance)**

There was no significant difference in macrophage expression between any of the three experiment arms (Figure 7.19). However, comparing CTLA4 vs. control, the p-value was near significant at 0.0655 for a reduction in macrophages in the CTLA4 arm.

## 7.3 Discussion

### 7.3.1 ATM inhibitor efficacy

In this chapter, I have investigated one method of CAF targeting, using the ATM inhibitor AZD0156 *in vitro* and in a mouse model of TNBC. *In vitro* the ATM inhibitor was found to exert some toxicity to fibroblasts, but was not dose-dependent. In an initial pilot study to evaluate *in vivo* toxicity in the 4T1 model, using the lower CAF:4T1 ratio of 6:1, there did not appear to be any effect on tumour growth. The recommended 20mg/kg dose did not appear to have any adverse toxic effects. Therefore, I went on to perform an experiment using the higher 10:1 ratio and comparing ATM inhibition with checkpoint inhibitor immunotherapy and control. ATM inhibition significantly reduced tumour growth compared to control. SMA expression was also significantly reduced compared to control in the ATMi-treated mice. SMA is generally considered to be a marker of CAFs, rather than normal fibroblasts, but it is unclear whether the ATMi is exhibiting a toxic effect on all fibroblasts, i.e. it is unclear whether its effect are CAF-specific. Unfortunately, the experiment was terminated at day 19 due to the tumours of the control mice exceeding ethical size limits, but there was a trend to inhibition of tumour growth in the CTLA-4 arm and this may have become significant with time. I would like to repeat this experiment using a smaller number of tumour cells in the hope that the experiment length could be increased to observe a more differential effect.

### 7.3.2 Mechanisms of anti-tumour effect by ATM inhibitor

CD8 expression was not increased in conjunction with the reduction in myofibroblastic activity found in ATMi-treated tumours. Therefore, although the ATMi was effective in reducing the myofibroblastic burden and anti-tumour effect, this did not appear to be mediated by influx of CD8 cells. It follows that, in this model, CAFs must alter the

microenvironment to support tumour progression via CD8-independent mechanisms, e.g. the production of cytokines and growth factors (e.g. by recruiting TAMs) or neoangiogenesis. The ATMi could also be working directly on the tumour cells or other non-CAF mechanisms; ATM has been shown to mediate PD-L1 expression, macrophage activation and release of cytokines (e.g. IL-8)(190, 304). Its regulation of IL-8 has been proposed as one mechanism by which ATM may promote cell migration and invasion(190).

Expression of the macrophage marker was not statistically lower in the treatment arms, compared to the control arm but was near-significant for CTLA4. As explained in the previous chapter, M2-TAMs are pro-tumorigenic. Both tumour cells and CAFs can help recruit and proliferate M2-TAMs via TGF-β secretion, in addition to other cytokines. TGF-β has also been shown to induce PD-L1 overexpression – the effect of TGF-β inhibitors in combination with anti-PD-L1 therapy is being explored in early-phase clinical trials(300). Therefore, the mechanism by which macrophages are reduced in the experimental arms here may be via a reduction of TGF-β in the TME as a result of CAF targeting.

It should be noted that in a previous study using a FAP vaccine to deplete CAFs, monotherapy was ineffective(180). The anti-tumour effect (and CD8 influx) was only seen when the vaccine was combined with chemotherapy. It might be that the tumour cell death from chemotherapy is required in combination with anti-CAF therapy in order to produce an effective anti-tumour response. In this way, the tumour microenvironment is more optimized to allow the effects of chemotherapy and subsequent immune control.

### 7.3.3 ATM in cancer

Ataxia telangiectasia is a rare autosomal recessive disorder, resulting from pathogenic mutations in the *ATM* gene. Women who carry a heterozygous *ATM* gene mutation have

an increased risk of breast cancer(305). It is therefore considered to be a breast cancer susceptibility gene. Clearly there is a paradox that mutations in genes involved in DDR can both promote and counteract neoplastic formation. DDR mutations can lead to the accumulation of oncogenic mutations and cell survival; however targeting the DDR pathway can induce sensitivity to DNA-damaging therapeutic agents. ATM inhibition may be effective in cancer treatment because cancer cells would be expected to have higher levels of replicative stress compared to normal cells and many have already acquired a DNA repair pathway defect early in neoplastic transformation (thereby leaving them reliant on ATM if it is intact)(306). Therefore targeting ATM could be potentially lethal to cancer cells while sparing normal cells. In contrast, having a globally defective ATM from birth results in neoplasm development following a somatic “second hit”. Here I have used an ATM inhibition to target CAFs; this is a novel use for a drug that has been developed with the aim of inhibiting cancer cells.

### 7.3.4 Checkpoint inhibition

Here we have shown that although there was a trend towards reduced tumour growth, there was no significant difference in tumour size between CTLA-4 and control groups. The 4T1 mouse model is thought to be poorly immunogenic and highly metastatic(310). Eradication of myeloid-derived suppressor cells and MUC1 vaccination have both been used successfully to improve the anti-tumour response of immunotherapy in the 4T1 model(310, 311). Tumour microenvironment focused approaches have also been used; one group used enzymatic degradation of hyaluronic acid to improve tumoural uptake of anti-PD-L1 in the 4T1 model, with a concomitant increase in CD8+ T cells(312). The mechanism by which hyaluronic acid promotes tumour growth is suggested to be associated with increased interstitial pressure and vascular collapse/hypoxia(312).

### 7.3.5 Limitations

There are a number of *in vitro* experiments that warrant repeating, e.g. unfortunately, the well containing extract from CAFs treated with ATM-inhibitor treatment in the Western lot experiment (Figure 7.4) was overloaded, making interpretation difficult. The *in vivo* experiments here have only been performed once, which is a major limitation. The lack of repetition of certain experiments was due to reaching the end of my available funding and time within the laboratory. As mentioned above, using a smaller number of tumour cells could be useful to allow extension of the experiment and a greater number of doses of therapy. Using a greater number of mice would allow more in-depth analysis of the tumours, e.g. flow cytometry, analysis of macrophage polarisation. Flow cytometry following tumour disaggregation would also allow an assessment of the spatial location of lymphocytes.

## 7.4 Summary

Here I have shown that ATM inhibition effectively reduces the growth of 4T1 tumours in Balb/c mice, with no change in CD8 influx but a reduction in macrophages. Given the promising results for ATM inhibition shown here, I would repeat the therapeutic experiment in order to verify the results and I would like to repeat the experiment with a reduced number of tumour cells. I would like to stain the murine tumours for CD31 to assess vasculature and I would also consider flow cytometry analysis. Assuming a positive result was found, I would proceed to an *in vivo* experiment combining the ATMi with chemotherapy, in order to determine whether a more effective response can be obtained with combination therapy. I would hope that by reducing the CAF burden, the immune response elicited by chemotherapy could be rendered more effective.

# Chapter 8 **Summary and future directions**

This work was designed to explore mechanisms that might underlie alteration in immune infiltrate in a subtype of breast cancer, in the hope that potential therapeutic avenues might be unearthed. Triple negative breast cancer (TNBC) is a particularly attractive target for immune modulation therapy as there are no existing targeted therapies and it is often characterised by an aggressive disease course. The POSH study was used throughout this work; a unique, large prospective cohort of women diagnosed with breast cancer prior to the age of 40, with a wealth of clinicopathological and survival data, in addition to genomic DNA and tumour tissue.

As part of my initial work exploring outcomes in young women with breast cancer, I undertook a detailed assessment of clinical factors associated with prognosis in the POSH cohort. The differences in clinical pattern and baseline features of *de novo* versus metastatic breast cancer remain unclear and this cohort offers an excellent opportunity to explore this. This data has now been published (Appendix D)(227). Here I showed that TNBC was a negative prognostic factor compared to non-TNBC, for both patients who presented with *de novo* metastatic breast cancer and those who developed metastatic disease later after an initial diagnosis.

I proceeded to examine prognostic markers in the cohort with early-stage TNBC. Amongst baseline clinicopathological features, a BMI classification of overweight was an independent negative prognostic feature for survival. The other prognostic factors were as anticipated: T-stage, N-stage, NPI score, LVI. Data showing a worse outcome for obese patients in the POSH cohort has been published previously, but was not evaluated in the TNBC cohort specifically(200). Further work to investigate the interaction between adiposity, inflammation and cancer is ongoing(237, 313). The data set available within POSH could be explored to investigate differences in the tumour microenvironment comparing tumours stratified by patient BMI. The 4T1 model could be used in mice fed

obesogenic diets vs. those on a normal diet, to explore changes at the tumour level in mice with high adiposity.

Host factors have not been adequately explored as a modulating factor in the immune response to cancer. In order to interrogate possible genetic factors influencing immunogenicity in patients with TNBC, I am undertaking the first GWAS study correlating TIL infiltrate with SNPs (Appendix C). As part of this work, I set up a study entitled "i-c-TNBC". Its purpose was the obtainment of fresh blood from patients' with TNBC (unselected age), for genetic analysis (Appendix A). However, a secondary purpose was to provide a cohort for validation work for the analyses performed here. The hypothesis for the GWAS study was that germline variation in immune pathway genes might alter immune signaling in the tumour microenvironment and subsequently affect lymphocyte activation and invasion. Pilot data has flagged up some possible SNPs of interest, for example in the *MERTK* gene. Whether there is an inherited predisposition to lymphocytic infiltrate is an interesting hypothesis that has not been investigated in this way before. As we have now gained funding to complete genotyping on a further 160 samples, report on a completed GWAS will follow and may generate further candidates for functional work. This will constitute the first GWAS assessment of SNP correlation with TILs in TNBC.

I then assessed a panel of molecular markers to assess for any correlation with survival in TNBC, using tissue microarrays. Only stromal TILs were found to be prognostic for overall survival. There is established evidence that TIL infiltrate is correlated with survival in TNBC, in both the neoadjuvant and adjuvant settings. However, there are only a small number of young patients included in previous work (88/647 in one of the largest studies)(45). I have validated this finding in a large cohort of patients aged less than 40 with TNBC naïve to chemotherapy treatment. Multivariable Cox regression analysis confirmed TILs to be an independent prognostic factor in this cohort. Using a ROC analysis with 3-year survival as the primary outcome, TIL percentage was second only to the number of pathological lymph nodes in terms of prognostic significance. Combining TIL score and the number of lymph nodes into a predictive score had a positive predictive value of +0.7603.

In order to determine potential features that may be associated with immunogenicity, I assessed for clinicopathological and molecular correlates of TILs. The TIL score was higher in patients with a germline *BRCA1* or *2* gene mutation. This has been shown previously, but in smaller numbers(122). Possible mechanisms are genomic instability and increased neoantigen load. The latter has been demonstrated in *BRCA*-mutated ovarian carcinoma using TCGA analysis(314). As discussed in section 1.5.4.4, the immunogenicity is variable even within *BRCA*-mutated cases, this is likely to be secondary to factors related to genomic instability(126). This supports efforts to utilize immunotherapy in patients with a *BRCA* mutation, but further stratification and/or combination therapy is likely to be required. Of note, there was no association between BMI and TIL score.

In order to further explore how the genetic characteristics of tumours associated with *BRCA* mutations might correlate with clinical features and/or prognosis, I have formed an international collaboration to facilitate sequencing of paired genomic (whole blood) DNA and tumour (FFPE) DNA from patients from the POSH cohort. This required local ethics and research approval (Appendix E). The team I have been collaborating with at the University of Pennsylvania, USA are international leaders in the field of BRCA1/2 genetics and genomics, with their recent work focusing on the somatic characterization of tumors associated with *BRCA1/2* germline mutations(126). To date, tumour and genomic DNA samples from 201 patients from the POSH cohort have been sent to Pennsylvania. Once this whole exome sequencing data is generated it will be a rich resource for future work.

Using full-face multiplex imaging, digital analysis was performed to analyse the spatial location of CD8 and SMA positive cells. Although none of the results were significant, likely due to the relatively few cases analysed (n=18), there was a trend for an increased number of CD8-SMA- cells and lower number of CD8+ cells inside the tumour borders for TIL-low tumours (relative to outside). It was hypothesised that the former are tumour cells or immunosuppressive cells; further multiplex staining would need to be performed to categorise these cells further. A greater number of cases will need to be included in

further multiplex analysis to determine whether these initial results are significant. It is also possible that further digital analysis using other parameters such as CD8 to tumour cell measurements could be utilised to provide further spatial data. Digital automated analysis of TMA expression was also found to be an effective alternative to pathologist scoring in this setting and could prove an efficient means of processing TMA data in future work.

I also utilised a bioinformatics approach to investigate correlates with lymphocytic infiltration in a publically available dataset. The approach of integrating genomic/transcriptonomic data with histopathological features has been used effectively before(218) and it is a rational approach to further understand the mechanisms underlying variable immunogenicity in TNBC. Interrogation of the TCGA database revealed a positive correlation between extracellular matrix (ECM) modules and lymphocyte depletion in TNBC. This highlighted the potential importance of cancer-associated fibroblasts (CAFs), as the primary ECM cell type, in TNBC biology. I therefore included smooth muscle actin (SMA; a marker of CAFs) amongst an immune panel of molecular markers.

All lymphocyte markers, in addition to MHC class 1 and PD-L1 (on lymphocytes) were positively prognostic for survival. In addition, PD-L1 was positively correlated with TILs. SMA significantly negatively correlated with TILs. Given this validated the TCGA analysis findings, I decided to investigate the possibility of targeting cancer-associated fibroblasts (CAFs) in TNBC.

I first evaluated the effect of TGF-β stimulation of murine normal breast fibroblasts *in vitro* and confirmed differentiation into the myofibroblastic phenotype. An initial *in vivo* experiment comparing fibroblast phenotypes in a co-injection model with 4T1 tumour cells showed a reduction in CD8 cells present with CAFs/TGF-β-activated fibroblasts but no change in tumour volume. As SMA was noted to be high in this model, the fibroblast:tumour ratio was increased. With this change, there was a significant increase

in tumour volume noted with CAFs/TGF-β-activated fibroblasts, with concomitant increase in SMA staining. There was a trend to increased macrophage infiltration in these two arms. Therefore, this shows that myofibroblasts augment tumour growth in this mouse model.

I wanted to assess the efficacy of a novel anti-CAF drug using the above model. There are no published studies evaluating the use of an ATM inhibitor to target myofibroblasts in cancer, but there is evidence of the importance of ATM in TGF-β-mediated fibrosis (see Section 7.1). First I assessed its effect on cells *in vitro*. There was some evidence of a reversal of the myofibroblastic phenotype for both CAFs/TGF-β-activated fibroblasts when treated with the ATM inhibitor, AZD0156. Gene knockdown of ATM confirmed reduced expression of the SMA protein. I proceeded to perform a therapeutic experiment utilising co-injection of CAFs with 4T1 tumour cells, comparing control to an immunotherapy arm and an ATM inhibitor arm. ATM inhibition was effective at reducing tumour growth *in vivo*. CD8 infiltration on the other hand was only increased in the immunotherapy arm, indicating that ATM inhibitor effect was independent of cytotoxic lymphocytes. However, SMA was markedly reduced in the ATM inhibited tumours, indicating that its effect was at least partly mediated via inhibition of CAFs. Of note there was a trend to a reduction in tumour-associated macrophages in both treatment arms. This therapeutic experiment bears repeating with additional work to assess for the effect on tumours; e.g. PD-L1 staining, flow cytometry with macrophage polarisation.

This work was designed to investigate the immunogenicity of TNBC tumours from a cohort of young women. It is generally accepted that an important stimulus for the immunogenicity of a tumour is the repertoire of mutated peptides present. However, it is increasingly being recognised that the microenvironment is of central importance, with the balance of stimulatory and suppressive signalling key in tipping the immunoediting equilibrium in favour of the tumour. Here I have shown that CAF targeting, via ATM inhibition, has potential as a therapeutic strategy in TNBC. Modulation of the microenvironment is likely to become a more common mechanism of anti-cancer agents. There is no doubt that combination therapy with novel agents and immunotherapy will

be necessary for some patients. As our understanding of the heterogeneity of tumours and their microenvironments grows, we will be able to select appropriate strategies to find the best combinations for our patients.

# Appendix A i-c-TNBC study protocol

## i-c-TNBC: immune recognition of triple negative breast cancer

Short title: i-c-TNBC

Protocol v2

**Chief Investigator:**

Dr Hayley McKenzie

CRUK Research Fellow

Cancer Research UK Centre

Somers Cancer Research Building, MP824

Southampton General Hospital

Tremona Road

Southampton

SO16 6YD

**Chief Investigator:**

Dr Ellen Copson

CRUK Associate Professor of Medical Oncology

Cancer Research UK Centre

Somers Cancer Research Building, MP824

Southampton General Hospital

Tremona Road

Southampton

SO16 6YD

**Principal Investigator:**

Dr Hayley McKenzie

Cancer Research UK Centre

Somers Cancer Research Building, MP824

Southampton General Hospital

Tremona Road

Southampton

SO16 6YD

**Co-investigators:**

Professor Anthony Williams

Professor Diana Eccles

Dr Edd James

**Study Sponsor:**

University Hospital Southampton NHS Foundation Trust

Tremona Road, Southampton

SO16 6YD

Telephone: 023 8120 8215

Email: R&Doffice@uhs.nhs.uk.

**Study Site:**

University Hospital Southampton NHS Foundation Trust

Tremona Road, Southampton

SO16 6YD

Tel: 023 8077 7222 (Switchboard)

1. Background

**An immune subtype of TNBC**

TNBC by definition is characterised by lack of expression of oestrogen, progesterone and HER2 receptors using immunohistochemistry and comprises 10-20% of breast cancers.

TNBC tends to be high grade and there is a greater chance of relapse within 5 years compared to other subtypes(315). There are no targeted treatments.

It is now accepted that increased lymphocytic infiltration at diagnosis is associated with decreased disease recurrence rates and improved overall survival in triple negative breast cancer(36, 43). It is also correlated with increased expression of immune and inflammatory markers(37, 78). Additionally, increased stromal TIL infiltration has been consistently associated with a complete response to neoadjuvant chemotherapy in TNBC(37, 316). It has therefore been suggested that TILs can be used as a biomarker for chemotherapy response and prognosis in TNBC. Some have suggested treatment should be stratified according to TIL status(29).

In every study using transcriptonomic and/or genomic data to subclassify breast cancer, an "immune-active" subtype has been identified, which confers a good prognosis. The mechanisms by which this subtype gains recognition of the immune system and resultant TIL infiltration are not clear.

**MHC Class I presentation**

Generation of an effective immune response is a cascade, which commences with antigen presentation. Tumour cells present epitopes generated from tumour-associated antigens (TAA) using MHC class I molecules, which are then recognised by cytotoxic T cells. Assuming immune recognition, a durable and effective T cell response then needs to be maintained. Self-antigens are usually ignored by T cells, whereas non-self and mutated proteins should generate an adaptive immune response. Some TAA may be tumour-specific neoantigens but most reported thus far are normal self-antigens (317). A limited repertoire of immunogenic antigens may be one reason why some tumours are not infiltrated by lymphocytes.

The generation of optimal peptides for class I presentation takes place in the cytoplasm, endoplastic reticulum (ER) and Golgi apparatus and requires a multitude of different molecules and proteins, knows as MHC class I antigen processing and presentation machinery (APM). Any defect in this pathway could affect the formation of optimum peptide fragments and subsequent immune recognition(318). Antigen presentation commences with proteolysis in the cytoplasm. Proteins, such as those generated by cancer cells, are cleaved into smaller fragments by the ubiquitin-proteasome pathway. A small number of peptides are then transported into the ER by the transporter associated with antigen processing (TAP).

ER aminopeptidase (ERAP1) performs a critical function in the final processing of MHC class ligands at this stage. On arrival in the ER, peptides are of varying lengths. Proteasomes can generate peptide fragments ranging from 2-25 residues long(138), although TAP preferentially transports those of 8-16 residues long(138). Class I molecules are very restrictive regarding the length of peptides that can fit into their groove, preferring those of 9 to 10 residues. Peptides therefore require further processing by N-terminal trimming in order to be of optimal length. ERAP1 appears to perform a unique function, acting as a “molecular ruler”, preferentially trimming peptides of 9-16 residues and generating optimal lengths for class I presentation(138). It can also destroy peptides by excess processing(140). In this way, the enzyme is well placed to influence the antigenic repertoire presented to T cells.

The *ERAP1* gene is encoded in chromosome 5q21 and is composed of 20 exons and 19 introns(319). *ERAP1* is highly polymorphic and unique allotypes continue to emerge; in one study 13 allotypes were reported and more than 70 polymorphisms have been documented (320). It is codominantly inherited and so a “normal” trimming allele may be able to compensate for a “low-trimming” allele, producing a normal phenotype. In mouse models, *ERAAP* (the murine aminopeptidase equivalent) has been shown to be vital for the generation of many antigenic epitopes and to influence the repertoire of peptides

produced(142).

**The role of ERAP1 in cancer**

Following on from evidence from genome-wide association studies, *ERAP1* has been most extensively studied in ankylosing spondylitis (AS). The functional intersection of ERAP1 and HLA-B27 underlies its importance in this disease. Variation in *ERAP1* is however likely to have impact on a wider set of diseases and its role in cancer immune surveillance is expanding.

In a murine colon carcinoma model, downregulation of ERAAP (the mouse equivalent of ERAP1/2) resulted in an increased number of antigenic epitopes and led to a greater cytotoxic response and tumour rejection(142). The conclusion drawn was that ERAAP normally overtrims TAA, such as GSW11, in this colon carcinoma model. They were also able to transiently downregulate ERAAP in this model using a small molecule inhibitor, paving the way for ERAP1 as a potential target for anticancer immunomodulation therapy.

Given the importance of cytotoxic and NK cell responses in the immune response to tumour cells, it has been hypothesised that naturally occurring polymorphic variants of *ERAP1* may have an influence on an individual's predisposition to certain cancers(139). This host effect may have most influence at the early stages of cellular transformation, with a certain *ERAP1* variant preventing engagement of the immune system and allowing further oncogenic mutations to occur.

The association of *ERAP1* haplotypes with certain cancers has been studied to a limited degree. One group identified specific ERAP1 polymorphisms as being associated with cervical carcinoma in a geographic subgroup in Indonesia (R-P at 127 and Q-E

at 730)(321). In a separate study, they investigated the association of these polymorphisms with outcome in cervical carcinoma and found variants at two loci that were significantly associated with decreased overall survival(322). They also correlated certain haplotypes with lymph node metastases. Given that SNPs have been found to have an effect on the expression of the ERAP1 protein, trimming activity and substrate specificity(139), it was hypothesised that the altered processing of the Human Papilloma Virus (such as protein E7) by certain ERAP1 haplotypes could lead to a change in the persistence of infection and the ability of transformed cells to evade the immune system. It raises the possibility that genetic variation in antigen processing could also change the genetic predisposition to non-viral induced cancer.

Given that TIL infiltration can be used as a surrogate marker of the immune response, investigating the impact of *ERAP1* variation first in the cancers for which this has a prognostic impact seems reasonable. In humans, so far efforts have focused on cervical carcinoma and head and neck cancer, both HPV-related. This is the first time the *ERAP1* genotype has been evaluated in patients with TNBC.

## 2. Aims and Objectives

This study aims to explore the association between ERAP1 genotype and TIL infiltration pattern in triple negative breast cancer

Primary Objectives:

To collect fresh blood samples from patients with recently diagnosed triple negative breast cancer (TNBC) in order to sequence both alleles of *ERAP1* in peripheral blood mononuclear cells.

To assess the degree of TIL infiltration; moderate/high or low) in TNBC tumours and investigate the association between certain single nucleotide polymorphisms

(SNPs) and degree of TIL infiltration will be investigated.

Secondary Objective:

To perform functional assays to determine the trimming function of these alleles and determine whether there is any association between the phenotype and TIL density.

## 3. Study design

This study is a non-interventional, prospective, observational cohort study intending to evaluate the influence of the *ERAP1* gene on lymphocyte infiltration in triple negative breast cancer. Patients will be recruited prospectively after they receive their diagnosis of invasive breast cancer and prior to the start of systemic anti-cancer therapy. Two additional samples of blood will be collected when they undergo venepuncture for routine blood tests prior to surgery or first chemotherapy treatment. The blood samples will be analysed in the Cancer Sciences department of the University of Southampton. Access to patients' histology slides for an assessment of TIL infiltration will be required; this will be performed on-site at the University of Southampton NHS Foundation Trust. Data regarding patients' diagnosis, treatment and clinical outcome will be recorded and anonymised (see Section 5). Access to excess tissue block material will be requested for future work, specifically related to *ERAP1*.

### 3.1 Study population

Recruitment will not be randomised. A maximum of 100 patients will be recruited over a period of three years.

### 3.2 Study plan

Upon recruitment, patients' will have two additional blood samples taken. This will be taken at the time of venesection for the purposes of pre-surgical assessment or prior to chemotherapy. Therefore no additional venepuncture will be necessary. One 10ml EDTA and one 6ml EDTA sample will be taken (a total of 16mls), to be stored as peripheral blood mononuclear cells (PBMCs), whole blood and serum. The samples will be processed in accordance with local protocol and stored on the day of collection in a research freezer. All samples will be anonymised and identified only a by a unique patient identification number. The sample number and date of receipt will also be recorded.

Following surgery for the primary tumour, access to diagnostic histopathological slides will be required for the purposes of TIL assessment. These will be reviewed on-site in the cellular pathology department of University Hospital Southampton under the supervision of a consultant histopathologist.

In addition, signed consent will be sought from study participants to permit future pathology review and ERAP1 related analysis (pending additional funding) of pathology tissue block specimens from breast surgery performed as part of the routine management of early breast cancer in patients participating in this study.

3.3 Patient recruitment

Potential study participants will be identified at the weekly breast multi-disciplinary meeting. Eligible women will be invited to participate in the study during their routine breast clinic appointment by their clinician. The clinician or research nurse will introduce the study and women who express an interest in being involved in the study will then be given a written patient information sheet and permission will be sought for the PI to subsequently contact them by telephone to discuss the study further should this be

necessary.

Patients who wish to participate will be asked to sign a written consent form by the PI, clinician or research nurse when they attend for routine venepuncture at either surgical pre-assessment or chemotherapy clinic.

All patients will be given a minimum of 24 hours to consider this study before they sign a consent form; however, should a patient specifically request to consent and participate in this study in a shorter time frame then this will be permitted for the patient's convenience. Approach to consent will follow local University of Southampton protocol.

3.4 Inclusion criteria

1. Patients diagnosed with triple negative breast cancer, defined as ER-negative (Quick score ≤3) , PR-negative (Quick score ≤3) and HER2-negative (IHC score 0 -1 or IHC 2 and FISH negative), according to local immunohistochemistry.

2. Aged 18 years or older.

3. Able to understand the reasoning for study and provide written informed consent.

3.5 Exclusion criteria

1. Aged less than 18 years.

2. Unwilling or unable to consent to inclusion within the study.

3.6 End of study participation and withdrawal of consent

Active participation will be limited to the single blood test performed following patient consent. Participants will be asked to consent to the research team continuing to track their clinical progress from case note review following the termination of their time in the study. Patients will be informed that they will be able to withdraw their consent to take part in the study at any time. At this time the blood sample held by the University of Southampton will be destroyed according to local protocol.

Consent will be sought for any remaining blood samples to be transferred to the Cancer Sciences tissue bank at the completion of this research study.

3.7. Collection of clinical and histopathological data

Information regarding patients' personal and histopathological characteristics will be recorded on study specific case report forms (CRF), linked to anonymised samples. Data collected will include: age, histological features of the surgical specimen, treatment administered, clinical outcome. All essential documents will be retained for a minimum of 15 years following the end of the study. Analytical data will be stored electronically on password protected data files by the investigators. The Chief Investigator and co-investigators will have access to the data for analyses. Patient confidentiality will be maintained by removing patient identifiable labels other than an assigned study specific number. No information that could identify a study participant will be released in any publication or report.

4. Processing of samples

Blood samples are taken primarily for the purposes of sequencing the *ERAP1* gene. This

gene is not considered high risk, does not currently have implications beyond the scope of this study and is not subject to guidance with regards to any additional treatment, investigations or familial testing. Further processing of blood may also be performed, related to immunological factors in TNBC.

The blood samples will be processed on the day of collection. Coordination between the phlebotomist and Chief Investigator will be required regarding the timing of venesection in order to ensure this occurs promptly. Once processed, samples will be stored at -80 degrees centigrade in a research freezer in the University of Southampton.

Histopathology slides will be reviewed for the purposes of TIL infiltration. Slides will not be removed from site. Should excess tissue blocks be required for further investigation, they may be moved to the University of Southampton for processing. Only material which is excess to clinical use will be used and the use of this material will have no effect on the patients' clinical management.

## 5. Ethics, governance and funding

This study will be performed subject to Research Ethics Committee (REC) and Health Research Authority (HRA) approval; local Research and Development approval will be gained. It will be conducted in accordance with International conference on harmonization of technical requirements for registration of pharmaceuticals for human use (ICH) Good Clinical Practice (GCP) guidelines.

This study is funded by Cancer Research UK, under a fellowship grant held by Hayley McKenzie. The local sponsor is University Hospital Southampton NHS Foundation Trust.

## 6. References

1. Le Du F, Eckhardt BL, Lim B, Litton JK, Moulder S, Meric-Bernstam F, et al. Is the future of personalized therapy in triple-negative breast cancer based on molecular subtype? Oncotarget. 2015;6(15):12890-908.

2. Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, et al. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. Ann Oncol. 2014;25(8):1544-50.

3. Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. J Clin Oncol. 2014;32(27):2959-66.

4. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res. 2008;14(16):5158-65.

5. Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. J Clin Oncol. 2015;33(9):983-91.

6. Miyashita M, Sasano H, Tamaki K, Chan M, Hirakawa H, Suzuki A, et al. Tumor-infiltrating CD8+ and FOXP3+ lymphocytes in triple-negative breast cancer: its correlation with pathological complete response to neoadjuvant chemotherapy. Breast Cancer Res Treat. 2014;148(3):525-34.

7. Savas P, Salgado R, Denkert C, Sotiriou C, Darcy PK, Smyth MJ, et al. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. Nat Rev Clin Oncol. 2016;13(4):228-41.

8. Gross S, Walden P. Immunosuppressive mechanisms in human tumors: why we still cannot cure cancer. Immunol Lett. 2008;116(1):7-14.

9. Leone P, Shin EC, Perosa F, Vacca A, Dammacco F, Racanelli V. MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. J Natl Cancer Inst. 2013;105(16):1172-87.

10. Chang SC, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. Proc Natl Acad Sci U S A. 2005;102(47):17107-12.

11. Reeves E, Edwards CJ, Elliott T, James E. Naturally occurring ERAP1 haplotypes encode functionally distinct alleles with fine substrate specificity. J Immunol. 2013;191(1):35-43.

12. Alvarez-Navarro C, Lopez de Castro JA. ERAP1 structure, function and pathogenetic role in ankylosing spondylitis and other MHC-associated diseases. Mol Immunol. 2014;57(1):12-21.

13. Reeves E, Colebatch-Bourn A, Elliott T, Edwards CJ, James E. Functionally distinct ERAP1 allotype combinations distinguish individuals with Ankylosing Spondylitis.

Proc Natl Acad Sci U S A. 2014;111(49):17594-9.

14. James E, Bailey I, Sugiyarto G, Elliott T. Induction of protective antitumor immunity through attenuation of ERAAP function. J Immunol. 2013;190(11):5839-46.

15. Stratikos E, Stamogiannos A, Zervoudi E, Fruci D. A role for naturally occurring alleles of endoplasmic reticulum aminopeptidases in tumor immunity and cancer predisposition. Front Oncol. 2014;4:363.

16. Mehta AM, Spaans VM, Mahendra NB, Osse EM, Vet JN, Purwoto G, et al. Differences in genetic variation in antigen-processing machinery components and association with cervical carcinoma risk in two Indonesian populations. Immunogenetics. 2015;67(5-6):267-75.

17. Mehta AM, Jordanova ES, Corver WE, van Wezel T, Uh HW, Kenter GG, et al. Single nucleotide polymorphisms in antigen processing machinery component ERAP1 significantly associate with clinical outcome in cervical carcinoma. Genes Chromosomes Cancer. 2009;48(5):410-8.

# Appendix B POSH metastatic study Statistical Analysis Plan

## Statistical Analysis Plan for the POSH study (Prospective Study of Outcomes in Sporadic versus Hereditary breast cancer) – Mo vs M1 paper

*[Please note this statistical analysis plan has been written in the past tense because it will form the basis of a paper]*

The headings used in this document come from the STROBE reporting guideline for observational studies (see http://www.strobe-statement.org/.

**Statistical Analysis Plan Version**

| Issue no | Revision History | Author | Date |
|---|---|---|---|
| 0.1 | First draft | Ellen Copson and Hayley McKenzie | 12th April 2018 |
| 0.2 | Updated following meeting with Ellen Copson, Hayley McKenzie, Tom Maishman and Laura Day | Tom Maishman and Laura Day | 12th July 2018 |

**To be approved and reviewed by:**

| | Name | Date |
|---|---|---|
| **To be reviewed by:** | Diana Eccles, Peter Simmonds and Ellen Copson | |
| **To be approved by:** | Diana Eccles | |

# 1. Introduction

## 1.1. Background / Rationale

Breast cancer is the most common neoplasm in women, with a yearly incidence of over 55,000 in the UK(323). When localized to the breast and lymph nodes it is often curable, but for the 3 to 6% who present with *de novo* metastatic breast cancer (dnMBC) it is not likely (7, 8). Overall, the median survival of those with metastatic breast cancer (MBC) is 2 to 3 years(9), although the range is wide, with some patients with ER+ or HER2+ disease living much longer.

Most studies to date have shown a longer survival time for those presenting with DNMBC compared to those who develop recurrent MBC (11-14). In one study this longer survival time was limited to patients with a metastasis-free interval of greater than 24 months; those with a shorter interval had a significantly worse prognosis(15). The phenotype of breast cancer for those with DNMBC is unclear. Some studies have reported more favourable pathological features(12), while others have reported an aggressive phenotype in those with *de novo* MBC(8). Data regarding clinical presentation has also yielded conflicting results; some reports suggest a similar pattern of metastatic disease(12), while others show a lower prevalence of brain metastases compared to those with recurrent disease(11, 15). Locoregional management in patients with DNMBC is an ongoing debate; results from randomised trials have been conflicting(324). Therefore, the clinicopathological presentation, prognosis and optimal treatment for patients with *de novo* metastatic breast cancer are not yet fully elucidated.

Women aged less than 40 at diagnosis are more likely to have breast cancer with adverse biological features, including higher grade, absence of hormone receptors, lymph node involvement and vascular invasion(196). We plan to characterise the clinical features, pattern of disease progression and survival of young breast cancer patients who present with metastatic disease, compared to those who later develop distant metastases in a large prospective cohort of young patients genotyped for germline BRCA1/ 2 and other high and moderate penetrance breast cancer susceptibility genes.

### 1.2. Objectives

This paper presents the results from analyses carried out on data collected from the POSH study.

To ensure clarity is provided in the objectives, the following definitions are provided below:

**Definitions:**

- dnMBC (de Novo Metastatic Breast Cancer)=M1 stage patients
- early24=M0 stage patients who have experienced a distant relapse within 24 months
- early12=M0 stage patients who have experienced a distant relapse within 12
- months
- early24_to_60=M0 stage patients who have experienced a distant relapse after 24 months but within 60 months
- early60+=M0 stage patients who have experienced a distant relapse after 60 months

The primary objective was to:

- To compare Overall Survival (OS), defined as time from date of diagnosis to death from any cause, of early12 with that of dnMBC patients.

Secondary objectives:

- To compare OS of early24 with that of dnMBC patients.
- To compare Post Distant Relapse Free Survival (PDRS), defined as time from date of distant metastases to death from any cause, early12 with that of dnMBC patients.
- To compare PDRS, defined as time from date of distant metastases to death from any cause, early24 with that of dnMBC patients.
- To describe patient and tumour characteristics (including Body Mass Index [BMI], BRCA status and sites of metastatic disease) and clinicopathological features of primary tumours in M distant groups:
    - dnMBC
    - early12
    - early24
    - early24_to_60
    - early60+
    - All patients

  (with formal comparisons between dnMBC v early12, and also between dnMBC v early24).

- dnMBC patients only: To compare OS in dnMBC patients who had breast conserving surgery (BCS) or mastectomy (or nodal surgery) vs those who had no surgery.
- 
- dnMBC patients only: To compare PDRS in dnMBC patients who had BCS or mastectomy (or nodal surgery) vs those who had no surgery.

- M0 stage patients only i.e. non-dnMBC: To assess the correlation between MFI (metastasis-free interval), defined as time between date of diagnosis and date of first distant relapse, and PDRS in M0 patients who later develop metastatic disease.

## 2. Methods

### a. Study Design

The POSH study is a prospective cohort study. The protocol for the study can be found in the following journal article http://www.biomedcentral.com/1471-2407/7/160.

### b. Setting

The POSH study recruited women from breast centres across England, Scotland, Wales and Northern Island between 1st June 2001 to 31st January 2008.

### c. Participants

The study recruited ≈3000 women aged 40 years or younger at breast cancer diagnosis. The women had to have been diagnosed with breast cancer between January 2000 and January 2008. In addition, 43 women aged 41-50 were included if they had a known BRCA1 or BRCA2 gene mutation and were diagnosed with invasive breast cancer within the study period but were excluded for this analysis. Women were excluded if they had a previous invasive malignancy (with the exception of non-melanomatous skin cancer), were not available for follow up or refused consent to retain diagnostic and follow up data. A total of 2977 women were included in the analysis population.

Clinical follow up data were obtained from the patient medical records by the clinical trials practitioner (CTP) at each recruiting centre. Data forms collecting information at diagnosis, 6 months, 12 months were completed by the CTP usually at 12 months from diagnosis. Annual data collection was continued from the date of definitive diagnosis until death, loss to follow up or until the end of the current phase of the study (June 2016).

d. **Variables**

| **Variable** | **Type of data / categories** | **Amount of missing data** | **Possible reasons for missing data** |
|---|---|---|---|
| **Primary outcome** | | | |
| Time to death (OS) , in years | Survival data<br><br>Date of death from any cause – Date of invasive breast cancer diagnosis | N/A, patients who haven't experienced an event will be censored at the date of their last follow up visit | N/A |
| **Secondary outcomes** | | | |
| Time from first relapse to death from any cause (PDRS) | Survival data<br><br>Date of death (or last follow-up if not died) – Date of distant relapse | N/A, patients who haven't experienced an event will be censored at the date of last follow-up. | N/A |
| **Candidate predictor(s)** | | | |

| Variable | Type of data / categories | Amount of missing data | Possible reasons for missing data |
|---|---|---|---|
| M stage | Multiple binary categories:<br>• deMBC vs. early12<br>• deMBC vs. early24<br><br>Also categorical:<br>• deMBC<br>• early12<br>• early24<br>• early24_to_60<br>• early60+ | 26 records | Consider Missing Completely at Random (MCAR) |
| Surgical Indicator | Binary (BCS or Mastectomy or Nodal surgery only vs. No surgery) | 0 records | N/A |
| **Potential confounders / effect modifiers** | | | |
| 1. Age at diagnosis, in years | Continuous, in years<br><br>In addition, Categorical:<br><br>18 to 25, 26 to 30, 31 to 35, 36 to 40 | 0 records | N/A |
| 2. Duration of follow-up, in years | Survival data (reverse Kaplan-Meier of OS)<br><br>Date of last follow-up – date of diagnosis.<br><br>Patients who die from any cause are censored at the date of last follow-up. | 0 records | N/A |

| Variable | Type of data / categories | Amount of missing data | Possible reasons for missing data |
|---|---|---|---|
| 3. Ethnicity | Categroical (Caucasian/White, Black, Asian, Other, or Missing/unknown) | 40 records | Consider MCAR |
| 4. Body mass index | Categorical (Underweight/Normal, Overweight, Obese) | 116 records | Patients who had not had chemotherapy were less likely to have their weight or height measured. In addition, some patients might have had chemotherapy privately and not had their weight or height recorded. Assume MAR. |
| 5. Family History (patients with at least one first or second degree relative with breast cancer) | Categorical<br>Yes, no, or missing/unknown. | 122 records | MCAR |
| 6. Presentation | Categorical<br>Symptomatic, screen-detected, other, or missing/unknown | 14 records | MCAR |

| Variable | Type of data / categories | Amount of missing data | Possible reasons for missing data |
|---|---|---|---|
| 7. Histological grade (Grade) | Categorical<br>1, 2, 3, or not graded/missing/unknown | 89 records | Consider MAR |
| 8. Histological type | Categorical<br>Ductal, lobular, ductal & lobular, mixed, medullary, metaplastic, other, unclassified, not graded, or missing/unknown | 50 records | Consider MAR |
| 9. Focality of cancer | Categorical<br>Multifocal, localised, or missing/unknown | 297 records | Consider MAR |
| 10. ER status | Categorical<br>Negative, positive, or missing/unknown | 18 records | MCAR |
| 11. HER2 status | Categorical<br>Negative, positive, or missing/unknown | 350 records | Missing when diagnosis predated routine testing. Potential bias towards missing in patients not experiencing disease recurrence. Consider MAR for this analysis. |
| 12. PR status | Categorical<br>Negative, positive, or missing/unknown | 584 records | Consider MAR |

| Variable | Type of data / categories | Amount of missing data | Possible reasons for missing data |
|---|---|---|---|
| 13. Pathological N stage (lymph node status) | Categorical<br>N0, N1 or missing/unknown | 29 records | No axillary surgery, no lymph nodes in resected specimen. Consider MAR |
| 14. Lymphovascular Invasion | Categorical<br>Absent, present, or missing/unknown | 229 records | Consider MAR |
| 15. Number of positive axillary lymph nodes | Continuous (integer) | 76 records | Consider MAR |
| 16. Maximum tumour diameter invasive, in cm | Continuous | 199 records | Consider MAR |
| 17. Maximum tumour diameter overall (including ductal carcinoma in-situ) (pathological), in mm | Continuous | 157 records | Consider MAR |
| 18. Maximum tumour diameter in-situ, in mm | Continuous | 2588 records | Consider MAR |
| 19. Pathological T stage (for patients receiving neo-adjuvant chemotherapy) | Categorical<br>T0, T1, T2, T3, T4, Tis, Tx, or missing/unknown | 119 records | MCAR |

| Variable | Type of data / categories | Amount of missing data | Possible reasons for missing data |
|---|---|---|---|
| 20. Chemotherapy timing | Categorical<br>Adjuvant, neo-adjuvant, palliative, or not applicable | 0 records | N/A |
| 21. Adjuvant trastuzumab | Categorical<br>Yes, no/missing/unknown | 2609 records | MAR |
| 22. Adjuvant Radiotherapy | Categorical<br>Yes, no/missing/unknown | 2380 records | MAR |
| 23. Hormone treatment | Categorical<br>Yes, no/missing/unknown | 1132 records | MAR |
| 24. Year of diagnosis | Categorical<br>2000, 2001… , 2008 | 0 records | N/A |
| 25. Patient status (for PDRS) | Categorical:<br>Alive following distant relapse, censored at last follow-up<br>Dead following distant relapse | 0 records | N/A |
| 26. Censored variable (for PDRS) | Binary (0,1)<br>0=Did not experienced death after distant relapse<br>1=Experienced death after distant relapse | 0 records | N/A |
| 27. Patient status (for OS) | Categorical:<br>Alive, censored at last follow-up<br>Dead | 0 records | N/A |

| Variable | Type of data / categories | Amount of missing data | Possible reasons for missing data |
|---|---|---|---|
| 28. Censored variable (for OS) | Binary (0,1)<br>0=Did not experienced OS event<br>1=Experienced OS event | 0 records | N/A |

**e. Data sources/measurement**

The tumour biopsy, definitive histopathological report, clinical and radiological reports were all submitted to the study. Pathological characteristics of the tumours were taken from the diagnostic and surgical histopathology report, or clinic letters if not available, clinical staging from the clinical and radiological reports or clinical letters where above not available. For patients treated with neoadjuvant chemotherapy radiological tumour size was used as in previous papers.

National death data were obtained for patients in the cohort from the Medical Research Information Service (MRIS).

ER, PR and HER2 data are taken from pathology reports. Scoring systems varied as expected across contributing hospitals. Positive and Negative categories are straightforward however borderline results exist in all three IHC categories and were classified as positive or negative. The borderline category was merged with positive for the purposes of these analyses.

HER2 data: There are concerns regarding the amount of missing HER2 data obtained. This is covered in more detail in the BRCA v1 SAP. Her2 was not routinely accessed or recorded at the start of the study. However if patients diagnosed prior to this date relapsed, then where the primary tumour was reassessed the information was added to the primary tumour data.

This paper presents the results of analyses conducted on follow up data available up until 26 June 2016.

### f. Representativeness

Clinical data for all patients were collected via standard clinical research forms which were completed from the clinical notes by the Clinical Trials Practitioner in each centre, and by copies of original pathology, scan reports and medical notes.

Recruiting bias: Possibly patients presenting with metastases were not invited (no evidence of this). Possibly patients who just had surgery for very small tumours were not recruited since we recruited mainly through oncology clinics.

Survival bias: Patients who were diagnosed with invasive breast carcinoma but died before consenting to the trial were not included in the trial. As a result, this could lead to a healthier sample of the UK population being analysed for the POSH cohort.

### g. Study Size

This is covered in the BMC paper.

### h. Statistical Methods

**Patients excluded from the analyses**

Patients were excluded from this analysis if they were 41 years of age or over at the date of invasive breast cancer diagnosis (43 patients) i.e. a patient born on 01-Jan-1960 would be included if she was diagnosed before 01-Jan-2001 and excluded if she was diagnosed on or after 01-Jan-2001. Patients were also excluded if they were diagnosed outside of the study period (n=1), did not have invasive cancer (n=72) or did not have any follow-up/primary tumour

information available (n=2). As a result, a total of 2977 patients were included in the analysis population.

**Primary Endpoint**

- Overall Survival (OS), where OS is defined as time from date of diagnosis of primary breast cancer to date of death from any cause.

**Secondary Endpoints**

The secondary endpoints are:

- Post Distant Relapse Free Survival (PDRS) defined as time from date of diagnosis of distant metastatic disease to date of death from any cause.

**Statistical analyses**

1. We described the baseline patient and tumour characteristics of the cohort by M categorisation:

   Patient characteristics:

   - Age at diagnosis, in years – median (range, IQR), n (%);
   - Age at diagnosis (18-25, 26-30, 31-35, 36-40) – n(%);
   - Ethnicity (Caucasian/White, Black, Asian, Other, or missing/unknown) – n(%);
   - Obesity (BMI) – median (range, IQR), n(%);
   - Family History (yes, no, or missing/unknown);
   - Presentation (Symptomatic, screen-detected, other, or missing/unknown) – n (%);
   - Genotype (BRCA1+, BRCA2+. TP53+ , other cancer susceptibility gene (CSG) mutation, sporadic)

   Tumour characteristics:

   - Histological grade (1, 2, 3, not graded/missing/unknown) – n (%);
   - Histological type (Ductal, lobular, ductal & lobular, mixed, medullary, metaplastic, other, unclassified, not graded, or missing/unknown) – n (%);
   - Surgical margins (0, >0 and <1, ≥1 and ≤5, >5, or missing/unknown) – n (%);
   - Extensive in situ (EIC) component (EIC Positive, EIC Negative, or missing/unknown) – n (%);
   - Lymphovascular invasion (Absent, Present, or missing/unknown) – n(%);
   - Number of positive axillary lymph nodes (0, 1-3, 4-9, 10+, or missing/unknown) – n(%);
   - Oestrogen receptor (ER) status (negative, positive, or missing/unknown) – n (%);
   - Progesterone receptor (PR) status (negative, positive, or missing/unknown) – n (%);
   - Human Epidermal growth factor receptor 2 (HER2) status (negative, positive or missing/unknown) – n (%).
   - Focality of cancer (multifocal, localised, or missing/unknown) – n (%);

- Pathological T stage (T0, T1, T2, T3, T4, Tis, Tx, or missing/unknown) – n(%);
- Pathological N stage (N0, N1, or missing/unknown) – n (%);
- Maximum diameter invasive tumour, in mm – median (range, IQR), n(%), missing/unknown – n(%).

2. We described the primary oncological treatment of the cohort by treatment type
    - Breast surgery median (definitive type: BCS/ mastectomy– n(%);
    - Axillary clearance/sample details (only if available at the time of analysis) (yes or no/missing/unknown) – n(%).
    - Palliative Chemotherapy – n(%);
    - Palliative trastuzumab (yes or no/missing/unknown) – n(%);
    - Palliative Radiotherapy (yes or no/missing/unknown) – n(%);
    - Hormone treatment (yes or no/missing/unknown) – n(%);

3. We produced Kaplan-Meier survival curves for OS and PDRS for the following groups and provided Hazard Ratios (except for the all patients group):
    - All patients;
    - dnMBCvs. early12
    - dnMBCvs. early24
    - dnMBCvs. early24_to_60
    - dnMBCvs. early60+

4. For deMBC patients only: We produced Kaplan-Meier survival curves for OS and PDRS for the following groups and provided Hazard Ratios:
    - Surery vs. no surgery.

5. For M0 stage patients only: We assessed the correlation between MFI and PDRS.

6. We fitted a multivariable model for OS and PDRS for the following covariates to compare dnMBCvs. early12 and dnMBCvs. early24, with the following covariates fitted:
    - Age at diagnosis, in years (fitted as a continuous covariate);
    - Body Mass Index (fitted as a categorical covariate [Underweight/Healthy, Overweight or Obese]);
    - Histological Grade (fitted as a categorical covariate [1, 2 or 3]);
    - Maximum invasive tumour size, in mm (fitted as a continuous covariate);
    - N stage (fitted as a binary covariate [N0 or N1]);
    - ER status (fitted as a binary covariate [Negative or Positive]);
    - HER2 status (fitted as a binary covariate [Negative or Positive]);
    - Ethnicity (fitted as a categorical covariate [Caucasian, Black or Asian]);
    - Surgery vs. no surgery.

Hazard Ratios

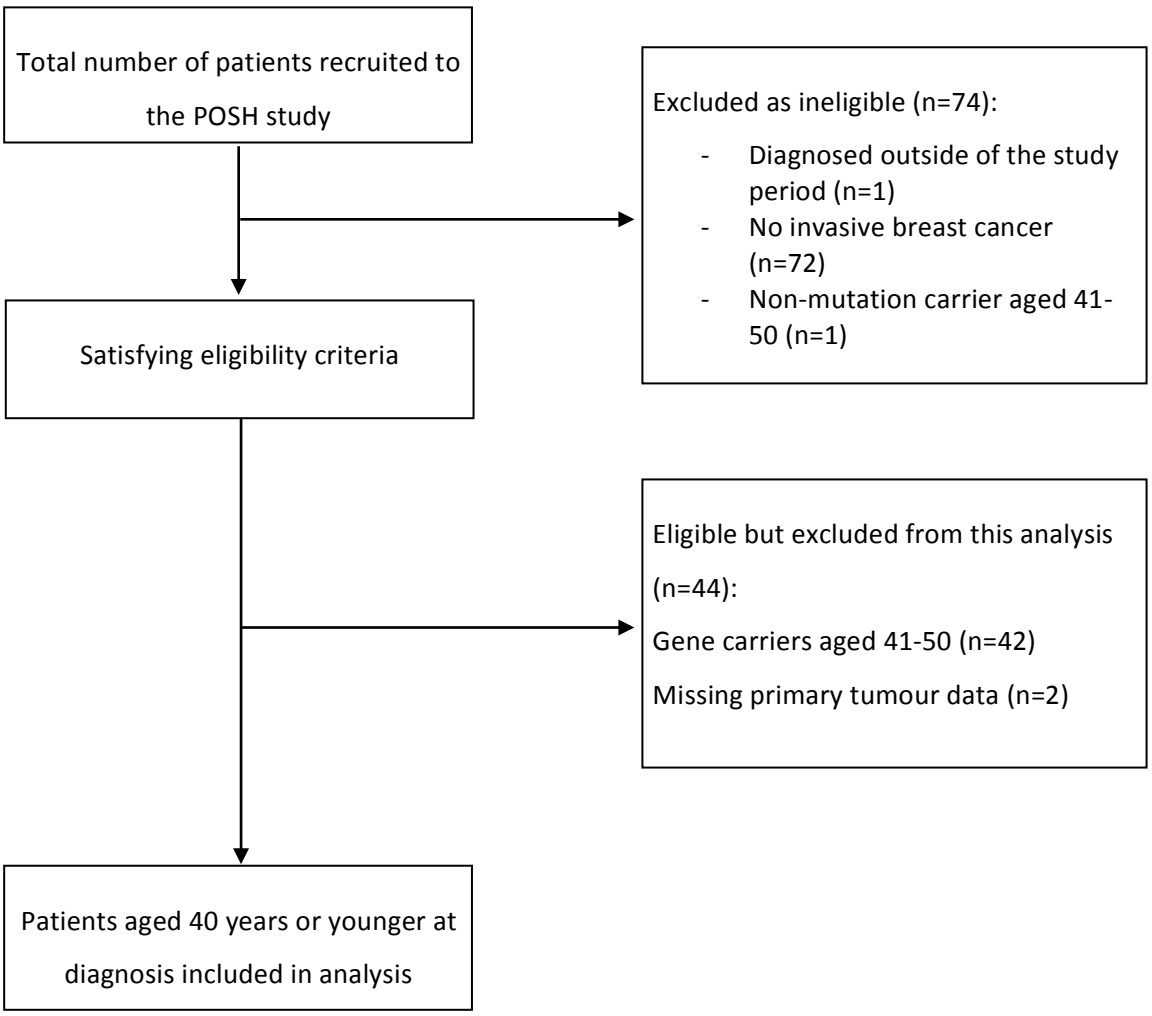
Evidence suggests that the effect of ER status on the HR for relapse changes over time

(Azzato, et al, 2009, Bellera et al, 2010)[1]. Indeed, this was evident after testing the proportional hazards assumption based on the Schoenfeld residuals and using the identity matrix for the time-scaling function[2]. This result provided strong evidence against the Cox proportional hazards assumption, which was also seen when plotting the scaled Schoenfeld residuals over time[2]. We therefore stratified any Cox models by ER status.

**Method used to handle missing data**

This was a complete case analysis.

**Consort Flow diagram**



# Appendix C GWAS Study

### *Introduction/background*

It is now well established that the amount of tumour-infiltrating lymphocytes (TIL) in the stroma or intratumoral compartment plays an important role in the efficacy of chemotherapy for early breast cancer[1-5]. In all breast cancer subtypes, the amount of TILs vary from zero in 10-20% of patients to lymphocyte predominant breast cancer (LPBC) with over 50% lymphocyte infiltrate which occurs in 5-10% of patients[2,6]. A complete pathological response (pCR), whereby residual cancer cells are absent in the resected breast tissue and axillary lymph nodes following neoadjuvant chemotherapy (NAC), is achieved in 40% of patients with LPBC[1,2]. In contrast, only 7% of patients without TILs have a pCR. The effect of TILs is particularly apparent in triple negative (TNBC) and HER2-positive ($HER2^+$) subtypes of breast cancer. The TNBC subtype is characterised by little or no tumour expression for oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2). $HER2^+$ cancers over-express the HER2 cell membrane receptor and may be either positive or negative for hormone receptors (ER and PR). Both TNBC and $HER2^+$ subtypes have a higher incidence of LPBC (20% in TNBC and 16% in $HER2^+$) and more TILs in general than other subtypesl[7]. In TNBC and $HER2^+$ subtypes, even incremental increases in the concentration of TILs are associated with an increase in pCR rate after NAC[1,6,8]. Additionally, high TIL concentrations are associated with improved survival (disease free survival and overall survival in the neoadjuvant and adjuvant setting) in both TNBC and $HER2^+$ subtypes and also in highly proliferative ER positive ($ER^+$) cancers[6,8-10].

The biological mechanisms that contribute to these variable levels of immune infiltration are poorly understood and very few studies have evaluated the effect of germline variation on TIL density. Recently, the association between immune metagene expression, which correlates with histologic TIL counts, and various molecular features (clonal heterogeneity, total mutation load, copy number variations (CNV), somatic mutation, and germline polymorphisms) was investigated in $ER^+$, $HER2^+$ and TNBC subtypes[11]. This study identified several genomic features that are associated with immune gene expression including germline single nucleotide polymorphisms (SNPs) in CFHR1 (Complement Factor H related 1), which inhibits the complement cascade[12], and MBP (Myelin Basic Protein), that can regulate T-helper 2 cells[13], in TNBC.

### *Hypothesis/aims*

The aims of the proposed study are to (i) identify germline SNPs that influence the level of immune infiltration in TNBC (ii) determine whether TIL associated SNPs influence survival and (iii) to explore the functional relevance of associated SNPs and implicated genes. Germline polymorphisms are associated with susceptibility to immune disorders and response to infectious agents[14-16]. Furthermore, germline SNPs in TGFBR2 (transforming growth factor beta receptor 2), which promotes proliferation of regulatory T cells (Treg) and immune evasion, have prognostic associations in patients with ER-negative breast cancer[17]. More recently, germline SNPs have been associated with low TIL concentrations in TNBC[11]. We therefore hypothesise that further germline variation in immunosuppression pathway genes is likely to contribute to variation in TIL concentrations via altered immune signalling in the tumour microenvironment (TME). We plan to explore this hypothesis by performing the first genome-wide association study (GWAS) of TIL concentrations in TNBC.

Identification of these germline factors will improve our understanding of the mechanisms that drive immune exclusion in breast cancer. They may also serve as biomarkers which could be used to predict TIL concentration, estimate response to chemotherapy and select patients who are more likely to respond to immunotherapies.

### *Plan of investigation*

### *Overview and samples*

To identify SNPs associated with the proportion of TILs we will conduct a two stage GWAS consisting of 292 patients at stage 1 and 191 independent patients at stage 2. All of

these patients will have a diagnosis of invasive TNBC. The stage 1 patients will be selected from the Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH)[18]. The POSH patients were recruited from UK hospitals between 2000 and 2007 and were aged forty years or younger at diagnosis. These early onset cases are particularly appropriate for this investigation since high lymphocyte infiltration is associated with young onset ages[19,20]. The stage 2 patients will be selected from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/) which is publicly available without limitations or restrictions.

***Quantification of TIL***

To date we have assessed stromal TILs in 210 patients from the POSH cohort using haematoxylin and eosin-stained sections. The remaining 82 stage 1 patients have been selected and will be scored using the same protocol between now and August 2018. The scoring method follows guidelines published by the International TILs Working Group[21]. In brief, all mononuclear cells in the stromal compartment and within the borders of the invasive tumour were evaluated and scored in 5% increments from 0 to 100%. The pathologist was blinded to clinical information. In the stage 2 samples, TIL concentrations will be inferred based on the expression level of 13 immune metagenes as previously described[11].

***Genotyping, quality control and imputation***

As part of a separate study, 132 of the stage 1 samples have already been genotyped using Illumina 660-Quad arrays[22]. The remaining 160 samples will be genotyped using Illumina Global screening arrays. Prior to the analysis of any genotypic data there will be rigorous quality control procedures. SNPs and samples with ≥10% missing genotypes and SNPs with minor allele frequencies less than 5% will be removed. SNPs with significant deviations from Hardy-Weinberg equilibrium (*P-value* ≤ $1x10^{-10}$) will also be excluded. We will undertake multi-dimensional scaling analyses using PLINK[23] to identify and remove any ancestry outliers when referenced against HapMap populations. Pairwise identity by state scores (IBS) will also be used to test for cryptic relatedness and samples with ≥86% IBS will be removed. Samples which exceed empirical thresholds of genome-wide heterozygosity may indicate contamination and will also be removed. Finally, checks will be performed for plate/batch calling effects, and to test the concordance of duplicate genotypes. To aid merging of the stage 1 samples that will be genotyped using different arrays and to increase the resolution of the stage-1 data, additional SNPs will be imputed using the Sanger imputation server (https://imputation.sanger.ac.uk/). The imputation sever uses EAGLE2 for imputation into the Haplotype Reference Consortium (r1.1) which is the largest reference panel of human haplotypes. The imputed genotypes will be quality controlled by excluding SNPs with a posterior probability less than 0.9, a minor allele frequency less than 5%, greater than 10% missing genotypes, or significant deviations from Hardy-Weinberg equilibrium (*P-value* ≤ $1x10^{-10}$).

***Stage 1 data analysis***

Logistic regression will be used to identify SNPs associated with the percentage of TILs. TILs will be analysed as a quantitative trait because there are no established cut-offs and because each increase in TIL concentration is associated with a further increase in the rate of pCR[1,21]. A Shapiro-Wilk test will be used to query the distribution of TIL scores. In the event of a non-normal distribution, TIL scores will be ranked and normalised using Blom's inverse normal transformation[24]. Significant SNPs will be selected for further stage 2 analyses based on significance alone and a combination of significance and biological relevance with respect to neighbouring genes in immunosuppression pathways.

***Stage 2 data analysis***

At stage 2, significant SNPs from stage 1 will be tested for association in 191 independent patients from TCGA. These patients have been genotyped using Genome-Wide Affymetrix SNP6 arrays. This genotypic data will be analysed using the same quality control and statistical methods as stage 1. A standard fixed effects inverse variance weighted meta-analysis will be used to combine evidence from stages 1 and 2 and to determine the final effect size and significance level.

***Power calculation***

To estimate the statistical strength of the study, R was used assuming a combined sample size of 483 patients, a genome-wide significance level of *P-value* = $5x10^{-8}$ and that the additive effect of an associated SNP explains 8% of the variance in TIL concentrations. According to these assumptions, the study is estimated to have 80% power to detect SNPs associated with the density of TIL infiltration. Although our sample size is relatively small for a GWAS it is sufficient to generate pilot data that will support further funding applications. Our focus on a particular breast cancer subtype (TNBC) will also help to increase power by reducing genetic heterogeneity. Furthermore, other GWAS with similar sample sizes and precisely defined phenotypes have been successful[25,26].

### *Survival analysis and functional inference*

To identify SNPs associated with overall survival (OS) and/or disease-free survival (DFS), we will conduct Cox regression analyses using GenABEL[27] and follow-up data from the POSH cohort. For SNPs associated with both TIL concentration and survival, we will use Mendelian randomisation[28] to examine the causal relationship and determine if the association with survival is due to the SNPs effect on TIL concentration. To explore the functional relevance of the regions associated with TIL concentration, HaploReg[29], RegulomeDB[30] and SeattleSeq[31] will be used to interrogate ENCODE data[32] and annotate the risk SNPs with respect to a range of functional motifs. To gain further functional insight, expression quantitative trait locus (eQTL) analysis will be performed using the Genotype-Tissue Expression (GTEx) portal[33] to query RNAseq data from breast mammary tissue in 183 samples with genotype data.

### *References*

1. Denkert, C. *et al.* Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol.* (2017). doi:10.1016/S1470-2045(17)30904-X

2. Denkert, C. *et al.* Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **28,** 105–113 (2010).

3. Ladoire, S. *et al.* Pathologic complete response to neoadjuvant chemotherapy of breast carcinoma is associated with the disappearance of tumor-infiltrating foxp3+ regulatory T cells. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **14,** 2413–2420 (2008).

4. Ono, M. *et al.* Tumor-infiltrating lymphocytes are correlated with response to neoadjuvant chemotherapy in triple-negative breast cancer. *Breast Cancer Res. Treat.* **132,** 793–805 (2012).

5. Issa-Nummer, Y. *et al.* Prospective validation of immunological infiltrate for prediction of response to neoadjuvant chemotherapy in HER2-negative breast cancer--a substudy of the neoadjuvant GeparQuinto trial. *PloS One* **8,** e79775 (2013).

6. Adams, S. *et al.* Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **32,** 2959–2966 (2014).

7. Stanton, S. E. & Disis, M. L. Clinical significance of tumor-infiltrating lymphocytes in breast cancer. *J. Immunother. Cancer* **4,** 59 (2016).

8. Dieci, M. V. *et al.* Prognostic and predictive value of tumor-infiltrating lymphocytes in two phase III randomized adjuvant breast cancer trials. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **26,** 1698–1704 (2015).

9. Loi, S. *et al.* Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **31,** 860–867 (2013).

10. Loi, S. *et al.* Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **25,** 1544–1550 (2014).

11. Safonov, A. *et al.* Immune Gene Expression Is Associated with Genomic Aberrations in Breast Cancer. *Cancer Res.* **77,** 3317–3324 (2017).

12. Goicoechea de Jorge, E. *et al.* Dimerization of complement factor H-related proteins modulates complement activation in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 4685–4690 (2013).

13. Katsara, M. *et al.* Altered peptide ligands of myelin basic protein ( MBP87-99 ) conjugated to reduced mannan modulate immune responses in mice. *Immunology* **128,** 521–533 (2009).

14. Dhiman, N. *et al.* Associations between SNPs in toll-like receptors and related intracellular signaling molecules and immune responses to measles vaccine: preliminary results. *Vaccine* **26,** 1731–1736 (2008).

15. Schott, E. *et al.* Association of TLR7 single nucleotide polymorphisms with chronic HCV-infection and response to interferon-a-based therapy. *J. Viral Hepat.* **15,** 71–78 (2008).

16. Lee, J. C. *et al.* Human SNP links differential outcomes in inflammatory and infectious disease to a FOXO3-regulated pathway. *Cell* **155,** 57–69 (2013).

17. Lei, J. *et al.* Assessment of variation in immunosuppressive pathway genes reveals TGFBR2 to be associated with prognosis of estrogen receptor-negative breast cancer after chemotherapy. *Breast Cancer Res. BCR* **17,** 18 (2015).

18. Eccles, D. *et al.* Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. *BMC Cancer* **7,** 160 (2007).

19. Ménard, S. *et al.* Lymphoid infiltration as a prognostic variable for early-onset breast carcinomas. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **3,** 817–819 (1997).

20. Thompson, N., Storr, S., Zhang, S. & Martin, S. Lymphovascular invasion: assessment and prognostic impact in melanoma and breast cancer. *Histol. Histopathol.* **30,** 1001–1009 (2015).

21. Salgado, R. *et al.* The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **26,** 259–271 (2015).

22. Kadalayil, L. *et al.* Germline variation in ADAMTSL1 is associated with prognosis following breast cancer treatment in young women. *Nat. Commun.* **8,** 1632 (2017).

23. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

24. Blom, G. *Statistical Estimates and Transformed Beta-Variables*. (Wiley, 1958).

25. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308,** 385–389 (2005).

26. Best, T. *et al.* Variants at 6q21 implicate PRDM1 in the etiology of therapy-induced second malignancies after Hodgkin's lymphoma. *Nat. Med.* **17,** 941–943 (2011).

27. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinforma. Oxf. Engl.* **23,** 1294–1296 (2007).

28. Swerdlow, D. I. *et al.* Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int. J. Epidemiol.* **45,** 1600–1616 (2016).

29. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44,** D877-881 (2016).

30. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22,** 1790–1797 (2012).

31. Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461,** 272–276 (2009).

32. de Souza, N. The ENCODE project. *Nat. Methods* **9,** 1046 (2012).

33. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).

**Strategy for further funding**

***Please provide a clear plan for how data from this research will pump-prime more substantive funding. Including the names of sponsors and collaborators and schemes to be targeted if possible:***

This proposal will generate pilot proof of principal results that will be used to support further funding applications to Breast Cancer Now (project grant deadline Friday 22nd June 2018) and Breast Cancer UK (deadline Autumn, 2018). These applications will seek funding for a much larger and higher statistically powered GWAS into the impact of germline polymorphisms on TIL concentrations in TNBC and additional breast cancer subtypes (including $HER2^+$ and $ER^+$) which are available from the POSH cohort.

The composition of TILs may also play an important role in mediating the response to chemotherapy (Emans et al 2012, Galon et al 2013) and could be influenced by germline SNPs. Approximately 75% of TILs are T cells, the remainder consists of macrophages, natural killer cells (NK) and dendritic cells (DC). These can be subdivided into cell types which restrict tumour growth (such as $CD4^+$ T-helper 1 cells (Th1), $CD8^+$ cytotoxic T-cells (CTL), NK cells, M1 macrophages and DC) and those that promote tumour growth and/or immunosuppression (such as $CD4^+$ T-helper 2 (Th2), $CD4^+$ forkhead box P3 ($FOXP3^+$), $CD4^+$ regulatory T-cells, M2 macrophages and myeloid-derived suppressor cells (MDSC)). This functional heterogeneity of infiltrating lymphocytes may explain why high TIL concentrations are not associated with improved survival in luminal-HER2 negative tumours[1]. In future work we will therefore use immunohistochemistry (IHC) to quantify TIL subtypes (CD4, CD8, FOXP3) and use these measures as covariates when testing for association with TIL.

In collaboration with the Breast Cancer Stratification project (BCAST: www.b-cast.eu) whole transcriptome sequencing is being performed in approximately 400 patients from the POSH cohort. We will use this data to test for associations between gene expression and TIL concentration and to perform eQTL analyses on the genes implicated by our GWAs.

To replicate our findings we will use publicly available datasets from Metabric (https://www.synapse.org/) and TCGA (https://cancergenome.nih.gov/).

# Appendix D British Journal of Cancer publication

McKenzie HS, Maishman T, Simmonds P, Durcan L, Group PS, Eccles D *et al.* Survival and disease characteristics of de novo versus recurrent metastatic breast cancer in a cohort of young patients. Br J Cancer, (2020) doi: 10.1038/s41416-020-0784-z (227).

BJC British Journal of Cancer
www.nature.com/bjc

ARTICLE
Clinical Study

# Survival and disease characteristics of de novo versus recurrent metastatic breast cancer in a cohort of young patients

Hayley S. McKenzie[1], Tom Maishman[2], Peter Simmonds[1], Lorraine Durcan[2] POSH Steering Group, Diana Eccles[1] and Ellen Copson[1]

**BACKGROUND:** It is not clear how the pathology, presentation and outcome for patients who present with de novo metastatic breast cancer (dnMBC) compare with those who later develop distant metastases. DnMBC is uncommon in younger patients. We describe these differences within a cohort of young patients in the United Kingdom.
**METHODS:** Women aged 40 years or younger with a first invasive breast cancer were recruited to the prospective POSH national cohort study. Baseline clinicopathological data were collected, with annual follow-up. Overall survival (OS) and post-distant relapse-free survival (PDRS) were assessed using Kaplan–Meier curves.
**RESULTS:** In total, 862 patients were diagnosed with metastatic disease. DnMBC prevalence was 2.6% (76/2977). Of those with initially localised disease, 27.1% (786/2901) subsequently developed a distant recurrence. Median follow-up was 11.00 years (95% CI 10.79–11.59). Patients who developed metastatic disease within 12 months had worse OS than dnMBC patients (HR 2.64; 1.84–3.77). For PDRS, dnMBC was better than all groups, including those who relapsed after 5 years. Of dnMBC patients, 1.3% had a g*BRCA1*, and 11.8% a g*BRCA2* mutation.
**CONCLUSIONS:** Young women with dnMBC have better PDRS than those who develop relapsed metastatic breast cancer. A gBRCA2 mutation was overrepresented in dnMBC.

*British Journal of Cancer* https://doi.org/10.1038/s41416-020-0784-z

## BACKGROUND

Breast cancer is the most common neoplasm in women, with over 55,000 new diagnoses per year in the United Kingdom.[1] The vast majority of patients present with disease localised to the breast and axillary lymph nodes, and are treated with the aim of cure, but for the 6–7% who present with de novo metastatic disease (dnMBC), treatment is usually with palliative intent.[1,2] Overall, the median survival of those with metastatic breast cancer (MBC) is 2–3 years,[3] although the range is wide, with some patients with ER + or HER2 + disease living much longer. Most MBC survival analyses are retrospective, with a median age of 53–65 (with <15% of participants being aged under 40).[4–6]

A number of studies to date have shown a longer survival time following diagnosis of metastases for those presenting with dnMBC, compared with those who later develop distant metastases after initial treatment for early breast cancer (recurrent MBC, rMBC).[5,7–9] In a retrospective multicentre study evaluating 815 consecutive patients with MBC in the Netherlands from 2007 to 2009, this was only true for rMBC patients with a metastasis-free interval (MFI) of <24 months.[6]

The phenotype of breast cancer for those with dnMBC is unclear. Compared with rMBC cases, more favourable pathological features have been reported, such as a lower frequency of triple-negative carcinomas.[5,7] However, more aggressive features have also been documented, such as larger tumours, and an increased frequency that is Grade 3.[5] However, with median follow-up length of less than 5 years, interpretation of these studies is limited by the omission of late ER + ve recurrences.

Data regarding clinical presentation have also yielded varying results. A higher prevalence of bone involvement in the dnMBC group at diagnosis has been reported in two studies: one reported an equal prevalence of brain metastases, and the other reported fewer brain metastases compared with those with rMBC.[6,7] Another study also found a lower prevalence of brain metastases, but a similar prevalence of bone involvement.[5] Locoregional management in patients with dnMBC is the subject of ongoing debate, as results from retrospective studies have been confounded by selection bias, and the results from randomised trials have been conflicting.[10]

Published studies on dnMBC have been limited by their retrospective nature (with the risk of survival bias), or by small patient numbers and short follow-up periods. None of them has complete germline *BRCA* status, or evaluated a specific age group. The incidence of breast cancer in young women (aged <40) is low, but increasing.[11] Young women are more likely to have breast cancer with adverse biological features, including higher grade, absence of hormone receptors, lymph node involvement and vascular invasion.[12] Young age has been consistently shown to be

[1]University Hospital Southampton NHS Foundation Trust, Southampton General Hospital, Southampton SO16 6YD, UK and [2]Southampton Clinical Trials Unit, Southampton General Hospital, Southampton SO16 6YD, UK
Correspondence: Hayley S. McKenzie (Hayley.mckenzie@uhs.nhs.uk)
Members of the POSH Steering Group are listed above Acknowledgements section.
These authors contributed equally: Diana Eccles, Ellen Copson

Received: 29 August 2019 Revised: 14 February 2020 Accepted: 25 February 2020
Published online: 31 March 2020

an adverse prognostic factor, with a higher risk of distant recurrence.[11] In addition, this group is less likely to have co-morbidities; they may tolerate chemotherapy and other treatments better than older patients. Although young women present more frequently with stage III disease, dnMBC is found infrequently (1% of those aged under 40 in one retrospective study).[13] The POSH study, a prospective observational study of almost 3000 patients aged 40 years or younger with a first diagnosis of invasive breast cancer,[12] provides a unique opportunity to study the natural history of dnMBC in young women. Patients were recruited between 2000 and 2008 in the United Kingdom. A wealth of clinicopathological data is available for these patients, including body mass index and ethnicity, and genotyping for germline *BRCA* mutation status has been performed on the vast majority (> 94%). This is an important variable to study as *BRCA* mutation status is increasingly being incorporated in decision-making regarding optimal treatment.[11] We aimed to characterise the clinical features, pattern of disease progression and survival of young breast cancer patients who present with metastatic disease, compared with those who later develop distant metastases, in a large prospective cohort genotyped for germline *BRCA1/2*.

## METHODS

Prospective outcomes in sporadic vs. hereditary breast cancer (POSH) are a multicentre prospective observational cohort study of young women diagnosed with breast cancer in the United Kingdom. The detailed study protocol was published in 2007.[14] The study received approval from the South West Multi-Centre Research Ethics Committee (MREC 00/6/69). Written informed consent was obtained from all participants.

### Patients

In total, 3021 female patients were recruited from 127 UK hospitals. Patients were eligible if they were diagnosed with an invasive breast cancer between January 1, 2000 and January 31, 2008, at an age of 40 years or younger. Patients were excluded if they had a previous invasive malignancy (excluding non-melanomatous skin cancer), Patients were consented within 12 months of initial diagnosis. All patients received treatment according to local protocols. Patients with confirmed distant metastatic disease at diagnosis (stage M1) according to the local site comprised the dnMBC cohort. Patients who initially had localised disease (stage M0), but developed distant metastatic disease within the follow-up period, (according to site reporting) comprised the rMBC cohort. Tissue diagnosis of metastatic disease was not mandated by the study. Patients without metastatic disease at any time were not included in this analysis.

### Data collection

Information regarding personal characteristics, tumour pathology, stage and treatment received was collected from medical records at study entry. Family history was collected by questionnaire. Pathology and imaging data were verified with copies of original reports. Follow-up data, including date and site of disease recurrence, were obtained from medical records at 6, 12 months and thereafter annually until death or loss to follow-up. Follow-up interval was determined according to local standards; no imaging or other investigation was mandated by this study, as it was observational. Patients were flagged in the National Health Service Medical Research Information Service to facilitate automatic notification of the date and cause of death. This study presents analyses conducted on follow-up data received until 26 June 2016.

### Biological testing

Oestrogen receptor (ER), progesterone receptor (PR) and HER2 receptor status of primary tumours were determined from routine diagnostic pathology reports. Hormone receptor concentrations equivalent to an Allred score of 3 or more were categorised as positive. Tissue microarray (TMA) immunohistochemical staining was used to supplement missing information regarding receptor status.

DNA for genotyping was extracted from whole-blood samples collected at recruitment. A multiplex amplicon-based library preparation system, Fluidigm Access Array (Fluidigm UK, Cambridge, UK) was used to sequence a panel of breast cancer susceptibility genes, including *BRCA1/2* and *TP53*. Illumina HiSeq2500 next-generation sequencing platform was utilised (Illumina, Little Chesterford, UK). If patients met current UK guidelines for genetic testing, multiplex ligation probe analysis was used to ensure that mutations consisting of large exonic deletions or duplications were not missed. Pathogenic variants were confirmed by Sanger sequencing. Those with variants of unknown significance were classified as *BRCA*-negative.

### Statistical methods

Statistical analyses were performed according to a pre-specified statistical analysis plan (Supplementary Information) as per STROBE guidelines.

The primary objective was the comparison of overall survival (OS) of patients with dnMBC with that of patients with rMBC with a MFI of less than 12 months (early12). OS was defined as the time from the date of diagnosis to death from any cause. MFI was defined as the time from the date of diagnosis to the date of the first distant relapse.

The secondary objectives included the comparison of OS and post-distant relapse survival (PDRS) of patients with dnMBC with that of patients with rMBC with a MFI of less than 24 months (early24). PDRS was defined as time from the date of diagnosis of the first distant metastases (date of diagnosis of primary tumour for patients with dnMBC) to death from any cause. Other secondary objectives included the comparison of PDRS of patients with dnMBC vs. early-12 patients, and the description of clinicopathological features in patients with dnMBC and those with rMBC in four cohorts (recurrent disease within 12 months, within 24 months, between 24 and 60 months and after 60 months). Patient and tumour characteristics included ethnicity, body mass index (BMI), germline *BRCA* status, first site of metastasis and primary tumour grading/receptor status. Time-to-event outcomes were described using Kaplan–Meier curves, and analysed using Cox regression models; stratified Cox models or flexible parametric survival models were used in cases where hazards were time-varying. All multivariable analyses were adjusted for age at diagnosis, BMI, grade, tumour size, pathological N stage, ethnicity and ER and HER2 tumour status. Further objectives included the comparison of OS of dnMBC patients who had surgery (breast-conserving surgery, nodal surgery only or mastectomy) vs. those who had no surgery and assessment of correlation between MFI and PDRS in rMBC patients using the survcorr command in R. Statistical analyses were carried out using Stata v15.1 and RStudio v1.1.456.

The study size and power calculations are discussed in the study protocol.[12]

## RESULTS

A total of 3021 eligible women were recruited to the POSH study. For this study, 44 women were excluded (42 were aged 41–50 years and 2 had missing primary tumour data). Of the 2977 women included, 862 (29.0%) were diagnosed with metastatic disease and comprise the analysis population. There were 76 women (2.6%) who presented with dnMBC. As of June 2016, the distant recurrence rate amongst the 2901 women with localised disease at presentation is 27.1% ($n = 786$). Median follow-up of the analysis population was 11.00 years (95% CI 10.79–11.59, $n = 862$).

Of patients with rMBC, 70 (8.9%) developed metastases within 12 months of diagnosis (early 12). There were 268 women (34.1%), who developed metastatic disease within 24 months of their first diagnosis (early 24), 360 (45.8%) within 24–60 months (early 24–60) and 158 (20.1%) after 60 months (late).

Baseline clinicopathological data

For the 862 women diagnosed with metastatic disease, clinicopathological data can be seen in Table 1. The proportion of patients that were very young (aged 30 or less) decreased with time to relapse amongst rMBC patients (18.6% for early 12, 6.3% for late relapse). The largest proportion of *BRCA1* mutation carriers was found in the early-24 group (8.6%; 23/268). The largest proportion of *BRCA2* mutation carriers was found in the dnMBC group (11.8%; 9/76); there was only one *BRCA1* mutation in this group (1.3%; 1/76).

On review of tumour characteristics, ER positivity was positively correlated with later time of relapse: 45.3% of early 24, 72.4% of early 24–60 and 84.2% of late-relapse cases (69.3% of dnMBC). HER2 positivity was the highest in the dnMBC cohort with 47.9% vs. 32.4% in the early-24 group, and 24.3% in the late-relapse group. The median maximum tumour size was the largest in the dnMBC group at 35 mm, compared with 25 mm in the late-relapse group. The proportion of cases with Grade 3 disease was inversely proportional with time to relapse amongst the rMBC group: 87.0% of early 12 and 47.4% of the late-relapse group (63.2% of dnMBC).

The early-12 group had a high proportion of adverse tumour characteristics, including Grade 3 disease (87.0%), LVI (76.9%) and node positivity (83.8%). The number of involved lymph nodes was more than 10 in a quarter of cases (25.0%), and over a third of cases were triple negative (34.3%).

Survival

Patients who relapsed within 12 months had a significantly worse OS than the dnMBC group (Fig. 1a), with a HR for death of 2.64 (1.84–3.77; $p = < 0.001$). For those who relapsed after 24 months, the OS varied over time, consistent with the delay from diagnosis to metastatic disease (clearly the HR for death at 2 years was very small). However, results from the time-varying regression model show that by 5 years, the risk of death for those who relapsed between 24 and 60 months was increased, compared with the dnMBC group, with a 5-year HR of 1.55 (1.10–2.18, $p = 0.013$) and 10-year HR 2.21 (1.02–4.77, $p = 0.044$) (Fig. 1b; Supplementary S1A). Similarly, the time-varying regression model shows that, for those who relapsed after 60 months, the risk of death at 5 years was very small, but compared with the dnMBC group, the 10-year HR was 1.74 (0.80–3.78, $p = 0.160$) (Fig. 1b; Supplementary S1B). There was longer PDRS for dnMBC compared with all other groups who developed metastases, including those with late relapse after 60 months (HR 2.67; 1.92–3.71, $p < 0.001$) (Fig. 2a, b). The hazard ratio for PDRS for early 12, compared with dnMBC, decreased over time (Fig. 2c).

We assessed for a correlation between time from initial diagnosis to metastatic relapse (metastasis-free interval, MFI) and PDRS, and found a very slight positive Rho correlation coefficient of 0.045 (95% CI: −0.023 to 0.113), suggesting that there is not a close correlation between these two factors.

A multivariable analysis was performed to assess for factors related to duration of survival in those with dnMBC vs. early12. For OS (Table 2), early-12 patients maintained a significantly worse OS compared with dnMBC after adjustment for other factors (HR 3.76; 2.22–6.38; $p < 0.001$). Positive nodes were found to be associated with significantly shorter survival (HR 2.29; 1.17–4.47; $p = 0.015$), whilst patients with HER2-positive tumours were at reduced risk of death (HR 0.500; 0.311–0.802; $p = 0.004$). Similar results were also found in the multivariable analyses for PDRS.

For early-24 patients (Table 3), PDRS was worse compared with dnMBC patients after adjustment for other factors at 2 and 5 years (HR 2.53; 1.50–4.27; $p < 0.001$, and HR 2.42; 1.39–4.22; $p = 0.0019$). Again, positive nodes were found to be a significant risk of earlier distant relapse (HR 1.42; 1.05–1.93; $p = 0.024$), whilst patients with HER2-positive tumours had longer survival (HR 0.66; 0.51–0.86; $p = 0.002$). ER-positive status was protective for disease relapse at 2 years compared with ER-negative (HR 0.50; 0.38–0.67; $p < 0.001$), but not at 5 or 10 years.

Sites of metastases

Regarding sites of metastases (at any time during disease course), patients in the dnMBC and early-12 groups were most likely to have widespread (bone, visceral and brain) disease, with 26.3% and 21.4%, respectively, compared with 13.6% in the late-relapse group. Patients with dnMBC had the highest prevalence of brain metastases (39.5%), which decreased with time to relapse (24.3% in the late-relapse group). The proportion with bone metastases correlated with time to relapse amongst those with rMBC: 54.3% of the early 12, up to 71.4% with late relapse (71.1% of the dnMBC group). Visceral metastases were equally prevalent throughout all groups.

When the first site of metastases was evaluated, bone-only or nodal-only disease at presentation was most common in the dnMBC group (30.3% and 15.8%, respectively). Visceral metastases at presentation of metastatic disease were less common in the dnMBC group (52.6%), and most common in the early 24–60 group (66.8%). Bony metastases at presentation were present in 50% of the early-12 group, increasing to 60.4% of the late- relapse group (57% of dnMBC). Brain metastases at presentation decreased with time to relapse amongst the rMBC cohort: 18.6% of early 12 and 8.7% of late relapses (1.3% of dnMBC).

Treatment

Amongst dnMBC patients, 65.8% (50/76) had local surgery. Survival was better in those who had surgery, with a univariable HR of 0.41 (0.24–0.68, $p = <0.001$) and 5-year OS of 44.6% (42.24–46.94) vs. 15.27% (7.85–24.97) (Fig. 3). Patients were treated with palliative cytotoxic chemotherapy in 98.7% of the dnMBC group vs. 71.4% in the early-12 group, and 70.3% in the late-relapse group. Palliative hormone therapy was also the highest in the dnMBC group (68.4%), whereas it was given in 28.6% of patients who relapsed within 12 months and 61.4% of patients with a late relapse. Palliative radiotherapy was also administered at the highest rate in the dnMBC group, with 76.3% of patients receiving it, compared with 60.0% in the early-12 group and 50.0% in the late-relapse group.

## DISCUSSION

This is the largest prospective study to evaluate metastatic disease in the young-onset breast cancer population. We have shown that young women who develop secondary metastatic disease, even if greater than 5 years after diagnosis, have shorter survival time following diagnosis of metastases, compared with those who present with de novo metastatic disease. When survival from initial diagnosis (OS) was compared, this was superior for those with dnMBC, compared with those who developed relapsed disease within 24 months.

In this study, nearly a third (27.1%) of women developed metastatic recurrence after presenting with localised disease. Only 2.6% of this cohort had metastatic disease at presentation, lower than the national (unselected age) estimate of 6–7% from Cancer Research UK.[1] Late stage at diagnosis is reported to be more common in women aged greater than 80, and so this likely contributes to the higher figure nationally.[1,13] It is also possible that there was an element of selection bias, as oncologists may have chosen to recruit patients with metastatic disease to an interventional study rather than an observational one (although participation in an interventional study did not exclude patients

**Table 1.** Demographic table by M-stage categories.

| | M1 (dnMBC) (*n* = 76) | M0 < 12 months (*n* = 70) | M0 < 24 months (*n* = 268) | M0 24–60 months (*n* = 360) | M0 60+ months (*n* = 158) | Total (*n* = 862) |
|---|---|---|---|---|---|---|
| *Age at diagnosis (years)* | | | | | | |
| 18–25 | 4 (5.3%) | 0 | 4 (1.5%) | 6 (1.7%) | 2 (1.3%) | 16 (1.9%) |
| 26–30 | 6 (7.9%) | 13 (18.6%) | 37 (13.8%) | 37 (10.3%) | 8 (5.1%) | 88 (10.2%) |
| 31–35 | 27 (35.5%) | 26 (37.1%) | 95 (35.4%) | 117 (32.5%) | 53 (33.5%) | 292 (33.9%) |
| 36–40 | 39 (51.3%) | 31 (44.3%) | 132 (49.3%) | 200 (55.6%) | 95 (60.1%) | 466 (54.1%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *Ethnicity* | | | | | | |
| Caucasian | 67 (88.2%) | 60 (85.7%) | 236 (89.7%) | 328 (91.6%) | 136 (86.6%) | 767 (89.8%) |
| Black | 6 (7.9%) | 7 (10.0%) | 17 (6.5%) | 18 (5.0%) | 13 (8.3%) | 54 (6.3%) |
| Asian | 3 (3.9%) | 3 (4.3%) | 8 (3.0%) | 10 (2.8%) | 7 (4.5%) | 28 (3.3%) |
| Other | 0 | 0 | 2 (0.8%) | 2 (0.6%) | 1 (0.6%) | 5 (0.6%) |
| Total | 76 (100%) | 70 (100%) | 263 (100%) | 358 (100%) | 157 (100%) | 854 (100%) |
| Missing | 0 | 0 | 5 (1.9%) | 2 (0.6%) | 1 (0.6%) | 8 (0.9%) |
| *BMI, categorical* | | | | | | |
| Underweight | 32 (45.1%) | 30 (44.8%) | 119 (45.4%) | 167 (47.2%) | 80 (51.3%) | 398 (47.2%) |
| Overweight | 21 (29.6%) | 21 (31.3%) | 85 (32.4%) | 105 (29.7%) | 38 (24.4%) | 249 (29.5%) |
| Obese | 18 (25.4%) | 16 (23.9%) | 58 (22.1%) | 82 (23.2%) | 38 (24.4%) | 196 (23.3%) |
| Total | 71 (100%) | 67 (100%) | 262 (100%) | 354 (100%) | 156 (100%) | 843 (100%) |
| Missing | 5 (6.6%) | 3 (4.3%) | 6 (2.2%) | 6 (1.7%) | 2 (1.3%) | 19 (2.2%) |
| *Family history* | | | | | | |
| No | 48 (68.6%) | 45 (67.2%) | 180 (68.7%) | 238 (68.0%) | 106 (68.8%) | 572 (68.4%) |
| Yes | 22 (31.4%) | 22 (32.8%) | 82 (31.3%) | 112 (32.0%) | 48 (31.2%) | 264 (31.6%) |
| Total | 70 (100%) | 67 (100%) | 262 (100%) | 350 (100%) | 154 (100%) | 836 (100%) |
| Missing | 6 (7.9%) | 3 (4.3%) | 6 (2.2%) | 10 (2.8%) | 4 (2.5%) | 26 (3.0%) |
| *Presentation* | | | | | | |
| Symptomatic | 76 (100.0%) | 70 (100.0%) | 264 (99.2%) | 360 (100.0%) | 155 (98.1%) | 855 (99.4%) |
| Screen-detected | 0 | 0 | 2 (0.8%) | 0 | 1 (0.6%) | 3 (0.3%) |
| Other | 0 | 0 | 0 | 0 | 2 (1.3%) | 2 (0.2%) |
| Total | 76 (100%) | 70 (100%) | 266 (100%) | 360 (100%) | 158 (100%) | 860 (100%) |
| Missing | 0 | 0 | 2 (0.7%) | 0 | 0 | 2 (0.2%) |
| *BRCA (BRCA1 or 2) status* | | | | | | |
| BRCA− | 66 (86.8%) | 62 (88.6%) | 237 (88.4%) | 325 (90.3%) | 138 (87.3%) | 766 (88.9%) |
| BRCA+ | 10 (13.2%) | 8 (11.4%) | 31 (11.6%) | 35 (9.7%) | 20 (12.7%) | 96 (11.1%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *BRCA1 status* | | | | | | |
| BRCA1− | 75 (98.7%) | 66 (94.3%) | 245 (91.4%) | 346 (96.1%) | 148 (93.7%) | 814 (94.4%) |
| BRCA1+ | 1 (1.3%) | 4 (5.7%) | 23 (8.6%) | 14 (3.9%) | 10 (6.3%) | 48 (5.6%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *BRCA2 status* | | | | | | |
| BRCA2− | 67 (88.2%) | 66 (94.3%) | 260 (97.0%) | 339 (94.2%) | 148 (93.7%) | 814 (94.4%) |
| BRCA2+ | 9 (11.8%) | 4 (5.7%) | 8 (3.0%) | 21 (5.8%) | 10 (6.3%) | 48 (5.6%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *TP53 status* | | | | | | |
| TP53− | 76 (100.0%) | 69 (98.6%) | 267 (99.6%) | 359 (99.7%) | 158 (100.0%) | 860 (99.8%) |
| TP53+ | 0 | 1 (1.4%) | 1 (0.4%) | 1 (0.3%) | 0 | 2 (0.2%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *Histological grade* | | | | | | |
| 1 | 2 (2.9%) | 0 | 1 (0.4%) | 10 (2.8%) | 6 (3.9%) | 19 (2.3%) |
| 2 | 23 (33.8%) | 9 (13.0%) | 42 (16.0%) | 123 (35.0%) | 74 (48.7%) | 262 (31.4%) |
| 3 | 43 (63.2%) | 60 (87.0%) | 220 (83.7%) | 218 (62.1%) | 72 (47.4%) | 553 (66.3%) |
| Total | 68 (100%) | 69 (100%) | 263 (100%) | 351 (100%) | 152 (100%) | 834 (100%) |
| Missing | 8 (10.5%) | 1 (1.4%) | 5 (1.9%) | 9 (2.5%) | 6 (3.8%) | 28 (3.2%) |
| *Histological type* | | | | | | |
| Ductal | 65 (86.7%) | 68 (97.1%) | 239 (89.8%) | 308 (86.8%) | 136 (87.2%) | 748 (87.8%) |
| Ductal and lobular | 4 (5.3%) | 2 (2.9%) | 7 (2.6%) | 12 (3.4%) | 4 (2.6%) | 27 (3.2%) |
| Lobular | 3 (4.0%) | 0 | 6 (2.3%) | 21 (5.9%) | 14 (9.0%) | 44 (5.2%) |
| Medullary | 0 | 0 | 2 (0.8%) | 2 (0.6%) | 0 | 4 (0.5%) |

**Table 1** continued

| | M1 (dnMBC) ($n = 76$) | M0 < 12 months ($n = 70$) | M0 < 24 months ($n = 268$) | M0 24–60 months ($n = 360$) | M0 60+ months ($n = 158$) | Total ($n = 862$) |
|---|---|---|---|---|---|---|
| Metaplastic | 0 | 0 | 4 (1.5%) | 0 | 0 | 4 (0.5%) |
| Mixed | 2 (2.7%) | 0 | 4 (1.5%) | 4 (1.1%) | 0 | 10 (1.2%) |
| Other | 1 (1.3%) | 0 | 4 (1.5%) | 8 (2.3%) | 2 (1.3%) | 15 (1.8%) |
| Total | 75 (100%) | 70 (100%) | 266 (100%) | 355 (100%) | 156 (100%) | 852 (100%) |
| Missing | 1 (1.3%) | 0 | 2 (0.7%) | 5 (1.4%) | 2 (1.3%) | 10 (1.2%) |
| *Surgical margins (mm)* | | | | | | |
| 0 | 5 (12.2%) | 6 (10.0%) | 27 (11.9%) | 33 (12.0%) | 15 (11.7%) | 80 (11.9%) |
| >0 to <1 | 0 | 3 (5.0%) | 7 (3.1%) | 3 (1.1%) | 3 (2.3%) | 13 (1.9%) |
| >=1 to <=5 | 19 (46.3%) | 36 (60.0%) | 115 (50.9%) | 155 (56.4%) | 66 (51.6%) | 355 (53.0%) |
| >5 | 17 (41.5%) | 15 (25.0%) | 77 (34.1%) | 84 (30.5%) | 44 (34.4%) | 222 (33.1%) |
| Total | 41 (100%) | 60 (100%) | 226 (100%) | 275 (100%) | 128 (100%) | 670 (100%) |
| Missing | 35 (46.1%) | 10 (14.3%) | 42 (15.7%) | 85 (23.6%) | 30 (19.0%) | 192 (22.3%) |
| *Lymphovascular invasion* | | | | | | |
| Absent | 18 (31.0%) | 15 (23.1%) | 68 (27.4%) | 121 (35.6%) | 60 (41.1%) | 267 (33.7%) |
| Present | 40 (69.0%) | 50 (76.9%) | 180 (72.6%) | 219 (64.4%) | 86 (58.9%) | 525 (66.3%) |
| Total | 58 (100%) | 65 (100%) | 248 (100%) | 340 (100%) | 146 (100%) | 792 (100%) |
| Missing | 18 (23.7%) | 5 (7.1%) | 20 (7.5%) | 20 (5.6%) | 12 (7.6%) | 70 (8.1%) |
| *Path T stage* | | | | | | |
| T0 | 1 (2.1%) | 1 (1.5%) | 7 (2.7%) | 7 (2.0%) | 3 (1.9%) | 18 (2.2%) |
| T1 | 16 (33.3%) | 16 (23.9%) | 79 (30.6%) | 136 (38.7%) | 68 (43.9%) | 299 (36.8%) |
| T2 | 22 (45.8%) | 34 (50.7%) | 128 (49.6%) | 168 (47.9%) | 70 (45.2%) | 388 (47.8%) |
| T3 | 9 (18.8%) | 14 (20.9%) | 41 (15.9%) | 38 (10.8%) | 14 (9.0%) | 102 (12.6%) |
| T4 | 0 | 2 (3.0%) | 3 (1.2%) | 2 (0.6%) | 0 | 5 (0.6%) |
| Total | 48 (100%) | 67 (100%) | 258 (100%) | 351 (100%) | 155 (100%) | 812 (100%) |
| Missing | 28 (36.8%) | 3 (4.3%) | 10 (3.7%) | 9 (2.5%) | 3 (1.9%) | 50 (5.8%) |
| *Max tumour size (invasive) (mm)* | | | | | | |
| Median | 35 | 32 | 30 | 27 | 25 | 28 |
| Range | 2–80 | 3–160 | 2–160 | 0–199 | .5–102 | 0–199 |
| IQR[a] | 18–49 | 25–60 | 20–47 | 18–41 | 18–40 | 19–43 |
| Missing | 29 (38.2%) | 5 (7.1%) | 18 (6.7%) | 21 (5.8%) | 6 (3.8%) | 74 (8.6%) |
| *No. of positive lymph nodes* | | | | | | |
| 0 | 11 (23.4%) | 11 (16.2%) | 66 (25.5%) | 96 (26.8%) | 49 (31.4%) | 222 (27.1%) |
| 1–3 | 12 (25.5%) | 21 (30.9%) | 89 (34.4%) | 128 (35.8%) | 70 (44.9%) | 299 (36.5%) |
| 4–9 | 11 (23.4%) | 19 (27.9%) | 52 (20.1%) | 86 (24.0%) | 30 (19.2%) | 179 (21.8%) |
| 10+ | 13 (27.7%) | 17 (25.0%) | 52 (20.1%) | 48 (13.4%) | 7 (4.5%) | 120 (14.6%) |
| Total | 47 (100%) | 68 (100%) | 259 (100%) | 358 (100%) | 156 (100%) | 820 (100%) |
| Missing | 29 (38.2%) | 2 (2.9%) | 9 (3.4%) | 2 (0.6%) | 2 (1.3%) | 42 (4.9%) |
| *ER status* | | | | | | |
| Negative | 23 (30.7%) | 41 (58.6%) | 146 (54.7%) | 99 (27.6%) | 25 (15.8%) | 293 (34.1%) |
| Positive | 52 (69.3%) | 29 (41.4%) | 121 (45.3%) | 260 (72.4%) | 133 (84.2%) | 566 (65.9%) |
| Total | 75 (100%) | 70 (100%) | 267 (100%) | 359 (100%) | 158 (100%) | 859 (100%) |
| Missing | 1 (1.3%) | 0 | 1 (0.4%) | 1 (0.3%) | 0 | 3 (0.3%) |
| *PR status* | | | | | | |
| Negative | 24 (40.7%) | 41 (66.1%) | 161 (70.6%) | 122 (41.9%) | 27 (21.8%) | 334 (47.6%) |
| Positive | 35 (59.3%) | 21 (33.9%) | 67 (29.4%) | 169 (58.1%) | 97 (78.2%) | 368 (52.4%) |
| Total | 59 (100%) | 62 (100%) | 228 (100%) | 291 (100%) | 124 (100%) | 702 (100%) |
| Missing | 17 (22.4%) | 8 (11.4%) | 40 (14.9%) | 69 (19.2%) | 34 (21.5%) | 160 (18.6%) |
| *HER2 status* | | | | | | |
| Negative | 38 (52.1%) | 44 (62.9%) | 177 (67.6%) | 224 (66.3%) | 106 (75.7%) | 545 (67.0%) |
| Positive | 35 (47.9%) | 26 (37.1%) | 85 (32.4%) | 114 (33.7%) | 34 (24.3%) | 268 (33.0%) |
| Total | 73 (100%) | 70 (100%) | 262 (100%) | 338 (100%) | 140 (100%) | 813 (100%) |
| Missing | 3 (3.9%) | 0 | 6 (2.2%) | 22 (6.1%) | 18 (11.4%) | 49 (5.7%) |
| *TNBC status*[b] | | | | | | |
| Not TNBC | 72 (94.7%) | 46 (65.7%) | 168 (62.7%) | 305 (84.7%) | 143 (90.5%) | 688 (79.8%) |
| TNBC | 4 (5.3%) | 24 (34.3%) | 100 (37.3%) | 55 (15.3%) | 15 (9.5%) | 174 (20.2%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| Missing | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 1** continued

| | M1 (dnMBC) (*n* = 76) | M0 < 12 months (*n* = 70) | M0 < 24 months (*n* = 268) | M0 24–60 months (*n* = 360) | M0 60+ months (*n* = 158) | Total (*n* = 862) |
|---|---|---|---|---|---|---|
| *Focality* | | | | | | |
| Localised | 30 (61.2%) | 40 (63.5%) | 161 (66.5%) | 202 (61.6%) | 89 (64.5%) | 482 (63.7%) |
| Multifocal | 19 (38.8%) | 23 (36.5%) | 81 (33.5%) | 126 (38.4%) | 49 (35.5%) | 275 (36.3%) |
| Total | 49 (100%) | 63 (100%) | 242 (100%) | 328 (100%) | 138 (100%) | 757 (100%) |
| Missing | 27 (35.5%) | 7 (10.0%) | 26 (9.7%) | 32 (8.9%) | 20 (12.7%) | 105 (12.2%) |
| *Surgical type* | | | | | | |
| BCS | 16 (21.1%) | 18 (25.7%) | 97 (36.2%) | 120 (33.3%) | 59 (37.3%) | 292 (33.9%) |
| Mastectomy | 33 (43.4%) | 50 (71.4%) | 166 (61.9%) | 238 (66.1%) | 98 (62.0%) | 535 (62.1%) |
| Nodal surgery only | 1 (1.3%) | 0 | 0 | 1 (0.3%) | 0 | 2 (0.2%) |
| None | 26 (34.2%) | 2 (2.9%) | 5 (1.9%) | 1 (0.3%) | 1 (0.6%) | 33 (3.8%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *Neoadjuvant chemotherapy*[c] | | | | | | |
| Yes | 0 | 28 (40.0%) | 87 (32.5%) | 84 (23.3%) | 26 (16.5%) | 197 (22.9%) |
| No | 76 (100%) | 42 (60%) | 181 (67.5%) | 276 (76.7%) | 132 (83.5%) | 665 (77.1%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *Adjuvant chemotherapy* | | | | | | |
| Yes | 0 | 40 (57.1%) | 173 (64.6%) | 256 (71.1%) | 120 (75.9%) | 549 (63.7%) |
| No | 76 (100%) | 30 (42.9%) | 95 (35.4%) | 104 (28.9%) | 38 (24.1%) | 313 (36.3%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *Palliative chemotherapy* | | | | | | |
| Yes | 75 (98.7%) | 50 (71.4%) | 204 (76.1%) | 301 (83.6%) | 111 (70.3%) | 691 (80.2%) |
| No | 1 (1.3%) | 20 (28.6%) | 64 (23.9%) | 59 (16.4%) | 47 (29.7%) | 171 (19.8%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *Palliative trastuzumab* | | | | | | |
| Yes | 12 (15.8%) | 6 (8.6%) | 19 (7.1%) | 36 (10.0%) | 10 (6.3%) | 77 (8.9%) |
| No | 64 (84.2%) | 64 (91.4%) | 249 (92.9%) | 324 (90.0%) | 148 (93.7%) | 785 (91.1%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *Palliative radiotherapy* | | | | | | |
| Yes | 58 (76.3%) | 42 (60.0%) | 155 (57.8%) | 205 (56.9%) | 79 (50.0%) | 497 (57.7%) |
| No | 18 (23.7%) | 28 (40.0%) | 113 (42.2%) | 155 (43.1%) | 79 (50.0%) | 365 (42.3%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *Hormone treatment* | | | | | | |
| Yes | 52 (68.4%) | 20 (28.6%) | 80 (29.9%) | 169 (46.9%) | 97 (61.4%) | 398 (46.2%) |
| No | 24 (31.6%) | 50 (71.4%) | 188 (70.1%) | 191 (53.1%) | 61 (38.6%) | 464 (53.8%) |
| Total | 76 (100%) | 70 (100%) | 268 (100%) | 360 (100%) | 158 (100%) | 862 (100%) |
| *Site of metastases at any time* | | | | | | |
| Bone | 8 (10.5%) | 7 (10.0%) | 31 (12.0%) | 31 (8.9%) | 14 (10.0%) | 84 (10.2%) |
| Bone–Brain | 4 (5.3%) | 4 (5.7%) | 7 (2.7%) | 13 (3.7%) | 2 (1.4%) | 26 (3.2%) |
| Bone–Visc[d] | 22 (28.9%) | 12 (17.1%) | 84 (32.4%) | 128 (36.8%) | 65 (46.4%) | 299 (36.3%) |
| Bone–Visc–Brain | 20 (26.3%) | 15 (21.4%) | 41 (15.8%) | 68 (19.5%) | 19 (13.6%) | 148 (18.0%) |
| Brain | 3 (3.9%) | 4 (5.7%) | 18 (6.9%) | 11 (3.2%) | 5 (3.6%) | 37 (4.5%) |
| Nodal | 4 (5.3%) | 3 (4.3%) | 8 (3.1%) | 15 (4.3%) | 10 (7.1%) | 37 (4.5%) |
| Visc | 12 (15.8%) | 21 (30.0%) | 53 (20.5%) | 55 (15.8%) | 17 (12.1%) | 137 (16.6%) |
| Visc–Brain | 3 (3.9%) | 4 (5.7%) | 17 (6.6%) | 27 (7.8%) | 8 (5.7%) | 55 (6.7%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 348 (100%) | 140 (100%) | 823 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 12 (3.3%) | 18 (11.4%) | 39 (4.5%) |
| *Site of first metastases* | | | | | | |
| Bone | 23 (30.3%) | 16 (22.9%) | 52 (20.1%) | 73 (21.1%) | 30 (21.7%) | 178 (21.7%) |
| Bone–Brain | 0 | 2 (2.9%) | 4 (1.5%) | 6 (1.7%) | 0 | 10 (1.2%) |
| Bone–Visc | 20 (26.3%) | 12 (17.1%) | 73 (28.2%) | 108 (31.2%) | 58 (42.0%) | 259 (31.6%) |
| Bone–Visc–Brain | 0 | 5 (7.1%) | 13 (5.0%) | 22 (6.4%) | 2 (1.4%) | 37 (4.5%) |
| Brain | 1 (1.3%) | 4 (5.7%) | 18 (6.9%) | 13 (3.8%) | 5 (3.6%) | 37 (4.5%) |
| Nodal | 12 (15.8%) | 4 (5.7%) | 17 (6.6%) | 23 (6.6%) | 15 (10.9%) | 67 (8.2%) |
| Visc | 20 (26.3%) | 25 (35.7%) | 73 (28.2%) | 88 (25.4%) | 23 (16.7%) | 204 (24.9%) |
| Visc–Brain | 0 | 2 (2.9%) | 9 (3.5%) | 13 (3.8%) | 5 (3.6%) | 27 (3.3%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 346 (100%) | 138 (100%) | 819 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 14 (3.9%) | 20 (12.7%) | 43 (5.0%) |

**Table 1** continued

| | M1 (dnMBC) ($n = 76$) | M0 < 12 months ($n = 70$) | M0 < 24 months ($n = 268$) | M0 24–60 months ($n = 360$) | M0 60+ months ($n = 158$) | Total ($n = 862$) |
|---|---|---|---|---|---|---|
| *Brain metastases at any time* | | | | | | |
| Yes | 30 (39.5%) | 27 (38.6%) | 83 (32.0%) | 119 (34.2%) | 34 (24.3%) | 266 (32.3%) |
| No | 46 (60.5%) | 43 (61.4%) | 176 (68.0%) | 229 (65.8%) | 106 (75.7%) | 557 (67.7%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 348 (100%) | 140 (100%) | 823 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 12 (3.3%) | 18 (11.4%) | 39 (4.5%) |
| *Bone metastases at any time* | | | | | | |
| Yes | 54 (71.1%) | 38 (54.3%) | 163 (62.9%) | 240 (69.0%) | 100 (71.4%) | 557 (67.7%) |
| No | 22 (28.9%) | 32 (45.7%) | 96 (37.1%) | 108 (31.0%) | 40 (28.6%) | 266 (32.3%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 348 (100%) | 140 (100%) | 823 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 12 (3.3%) | 18 (11.4%) | 39 (4.5%) |
| *Visceral metastases at any time* | | | | | | |
| Yes | 57 (75.0%) | 52 (74.3%) | 195 (75.3%) | 278 (79.9%) | 109 (77.9%) | 639 (77.6%) |
| No | 19 (25.0%) | 18 (25.7%) | 64 (24.7%) | 70 (20.1%) | 31 (22.1%) | 184 (22.4%) |
| Total | 76 (100%) | 70 (100%) | 259 (100%) | 348 (100%) | 140 (100%) | 823 (100%) |
| Missing | 0 | 0 | 9 (3.4%) | 12 (3.3%) | 18 (11.4%) | 39 (4.5%) |

[a]IQR, interquartile range.
[b]TNBC, triple-negative breast cancer.
[c]Chemotherapy refers to cytotoxic chemotherapy only, not hormone, antibody or other therapy.
[d]Visc, visceral metastatic disease.

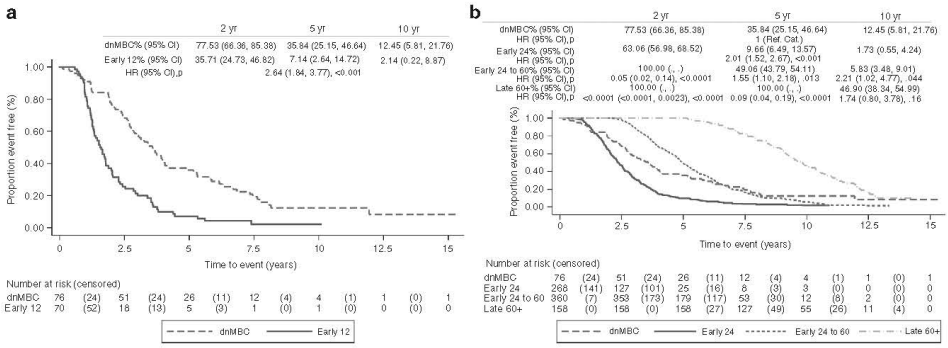



**Fig. 1 Kaplan-Meier plots of overall survival. a** OS for dnMBC vs. early 12, reference category: dnMBC. **b** OS for dnMBC vs. early 24, early 24–60 and late60 +, reference category: dnMBC.

from this study). However, a retrospective Swedish study found that only 1% of patients aged less than 40 presented with metastases, with this figure increasing successively in each age cohort (up to 10% for those aged greater than 80).[13] In other retrospective studies of women in the same age group, the de novo rate was 3.0–3.9%, not dissimilar to what is reported here.[15–17]

With regard to the identification of de novo disease, at the time of recruitment, further preoperative imaging would only have been performed if the patient had symptoms suggestive of metastatic disease, or possibly because of clinically positive axillary nodes or a large primary tumour. The 2009 National Institute of Clinical Excellence guidelines advised that patients with early breast cancer should only undergo staging for metastatic disease in the presence of symptoms.[18] CT would not have routinely been used in all centres; screening for occult metastases may instead have involved chest radiographs, liver ultrasound and bone scintigraphy.[19] Therefore, women diagnosed with de novo disease at the time of the POSH study are more likely to have had adverse tumour features clinically or concerning symptoms. Given that we have shown younger patients to have a high rate of node positivity and more advanced T stage, in addition to ER negativity, it is possible that they were more likely to have baseline imaging. At the present time, there is no difference in recommended staging or follow-up for younger patients.[11] One retrospective study found that a baseline PET/CT scan upgraded 15% of young, asymptomatic patients with early-stage breast cancer to stage IV.[20] Given the better survival for de novo patients here, compared with those who relapsed within 24 months, and the more adverse biology in young patients, age should be incorporated into clinicians' decision-making with regard to baseline imaging. A randomised controlled trial would be required to identify whether routine imaging for metastatic disease at baseline would improve survival for young patients.

A third of women who developed metastatic disease (34.1%) relapsed within 24 months of diagnosis. This group had a

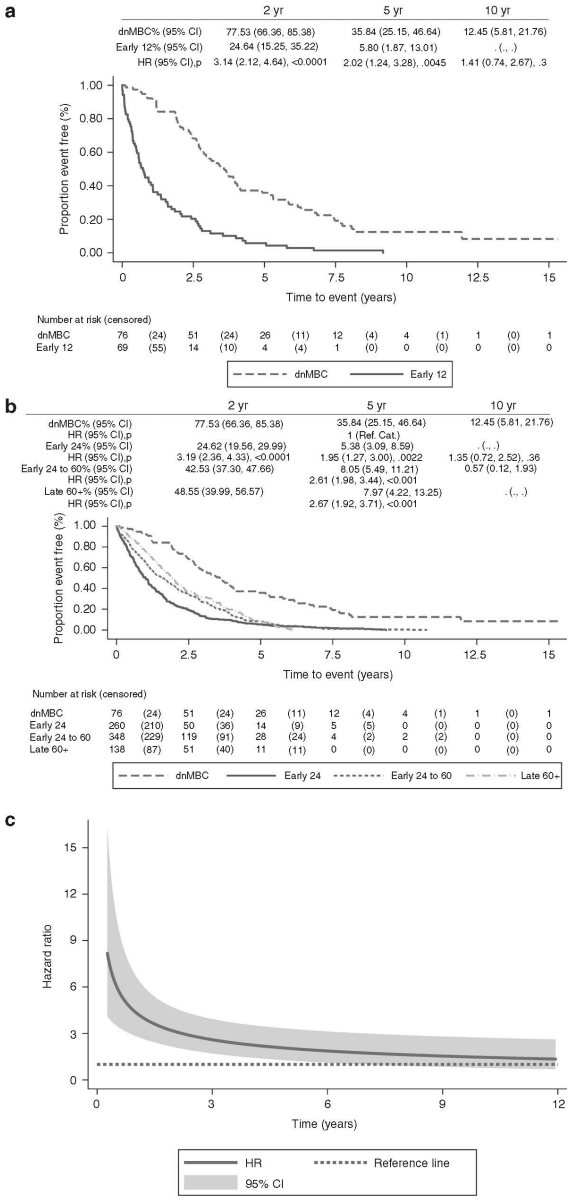



**Fig. 2 Kaplan-Meier plots of post-distant relapse-free survival. a** PDRS for dnMBC vs. early 12, reference category: dnMBC. **b** PDRS for dnMBC vs. early 24, early 24–60 and early 60 +, reference category: dnMBC. **c** Time-varying HR for PDRS for dnMBC vs. early12, reference category: dnMBC.

**Table 2.** MVA[a] for OS and PDRS for dnMBC vs. early 12.

| Covariate | OS HR (95% CI), *p* value | PDRS |
|---|---|---|
| *DENOVO* | | |
| dnMBC | 1 (Ref. category) | 1 (Ref. category) |
| Early 12 | 3.76 (2.22–6.38), <0.001 | 5.12 (2.95–8.87), <0.001 |
| Age at diagnosis (years) | 0.97 (0.91–1.03), 0.295 | 0.97 (0.91–1.03), 0.355 |
| *BMI* | | |
| BMI < 25 | 1 (Ref. category) | 1 (Ref. category) |
| 25 ≤ BMI < 30 | 1.31 (0.76–2.25), $p - 0.333$ | 1.27 (0.74–2.20), $p - 0.387$ |
| BMI ≥ 30 | 1.30 (0.75–2.27), $p - 0.355$ | 1.21 (0.69–2.12), $p - 0.498$ |
| *Grade* | | |
| 1 | 1 (Ref. category) | 1 (Ref. category) |
| 2 | 7.78 (0.69–87.14), $p - 0.096$ | 7.14 (0.65–78.91), $p - 0.109$ |
| 3 | 5.47 (0.54–55.22), $p - 0.150$ | 5.13 (0.51–51.75), $p - 0.165$ |
| Max invasive size (mm) | 1.00 (0.99–1.01), $p - 0.997$ | 1.00 (0.99–1.01), $p - 0.900$ |
| *N stage* | | |
| N0 | 1 (Ref. category) | 1 (Ref. category) |
| N1 | 2.29 (1.17–4.47), $p - 0.015$ | 2.42 (1.23–4.77), $p - 0.011$ |
| *HER2 status* | | |
| Negative | 1 (Ref. category) | 1 (Ref. category) |
| Positive | 0.50 (0.31–0.80), $p - 0.004$ | 0.48 (0.30–0.78), $p - 0.003$ |

[a]MVA Cox model stratified by ER status and ethnicity due to time-varying hazards.

**Table 3.** MVA[a] for PDRS for dnMBC vs. early 24.

| Covariate | HR (95% CI), *p* value |
|---|---|
| *DENOVO (time-varying)* | |
| dnMBC | 1 (Ref. category) |
| Early 24 (at 2 years) | 2.53 (1.50–4.27), 0.00052 |
| Early 24 (at 5 years) | 2.42 (1.39–4.22), 0.0019 |
| Early 24 (at 10 years) | 3.88 (1.07–14.07), 0.039 |
| Age at diagnosis (years) | 1.01 (0.98–1.05), 0.458 |
| *BMI* | |
| BMI < 25 | 1 (Ref. category) |
| 25 ≤ BMI < 30 | 1.14 (0.84–1.53), $p = 0.403$ |
| BMI ≥ 30 | 1.19 (0.86–1.65), $p = 0.301$ |
| *Grade* | |
| 1 | 1 (Ref. category) |
| 2 | 1.09 (0.24–4.89), $p = 0.912$ |
| 3 | 1.03 (0.23–4.51), $p = 0.971$ |
| Max invasive size (mm) | 1.00 (0.997–1.01), $p = 0.384$ |
| *N stage* | |
| N0 | 1 (Ref. category) |
| N1 | 1.42 (1.05–1.93), $p = 0.024$ |
| *HER2 status* | |
| Negative | 1 (Ref. category) |
| Positive | 0.66 (0.51–0.86), $p = 0.002$ |
| *Ethnicity* | |
| White/Caucasian | 1 (Ref. category) |
| Black | 0.61 (0.34–1.12), $p = 0.114$ |
| Asian | 1.63 (0.74–3.61), $p = 0.225$ |
| Other | 0.43 (0.10–1.77), $p = 0.240$ |
| *ER status (time-varying)* | |
| Negative | 1 (Ref. category) |
| Positive (at 2 years) | 0.50 (0.38–0.67), $p < 0.0001$ |
| Positive (at 5 years) | 0.79 (0.48–1.29), $p = 0.350$ |
| Positive (at 10 years) | 1.30 (0.52–3.25), $p = 0.590$ |

[a]Stpm2 model (Flexible parametric survival model). Time varying for dnMBC vs. early24 and ER status. However, no stratification of ethnicity incorporated into this model as non-time-varying.

significantly worse OS compared with those who presented with metastatic disease. A review of long-term survivorship with metastatic disease shows 10-year OS rates of 12.45% (95% CI 5.81–21.76) for dnMBC patients, but only 1.73% (0.55–4.24) for patients who relapsed within 24 months (Fig. 1b). For PDRS, outcomes were better for de novo disease compared with all relapsed groups, including those who relapsed late (after 5 years). Whether the improved outcomes for dnMBC are due to inherently more favourable biology or different treatment approaches is unclear. Adjuvant therapy is likely to drive subclonal diversification, resulting in mutations that confer resistance to further cytotoxic/hormonal treatment. Women with de novo disease may also have been treated more aggressively, e.g., with local therapy for oligometastatic disease. It might be assumed that late-relapsing disease (84.2% of whom are ER-positive) is inherently indolent, but this study challenges that assumption. Either the metastatic clones present remain dormant until acquiring sufficient oncogenic drivers, or the clones are present from diagnosis, but are suppressed by adjuvant hormone therapy. Further genomic work to understand the subclonal architecture, and heterogeneity of primary and metastatic sites, is required, in order to guide treatment strategies in the era of personalised medicine.

The de novo cohort had a remarkably high prevalence of HER2-positive tumours (47.9%). The number of HER2 + cases in this group may contribute to the improved survival seen with dnMBC, given previous reports of long-term responses to trastuzumab in a proportion of patients with HER2 + MBC.[21] However, not all HER2 + patients received trastuzumab, reflective of the era during which POSH was recruited. Trastuzumab was approved for use in the adjuvant setting in 2005, 3 years after the trial began recruitment. In general, baseline tumour characteristics were adverse in the early-relapse group, with the greatest proportion of T2/3, node-positive and LVI tumours. These features may account for the worse prognosis in this group. The late-relapse group was marked by ER positivity (84.2% vs. 41.4% in the early-12 group), node negativity (31.4% vs. 16.2%) and smaller median tumour size (25 mm vs. 32 mm).

It has been hypothesised that there is a different pattern of metastatic spread between patients with primary and secondary metastatic breast cancer. In fact, the de novo group had the highest proportion of widespread (bone, visceral and brain) metastases during their disease course (26.3%), although this may be reflective of their longer survival and the resultant time to allow dissemination. In addition, their widespread metastases may have produced symptoms that resulted in their de novo disease being detected with imaging at diagnosis. The prevalence of HER2 positivity in this group may also account for this. The late-relapse group was the least likely to develop brain metastases during the course of their disease (24.3%); proportions were similar for the dnMBC and early-12 groups (39.5 and 38.6%, respectively). However, only one of the dnMBC patients (1.3%) had brain metastases at diagnosis, compared with 18.6% of the early-12 group. This may reflect clinicians being more likely to perform a baseline CT brain in patients with recurrent, as opposed to newly diagnosed, breast cancer. Young age has previously been

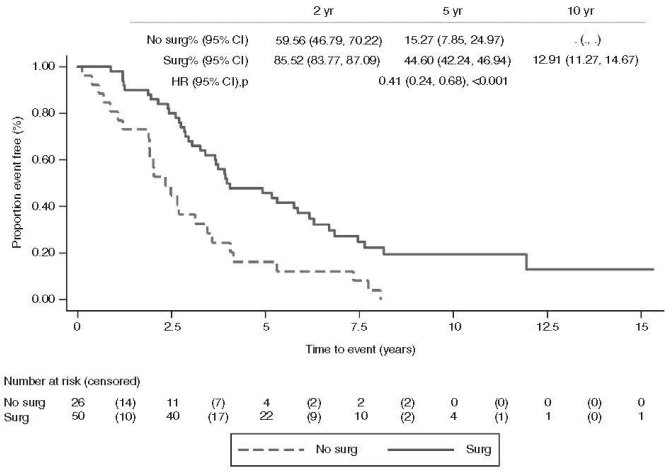



**Fig. 3 Kaplan-Meier plot of overall survival by surgical category.** OS by surgical type (dnMBC patients only); reference category: no surgery.

associated with an increased risk of brain metastases; the prevalence in this entire cohort was nearly a third (32.3%).[11] Clinicians should be vigilant for central nervous system symptoms in young women during follow-up for breast cancer. There is no clear evidence that the sites of distant disease explain the differing prognosis between relapse categories; it seems more likely that the underlying biology influences metastatic sites, which determines whether or not a patient presents with de novo disease.

This cohort is unique not only for its age but also for completeness of *BRCA* gene mutation testing. It was notable therefore that a relatively large proportion of patients with de novo disease (11.8%) had a *BRCA2* mutation, whereas just one (1.3%) had a *BRCA1* mutation. Although the 69.3% ER-positivity rate in this group may explain this to some extent, the ER positivity was higher in the early-24–60 and late-relapse groups with a lower BRCA2 prevalence (5.8% and 6.3%, respectively). Across the cohort of 862 patients with metastatic disease, the BRCA2 mutation rate was 5.6%; the rate was 5.0% across the POSH cohort as a whole (excluding dnMBC patients).[22] The reason for such a large proportion of dnMBC cases having a *BRCA2* mutation is unclear; it is possible that a family history of breast and ovarian cancer in *BRCA2* mutation carriers meant that they were more vigilant regarding symptoms of metastatic disease. In addition, 57% of the dnMBC cohort had bone metastases at presentation (in common with 69.3% being ER + ve, the common phenotype arising from a *BRCA2* mutation); perhaps, bone pain in a young woman is a red- flag symptom that resulted in early imaging. This might enrich the dnMBC with *BRCA2* mutation carriers. Our results would suggest that further studies using *BRCA* germline testing in young women with dnMBC are warranted.

Primary surgery in patients with de novo metastatic breast cancer remains a debated issue, with decisions made on a case-by-case basis. In this study, for patients with dnMBC, there was an improved survival for those who had surgery ($n = 50$) vs. those who did not ($n = 26$), with a univariable HR of 0.41 ($p = <0.001$). This outcome is susceptible to selection bias as locoregional treatment was presumably more likely to occur if disease was apparently relatively indolent, less widespread and the patient's performance status was good. A randomised prospective trial to address this issue in the modern era is required.

The groups least likely to receive palliative chemotherapy were the early-12 and late-relapse groups (71.4% and 70.3%, respectively). In the early-12 group, this may be related to lack of recovery from toxicity following recent adjuvant/neoadjuvant chemotherapy, radiotherapy and surgery. Physicians may also have chosen not to treat refractory disease with further systemic treatment. Amongst the late-relapse group, 84.2% of whom were ER-positive, most patients received hormone therapy (61.4%). This is likely to reflect the burden and distribution of metastatic disease. It may also be that the late relapses were perceived as more indolent. They experienced less brain metastases compared with the dnMBC group (24.3% vs. 39.5%), but a similar proportion of visceral metastases (77.9% vs. 75.0%). In fact, the poor PDRS of the late-relapse group compared with dnMBC would indicate that alternative treatment strategies are needed for this cohort.

The potential limitations of this study include its age. As a result, there have been changes in systemic options available, although most patients in this study were treated with anthracycline +/− taxane chemotherapy, and approximately half of them received trastuzumab if HER2 + .[12] There is an increasingly proactive approach to staging investigations, including the use of advanced imaging, such as positron emission tomography. This may mean that more patients are now diagnosed with occult metastases at presentation, affecting the characteristics of the de novo group. Finally, we cannot rule out a degree of selection bias during POSH recruitment. However, as detailed previously, POSH participants recruited from England represented 23% of the available population during the recruitment period, and comparison with cancer registry data confirmed that the cohort is representative of the wider population.[22] The strengths of this study include the large cohort size and complete germline *BRCA* testing. There are few missing data (with HER2 status missing in only 5.7% of cases) and long follow-up, with only a small number of patients lost to follow-up.

## CONCLUSION

This is the first report to describe patterns of metastatic disease in a large prospective cohort of young-onset breast cancer patients, with a long follow-up period and complete *BRCA* germline testing.

We have shown that women aged 40 or less with de novo metastatic breast cancer have better survival following the onset of metastatic disease than those who develop secondary breast cancer. Despite more favourable baseline tumour characteristics, patients who developed late-onset metastatic disease had a worse PDRS than de novo patients, indicating that chemotherapy-resistant clones and/or perceived poor fitness due to prior therapies have a significant impact on prognosis. A notable proportion of women with dnMBC had a germline *BRCA2* mutation; this has not previously been highlighted in the literature.

## THE POSH STEERING GROUP

Douglas G. Altman[3], Louise Jones[4], Gareth Evans[5], Alastair M. Thompson[6], Paul Pharaoh[7], Andrew Hanby[8], Sunil Lakhani[9], Ros Eeles[10], Fiona J. Gilbert[11], Hisham Hamed[12], Shirley Hodgson[13], Peter Simmonds[14] and Diana Eccles[15]

[3]Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK; [4]Tumour Biology Department, Institute of Cancer, Barts and The London School of Medicine & Dentistry, London, UK; [5]Genomic Medicine, Division of Evolution and Genomic Sciences, University of Manchester MAHSC, St Mary's Hospital, Manchester, UK; [6]University of Texas MD Anderson Cancer Center, Houston, TX, USA; [7]Public Health and Primary Care, University of Cambridge, Cambridge, UK; [8]Department of Pathology, University of Leeds, Faculty of Medicine, Leeds, UK; [9]Discipline of Molecular & Cellular Pathology, Faculty of Medicine, University of Queensland, The Royal Brisbane & Women's Hospital, Brisbane, QLD, Australia; [10]Institute of Cancer Research, London, UK; [11]Department of Radiology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK; [12]Guy's & St Thomas' Hospital, London, UK; [13]St George's Hospital, University of London, London, UK; [14]University Hospital Southampton NHS Foundation Trust, Southampton, UK and [15]Cancer Sciences Academic Unit, Faculty of Medicine, University of Southampton, Southampton, UK

## ACKNOWLEDGEMENTS

We are grateful to all the patients, clinicians, research staff at the National Cancer Research Institute Clinical Research Network and the POSH research team who made this study possible. We also thank the POSH study steering group. Sample handling was facilitated by Southampton CRUK Centre tissue bank and Southampton University Faculty of Medicine DNA bank (Southampton, UK) and the Barts Cancer Research Centre (London, UK). DNA sequencing for the whole cohort took place in the Strangeways research laboratories (Cambridge, UK), and validation, Sanger sequencing and multiplex ligation-dependent probe amplification was done in the Wessex Regional Genetics Laboratories (Wessex, UK). IT support, histopathology image storage and reporting software were developed and supported by the University of Southampton Clinical Informatics Support team.

## AUTHOR CONTRIBUTIONS

H.M., D.E., E.C. and T.M. participated in the study conception and design. H.M., D.E., E. C. and T.M. wrote the paper. H.M., L.D. and T.M. performed data analysis and interpretation. All authors approved the final draft.

## ADDITIONAL INFORMATION

**Ethics approval and consent to participate** Written informed consent was obtained from all participants, and the study was performed in accordance with the Declaration of Helsinki. Ethical approval was granted in 2000 (MREC 00/6/69), and the study was approved for recruitment as part of the UK National Cancer Research Network (NCRN) portfolio in 2002, subsequently the NIHR portfolio.

**Consent to publish** None.

**Data availability** The data sets generated and/or analysed during this study are not publicly available.

**Competing interests** ERC declares honoraria from Roche, Astra-Zeneca, Pfizer, Lilly and Nanostring. DME declares honoraria from Astra-Zeneca and Pierre Fabre. The remaining authors declare no competing interests.

**Funding information** Funding over 18 years has been provided by the Wessex Cancer Trust, Cancer Research UK (C1275/A7572, C22524, A11699 and A19187) and Breast Cancer Now (2005Nov53). HM is funded by Cancer Research UK.

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41416-020-0784-z.



## REFERENCES

1. UK CR. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive. Accessed June 2019.
2. Cortesi, L., Toss, A., Cirilli, C., Marcheselli, L., Braghiroli, B., Sebastiani, F. et al. Twenty-years experience with de novo metastatic breast cancer. *Int J. Cancer* **137**, 1417–1426 (2015).
3. Lin, N. U., Thomssen, C., Cardoso, F., Cameron, D., Cufer, T., Fallowfield, L. et al. International guidelines for management of metastatic breast cancer (MBC) from the European School of Oncology (ESO)-MBC Task Force: surveillance, staging, and evaluation of patients with early-stage and metastatic breast cancer. *Breast* **22**, 203–210 (2013).
4. Chia, S. K., Speers, C. H., D'Yachkova, Y., Kang, A., Malfair-Taylor, S., Barnett, J. et al. The impact of new chemotherapeutic and hormone agents on survival in a population-based cohort of women with metastatic breast cancer. *Cancer* **110**, 973–979 (2007).
5. Malmgren, J. A., Mayer, M., Atwood, M. K. & Kaplan, H. G. Differential presentation and survival of de novo and recurrent metastatic breast cancer over time: 1990-2010. *Breast Cancer Res. Treat.* **167**, 579–590 (2018).
6. Lobbezoo, D. J., van Kampen, R. J., Voogd, A. C., Dercksen, M. W., van den Berkmortel, F., Smilde, T. J. et al. Prognosis of metastatic breast cancer: are there differences between patients with de novo and recurrent metastatic breast cancer? *Br. J. Cancer* **112**, 1445–1451 (2015).
7. Guth, U., Magaton, I., Huang, D. J., Fisher, R., Schotzau, A. & Vetter, M. Primary and secondary distant metastatic breast cancer: two sides of the same coin. *Breast* **23**, 26–32 (2014).
8. Dawood, S., Broglio, K., Ensor, J., Hortobagyi, G. N. & Giordano, S. H. Survival differences among women with de novo stage IV and relapsed breast cancer. *Ann. Oncol.* **21**, 2169–2174 (2010).
9. den Brok, W. D., Speers, C. H., Gondara, L., Baxter, E., Tyldesley, S. K. & Lohrisch, C. A. Survival with metastatic breast cancer based on initial presentation, de novo versus relapsed. *Breast Cancer Res. Treat.* **161**, 549–556 (2017).
10. Poggio, F., Lambertini, M. & de Azambuja, E. Surgery of the primary tumour in patients presenting with de novo metastatic breast cancer: to do or not to do? *ESMO Open* **3**, e000324 (2018).
11. Cardoso, F., Loibl, S., Pagani, O., Graziottin, A., Panizza, P., Martincich, L. et al. The European Society of Breast Cancer Specialists recommendations for the management of young women with breast cancer. *Eur. J. Cancer* **48**, 3355–3377 (2012).
12. Copson, E., Eccles, B., Maishman, T., Gerty, S., Stanton, L., Cutress, R. I. et al. Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study. *J. Natl Cancer Inst.* **105**, 978–988 (2013).
13. Brandt, J., Garne, J. P., Tengrup, I. & Manjer, J. Age at diagnosis in relation to survival following breast cancer: a cohort study. *World J. Surg. Oncol.* **13**, 33 (2015).
14. Eccles, D., Gerty, S., Simmonds, P., Hammond, V., Ennis, S., Altman, D. G. et al. Prospective study of outcomes in sporadic versus hereditary breast cancer (POSH): study protocol. *BMC Cancer* **7**, 160 (2007).
15. Fredholm, H., Eaker, S., Frisell, J., Holmberg, L., Fredriksson, I. & Lindman, H. Breast cancer in young women: poor survival despite intensive treatment. *PLoS ONE* **4**, e7695 (2009).
16. Gnerlich, J. L., Deshpande, A. D., Jeffe, D. B., Sweet, A., White, N. & Margenthaler, J. A. Elevated breast cancer mortality in women younger than age 40 years compared with older women is attributed to poorer survival in early-stage disease. *J. Am. Coll. Surg.* **208**, 341–347 (2009).
17. Kheirelseid, E. H., Boggs, J. M., Curran, C., Glynn, R. W., Dooley, C., Sweeney, K. J. et al. Younger age as a prognostic indicator in breast cancer: a cohort study. *BMC Cancer* **11**, 383 (2011).
18. Chand, N., Cutress, R. I., Oeppen, R. S. & Agrawal, A. Staging investigations in breast cancer: collective opinion of UK breast surgeons. *Int. J. Breast Cancer* **2013**, 506172 (2013).

19. Barrett, T., Bowden, D. J., Greenberg, D. C., Brown, C. H., Wishart, G. C. & Britton, P. D. Radiological staging in breast cancer: which asymptomatic patients to image and how. *Br. J. Cancer* **101**, 1522–1528 (2009).
20. Riedl, C. C., Slobod, E., Jochelson, M., Morrow, M., Goldman, D. A., Gonen, M. et al. Retrospective analysis of 18F-FDG PET/CT for staging asymptomatic breast cancer patients younger than 40 years. *J. Nucl. Med.* **55**, 1578–1583 (2014).
21. Murthy, P., Kidwell, K. M., Schott, A. F., Merajver, S. D., Griggs, J. J., Smerage, J. D. et al. Clinical predictors of long-term survival in HER2-positive metastatic breast cancer. *Breast Cancer Res. Treat.* **155**, 589–595 (2016).
22. Copson, E. R., Maishman, T. C., Tapper, W. J., Cutress, R. I., Greville-Heygate, S., Altman, D. G. et al. Germline BRCA mutation and outcome in young-onset breast cancer (POSH): a prospective cohort study. *Lancet Oncol.* **19**, 169–180 (2018).

# Appendix E BRCA ExomeSeq collaboration proposal

The BRCA ExomeSeq project is a collaboration between the POSH group here at Southampton (including Chief and Principal Investigators Diana Eccles and Ellen Copson) and long standing collaborating clinicians scientists Professors Katherine Nathanson and Susan Domchek of the Basser Center for BRCA at the University of Pennsylvania, USA. In order to facilitate the transfer of tissue material from Southampton to Pennsylvania, an MTA needs to be in place.

The BRCA ExomeSeq project is covered by the POSH protocol and consent form and therefore covered by the current POSH ethics approval which was approved in 2001 by a multicentre research ethics committee predating the introduction of IRAS (MREC 00/6/69); it. The documents are attached to this ERGO application with relevant sections highlighted. In particular it should be noted that all patients specifically consented to the use of their surplus breast cancer tissue for further analysis. It should also be noted that over a third of recruited patients are no longer alive and that patient representatives and lay representatives with legal background consider that it would be both impractical and potentially distressing to return to participants to request permission for a specific related research study like this when the principle is already covered by the current consent.

In brief, the project will involve sending a sample of formalin-fixed paraffin embedded tumour tissue and a matched DNA sample extracted from blood for cases identified within the POSH cohort where there has been a confirmed germline BRCA mutation (approximately 200-220 cases in total). The team at Pennsylvania will extract DNA from tumour tissue and sequence tumour and blood DNA pairs allowing analysis of the

acquired genetic alterations in the tumour. The pattern of tumour genetic alterations will then be correlated with a minimal dataset relating to tumour characteristics and treatment outcomes. The sequence data generated in Pennsylvania will be returned to the POSH group (as with previously completed MTAs) and will enable further research opportunities for the University of Southampton within the field of breast cancer biology and homologous repair deficiency. Dr. Nathanson has extensive experience with this type of tumour sequencing and this collaboration is considered to be of great value to the POSH group in pursuing high impact research outputs therefore of benefit to the University of Southampton.

It should be noted that in the POSH protocol, the main purpose of the study was to be a "first class biological resource that will allow a number of key research questions about hereditary breast cancer to be addressed" and this current project plans to utilize this resource with this exact aim in mind.

Recruitment to the POSH study began in 2000 and therefore predates some requirements now in place as part of ethics and research governance. With regards to data sharing, I have attached the most recent POSH data management policy, which was updated in May 2018. This was written by Ellen Copson as part of a Cancer Research UK grant application. Please note Section 4.1 "Mechanisms for sharing" which includes details on previous collaborations using data from POSH.

**Hayley McKenzie**

**Cancer Research UK Clinical Research Fellow, PhD student**

**Cancer Sciences, University of Southampton**

**February 2019**

# List of References

1. UK CR. [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive.

2. Waks AG, Winer EP. Breast Cancer Treatment: A Review. JAMA. 2019;321(3):288-300.

3. Cancer Research UK [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/survival.

4. Allred DC. Issues and updates: evaluating estrogen receptor-alpha, progesterone receptor, and HER2 in breast cancer. Mod Pathol. 2010;23 Suppl 2:S52-9.

5. Rakha EA, Pinder SE, Bartlett JM, Ibrahim M, Starczynski J, Carder PJ, et al. Updated UK Recommendations for HER2 assessment in breast cancer. J Clin Pathol. 2015;68(2):93-9.

6. Colleoni M, Sun Z, Price KN, Karlsson P, Forbes JF, Thurlimann B, et al. Annual Hazard Rates of Recurrence for Breast Cancer During 24 Years of Follow-Up: Results From the International Breast Cancer Study Group Trials I to V. J Clin Oncol. 2016;34(9):927-35.

7. Lee T, Isaacs C. Treatment of primary breast tumors in de novo metastatic breast cancer. Clin Adv Hematol Oncol. 2014;12(12):820-7.

8. Cortesi L, Toss A, Cirilli C, Marcheselli L, Braghiroli B, Sebastiani F, et al. Twenty-years experience with de novo metastatic breast cancer. Int J Cancer. 2015;137(6):1417-26.

9. Lin NU, Thomssen C, Cardoso F, Cameron D, Cufer T, Fallowfield L, et al. International guidelines for management of metastatic breast cancer (MBC) from the European School of Oncology (ESO)-MBC Task Force: Surveillance, staging, and evaluation of patients with early-stage and metastatic breast cancer. Breast. 2013;22(3):203-10.

10. Murthy P, Kidwell KM, Schott AF, Merajver SD, Griggs JJ, Smerage JD, et al. Clinical predictors of long-term survival in HER2-positive metastatic breast cancer. Breast Cancer Res Treat. 2016;155(3):589-95.

11. Malmgren JA, Mayer M, Atwood MK, Kaplan HG. Differential presentation and survival of de novo and recurrent metastatic breast cancer over time: 1990-2010. Breast Cancer Res Treat. 2018;167(2):579-90.

12. Guth U, Magaton I, Huang DJ, Fisher R, Schotzau A, Vetter M. Primary and secondary distant metastatic breast cancer: two sides of the same coin. Breast. 2014;23(1):26-32.

13. Dawood S, Broglio K, Ensor J, Hortobagyi GN, Giordano SH. Survival differences among women with de novo stage IV and relapsed breast cancer. Ann Oncol. 2010;21(11):2169-74.

14. den Brok WD, Speers CH, Gondara L, Baxter E, Tyldesley SK, Lohrisch CA.

Survival with metastatic breast cancer based on initial presentation, de novo versus relapsed. Breast Cancer Res Treat. 2017;161(3):549-56.

15. Lobbezoo DJ, van Kampen RJ, Voogd AC, Dercksen MW, van den Berkmortel F, Smilde TJ, et al. Prognosis of metastatic breast cancer: are there differences between patients with de novo and recurrent metastatic breast cancer? Br J Cancer. 2015;112(9):1445-51.

16. Chia SK, Speers CH, D'Yachkova Y, Kang A, Malfair-Taylor S, Barnett J, et al. The impact of new chemotherapeutic and hormone agents on survival in a population-based cohort of women with metastatic breast cancer. Cancer. 2007;110(5):973-9.

17. Cardoso F, Loibl S, Pagani O, Graziottin A, Panizza P, Martincich L, et al. The European Society of Breast Cancer Specialists recommendations for the management of young women with breast cancer. Eur J Cancer. 2012;48(18):3355-77.

18. Parish CR. Cancer immunotherapy: the past, the present and the future. Immunol Cell Biol. 2003;81(2):106-13.

19. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.

20. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.

21. MacPherson Ga, Austyn JMa. Exploring immunology : concepts and evidence.

22. Schreiber TH, Raez L, Rosenblatt JD, Podack ER. Tumor immunogenicity and responsiveness to cancer vaccine therapy: the state of the art. Semin Immunol. 2010;22(3):105-12.

23. Blankenstein T, Coulie PG, Gilboa E, Jaffee EM. The determinants of tumour immunogenicity. Nat Rev Cancer. 2012;12(4):307-13.

24. Spranger S, Luke JJ, Bao R, Zha Y, Hernandez KM, Li Y, et al. Density of immunogenic antigens does not explain the presence or absence of the T-cell-inflamed tumor microenvironment in melanoma. Proc Natl Acad Sci U S A. 2016;113(48):E7759-E68.

25. Owen JAa, Punt Ja, Stranford SAa, Kuby J, Kindt TJKi. Kuby immunology. Seventh edition, International edition / Judith A. Owen, Jenni Punt, Sharon A. Stranford ; with contributions by Patricia P. Jones. ed.

26. Matsumoto H, Koo SL, Dent R, Tan PH, Iqbal J. Role of inflammatory infiltrates in triple negative breast cancer. J Clin Pathol. 2015;68(7):506-10.

27. Dushyanthen S, Beavis PA, Savas P, Teo ZL, Zhou C, Mansour M, et al. Relevance of tumor-infiltrating lymphocytes in breast cancer. BMC Med. 2015;13(1):202.

28. Vinay DS, Ryan EP, Pawelec G, Talib WH, Stagg J, Elkord E, et al. Immune evasion in cancer: Mechanistic basis and therapeutic strategies. Semin Cancer Biol. 2015;35 Suppl:S185-S98.

29. Savas P, Salgado R, Denkert C, Sotiriou C, Darcy PK, Smyth MJ, et al. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. Nat Rev Clin Oncol.

2016;13(4):228-41.

30. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, et al. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. J Clin Oncol. 2011;29(15):1949-55.

31. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann Oncol. 2015;26(2):259-71.

32. Dieci MV, Radosevic-Robin N, Fineberg S, van den Eynden G, Ternes N, Penault-Llorca F, et al. Update on tumor-infiltrating lymphocytes (TILs) in breast cancer, including recommendations to assess TILs in residual disease after neoadjuvant therapy and in carcinoma in situ: A report of the International Immuno-Oncology Biomarker Working Group on Breast Cancer. Semin Cancer Biol. 2017.

33. Aaltomaa S, Lipponen P, Eskelinen M, Kosma VM, Marin S, Alhava E, et al. Lymphocyte infiltrates as a prognostic variable in female breast cancer. Eur J Cancer. 1992;28A(4-5):859-64.

34. Baker K, Lachapelle J, Zlobec I, Bismar TA, Terracciano L, Foulkes WD. Prognostic significance of CD8+ T lymphocytes in breast cancer depends upon both oestrogen receptor status and histological grade. Histopathology. 2011;58(7):1107-16.

35. Liu S, Lachapelle J, Leung S, Gao D, Foulkes WD, Nielsen TO. CD8+ lymphocyte infiltration is an independent favorable prognostic indicator in basal-like breast cancer. Breast Cancer Res. 2012;14(2):R48.

36. Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. J Clin Oncol. 2014;32(27):2959-66.

37. Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. J Clin Oncol. 2015;33(9):983-91.

38. Bianchini G, Gianni L. The immune system and response to HER2-targeted treatment in breast cancer. The Lancet Oncology. 2014;15(2):e58-e68.

39. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. Lancet Oncol. 2017.

40. Dieci MV, Criscitiello C, Goubar A, Viale G, Conte P, Guarneri V, et al. Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: a retrospective multicenter study. Ann Oncol. 2014;25(3):611-8.

41. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of

docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. J Clin Oncol. 2013;31(7):860-7.

42. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol. 2016;13(11):674-90.

43. Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, et al. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. Ann Oncol. 2014;25(8):1544-50.

44. Dieci MV, Mathieu MC, Guarneri V, Conte P, Delaloge S, Andre F, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in two phase III randomized adjuvant breast cancer trials. Ann Oncol. 2015;26(8):1698-704.

45. Pruneri G, Gray KP, Vingiani A, Viale G, Curigliano G, Criscitiello C, et al. Tumor-infiltrating lymphocytes (TILs) are a powerful prognostic marker in patients with triple-negative breast cancer enrolled in the IBCSG phase III randomized clinical trial 22-00. Breast Cancer Res Treat. 2016;158(2):323-31.

46. Demaria S, Volm MD, Shapiro RL, Yee HT, Oratz R, Formenti SC, et al. Development of tumor-infiltrating lymphocytes in breast cancer after neoadjuvant paclitaxel chemotherapy. Clin Cancer Res. 2001;7(10):3025-30.

47. Ladoire S, Arnould L, Apetoh L, Coudert B, Martin F, Chauffert B, et al. Pathologic complete response to neoadjuvant chemotherapy of breast carcinoma is associated with the disappearance of tumor-infiltrating foxp3+ regulatory T cells. Clin Cancer Res. 2008;14(8):2413-20.

48. Conforti R, Ma Y, Morel Y, Paturel C, Terme M, Viaud S, et al. Opposing effects of toll-like receptor (TLR3) signaling in tumors can be therapeutically uncoupled to optimize the anticancer efficacy of TLR3 ligands. Cancer Res. 2010;70(2):490-500.

49. Huang Z, Peng S, Knoff J, Lee SY, Yang B, Wu TC, et al. Combination of proteasome and HDAC inhibitor enhances HPV16 E7-specific CD8+ T cell immune response and antitumor effects in a preclinical cervical cancer model. J Biomed Sci. 2015;22:7.

50. Woan KV, Lienlaf M, Perez-Villaroel P, Lee C, Cheng F, Knox T, et al. Targeting histone deacetylase 6 mediates a dual anti-melanoma effect: Enhanced antitumor immunity and impaired cell proliferation. Mol Oncol. 2015;9(7):1447-57.

51. Ladoire S, Mignot G, Dabakuyo S, Arnould L, Apetoh L, Rebe C, et al. In situ immune response after neoadjuvant chemotherapy for breast cancer predicts survival. J Pathol. 2011;224(3):389-400.

52. DeNardo DG, Brennan DJ, Rexhepaj E, Ruffell B, Shiao SL, Madden SF, et al. Leukocyte complexity predicts breast cancer survival and functionally regulates response to chemotherapy. Cancer Discov. 2011;1(1):54-67.

53. Tan W, Zhang W, Strasner A, Grivennikov S, Cheng JQ, Hoffman RM, et al. Tumour-infiltrating regulatory T cells stimulate mammary cancer metastasis through RANKL-RANK signalling. Nature. 2011;470(7335):548-53.

54. Bates GJ, Fox SB, Han C, Leek RD, Garcia JF, Harris AL, et al. Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk of late relapse. J Clin Oncol. 2006;24(34):5373-80.

55. West NR, Kost SE, Martin SD, Milne K, Deleeuw RJ, Nelson BH, et al. Tumour-infiltrating FOXP3(+) lymphocytes are associated with cytotoxic immune responses and good clinical outcome in oestrogen receptor-negative breast cancer. Br J Cancer. 2013;108(1):155-62.

56. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Lee AH, Ellis IO, et al. An evaluation of the clinical significance of FOXP3+ infiltrating cells in human breast cancer. Breast Cancer Res Treat. 2011;127(1):99-108.

57. Rech AJ, Mick R, Martin S, Recio A, Aqui NA, Powell DJ, Jr., et al. CD25 blockade depletes and selectively reprograms regulatory T cells in concert with immunotherapy in cancer patients. Sci Transl Med. 2012;4(134):134ra62.

58. Park JH, Jonas SF, Bataillon G, Criscitiello C, Salgado R, Loi S, et al. Prognostic value of tumor-infiltrating lymphocytes in patients with early-stage triple-negative breast cancers (TNBC) who did not receive adjuvant chemotherapy. Ann Oncol. 2019.

59. Leon-Ferre RA, Polley MY, Liu H, Gilbert JA, Cafourek V, Hillman DW, et al. Impact of histopathology, tumor-infiltrating lymphocytes, and adjuvant chemotherapy on prognosis of triple-negative breast cancer. Breast Cancer Res Treat. 2018;167(1):89-99.

60. Mori H, Kubo M, Yamaguchi R, Nishimura R, Osako T, Arima N, et al. The combination of PD-L1 expression and decreased tumor-infiltrating lymphocytes is associated with a poor prognosis in triple-negative breast cancer. Oncotarget. 2017.

61. Ruan M, Tian T, Rao J, Xu X, Yu B, Yang W, et al. Predictive value of tumor-infiltrating lymphocytes to pathological complete response in neoadjuvant treated triple-negative breast cancers. Diagn Pathol. 2018;13(1):66.

62. Telli ML, Chu C, Badve SS, Vinayak S, Silver DP, Isakoff SJ, et al. Association of Tumor Infiltrating Lymphocytes with Homologous Recombination Deficiency and BRCA1/2 Status in Patients with Early Triple-Negative Breast Cancer: A Pooled Analysis. Clin Cancer Res. 2019.

63. Ochi T, Bianchini G, Ando M, Nozaki F, Kobayashi D, Criscitiello C, et al. Predictive and prognostic value of stromal tumour-infiltrating lymphocytes before and after neoadjuvant therapy in triple negative and HER2-positive breast cancer. Eur J Cancer. 2019;118:41-8.

64. Criscitiello C, Azim HA, Jr., Schouten PC, Linn SC, Sotiriou C. Understanding the biology of triple-negative breast cancer. Ann Oncol. 2012;23 Suppl 6:vi13-8.

65. Stagg J, Allard B. Immunotherapeutic approaches in triple-negative breast cancer: latest research and clinical prospects. Ther Adv Med Oncol. 2013;5(3):169-81.

66. Turk AA, Wisinski KB. PARP inhibitors in breast cancer: Bringing synthetic lethality to the bedside. Cancer. 2018;124(12):2498-506.

67. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al.

Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. The Lancet. 2014;384(9938):164-72.

68. von Minckwitz G, Untch M, Nuesch E, Loibl S, Kaufmann M, Kummel S, et al. Impact of treatment characteristics on response of different breast cancer phenotypes: pooled analysis of the German neo-adjuvant chemotherapy trials. Breast Cancer Res Treat. 2011;125(1):145-56.

69. Sharma P. Biology and Management of Patients With Triple-Negative Breast Cancer. Oncologist. 2016;21(9):1050-62.

70. Poggio F, Bruzzone M, Ceppi M, Ponde NF, La Valle G, Del Mastro L, et al. Platinum-based neoadjuvant chemotherapy in triple-negative breast cancer: a systematic review and meta-analysis. Ann Oncol. 2018;29(7):1497-508.

71. Santarpia L, Bottai G, Kelly CM, Gyorffy B, Szekely B, Pusztai L. Deciphering and Targeting Oncogenic Mutations and Pathways in Breast Cancer. Oncologist. 2016;21(9):1063-78.

72. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.

73. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70.

74. Badve S, Dabbs DJ, Schnitt SJ, Baehner FL, Decker T, Eusebi V, et al. Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. Mod Pathol. 2011;24(2):157-67.

75. De Summa S, Pinto R, Sambiasi D, Petriella D, Paradiso V, Paradiso A, et al. BRCAness: a deeper insight into basal-like breast tumors. Ann Oncol. 2013;24 Suppl 8:viii13-viii21.

76. Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Peterse H, Bartelink H, et al. Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. Breast Cancer Res. 2007;9(5):R65.

77. Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer. 2016;16(2):110-20.

78. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res. 2008;14(16):5158-65.

79. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol. 2007;8(8):R157.

80. Callari M, Cappelletti V, D'Aiuto F, Musella V, Lembo A, Petel F, et al. Subtype-Specific Metagene-Based Prediction of Outcome after Neoadjuvant and Adjuvant Treatment in Breast Cancer. Clin Cancer Res. 2016;22(2):337-45.

81. Carey L, Winer E, Viale G, Cameron D, Gianni L. Triple-negative breast cancer: disease entity or title of convenience? Nat Rev Clin Oncol. 2010;7(12):683-92.

82. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346-52.

83. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121(7):2750-67.

84. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res. 2015;21(7):1688-98.

85. Gerratana L, Basile D, Buono G, De Placido S, Giuliano M, Minichillo S, et al. Androgen receptor in triple negative breast cancer: A potential target for the targetless subtype. Cancer Treat Rev. 2018;68:102-10.

86. Abraham JE VA-L, Qian W, Machin A, Grybowicz L, Thomas S, Weiss M, Harvey C, McAdam K, Hughes-Davies L, Roberts A, Provenzano E, Pinilla K, Roylance R, Copson E, Armstrong A, McMurtry E, Tischkowitz M, Earl HM. Abstract OT3-01-02: PARTNERING / PARTNER : Phase II sub-study to establish if the addition of combinations of new agents (olaparib, cell cycle and immune checkpoint inhibitors) can improve the rate of pathological complete response (pCR) and minimal residual disease (MRD) in triple negative breast cancer (TNBC) and / or germline BRCA mutated (gBRCAm) patients with evidence of residual disease after PARTNER therapy. Cancer Res. 2019;79:4 Suppl.

87. Chan TY, M.; Jaffee, E.; Swanton, C.; Quezada, S.; Stenzinger, A.; Peters, S. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Ann Oncol. 2019;30(1):44-56.

88. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012;486(7403):395-9.

89. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012;486(7403):405-9.

90. Hu X, Stern HM, Ge L, O'Brien C, Haydu L, Honchell CD, et al. Genetic alterations and oncogenic pathways associated with breast cancer subtypes. Mol Cancer Res. 2009;7(4):511-22.

91. Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci U S A. 2008;105(6):2070-5.

92. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell. 2013;152(4):714-26.

93. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47-54.

94. Schmid P, Abraham J, Chan S, Wheatley D, Brunt AM, Nemsadze G, et al. Capivasertib Plus Paclitaxel Versus Placebo Plus Paclitaxel As First-Line Therapy for Metastatic Triple-Negative Breast Cancer: The PAKT Trial. J Clin Oncol. 2020;38(5):423-33.

95. Massihnia D, Galvano A, Fanale D, Perez A, Castiglia M, Incorvaia L, et al. Triple negative breast cancer: shedding light onto the role of pi3k/akt/mtor pathway. Oncotarget. 2016;7(37):60712-22.

96. Lee JJ, Loh K, Yap YS. PI3K/Akt/mTOR inhibitors in breast cancer. Cancer Biol Med. 2015;12(4):342-54.

97. Lips EH, Michaut M, Hoogstraat M, Mulder L, Besselink NJ, Koudijs MJ, et al. Next generation sequencing of triple negative breast cancer to find predictors for chemotherapy response. Breast Cancer Res. 2015;17(1):134.

98. Saravia CH, Flores C, Schwarz LJ, Bravo L, Zavaleta J, Araujo J, et al. Patterns of Mutation Enrichment in Metastatic Triple-Negative Breast Cancer. Clinical Medicine Insights: Oncology. 2019;13:1179554919868482.

99. Yu KD, Zhu R, Zhan M, Rodriguez AA, Yang W, Wong S, et al. Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer. Clin Cancer Res. 2013;19(10):2723-33.

100. Balko JM, Giltnane JM, Wang K, Schwarz LJ, Young CD, Cook RS, et al. Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. Cancer Discov. 2014;4(2):232-45.

101. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. Nat Genet. 2016;48(10):1119-30.

102. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature. 2014;512(7513):155-60.

103. Vollebergh MA, Klijn C, Schouten PC, Wesseling J, Israeli D, Ylstra B, et al. Lack of genomic heterogeneity at high-resolution aCGH between primary breast cancers and their paired lymph node metastases. PLoS One. 2014;9(8):e103177.

104. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010;464(7291):999-1005.

105. Hoadley KA, Siegel MB, Kanchi KL, Miller CA, Ding L, Zhao W, et al. Tumor Evolution in Two Patients with Basal-like Breast Cancer: A Retrospective Genomics Study of Multiple Metastases. PLoS Med. 2016;13(12):e1002174.

106. Savas P, Teo ZL, Lefevre C, Flensburg C, Caramia F, Alsop K, et al. The Subclonal Architecture of Metastatic Breast Cancer: Results from a Prospective Community-Based Rapid Autopsy Program "CASCADE". PLoS Med. 2016;13(12):e1002204.

107. Massink MP, Kooi IE, van Mil SE, Jordanova ES, Ameziane N, Dorsman JC, et al. Proper genomic profiling of (BRCA1-mutated) basal-like breast carcinomas requires prior removal of tumor infiltrating lymphocytes. Mol Oncol. 2015;9(4):877-88.

108. Stoppa-Lyonnet D. The biological effects and clinical implications of BRCA mutations: where do we go from here? Eur J Hum Genet. 2016;24 Suppl 1:S3-9.

109. Savage KI, Harkin DP. BRCA1, a 'complex' protein involved in the maintenance of genomic stability. FEBS J. 2015;282(4):630-46.

110. Lips EH, Mulder L, Oonk A, van der Kolk LE, Hogervorst FB, Imholz AL, et al. Triple-negative breast cancer: BRCAness and concordance of clinical features with BRCA1-mutation carriers. Br J Cancer. 2013;108(10):2172-7.

111. Prat A, Cruz C, Hoadley KA, Diez O, Perou CM, Balmana J. Molecular features of the basal-like breast cancer subtype based on BRCA1 mutation status. Breast Cancer Res Treat. 2014;147(1):185-91.

112. Mylavarapu S, Das A, Roy M. Role of BRCA Mutations in the Modulation of Response to Platinum Therapy. Front Oncol. 2018;8:16.

113. Tutt A, Ellis P, Kilburn L, Gilett C, Pinder S, Abraham J, et al. The TNT trial: A randomized phase III trial of carboplatin (C) compared with docetaxel (D) for patients with metastatic or recurrent locally advanced triple negative or BRCA1/2 breast cancer (CRUK/07/012). Cancer Research. 2015;75.

114. Tutt A, Tovey H, Cheang MCU, Kernaghan S, Kilburn L, Gazinska P, et al. Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial. Nat Med. 2018;24(5):628-37.

115. Hahnen E, Lederer B, Hauke J, Loibl S, Krober S, Schneeweiss A, et al. Germline Mutation Status, Pathological Complete Response, and Disease-Free Survival in Triple-Negative Breast Cancer: Secondary Analysis of the GeparSixto Randomized Clinical Trial. JAMA Oncol. 2017;3(10):1378-85.

116. Pujade-Lauraine E, Ledermann JA, Selle F, Gebski V, Penson RT, Oza AM, et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. Lancet Oncol. 2017;18(9):1274-84.

117. Tucker H, Charles Z, Robertson J, Adam J. NICE guidance on olaparib for maintenance treatment of patients with relapsed, platinum-sensitive, BRCA mutation-positive ovarian cancer. Lancet Oncol. 2016;17(3):277-8.

118. Robson M, Im SA, Senkus E, Xu B, Domchek SM, Masuda N, et al. Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. N Engl J Med. 2017;377(6):523-33.

119. Patsouris A, Vicier C, Campion L, Gouraud W, Jimenez M, Pezzella V, et al. An open-label, phase II study of rucaparib, a PARP inhibitor, in HER2-metastatic breast cancer patients with high genomic loss of heterozygosity: RUBY. Journal of Clinical Oncology. 2017;35.

120. Niraparib for maintenance treatment of relapsed, platinum-sensitive ovarian, fallopian tube and peritoneal cancer: Nice Institute for Health and Care Excellence; 2018 [Available from: https://www.nice.org.uk/guidance/gid-ta10135/documents/appraisal-consultation-document.

121. Westdorp H, Fennemann FL, Weren RD, Bisseling TM, Ligtenberg MJ, Figdor CG, et al. Opportunities for immunotherapy in microsatellite instable colorectal cancer. Cancer Immunol Immunother. 2016;65(10):1249-59.

122. Nolan E, Savas P, Policheni AN, Darcy PK, Vaillant F, Mintoff CP, et al. Combined immune checkpoint blockade as a therapeutic strategy for BRCA1-mutated breast cancer. Sci Transl Med. 2017;9(393).

123. Vikas P, Borcherding N, Chennamadhavuni A, Garje R. Therapeutic Potential of Combining PARP Inhibitor and Immunotherapy in Solid Tumors. Front Oncol. 2020;10:570.

124. Solinas C, Marcoux D, Garaud S, Vitoria JR, Van den Eynden G, de Wind A, et al. BRCA gene mutations do not shape the extent and organization of tumor infiltrating lymphocytes in triple negative breast cancer. Cancer Lett. 2019;450:88-97.

125. Pantelidou C, Sonzogni O, De Oliveria Taveira M, Mehta AK, Kothari A, Wang D, et al. PARP Inhibitor Efficacy Depends on CD8<sup>+</sup> T-cell Recruitment via Intratumoral STING Pathway Activation in BRCA-Deficient Models of Triple-Negative Breast Cancer. Cancer Discovery. 2019;9(6):722-37.

126. Kraya AA, Maxwell KN, Wubbenhorst B, Wenz BM, Pluta J, Rech AJ, et al. Genomic Signatures Predict the Immunogenicity of BRCA-Deficient Breast Cancer. Clin Cancer Res. 2019;25(14):4363-74.

127. Pier G. Antigens, Antigenicity, and Immunogenicity. Immunology, Infection, and Immunity2004. p. 185-205.

128. Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB, et al. MHC-I Genotype Restricts the Oncogenic Mutational Landscape. Cell. 2017;171(6):1272-83 e15.

129. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat Biotechnol. 2015;33(11):1152-8.

130. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell. 2017;171(6):1259-71 e11.

131. Forero A, Li Y, Chen D, Grizzle WE, Updike KL, Merz ND, et al. Expression of the MHC Class II Pathway in Triple-Negative Breast Cancer Tumor Cells Is Associated with a Good Prognosis and Infiltrating Lymphocytes. Cancer Immunol Res. 2016;4(5):390-9.

132. Park IA, Hwang SH, Song IH, Heo SH, Kim YA, Bang WS, et al. Expression of the MHC class II in triple-negative breast cancer is associated with tumor-infiltrating lymphocytes and interferon signaling. PLoS One. 2017;12(8):e0182786.

133. Pedersen MH, Hood BL, Beck HC, Conrads TP, Ditzel HJ, Leth-Larsen R. Downregulation of antigen presentation-associated pathway proteins is linked to poor outcome in triple-negative breast cancer patient tumors. Oncoimmunology. 2017;6(5):e1305531.

134. Melsted WN, Johansen LL, Lock- Andersen J, Behrendt N, Eriksen JO, Bzorek

M, et al. HLA class Ia and Ib molecules and FOXP3+ TILs in relation to the prognosis of malignant melanoma patients. Clin Immunol. 2017;183:191-7.

135. Simpson JA, Al-Attar A, Watson NF, Scholefield JH, Ilyas M, Durrant LG. Intratumoral T cell infiltration, MHC class I and STAT1 as biomarkers of good prognosis in colorectal cancer. Gut. 2010;59(7):926-33.

136. Zhang J, Xu Z, Zhou X, Zhang H, Yang N, Wu Y, et al. Loss of expression of MHC class I-related chain A (MICA) is a frequent event and predicts poor survival in patients with hepatocellular carcinoma. Int J Clin Exp Pathol. 2014;7(6):3123-31.

137. Braunlein E, Krackhardt AM. Tools to define the melanoma-associated immunopeptidome. Immunology. 2017;152(4):536-44.

138. Chang SC, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. Proc Natl Acad Sci U S A. 2005;102(47):17107-12.

139. Stratikos E, Stamogiannos A, Zervoudi E, Fruci D. A role for naturally occurring alleles of endoplasmic reticulum aminopeptidases in tumor immunity and cancer pre-disposition. Front Oncol. 2014;4:363.

140. Reeves E, Edwards CJ, Elliott T, James E. Naturally occurring ERAP1 haplotypes encode functionally distinct alleles with fine substrate specificity. J Immunol. 2013;191(1):35-43.

141. Hammer GE, Gonzalez F, James E, Nolla H, Shastri N. In the absence of aminopeptidase ERAAP, MHC class I molecules present many unstable and highly immunogenic peptides. Nat Immunol. 2007;8(1):101-8.

142. James E, Bailey I, Sugiyarto G, Elliott T. Induction of protective antitumor immunity through attenuation of ERAAP function. J Immunol. 2013;190(11):5839-46.

143. Wang B, Niu D, Lai L, Ren EC. p53 increases MHC class I expression by upregulating the endoplasmic reticulum aminopeptidase ERAP1. Nat Commun. 2013;4:2359.

144. Vonderheide RH, Domchek SM, Clark AS. Immunotherapy for Breast Cancer: What Are We Missing? Clin Cancer Res. 2017;23(11):2640-6.

145. Brown SD, Warren RL, Gibb EA, Martin SD, Spinelli JJ, Nelson BH, et al. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res. 2014;24(5):743-50.

146. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med. 2014;371(23):2189-99.

147. Dituri F, Mazzocca A, Giannelli G, Antonaci S. PI3K functions in cancer progression, anticancer immunity and immune evasion by tumors. Clin Dev Immunol. 2011;2011:947858.

148. Safonov A, Jiang T, Bianchini G, Gyorffy B, Karn T, Hatzis C, et al. Immune Gene

Expression Is Associated with Genomic Aberrations in Breast Cancer. Cancer Res. 2017;77(12):3317-24.

149. Yeong J, Lim JCT, Lee B, Li H, Ong CCH, Thike AA, et al. Prognostic value of CD8 + PD-1+ immune infiltrates and PDCD1 gene expression in triple negative breast cancer. J Immunother Cancer. 2019;7(1):34.

150. Frydenlund N, Mahalingam M. PD-L1 and immune escape: insights from melanoma and other lineage-unrelated malignancies. Hum Pathol. 2017;66:13-33.

151. Wimberly H, Brown JR, Schalper K, Haack H, Silver MR, Nixon C, et al. PD-L1 Expression Correlates with Tumor-Infiltrating Lymphocytes and Response to Neoadjuvant Chemotherapy in Breast Cancer. Cancer Immunol Res. 2015;3(4):326-32.

152. Kitano A, Ono M, Yoshida M, Noguchi E, Shimomura A, Shimoi T, et al. Tumour-infiltrating lymphocytes are correlated with higher expression levels of PD-1 and PD-L1 in early breast cancer. ESMO Open. 2017;2(2):e000150.

153. Bae SB, Cho HD, Oh MH, Lee JH, Jang SH, Hong SA, et al. Expression of Programmed Death Receptor Ligand 1 with High Tumor-Infiltrating Lymphocytes Is Associated with Better Prognosis in Breast Cancer. J Breast Cancer. 2016;19(3):242-51.

154. Beckers RK, Selinger CI, Vilain R, Madore J, Wilmott JS, Harvey K, et al. Programmed death ligand 1 expression in triple-negative breast cancer is associated with tumour-infiltrating lymphocytes and improved outcome. Histopathology. 2016;69(1):25-34.

155. Egelston CA, Avalos C, Tu TY, Simons DL, Jimenez G, Jung JY, et al. Human breast tumor-infiltrating CD8(+) T cells retain polyfunctionality despite PD-1 expression. Nat Commun. 2018;9(1):4297.

156. Schmid P, Adams S, Rugo HS, Schneeweiss A, Barrios CH, Iwata H, et al. Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. N Engl J Med. 2018;379(22):2108-21.

157. Vikas P, Borcherding N, Zhang W. The clinical promise of immunotherapy in triple-negative breast cancer. Cancer Manag Res. 2018;10:6823-33.

158. Kwa MJ, Adams S. Checkpoint inhibitors in triple-negative breast cancer (TNBC): Where to go from here. Cancer. 2018.

159. Garcia-Teijido P, Cabal ML, Fernandez IP, Perez YF. Tumor-Infiltrating Lymphocytes in Triple Negative Breast Cancer: The Future of Immune Targeting. Clin Med Insights Oncol. 2016;10(Suppl 1):31-9.

160. Loi S, Adams S, Schmid P, Cortes J, Cescon DW, Winer EP, et al. Relationship between tumor infiltrating lymphocyte (TIL) levels and response to pembrolizumab (pembro) in metastatic triple-negative breast cancer (mTNBC): Results from KEYNOTE-086. Annals of Oncology. 2017;28.

161. Schmid P, Salgado R, Park YH, Munoz-Couselo E, Kim SB, Sohn J, et al. 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. Ann Oncol. 2020;31(5):569-81.

162. P. Schmid JC, R. Dent, L. Pusztai, H.L. McArthur, S. Kuemmel, J. Bergh, C. Denkert, Y.H. Park, R. Hui, N. Harbeck, M. Takahashi, T. Foukakis, P.A. Fasching, F. Cardoso, L. Jia, V. Karantza, J. Zhao, G. Aktan, J. O'Shaughnessy. KEYNOTE-522: Phase 3 study of pembrolizumab (pembro) + chemotherapy (chemo) vs placebo (pbo) + chemo as neoadjuvant treatment. ESMO 2019 Congress: Annals of Oncology; 2019. p. v851-v934.

163. Milani A, Sangiolo D, Aglietta M, Valabrega G. Recent advances in the development of breast cancer vaccines. Breast Cancer (Dove Med Press). 2014;6:159-68.

164. June CH, O'Connor RS, Kawalekar OU, Ghassemi S, Milone MC. CAR T cell immunotherapy for human cancer. Science. 2018;359(6382):1361-5.

165. Luen SJ, Savas P, Fox SB, Salgado R, Loi S. Tumour-infiltrating lymphocytes and the emerging role of immunotherapy in breast cancer. Pathology. 2017;49(2):141-55.

166. Allaoui R, Bergenfelz C, Mohlin S, Hagerling C, Salari K, Werb Z, et al. Cancer-associated fibroblast-secreted CXCL16 attracts monocytes to promote stroma activation in triple-negative breast cancers. Nat Commun. 2016;7:13050.

167. Lakins MA, Ghorani E, Munir H, Martins CP, Shields JD. Cancer-associated fibroblasts induce antigen-specific deletion of CD8 (+) T Cells to protect tumour cells. Nat Commun. 2018;9(1):948.

168. Roswall P, Bocci M, Bartoschek M, Li H, Kristiansen G, Jansson S, et al. Microenvironmental control of breast cancer subtype elicited through paracrine platelet-derived growth factor-CC signaling. Nature Medicine. 2018;24:463.

169. Bareche Y, Buisseret L, Gruosso T, Girard E, Venet D, Dupont F, et al. Unraveling triple-negative breast cancer tumor microenvironment heterogeneity: towards an optimized treatment approach. J Natl Cancer Inst. 2019.

170. Gruosso T, Gigoux M, Manem VSK, Bertos N, Zuo D, Perlitch I, et al. Spatially distinct tumor immune microenvironments stratify triple-negative breast cancers. J Clin Invest. 2019;129(4):1785-800.

171. Buchsbaum RJ, Oh SY. Breast Cancer-Associated Fibroblasts: Where We Are and Where We Need to Go. Cancers (Basel). 2016;8(2).

172. Kalluri R, Zeisberg M. Fibroblasts in cancer. Nat Rev Cancer. 2006;6(5):392-401.

173. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, et al. Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. Cancer Cell. 2018;33(3):463-79 e10.

174. Shu DY, Lovicu FJ. Myofibroblast transdifferentiation: The dark force in ocular wound healing and fibrosis. Prog Retin Eye Res. 2017;60:44-65.

175. Takahashi H, Sakakura K, Kudo T, Toyoda M, Kaira K, Oyama T, et al. Cancer-associated fibroblasts promote an immunosuppressive microenvironment through the induction and accumulation of protumoral macrophages. Oncotarget. 2017;8(5):8633-47.

176. Bartoschek M, Oskolkov N, Bocci M, Lovrot J, Larsson C, Sommarin M, et al. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts

revealed by single cell RNA sequencing. Nat Commun. 2018;9(1):5150.

177. Jansson S, Aaltonen K, Bendahl PO, Falck AK, Karlsson M, Pietras K, et al. The PDGF pathway in breast cancer is linked to tumour aggressiveness, triple-negative subtype and early recurrence. Breast Cancer Res Treat. 2018;169(2):231-41.

178. Calon A, Tauriello DV, Batlle E. TGF-beta in CAF-mediated tumor growth and metastasis. Semin Cancer Biol. 2014;25:15-22.

179. Kraman M, Bambrough PJ, Arnold JN, Roberts EW, Magiera L, Jones JO, et al. Suppression of antitumor immunity by stromal cells expressing fibroblast activation protein-alpha. Science. 2010;330(6005):827-30.

180. Liao D, Luo Y, Markowitz D, Xiang R, Reisfeld RA. Cancer associated fibroblasts promote tumor growth and metastasis by modulating the tumor immune microenvironment in a 4T1 murine breast cancer model. PLoS One. 2009;4(11):e7965.

181. Hofheinz RD, al-Batran SE, Hartmann F, Hartung G, Jager D, Renner C, et al. Stromal antigen targeting by a humanised monoclonal antibody: an early phase II trial of sibrotuzumab in patients with metastatic colorectal cancer. Onkologie. 2003;26(1):44-8.

182. Lacouture ME, Morris JC, Lawrence DP, Tan AR, Olencki TE, Shapiro GI, et al. Cutaneous keratoacanthomas/squamous cell carcinomas associated with neutralization of transforming growth factor beta by the monoclonal antibody fresolimumab (GC1008). Cancer Immunol Immunother. 2015;64(4):437-46.

183. Hanley CJ, Mellone M, Ford K, Thirdborough SM, Mellows T, Frampton SJ, et al. Targeting the Myofibroblastic Cancer-Associated Fibroblast Phenotype Through Inhibition of NOX4. J Natl Cancer Inst. 2018;110(1).

184. Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. Nat Rev Drug Discov. 2019;18(2):99-115.

185. Calon A, Espinet E, Palomo-Ponce S, Tauriello DV, Iglesias M, Cespedes MV, et al. Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. Cancer Cell. 2012;22(5):571-84.

186. Feig C, Jones JO, Kraman M, Wells RJ, Deonarine A, Chan DS, et al. Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. Proc Natl Acad Sci U S A. 2013;110(50):20212-7.

187. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018;554(7693):544-8.

188. Tauriello DVF, Palomo-Ponce S, Stork D, Berenguer-Llergo A, Badia-Ramentol J, Iglesias M, et al. TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. Nature. 2018;554(7693):538-43.

189. Marechal A, Zou L. DNA damage sensing by the ATM and ATR kinases. Cold Spring Harb Perspect Biol. 2013;5(9).

190. Chen WT, Ebelt ND, Stracker TH, Xhemalce B, Van Den Berg CL, Miller KM.

ATM regulation of IL-8 links oxidative stress to cancer cell migration and invasion. Elife. 2015;4.

191. Carrassa L, Damia G. DNA damage response inhibitors: Mechanisms and potential applications in cancer therapy. Cancer Treat Rev. 2017;60:139-51.

192. Overstreet JM, Samarakoon R, Cardona-Grau D, Goldschmeding R, Higgins PJ. Tumor suppressor ataxia telangiectasia mutated functions downstream of TGF-beta1 in orchestrating profibrotic responses. FASEB J. 2015;29(4):1258-68.

193. Guo Z, Kozlov S, Lavin MF, Person MD, Paull TT. ATM activation by oxidative stress. Science. 2010;330(6003):517-21.

194. Svegliati S, Marrone G, Pezone A, Spadoni T, Grieco A, Moroncini G, et al. Oxidative DNA damage induces the ATM-mediated transcriptional suppression of the Wnt inhibitor WIF-1 in systemic sclerosis and fibrosis. Sci Signal. 2014;7(341):ra84.

195. Mellone M, Hanley CJ, Thirdborough S, Mellows T, Garcia E, Woo J, et al. Induction of fibroblast senescence generates a non-fibrogenic myofibroblast phenotype that differentially impacts on cancer prognosis. Aging (Albany NY). 2016;9(1):114-32.

196. Copson E, Eccles B, Maishman T, Gerty S, Stanton L, Cutress RI, et al. Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study. J Natl Cancer Inst. 2013;105(13):978-88.

197. Eccles D, Gerty S, Simmonds P, Hammond V, Ennis S, Altman DG, et al. Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. BMC Cancer. 2007;7:160.

198. Copson E, Maishman T, Gerty S, Eccles B, Stanton L, Cutress RI, et al. Ethnicity and outcome of young breast cancer patients in the United Kingdom: the POSH study. Br J Cancer. 2014;110(1):230-41.

199. Eccles BK, Copson ER, Cutress RI, Maishman T, Altman DG, Simmonds P, et al. Family history and outcome of young patients with breast cancer in the UK (POSH study). Br J Surg. 2015;102(8):924-35.

200. Copson ER, Cutress RI, Maishman T, Eccles BK, Gerty S, Stanton L, et al. Obesity and the outcome of young breast cancer patients in the UK: the POSH study. Ann Oncol. 2015;26(1):101-12.

201. Copson ER, Maishman TC, Tapper WJ, Cutress RI, Greville-Heygate S, Altman DG, et al. Germline BRCA mutation and outcome in young-onset breast cancer (POSH): a prospective cohort study. Lancet Oncol. 2018.

202. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res. 2007;13(15 Pt 1):4429-34.

203. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344-9.

204. Pogoda K, Niwinska A, Murawska M, Pienkowski T. Analysis of pattern, time and risk factors influencing recurrence in triple-negative breast cancer patients. Med Oncol. 2013;30(1):388.

205. Lee SK, Bae SY, Lee JH, Lee HC, Yi H, Kil WH, et al. Distinguishing Low-Risk Luminal A Breast Cancer Subtypes with Ki-67 and p53 Is More Predictive of Long-Term Survival. PLoS One. 2015;10(8):e0124658.

206. Levva S, Kotoula V, Kostopoulos I, Manousou K, Papadimitriou C, Papadopoulou K, et al. Prognostic Evaluation of Epidermal Growth Factor Receptor (EGFR) Genotype and Phenotype Parameters in Triple-negative Breast Cancers. Cancer Genomics Proteomics. 2017;14(3):181-95.

207. Alsner J, Jensen V, Kyndi M, Offersen BV, Vu P, Borresen-Dale AL, et al. A comparison between p53 accumulation determined by immunohistochemistry and TP53 mutations as prognostic variables in tumours from breast cancer patients. Acta Oncol. 2008;47(4):600-7.

208. Bustreo S, Osella-Abate S, Cassoni P, Donadio M, Airoldi M, Pedani F, et al. Optimal Ki67 cut-off for luminal breast cancer prognostic evaluation: a large case series study with a long-term follow-up. Breast Cancer Res Treat. 2016;157(2):363-71.

209. Muftah AA, Aleskandarany MA, Al-Kaabi MM, Sonbul SN, Diez-Rodriguez M, Nolan CC, et al. Ki67 expression in invasive breast cancer: the use of tissue microarrays compared with whole tissue sections. Breast Cancer Res Treat. 2017;164(2):341-8.

210. Ohara M, Matsuura K, Akimoto E, Noma M, Doi M, Nishizaka T, et al. Prognostic value of Ki67 and p53 in patients with estrogen receptor-positive and human epidermal growth factor receptor 2-negative breast cancer: Validation of the cut-off value of the Ki67 labeling index as a predictive factor. Mol Clin Oncol. 2016;4(4):648-54.

211. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res. 2008;14(5):1368-76.

212. Banerjee S, Reis-Filho JS, Ashley S, Steele D, Ashworth A, Lakhani SR, et al. Basal-like breast carcinomas: clinical outcome and response to chemotherapy. J Clin Pathol. 2006;59(7):729-35.

213. Ricardo S, Vieira AF, Gerhard R, Leitao D, Pinto R, Cameselle-Teijeiro JF, et al. Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype. J Clin Pathol. 2011;64(11):937-46.

214. Resetkova E, Reis-Filho JS, Jain RK, Mehta R, Thorat MA, Nakshatri H, et al. Prognostic impact of ALDH1 in breast cancer: a story of stem cells and tumor microenvironment. Breast Cancer Res Treat. 2010;123(1):97-108.

215. Viale G, Rotmensz N, Maisonneuve P, Bottiglieri L, Montagna E, Luini A, et al. Invasive ductal carcinoma of the breast with the "triple-negative" phenotype: prognostic implications of EGFR immunoreactivity. Breast Cancer Res Treat. 2009;116(2):317-28.

216. Yue Y, Astvatsaturyan K, Cui X, Zhang X, Fraass B, Bose S. Stratification of Prognosis

of Triple-Negative Breast Cancer Patients Using Combinatorial Biomarkers. PLoS One. 2016;11(3):e0149661.

217. Scaglioni D, Ellis M, Catapano F, Torelli S, Chambers D, Feng L, et al. A high-throughput digital script for multiplexed immunofluorescent analysis and quantification of sarcolemmal and sarcomeric proteins in muscular dystrophies. Acta Neuropathol Commun. 2020;8(1):53.

218. Ottensmeier CH, Perry KL, Harden EL, Stasakova J, Jenei V, Fleming J, et al. Upregulated Glucose Metabolism Correlates Inversely with CD8+ T-cell Infiltration and Survival in Squamous Cell Carcinoma. Cancer Res. 2016;76(14):4136-48.

219. Heng TS, Painter MW, Immunological Genome Project C. The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol. 2008;9(10):1091-4.

220. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18(1):220.

221. Pulaski BA, Ostrand-Rosenberg S. Mouse 4T1 breast tumor model. Curr Protoc Immunol. 2001;Chapter 20:Unit 20 2.

222. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671-5.

223. Sapi J, Kovacs L, Drexler DA, Kocsis P, Gajari D, Sapi Z. Tumor Volume Estimation and Quasi-Continuous Administration for Most Effective Bevacizumab Therapy. PLoS One. 2015;10(11):e0142190.

224. Liedtke C, Hess KR, Karn T, Rody A, Kiesel L, Hortobagyi GN, et al. The prognostic impact of age in patients with triple-negative breast cancer. Breast Cancer Res Treat. 2013;138(2):591-9.

225. Brandt J, Garne JP, Tengrup I, Manjer J. Age at diagnosis in relation to survival following breast cancer: a cohort study. World J Surg Oncol. 2015;13:33.

226. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Res Treat. 1992;22(3):207-19.

227. McKenzie HS, Maishman T, Simmonds P, Durcan L, Group PS, Eccles D, et al. Survival and disease characteristics of de novo versus recurrent metastatic breast cancer in a cohort of young patients. Br J Cancer. 2020.

228. Service NCRaA. Stage breakdown by CCG 2019 [updated February 2019. Available from: http://www.ncin.org.uk/publications/survival_by_stage.

229. SEER Cancer Statistics Review [Internet]. 1975-2012. Available from: https://seer.cancer.gov/archive/csr/1975_2012/.

230. Fredholm H, Eaker S, Frisell J, Holmberg L, Fredriksson I, Lindman H. Breast cancer in young women: poor survival despite intensive treatment. PLoS One. 2009;4(11):e7695.

231. Gnerlich JL, Deshpande AD, Jeffe DB, Sweet A, White N, Margenthaler JA. Elevated

breast cancer mortality in women younger than age 40 years compared with older women is attributed to poorer survival in early-stage disease. J Am Coll Surg. 2009;208(3):341-7.

232. Kheirelseid EH, Boggs JM, Curran C, Glynn RW, Dooley C, Sweeney KJ, et al. Younger age as a prognostic indicator in breast cancer: a cohort study. BMC Cancer. 2011;11:383.

233. Wittersheim M, Buttner R, Markiefka B. Genotype/Phenotype correlations in patients with hereditary breast cancer. Breast Care (Basel). 2015;10(1):22-6.

234. Loman N, Johannsson O, Bendahl P, Dahl N, Einbeigi Z, Gerdes A, et al. Prognosis and clinical presentation of BRCA2-associated breast cancer. Eur J Cancer. 2000;36(11):1365-73.

235. Chan DS, Vieira AR, Aune D, Bandera EV, Greenwood DC, McTiernan A, et al. Body mass index and survival in women with breast cancer-systematic literature review and meta-analysis of 82 follow-up studies. Ann Oncol. 2014;25(10):1901-14.

236. Turkoz FP, Solak M, Petekkaya I, Keskin O, Kertmen N, Sarici F, et al. The prognostic impact of obesity on molecular subtypes of breast cancer in premenopausal women. J BUON. 2013;18(2):335-41.

237. Rose DP, Gracheck PJ, Vona-Davis L. The Interactions of Obesity, Inflammation and Insulin Resistance in Breast Cancer. Cancers (Basel). 2015;7(4):2147-68.

238. Early Breast Cancer Trialists' Collaborative G, Darby S, McGale P, Correa C, Taylor C, Arriagada R, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. Lancet. 2011;378(9804):1707-16.

239. Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. Transl Cancer Res. 2015;4(3):256-69.

240. Silwal-Pandit L, Vollan HK, Chin SF, Rueda OM, McKinney S, Osako T, et al. TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. Clin Cancer Res. 2014;20(13):3569-80.

241. Penault-Llorca F, Radosevic-Robin N. Ki67 assessment in breast cancer: an update. Pathology. 2017;49(2):166-71.

242. Abdelrahman AE, Rashed HE, Abdelgawad M, Abdelhamid MI. Prognostic impact of EGFR and cytokeratin 5/6 immunohistochemical expression in triple-negative breast cancer. Ann Diagn Pathol. 2017;28:43-53.

243. Adamo B, Ricciardi GRR, Ieni A, Franchina T, Fazzari C, Sano MV, et al. The prognostic significance of combined androgen receptor, E-Cadherin, Ki67 and CK5/6 expression in patients with triple negative breast cancer. Oncotarget. 2017;8(44):76974-86.

244. Pan Y, Yuan Y, Liu G, Wei Y. P53 and Ki-67 as prognostic markers in triple-negative breast cancer patients. PLoS One. 2017;12(2):e0172324.

245. Lee JS, Kim WG. Potential protective effect of ALDH-1 stromal expression

against early recurrence of operable breast cancers. Stem Cell Investig. 2018;5:43.

246. Anders CK, Acharya CR, Hsu DS, Broadwater G, Garman K, Foekens JA, et al. Age-specific differences in oncogenic pathway deregulation seen in human breast tumors. PLoS One. 2008;3(1):e1373.

247. Kim HK, Park KH, Kim Y, Park SE, Lee HS, Lim SW, et al. Discordance of the PAM50 Intrinsic Subtypes Compared with Immunohistochemistry-Based Surrogate in Breast Cancer Patients: Potential Implication of Genomic Alterations of Discordance. Cancer Res Treat. 2019;51(2):737-47.

248. Elsawaf Z, Sinn HP, Rom J, Bermejo JL, Schneeweiss A, Aulmann S. Biological subtypes of triple-negative breast cancer are associated with distinct morphological changes and clinical behaviour. Breast. 2013;22(5):986-92.

249. Ilie SM, Bacinschi XE, Botnariuc I, Anghel RM. Potential clinically useful prognostic biomarkers in triple-negative breast cancer: preliminary results of a retrospective analysis. Breast Cancer (Dove Med Press). 2018;10:177-94.

250. Wang W, Wu J, Zhang P, Fei X, Zong Y, Chen X, et al. Prognostic and predictive value of Ki-67 in triple-negative breast cancer. Oncotarget. 2016;7(21):31079-87.

251. de Beca FF, Caetano P, Gerhard R, Alvarenga CA, Gomes M, Paredes J, et al. Cancer stem cells markers CD44, CD24 and ALDH1 in breast cancer special histological types. J Clin Pathol. 2013;66(3):187-91.

252. Tsang JY, Huang YH, Luo MH, Ni YB, Chan SK, Lui PC, et al. Cancer stem cell markers are associated with adverse biomarker profiles and molecular subtypes of breast cancer. Breast Cancer Res Treat. 2012;136(2):407-17.

253. Demir H, Dulgar O, Gulle BT, Turna H, Ilvan S. Prognostic value of aldehyde dehydrogenase 1 (ALDH1) in invasive breast carcinomas. Bosn J Basic Med Sci. 2018;18(4):313-9.

254. Collina F, Di Bonito M, Li Bergolis V, De Laurentiis M, Vitagliano C, Cerrone M, et al. Prognostic Value of Cancer Stem Cells Markers in Triple-Negative Breast Cancer. Biomed Res Int. 2015;2015:158682.

255. Kida K, Ishikawa T, Yamada A, Shimada K, Narui K, Sugae S, et al. Effect of ALDH1 on prognosis and chemoresistance by breast cancer subtype. Breast Cancer Res Treat. 2016;156(2):261-9.

256. Ma F, Li H, Li Y, Ding X, Wang H, Fan Y, et al. Aldehyde dehydrogenase 1 (ALDH1) expression is an independent prognostic factor in triple negative breast cancer (TNBC). Medicine (Baltimore). 2017;96(14):e6561.

257. Yang F, Cao L, Sun Z, Jin J, Fang H, Zhang W, et al. Evaluation of Breast Cancer Stem Cells and Intratumor Stemness Heterogeneity in Triple-negative Breast Cancer as Prognostic Factors. Int J Biol Sci. 2016;12(12):1568-77.

258. Jin MS, Park IA, Kim JY, Chung YR, Im SA, Lee KH, et al. New insight on the biological role of p53 protein as a tumor suppressor: re-evaluation of its clinical significance in triple-negative breast cancer. Tumour Biol. 2016;37(8):11017-24.

259. Kim MC, Choi JE, Lee SJ, Bae YK. Coexistent Loss of the Expressions of BRCA1 and p53 Predicts Poor Prognosis in Triple-Negative Breast Cancer. Ann Surg Oncol. 2016;23(11):3524-30.

260. Louhichi T, Ziadi S, Saad H, Dhiab MB, Mestiri S, Trimeche M. Clinicopathological significance of cancer stem cell markers CD44 and ALDH1 expression in breast cancer. Breast Cancer. 2018;25(6):698-705.

261. Kim S, Moon BI, Lim W, Park S, Cho MS, Sung SH. Feasibility of Classification of Triple Negative Breast Cancer by Immunohistochemical Surrogate Markers. Clin Breast Cancer. 2018;18(5):e1123-e32.

262. Prat A, Pineda E, Adamo B, Galvan P, Fernandez A, Gaba L, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast. 2015;24 Suppl 2:S26-35.

263. Lakhani SR, Jacquemier J, Sloane JP, Gusterson BA, Anderson TJ, van de Vijver MJ, et al. Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. J Natl Cancer Inst. 1998;90(15):1138-45.

264. Jiang D, Gao Z, Cai Z, Wang M, He J. Clinicopathological and prognostic significance of FOXP3+ tumor infiltrating lymphocytes in patients with breast cancer: a meta-analysis. BMC Cancer. 2015;15:727.

265. Steele KE, Brown C. Multiplex Immunohistochemistry for Image Analysis of Tertiary Lymphoid Structures in Cancer. Methods Mol Biol. 2018;1845:87-98.

266. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. Immunity. 2018;48(4):812-30 e14.

267. Pawelec G. Does patient age influence anti-cancer immunity? Semin Immunopathol. 2019;41(1):125-31.

268. Shao H, Ou Y, Wang T, Shen H, Wu F, Zhang W, et al. Differences in TCR-Vbeta repertoire and effector phenotype between tumor infiltrating lymphocytes and peripheral blood lymphocytes increase with age. PLoS One. 2014;9(7):e102327.

269. Yoshida K, Cologne JB, Cordova K, Misumi M, Yamaoka M, Kyoizumi S, et al. Aging-related changes in human T-cell repertoire over 20years delineated by deep sequencing of peripheral T-cell receptors. Exp Gerontol. 2017;96:29-37.

270. Jergovic M, Smithey MJ, Nikolich-Zugich J. Intrinsic and extrinsic contributors to defective CD8+ T cell responses with aging. Exp Gerontol. 2018;105:140-5.

271. Castro F, Leal B, Denny A, Bahar R, Lampkin S, Reddick R, et al. Vaccination with Mage-b DNA induces CD8 T-cell responses at young but not old age in mice with metastatic breast cancer. Br J Cancer. 2009;101(8):1329-37.

272. Liu F, Lang R, Zhao J, Zhang X, Pringle GA, Fan Y, et al. CD8(+) cytotoxic T cell and FOXP3(+) regulatory T cell infiltration in relation to breast cancer survival and molecular subtypes. Breast Cancer Res Treat. 2011;130(2):645-55.

273. Yeong J, Thike AA, Lim JC, Lee B, Li H, Wong SC, et al. Higher densities of Foxp3(+) regulatory T cells are associated with better prognosis in triple-negative breast

cancer. Breast Cancer Res Treat. 2017;163(1):21-35.

274. Salama P, Phillips M, Grieu F, Morris M, Zeps N, Joseph D, et al. Tumor-infiltrating FOXP3+ T regulatory cells show strong prognostic significance in colorectal cancer. J Clin Oncol. 2009;27(2):186-92.

275. Ladoire S, Martin F, Ghiringhelli F. Prognostic role of FOXP3+ regulatory T cells infiltrating human carcinomas: the paradox of colorectal cancer. Cancer Immunol Immunother. 2011;60(7):909-18.

276. Saito T, Nishikawa H, Wada H, Nagano Y, Sugiyama D, Atarashi K, et al. Two FOXP3(+)CD4(+) T cell subpopulations distinctly control the prognosis of colorectal cancers. Nat Med. 2016;22(6):679-84.

277. Ionescu DN, Downes MR, Christofides A, Tsao MS. Harmonization of PD-L1 testing in oncology: a Canadian pathology perspective. Curr Oncol. 2018;25(3):e209-e16.

278. Festino L, Botti G, Lorigan P, Masucci GV, Hipp JD, Horak CE, et al. Cancer Treatment with Anti-PD-1/PD-L1 Agents: Is PD-L1 Expression a Biomarker for Patient Selection? Drugs. 2016;76(9):925-45.

279. Koomen BM, Badrising SK, van den Heuvel MM, Willems SM. Comparability of PD-L1 immunohistochemistry assays for non-small cell lung cancer: a systematic review. Histopathology. 2019.

280. Volonte A, Di Tomaso T, Spinelli M, Todaro M, Sanvito F, Albarello L, et al. Cancer-initiating cells from colorectal cancer patients escape from T cell-mediated immunosurveillance in vitro through membrane-bound IL-4. J Immunol. 2014;192(1):523-32.

281. Masili-Oku SM, Almeida BGL, Bacchi CE, Filassi JR, Baracat EC, Carvalho FM. Lymphocyte-predominant triple-negative breast carcinomas in premenopausal patients: Lower expression of basal immunohistochemical markers. Breast. 2017;31:34-9.

282. Lee M, Park IA, Heo SH, Kim YA, Gong G, Lee HJ. Association between p53 Expression and Amount of Tumor-Infiltrating Lymphocytes in Triple-Negative Breast Cancer. J Pathol Transl Med. 2019;53(3):180-7.

283. Griffin J, Treanor D. Digital pathology in clinical use: where are we now and what is holding us back? Histopathology. 2017;70(1):134-45.

284. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol. 2010;28(16):2784-95.

285. Jensen K, Krusenstjerna-Hafstrom R, Lohse J, Petersen KH, Derand H. A novel quantitative immunohistochemistry method for precise protein measurements directly in formalin-fixed, paraffin-embedded specimens: analytical performance measuring HER2. Mod Pathol. 2017;30(2):180-93.

286. Paulsson J, Micke P. Prognostic relevance of cancer-associated fibroblasts in human cancer. Semin Cancer Biol. 2014;25:61-8.

287. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med. 2008;14(5):518-27.

288. Hu B, Wu Z, Phan SH. Smad3 mediates transforming growth factor-beta-induced alpha-smooth muscle actin expression. Am J Respir Cell Mol Biol. 2003;29(3 Pt 1):397-404.

289. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as treatment targets in oncology. Nat Rev Clin Oncol. 2017;14(7):399-416.

290. Wynn TA, Barron L. Macrophages: master regulators of inflammation and fibrosis. Semin Liver Dis. 2010;30(3):245-57.

291. Qiu SQ, Waaijer SJH, Zwager MC, de Vries EGE, van der Vegt B, Schroder CP. Tumor-associated macrophages in breast cancer: Innocent bystander or important player? Cancer Treat Rev. 2018;70:178-89.

292. Arroyo-Crespo JJ, Arminan A, Charbonnier D, Deladriere C, Palomino-Schatzlein M, Lamas-Domingo R, et al. Characterization of triple-negative breast cancer preclinical models provides functional evidence of metastatic progression. Int J Cancer. 2019;145(8):2267-81.

293. Stuelten CH, DaCosta Byfield S, Arany PR, Karpova TS, Stetler-Stevenson WG, Roberts AB. Breast cancer cells induce stromal fibroblasts to express MMP-9 via secretion of TNF-alpha and TGF-beta. J Cell Sci. 2005;118(Pt 10):2143-53.

294. Stuelten CH, Busch JI, Tang B, Flanders KC, Oshima A, Sutton E, et al. Transient tumor-fibroblast interactions increase tumor cell malignancy by a TGF-Beta mediated mechanism in a mouse xenograft model of breast cancer. PLoS One. 2010;5(3):e9832.

295. Evans A, Sim YT, Pourreyron C, Thompson A, Jordan L, Fleming D, et al. Pre-operative stromal stiffness measured by shear wave elastography is independently associated with breast cancer-specific survival. Breast Cancer Res Treat. 2018;171(2):383-9.

296. Houthuijzen JM, Jonkers J. Cancer-associated fibroblasts as key regulators of the breast cancer tumor microenvironment. Cancer Metastasis Rev. 2018;37(4):577-97.

297. Vaughan MB, Howard EW, Tomasek JJ. Transforming growth factor-beta1 promotes the morphological and functional differentiation of the myofibroblast. Exp Cell Res. 2000;257(1):180-9.

298. Al-Rakan MA, Colak D, Hendrayani SF, Al-Bakheet A, Al-Mohanna FH, Kaya N, et al. Breast stromal fibroblasts from histologically normal surgical margins are pro-carcinogenic. J Pathol. 2013;231(4):457-65.

299. Coppe JP, Desprez PY, Krtolica A, Campisi J. The senescence-associated secretory phenotype: the dark side of tumor suppression. Annu Rev Pathol. 2010;5:99-118.

300. Santoni M, Romagnoli E, Saladino T, Foghini L, Guarino S, Capponi M, et al. Triple negative breast cancer: Key role of Tumor-Associated Macrophages in regulating the activity of anti-PD-1/PD-L1 agents. Biochim Biophys Acta. 2017;1869(1):78-84.

301. Zhou J, Wang XH, Zhao YX, Chen C, Xu XY, Sun Q, et al. Cancer-Associated Fibroblasts Correlate with Tumor-Associated Macrophages Infiltration and Lymphatic Metastasis in Triple Negative Breast Cancer Patients. J Cancer. 2018;9(24):4635-41.

302. Riches LC, Trinidad AG, Hughes G, Jones GN, Hughes AM, Thomason AG, et al. Pharmacology of the ATM inhibitor AZD0156: potentiation of irradiation and olaparib responses pre-clinically. Mol Cancer Ther. 2019.

303. Sun T, Zhang W, Li Y, Jin Z, Du Y, Tian J, et al. Combination Immunotherapy with Cytotoxic T-Lymphocyte–Associated Antigen-4 and Programmed Death Protein-1 Inhibitors Prevents Postoperative Breast Tumor Recurrence and Metastasis. Molecular Cancer Therapeutics. 2020;19(3):802.

304. Jin MH, Oh DY. ATM in DNA repair in cancer. Pharmacol Ther. 2019:107391.

305. van Os NJ, Roeleveld N, Weemaes CM, Jongmans MC, Janssens GO, Taylor AM, et al. Health risks for ataxia-telangiectasia mutated heterozygotes: a systematic review, meta-analysis and evidence-based guideline. Clin Genet. 2016;90(2):105-17.

306. Choi S, Gamper AM, White JS, Bakkenist CJ. Inhibition of ATM kinase activity does not phenocopy ATM protein disruption: implications for the clinical utility of ATM kinase inhibitors. Cell Cycle. 2010;9(20):4052-7.

307. Mak JPY, Ma HT, Poon RYC. Synergism between ATM and PARP1 inhibition involves DNA damage and abrogating the G2 DNA damage checkpoint. Mol Cancer Ther. 2019.

308. Bang YJ, Im SA, Lee KW, Cho JY, Song EK, Lee KH, et al. Randomized, Double-Blind Phase II Trial With Prospective Classification by ATM Protein Level to Evaluate the Efficacy and Tolerability of Olaparib Plus Paclitaxel in Patients With Recurrent or Metastatic Gastric Cancer. J Clin Oncol. 2015;33(33):3858-65.

309. Mateo J, Carreira S, Sandhu S, Miranda S, Mossop H, Perez-Lopez R, et al. DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. N Engl J Med. 2015;373(18):1697-708.

310. Kim K, Skora AD, Li Z, Liu Q, Tam AJ, Blosser RL, et al. Eradication of metastatic mouse cancers resistant to immune checkpoint blockade by suppression of myeloid-derived cells. Proc Natl Acad Sci U S A. 2014;111(32):11774-9.

311. Liu L, Wang Y, Miao L, Liu Q, Musetti S, Li J, et al. Combination Immunotherapy of MUC1 mRNA Nano-vaccine and CTLA-4 Blockade Effectively Inhibits Growth of Triple Negative Breast Cancer. Mol Ther. 2018;26(1):45-55.

312. Clift R, Souratha J, Garrovillo SA, Zimmerman S, Blouw B. Remodeling the Tumor Microenvironment Sensitizes Breast Tumors to Anti-Programmed Death-Ligand 1 Immunotherapy. Cancer Res. 2019;79(16):4149-59.

313. Delort L, Rossary A, Farges MC, Vasson MP, Caldefie-Chezet F. Leptin, adipocytes and breast cancer: Focus on inflammation and anti-tumor immunity. Life Sci. 2015;140:37-48.

314. Strickland KC, Howitt BE, Shukla SA, Rodig S, Ritterhouse LL, Liu JF, et al. Association and prognostic significance of BRCA1/2- mutation status with neoantigen load,

number of tumor-infiltrating lymphocytes and expression of PD-1/PD-L1 in high grade serous ovarian cancer. Oncotarget. 2016;7(12):13587-98.

315. Le Du F, Eckhardt BL, Lim B, Litton JK, Moulder S, Meric-Bernstam F, et al. Is the future of personalized therapy in triple-negative breast cancer based on molecular subtype? Oncotarget. 2015;6(15):12890-908.

316. Miyashita M, Sasano H, Tamaki K, Chan M, Hirakawa H, Suzuki A, et al. Tumor-infiltrating CD8+ and FOXP3+ lymphocytes in triple-negative breast cancer: its correlation with pathological complete response to neoadjuvant chemotherapy. Breast Cancer Res Treat. 2014;148(3):525-34.

317. Gross S, Walden P. Immunosuppressive mechanisms in human tumors: why we still cannot cure cancer. Immunol Lett. 2008;116(1):7-14.

318. Leone P, Shin EC, Perosa F, Vacca A, Dammacco F, Racanelli V. MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. J Natl Cancer Inst. 2013;105(16):1172-87.

319. Alvarez-Navarro C, Lopez de Castro JA. ERAP1 structure, function and pathogenetic role in ankylosing spondylitis and other MHC-associated diseases. Mol Immunol. 2014;57(1):12-21.

320. Reeves E, Colebatch-Bourn A, Elliott T, Edwards CJ, James E. Functionally distinct ERAP1 allotype combinations distinguish individuals with Ankylosing Spondylitis. Proc Natl Acad Sci U S A. 2014;111(49):17594-9.

321. Mehta AM, Spaans VM, Mahendra NB, Osse EM, Vet JN, Purwoto G, et al. Differences in genetic variation in antigen-processing machinery components and association with cervical carcinoma risk in two Indonesian populations. Immunogenetics. 2015;67(5-6):267-75.

322. Mehta AM, Jordanova ES, Corver WE, van Wezel T, Uh HW, Kenter GG, et al. Single nucleotide polymorphisms in antigen processing machinery component ERAP1 significantly associate with clinical outcome in cervical carcinoma. Genes Chromosomes Cancer. 2009;48(5):410-8.

323. Cancer Research UK [Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive.

324. Poggio F, Lambertini M, de Azambuja E. Surgery of the primary tumour in patients presenting with de novo metastatic breast cancer: to do or not to do? ESMO Open. 2018;3(1):e000324.